Classification and Characterization of Emotional Body Expression in Daily Actions

Nesrine Fourati

2015-ENST-0050

EDITE - ED 130

Doctorat ParisTech

THÈSE

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Traitement du Signal et des Images »

présentée et soutenue publiquement par

Nesrine FOURATI

le 16 Septembre 2015

Classification et Caractérisation de l'Expression Corporelle des Emotions dans des Actions Quotidiennes

Directeur de thèse : **Catherine Pelachaud**

Jury

Mme Nadia Berthouze, Professeur, UCLIC, University College London	Rapporteur
M. Benoît Bardy, Professeur, EuroMov, Université de Montpellier 1	Rapporteur
M. Frédéric Bevilacqua, Chercheur-HDR, STMS IRCAM CNRS, Université Paris 6	Examinateur
Mme Laurence Likforman, Maître de Conférence - HDR, LTCI, Télécom-ParisTech	Examinateur
M. Franck Multon, Professeur, IRISA, Université de Rennes 2	Examinateur

TELECOM ParisTech
école de l'Institut Mines-Télécom - membre de ParisTech
46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

T H È S E

To Kaïs

Remerciements

Je voudrais d'abord exprimer ma profonde reconnaissance à ma directrice de thèse Catherine Pelachaud pour sa confiance, son encadrement enrichissant ainsi que son soutien inconditionné. Aussi, l'attention et le temps qu'elle m'a accordés, les corrections qu'elle m'a conseillées et les encouragements qu'elle n'a cessé de me prodiguer ont été des facteurs clés de succès de ma thèse.

J'ai eu aussi l'opportunité de travailler dans l'équipe Greta autour d'une ambiance remarquable. J'ai été marquée par l'accueil chaleureux qu'ils m'ont réservé dès mon arrivée; avec Andre-Marie Pez, Pierre Philippe, Yu Ding, Jing Huang, Ken Prepin, Radoslaw Niewiadomski, Sylwia Julia Hyniewska et Quoc Anh Le. Je tiens à remercier l'ingénieur Andre-Marie Pez, un des sachants Greta, pour son aide sur la partie développement. Je remercie également le docteur Jing Huang pour sa disponibilité et son aide technique tout au long de mes travaux de thèse. Je remercie aussi Radoslaw Niewiadomski et Ken Prepin pour leurs soutiens et encouragements ainsi que Sylwia Julia Hyniewska pour ses conseils précieux. Par ailleurs, je tiens à remercier Magalie Ochs, pour ses conseils et son soutien: son énergie et son enthousiasme m'ont aussi toujours motivé. Je tiens aussi à remercier particulièrement Mathieu Chollet, qui a démarré sa thèse le même jour que moi et qui m'a accompagné tout au long de mon parcours de thèse. Avec Mathieu, mon frère de thèse, nous avons fêté l'anniversaire de nos thèses respectives chaque année. Peu après mon arrivée à Télécom, Brian Ravenet et Florian Pecune ont aussi intégré notre équipe. Ce fut un plaisir de pouvoir travailler autour d'un environnement stimulant et d'avoir partagé les angoisses respectives à chaque doctorant. Nous avons eu aussi l'occasion de faire plusieurs échanges constructifs qui m'ont, personnellement, beaucoup aidé à surmonter plusieurs difficultés rencontrées à travers les différentes étapes de ma thèse.

Tout au long de mon parcours à Télécom, j'ai pu aussi rencontrer, collaborer ou partager divers discussions avec plusieurs collègues: Angelo Cafaro (je tiens à saluer son attitude positive, sa gentillesse et son dynamisme incessant), Sabrina Campano (je tiens aussi à saluer sa sympathie et à la remercier pour son aide précieuse), Chloé Clavel (je tiens à la remercier également pour ses encouragements et son soutien), Brice Donval, Nadine Glas (ma chère voisine de bureau) et Caroline Langlet. J'adresse en particulier mes remerciements à Caroline pour son encouragement et son soutien, pour toutes les discussions agréables que nous avons pu avoir ensemble, et pour les moments inoubliables et exceptionnels partagés lors de nos déplacements à Reykjavik et à Xi'an.

Je souhaite également exprimer ma profonde gratitude à tous les acteurs qui ont participé à la collection de la base de données Emilya. Je remercie également Janina Breidaka qui a été recrutée dans le cadre des post-traitements de la base de données, et qui a apporté une aide considérable au traitement manuel d'une bonne partie des données. Une mention spéciale pour le directeur professionnel Robert Bennet qui a formé les acteurs, ainsi que pour Gilbert Pappalia et Giovanni Nolli qui m'ont beaucoup aidé sur les aspects logistique et technique.

Je dois ma réussite en partie à une famille qui n'a cessé de me soutenir et m'encourager; ma mère et mon père, ma petite sœur, mon grand frère, et enfin ma belle-famille qui m'a aussi beaucoup soutenu et particulièrement ma belle-mère et ma belle-sœur. Je tiens aussi à remercier mes amis, particulièrement Nouha Sallem, Wiem Ghazel, Amira Makhlouf, Emna Ben Ayed, Hazar Mliki et Wael Hachicha pour leurs encouragements permenants. Et enfin un grand merci à tous mes amis à Paris pour leurs soutiens continus.

Enfin, je ne pourrais jamais exprimer assez de gratitude pour mon bien-aimé Kaïs qui, malgré la tension et les difficultés rencontrées pendant ma thèse, n'a pas hésité à faire de son mieux pour me motiver, me soutenir et m'encourager. Les mots ne peuvent pas exprimer suffisamment ma reconnaissance pour son aide et pour sa présence.

Résumé

Cette thèse se situe dans le domaine de l'informatique affective. Le domaine de l'informatique affective combine plusieurs disciplines, inclus la psychologie et l'informatique pour étudier la reconnaissance, l'interprétation et la synthèse des états émotionnels. La reconnaissance des émotions à partir des expressions faciales et corporelles a été largement explorée en informatique affective afin de développer des machines "intelligentes", "affectives" et "interactives" pour différentes applications comme l'éducation, les soins de santé, les jeux et le divertissement. L'analyse de l'expression corporelle des émotions dans des actions quotidiennes est basée sur des paramètres d'expressivité qui décrivent la posture, le mouvement ou les deux (paramètres spatiotemporels). Par contre, la plupart des études qui s'intéressent à l'analyse des expressions corporelles des émotions dans des actions quotidiennes se limitent à un nombre d'émotions et d'actions très limités. Ces études se focalisent généralement sur des paramètres corporels qui dépendent de l'action comme par exemple les paramètres qui décrivent le mouvement de la flexion du coude durant l'action de frapper à la porte. En effet, l'effet du type de mouvement (e.g. marche, frapper à la porte) sur la reconnaissance et la caractérisation de l'expression corporelle de l'émotion reste peu clair. Il est temps de dépasser les systèmes de reconnaissance de l'expression corporelle des émotions à partir d'une action en particulier. Ceci aidera à mieux généraliser l'expression corporelle des émotions dans différents types de mouvement. Dans cette thèse, nous explorons la classification et la caractérisation de l'expression corporelle des émotions dans différentes actions quotidiennes. Huit émotions (Fierté, Joie, Tristesse, Honte, Colère, Peut panique, Anxiété, et Neutre) et 7 actions (Marcher, Marcher avec un objet dans la main, Frapper à la porte, S'asseoir, Déplacer une pile de livres, Ramasser et Jeter un objet) sont considérés. Nous explorons la classification et la caractérisation de l'expression corporelle des émotions en se basant sur les données de capture de mouvement 3D.

Ce travail de thèse peut être résumé en quatre étapes principales. Premièrement, nous avons proposé un système d'annotation multi-niveaux pour décrire le mouvement corporel expressif dans différentes actions. Deuxièmement, nous avons enregistré une base de données de l'expression corporelle des émotions dans des actions quotidiennes. Cette base de données constitue un large corpus de comportements expressifs considérant l'expression de 8 émotions dans 7 actions quotidiennes, combinant à la fois les données audio-visuelle et les données de capture de mouvement et donnant lieu à plus que 8000 séquences de mouvement expressifs. Troisièmement, nous avons exploré la classification des émotions en se basant sur notre système d'annotation multi-niveaux. L'approche des forêts aléatoires est utilisée pour cette fin. L'utilisation des forêts aléatoires dans notre travail a un double objectif : 1) la fiabilité du modèle de classification, et 2) la possibilité de

sélectionner un sous-ensemble de paramètres pertinents en se basant sur la mesure d'importance retournée par le modèle. Nous avons aussi comparé la classification automatique des émotions avec la perception humaine des émotions exprimées dans différentes actions. Finalement, nous avons extrait les paramètres les plus pertinents qui retiennent l'expressivité du mouvement en se basant sur la mesure d'importance retournée par le modèle des forêts aléatoires. Nous avons utilisé ce sous-ensemble de paramètres pour explorer la caractérisation de l'expression corporelle des émotions dans différentes actions quotidiennes. Un modèle d'arbre de décision a été utilisé pour cette fin.

Summary

This thesis falls within the framework of Affective Computing. Affective Computing combines several disciplines including psychology and computer science to study the recognition, the interpretation and the synthesis of affect. The recognition of emotions from facial and bodily expressions has been recently widely explored in affective computing field in order to build "intelligent", "affective" and "interactive" machines in a number of areas such as education, health care, games and entertainment. The analysis of bodily expression of emotions in daily actions is based on expressive body cues that describe body form, body motion or both (spatiotemporal features). However, most of the previous studies that analyzed bodily expression of emotions in daily actions focused on a limited range of emotional labels, actions and body cues. They mostly focused on action-depended body cues such as the elbow flexion features in Knocking action. Indeed, the effect of the performed movement task (e.g. walking, knocking at the door) on the recognition and the characterization of emotions expression remain unclear. There is a need to go beyond action-dependent emotion recognition and characterization to provide better insights into the expression of emotions in body movement. In this thesis, we explore the classification and the characterization of emotional body expression in different daily actions. Eight emotions (Pride, Joy, Sadness, Shame, Anger, Panic Fear, Anxiety and Neutral) and seven actions (Simple Walking, Walking with an object in hand, Knocking, Sitting Down, Moving Books, Lifting and Throwing) are considered for this purpose. We explore the classification and the characterization of emotional body expressions based on 3D motion-capture data.

The work conducted in this thesis can be summarized into four main steps. Firstly, we proposed a multi-level body movement notation system that allows the description of expressive body movement across various body actions. Secondly, we collected a new database of emotional body expression in daily actions. This database constitutes a large repository of bodily expression of emotions including the expression of 8 emotions in 7 actions, combining video and motion capture recordings and resulting in more than 8000 sequences of expressive behaviors. Thirdly, we explored the classification of emotions based on our multi-level body movement notation system. Random Forest approach is used for this purpose. The advantage of using Random Forest approach in our work is double-fold: 1) reliability of the classification model and 2) possibility to select a subset of relevant features based on their relevance measures. We also compared the automatic classification of emotions with human perception of emotions expressed in different actions. Finally, we extracted the most relevant features that capture the expressive content of the motion based on the relevance measure of features returned by the Random Forest model. We used this subset of features to explore the characterization of emotional

body expression across different actions. A Decision Tree model was used for this purpose.

Contents

Résumé Long

Introduction

L'étude de l'expression faciale et corporelle des émotions a été largement explorée en informatique affective afin de développer des machines "intelligentes", "affectives" et "interactives" pour différentes applications. Bien que les études qui portent sur la reconnaissance et la synthèse des expressions faciales aient été largement répandues dans la littérature, l'étude de l'expression corporelle a été relativement limitée. Ceci est probablement dû au fait que l'expression corporelle a été généralement considérée comme étant liée à l'intensité de l'émotion. Des études antérieures ont montré la capacité des humains à reconnaître les émotions exprimées seulement avec des mouvements corporels.

L'étude de l'expression corporelle d'émotion trouve son intérêt dans différentes applications. Par exemple, l'expression corporelle joue un rôle primordial pour la crédibilité de l'expressivité des acteurs dans le monde de théâtre et de cinéma. Elle joue énormément aussi sur la crédibilité d'un agent virtuelle lors d'une interaction homme machine [Demulier et al., 2014] [Huang and Pelachaud, 2012]. La génération de l'expression d'émotion a été aussi employée dans les émoticônes qui font de plus en plus partie de nos discussions écrites. Plusieurs autres applications intègrent des systèmes de reconnaissance automatique de l'expression corporelle d'émotion comme dans l'éducation [Mota and Picard, 2003], pour détecter l'état émotionnel des étudiants, dans les jeux vidéo pour détecter l'état émotionnel des joueurs [Kleinsmith et al., 2011] ou dans la e-santé pour détecter l'état émotionnel des patients [Aung et al., 2015] comme la dépression et les émotions déclenchées par la douleur chronique.

L'expression corporelle des émotions peut être effectuée de manière explicite ou implicite. L'expression explicite se traduit par la communication des émotions à travers des gestes spécifiques [Dael et al., 2011]. L'expression implicite implique la variation de la manière avec laquelle on effectue le même mouvement selon l'état émotionnel exprimé [Roether et al., 2009] [Hicheur et al., 2013] [Gross et al., 2010]. Dans les travaux de cette thèse, nous nous focalisons sur l'expression implicite, et particulièrement sur l'expression d'émotion durant des actions quotidiennes.

L'analyse de l'expression corporelle des émotions dans des actions quotidiennes est basée sur des paramètres d'expressivité qui décrivent la posture, le mouvement ou les deux (paramètres spatiotemporels). Par contre, la plupart des études qui s'intéressent à l'analyse des expressions corporelles des émotions dans des actions quotidiennes se limitent à un nombre d'émotions et d'actions très limités [Roether et al., 2009] [Hicheur et al., 2013] [Gross et al., 2010]. Ces études se focalisent généralement sur des paramètres corporels intrinsèques à l'action comme par ex-

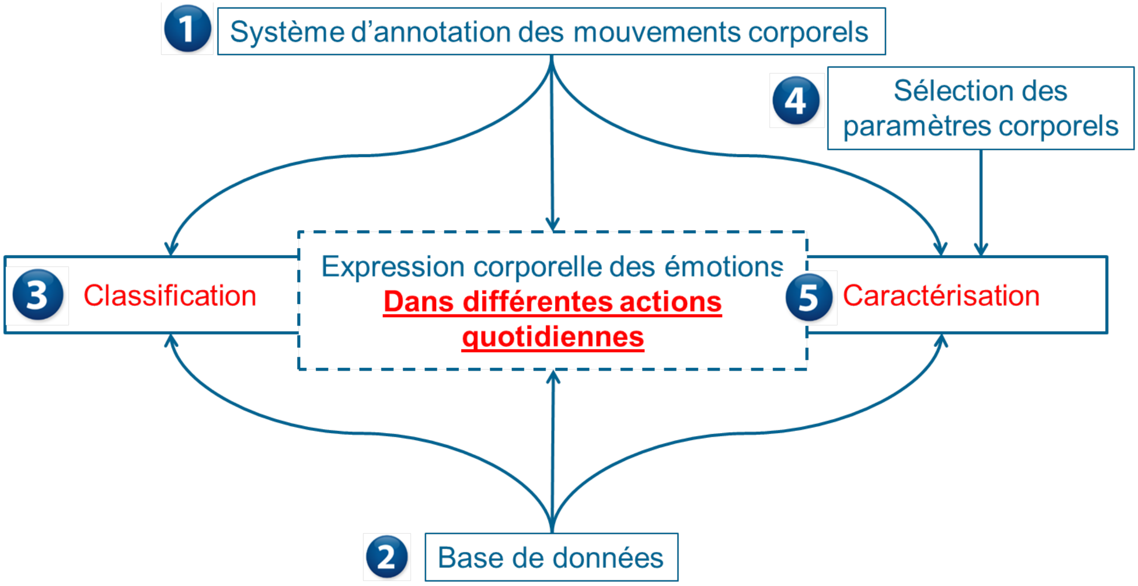


Figure 1: Les étapes de notre méthodologie.

emple les paramètres qui décrivent le mouvement de la flexion du coude durant l'action de frapper à la porte [Bernhardt and Robinson, 2007]. En effet, l'effet de l'action (e.g. marcher, frapper à la porte) sur la reconnaissance et la caractérisation de l'expression corporelle de l'émotion reste peu clair. Il serait important d'étudier l'expression corporelle des émotions dans différentes actions afin de mieux généraliser l'expression corporelle des émotions. Dans cette thèse, nous explorons la classification et la caractérisation de l'expression corporelle des émotions dans différentes actions quotidiennes. Huit émotions (Fierté, Joie, Tristesse, Honte, Colère, Peut panique, Anxiété, et Neutre) et 7 actions (Marcher, Marcher avec un objet dans la main, Frapper à la porte, S'asseoir, Déplacer une pile de livres, Ramasser et Jeter un objet) sont considérés. Nous explorons la classification et la caractérisation de l'expression corporelle des émotions en se basant sur les données de capture de mouvement 3D.

Questions de recherche

Les questions de recherche principalement abordées dans cette thèse sont les suivantes ; Peut-on percevoir, classifier et caractériser une émotion exprimée à travers le mouvement corporel de la même manière dans différentes actions quotidiennes ? Quel est l'effet de l'action sur la perception, la classification et la caractérisation de l'expression d'une émotion ? Nous souhaitons aussi trouver les paramètres expressifs qui caractérisent l'expression d'émotion dans différentes actions quotidiennes.

Méthodologie

Pour aborder ces questions de recherche, différents besoins ont été identifiés. En premier lieu, nous avons identifié le besoin d'un système de codage permettant la description du mouvement expressif. En deuxième lieu, nous avons identifié le besoin d'une base de données de comportements corporels expressifs. La spécification des paramètres corporels pour décrire le mouvement est utile pour classifier les expressions émotionnelles corporelles et pour explorer les caractéristiques intrinsèques de l'expression d'émotion.

Notre méthodologie est décrite dans la Figure 1. Elle repose sur les étapes suivantes: d'abord nous avons proposé un système d'annotation permettant de capturer l'expressivité corporelle. Ensuite Nous avons enregistré une base de données de l'expression corporelle. Nous avons ensuite étudié la classification des émotions dans chacune des actions. Enfin, nous avons étudié les paramètres corporels qui contribuent le plus à la classification correcte des émotions et nous nous sommes basés sur ces paramètres pour étudier la caractérisation de l'expression corporelle. Dans les sections suivantes, nous illustrons chacune de ces étapes.

Système d'annotation des mouvements corporels

Nous avons identifié 5 principaux niveaux de descriptions utilisées généralement pour la description du mouvement expressives. Le niveau anatomique (i.e. la décomposition du corps en différentes entités come le haut ou le bas du corps), le niveau directionnel (i.e. la direction dans laquelle on effectue le mouvement), le niveau qui distingue la description de la forme et de la dynamique du mouvement, le niveau spatial (i.e. on peut décrire le mouvement du corps dans l'espace personnel qui englobe le corps ou bien on peut décrire le déplacement du corps dans l'espace qui l'entoure), et enfin le niveau fonctionnel (i.e. décrire les gestes selon l'aspect sémantique).

Différent systèmes d'annotation ont été proposés en littérature en se basant sur ces niveau de descriptions. On trouve par exemple le système BAP proposé par Nele Dael et ses collègues en 2012 [Dael et al., 2012]. Ce système est dédié pour la description de l'expression explicite des émotions. Il décrit de manière détaillée les unités de posture et des actions. On trouve aussi les systèmes de codage qui sont spécifiques à la description de posture à travers des mesures de distance [Kleinsmith and Berthouze, 2007]. Le système SPAFF [Coan and Gottman, 2007] décrit à la fois les expressions faciales et corporelles. Le système LABAN [Laban, 1988] a été d'abord proposé pour décrire les mouvements des danseurs. Il a été récemment adopté pour décrire l'expression corporelle des émotions. D'autres études se focalisaient sur des paramètres définis par rapport à des contraintes particuliers comme le type de données ou le type de mouvement [Mota and Picard, 2003],[Roether et al., 2009]. Par contre, ce type de paramètres ne peut pas être utilisé dans différentes actions. D'autres études en psychologie [Wallbott, 1998, Meijer, 1989] proposent d'étudier

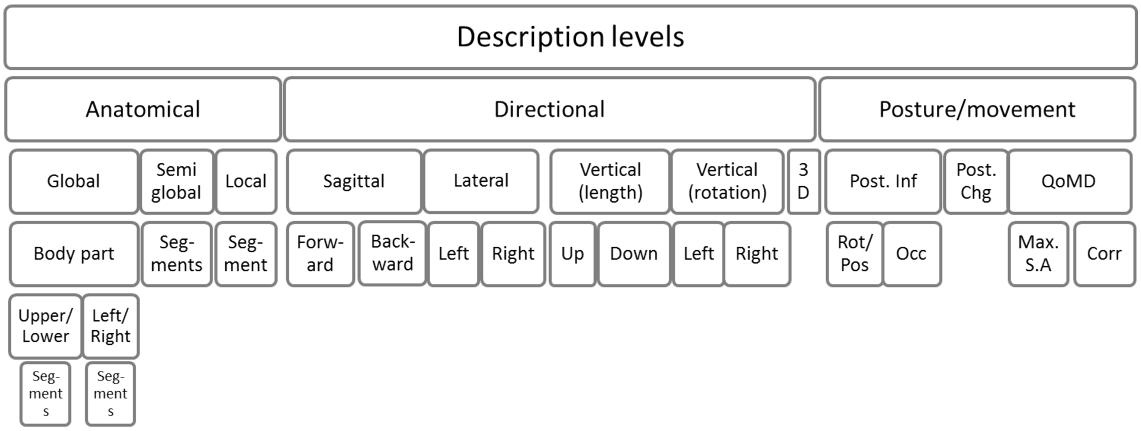


Figure 2: Les niveaux de descriptions utilisés pour définir notre système d'annotation multi-niveau.

l'expression d'émotion avec des paramètres corporels de haut niveau. Cependant, l'implémentation de ces paramètres peut être différente d'une étude à une autre. Les systèmes d'annotation utilisés pour décrire l'expression d'émotion fournissent ou bien une description détaillée de la posture tout en négligeant la description de la qualité de mouvement, ou bien une description de la qualité de mouvement sans pour autant fournir une quantification détaillée des paramètres corporels.

De ce fait, nous avons proposé un système d'annotation multi-niveaux qui regroupe les avantages des systèmes d'annotation existants. Ce système repose essentiellement sur trois niveaux de description : Anatomique, Directionnel et Posture/ Mouvement (voir Figure 2). Nous considérons que le niveau fonctionnel est plus adapté pour la description de l'expression d'émotion explicite, étant donné que celle-ci se définit par des gestes explicites bien défini. Dans ce travail de thèse, nous nous intéressons seulement à l'expression implicite, c'est-à-dire à la manière avec laquelle le mouvement est effectué.

Notre système d'annotation multi-niveaux est décrit dans la figure 2. A l'inverse de plusieurs autres études qui se focalisent sur la description des segments du corps principalement impliqués dans le mouvement, notre système d'annotation permet de décrire le mouvement du corps entier. De plus, nous considérons différentes directions du mouvement et nous proposons de décrire aussi bien la forme que la dynamique du mouvement. Ce système d'annotation multi-niveaux est utilisé dans ce travail de thèse pour définir un ensemble de paramètres corporels, qui vont être plus tard utilisé pour décrire un mouvement corporel expressif. Nous avons proposé un ensemble de 114 paramètres. La description de chaque paramètre se fait à l'aide de la communication entre plusieurs niveaux de descriptions. Par exemple, la distance 3D entre les pieds se définit sur la dimension anatomique par la relation entre deux segments du corps, qui sont les pieds. Ce paramètre se définit sur le niveau directionnel par une distance 3D. Enfin la distance entre les pieds se définit sur la dernière dimension par 4 paramètres; 1) un paramètre de posture (e.g. la distance maximale entre les pieds), 2) un paramètre de changement de posture qui est l'écart type de mouvement et enfin 3) deux paramètres de la qualité de mouvement qui sont la vitesse et l'accélération du mouvement. En total nous avons proposé 114 paramètres qui se définissent de manière similaire par la combinaison de ces différents niveaux de descriptions. Nous présentons ces paramètres de manière détaillé dans l'annexe E.

Ainsi, nous avons proposé un système de codage multi-niveaux permettant de fournir une description détaillée et structurée de la posture et la qualité du mouvement corporel. Nous avons proposé un ensemble de 114 paramètres corporels. Après avoir proposé la description des mouvements, la deuxième étape de ce travail de thèse consiste à collecter une nouvelle base de données.

Collection d'une base de données

Figure 3 résume la littérature sur l'enregistrement des expressions corporelles des émotions. L'enregistrement des comportements corporels émotionnels englobe deux aspects : l'expression d'émotion et le mouvement corporel (voir Figure 3).

L'expression d'émotion peut être "spontanée" ou "actée". Un des problèmes les plus communs est de pouvoir enregistrer et étudier l'expression "spontanée" de l'émotion (dites aussi "naturelle"). Pour pouvoir enregistrer des expressions naturelles, certaines recherches se focalisent sur l'expression des joueurs durant un jeu vidéo [Kleinsmith et al., 2011]. Néanmoins, il est difficile de capturer l'expression spontanée d'une émotion donnée durant une certaine action. Pour avoir un bon compromis entre l'aspect "acté" et l'aspect naturel de l'expression, il existe des méthodes d'inductions qui permettent d'aider l'acteur à imaginer un scénario qui décrit une certaine séquence d'événement. L'émotion est donc induite par le contexte donné par le scénario.

L'enregistrement du mouvement corporel effectué dans les bases de données existantes repose généralement sur une caméra (donnant lieu à des données audiovisuelle) ou bien sur un système de capture de mouvement 3D (voir Figure 3). La première approche favorise le côté naturel de l'expression étant donné qu'elle n'impose pas de contrainte physique à la personne. Cependant, plusieurs traitements sont nécessaires pour pouvoir extraire une description détaillée du mouvement corporel à partir des données audio-visuelles. Les systèmes de capture de mouvement 3D fournissent des données plus précises favorisant la description détaillée du mouvement corporel. Néanmoins, ils réduisent l'aspect naturel de l'expression étant donné que l'acteur est obligé de porter une combinaison spécifique.

Les bases de données existantes qui reposent sur l'utilisation d'un système de capture de mouvement 3D se limitent généralement à un nombre d'actions et d'émotions limité [Ma et al., 2006] [Kleinsmith et al., 2006b]. Certaines bases de données contiennent à la fois des vidéos et des données de capture de mouvements 3D, mais elles sont également limitées à un spectre restreint d'actions et d'émotions [Gross et al.,

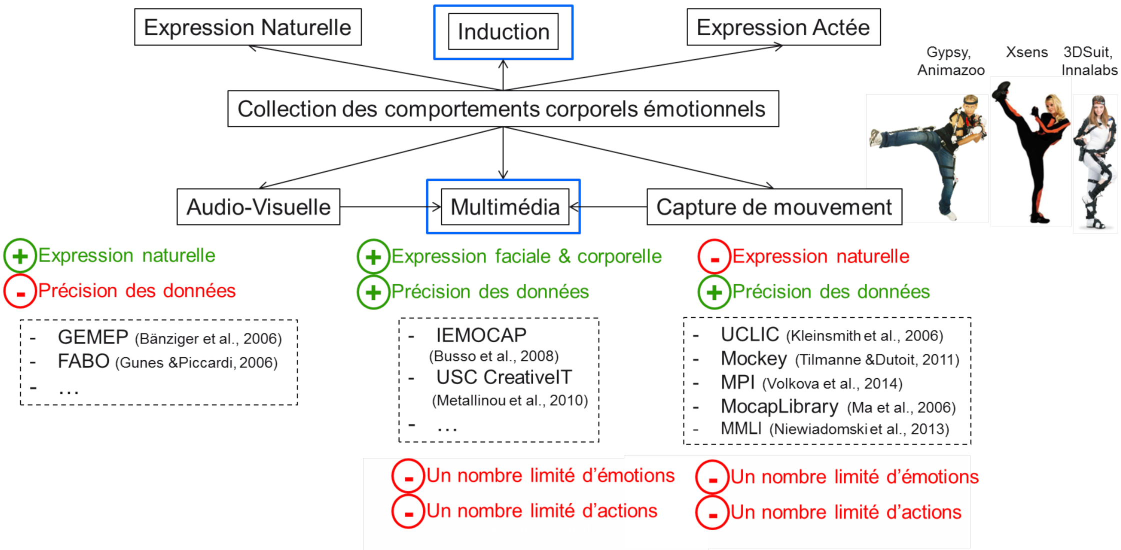


Figure 3: L'enregistrement des comportements corporels émotionnels dans la littérature.

2010].

De ce fait, nous avons proposé d'enregistrer une nouvelle base de données qui se base sur une technique d'induction, qui implique les deux types d'enregistrement discutés, et qui repose sur un large spectre d'actions et d'émotions en comparaison avec les études précédents. Cette base de données s'appelle Emilya pour "EMotional body expression In daiLY Actions". 11 acteurs ont participé à la collection de cette base de données. Ils ont tous reçu 7 séances de formation avec un directeur professionnel américain. Le rôle du directeur professionnel pendant les séances de formation était de s'assurer que les acteurs n'exagèrent pas leurs comportements émotionnels afin d'éviter les expressions stéréotypées. Son rôle a été aussi d'aider les acteurs à se focaliser sur l'expression corporelle et non pas faciale.

Nous avons utilisé deux types d'enregistrement: un système de capture de mouvement inertiel qui est Xsens [Roetenberg et al., 2009] et 4 caméras synchronisées pour capturer les données audio-visuelle. Figure 4 illustre le protocole de synchronisation des données que nous avons utilisé. Un générateur référence a été utilisé pour émettre un code temporel. Ce dernier a été transmis à toutes les caméras ainsi qu'au logiciel (MVN Pro) d'enregistrements des données de capture de mouvements fourni avec le système Xsens. Ainsi, les données audio-visuelles et les données de capture de mouvement ont été synchronisées. Figure 5 montre l'extraction de la même posture d'une vidéo et d'un fichier de capture de mouvement.

Nous avons demandé aux acteurs de jouer l'expression de 8 émotions qui sont la Joie, la Colère, la Tristesse, la Peur panique, la Fierté, l'Anxiété, la Honte et enfin la Neutre. Il a été montré que l'expression de ces émotions peut durer un moment, ce qui favorise l'étude de l'expression corporelle durant la performance d'une

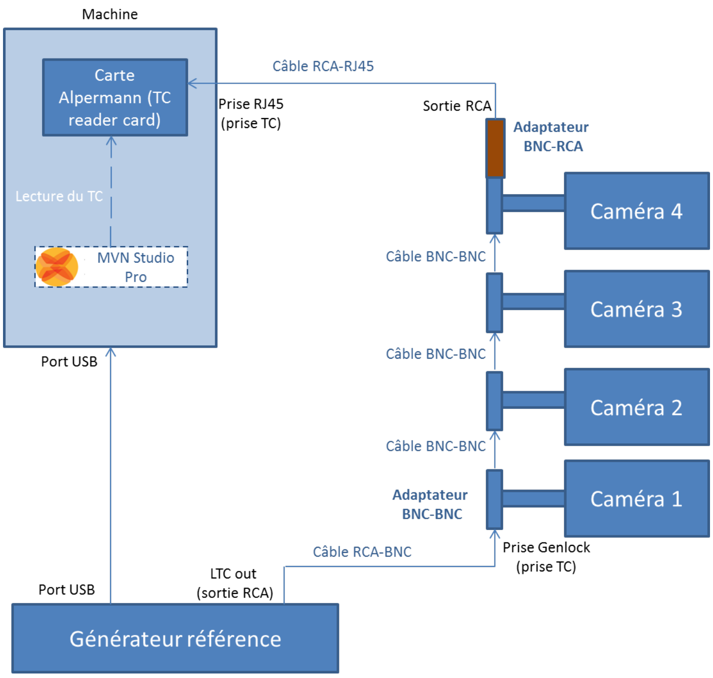


Figure 4: Le protocole de synchronisation des données de capture de mouvement et les données audio-visuelles que nous avons utilisé pour enregistrer les expressions corporelles dans la base de données Emilya.

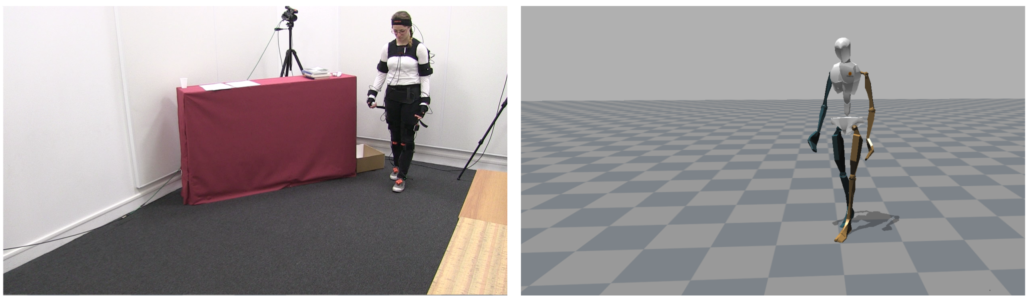


Figure 5: Extraction de la même posture d'une vidéo et d'un fichier de capture de mouvement.

action [Dael et al., 2011] [Montepare et al., 1987]. Nous avons utilisé une méthode d'induction à base de scénario pour essayer d'induire l'expression d'émotion. Nous avons utilisé 3 scénarios pour chaque émotion. Ces scénarios ont été validés dans d'autres travaux en psychologie. A titre d'exemple, un scénario de joie est "Mon oncle arrive pour mon anniversaire. Il m'a demandé de regarder par la fenêtre. Son cadeau est devant la maison : la voiture que nous avons toujours rêvée d'avoir". Donc nous avons demandé aux acteurs de relire le scénario spécifique à l'émotion, d'imaginer la situation et d'exprimer l'émotion durant les actions quotidiennes. Nous avons considéré 7 actions quotidiennes ; 3 actions ont été choisi pour étudier l'expression d'émotion corporelles dans des actions impliquant le corps entier comme s'assoir, marcher et marcher avec quelque chose dans la main, 2 actions pour étudier les mouvements répétitifs comme frapper à la porte et déplacer une pile de livres de droite à gauche, et 2 autres actions pour étudier les mouvements impliquant principalement un seul segment du corps comme ramasser et jeter un objet. Nous avons demandé aux acteurs de répéter chaque action 4 fois pour capturer un large ensemble de données.

Le directeur professionnel a été aussi invité pour une séance d'enregistrement. Nous avons obtenu 8206 fichiers de capture de mouvement représentant 12 acteurs, 8 émotions, 3 scénarios, 7 actions et 4 répétitions pour chacune, et plus que 8000 fichiers vidéo pour les 4 points de vue. Dans cette thèse, nous avons principalement utilisé les données de capture de mouvement 3D.

La prochaine étape de ces travaux de thèse consiste à étudier la perception des expression corporelles afin de valider la base de données Emilya. Cette étape permet de s'assurer que les émotions qui ont été exprimées par les acteurs peuvent êtres perçues comme telles. Nous souhaitons aussi étudier la caractérisation de ces expressions corporelles à travers l'évaluation de quelques caractéristiques corporelles comme la vitesse et l'amplitude de mouvement.

Validation de la base de données Emilya

La validation de la base de données repose sur une étude perceptive. Cette étude consiste à demander à des participants d'évaluer des séquences de mouvement expressif. La base de données Emilya contient plus que 8200 fichiers de capture de mouvement. L'évaluation de tous les fichiers serait couteuse en termes de temps. Nous avons donc choisi de valider un ensemble de 664 séquences qui correspondent à une séquence par acteur, émotion et action.

En utilisant les vidéos acquises par les caméras (i.e. les données audio-visuelles), la perception de l'expression corporelle peut être biaisée par plusieurs facteurs comme les expressions faciales et le genre de l'acteur. Afin d'éviter ce biais, nous avons utilisé les données de capture de mouvement pour créer les stimuli nécessaires pour notre étude perceptive. En effet, les données de capture de mouvement 3D ont été reproduis sur un avatar virtuel (voir Figure 6). Nous avons développé une approche permettant de créer automatiquement les stimuli tout en contrôlant de

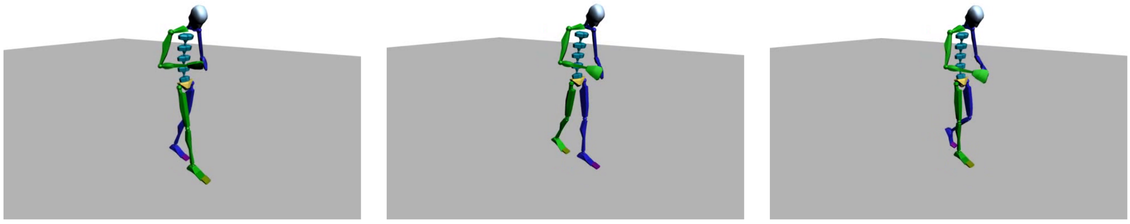


Figure 6: Un exemple de stimuli crée par le logiciel 3Ds Max : trois captures d'écran correspondantes à trois postures prises d'une séquence de marche expressive.

façon automatique le point de vue de l'avatar et le déplacement de la caméra dans le but d'éviter l'influence de ces facteurs sur la perception de l'expression corporelle. MAXScript (le langage de script de 3DS MAX) a été utilisé pour cette fin.

Après avoir créer les stimuli, nous avons utilisé la plateforme Amazon pour avoir accès à plusieurs participants. 1008 personnes ont participé à notre étude perceptive en ligne. Chaque participant a été demandé d'évaluer 16 vidéos, et chaque vidéo a été évaluée 24 fois. Nous avons demandé aux participants d'évaluer la reconnaissance de l'émotion exprimée par l'avatar ainsi que la caractérisation du mouvement à travers 8 paramètres d'expressivité (voir Figure 7).

En ce qui concerne la reconnaissance de l'émotion, nous avons demandé aux participants d'annoter sur une échelle de Likert la présence de chacune des 8 émotions (voir Figure 7). Les participants pouvaient aussi proposer un autre label d'émotion ou sélectionner le bouton “ je ne sais pas ”.

En se basant sur l'approche de l'émotion reconnue de manière plus fréquente, nous avons quantifié le taux de reconnaissance de l'émotion ainsi que la matrice de confusion (voir Figure 8). Le taux de reconnaissance des émotions est de l'ordre de 41%. En examinant les confusions de perception des émotions, nous avons remarqué que les participants ont pu reconnaitre l'expression de l'anxiété, la joie, tristesse, la colère et la neutre (voir Figure 8). Nous avons également remarqué que la fierté a été confus avec la neutre, la peur panique avec l'anxiété et la honte avec la tristesse (voir Figure 8).

Ces confusions peuvent être expliquées par la similarité des comportements corporels de ces expressions, mais aussi par le manque des facteurs contextuels. En effet, les participants ont évalué les expressions corporelles sans avoir recours aux scénarios qui ont été utilisé pour induire les émotions. La confusion entre certaines émotions (e.g. entre la neutre et la fierté, la honte et la tristesse) peut être aussi due à l'absence d'autres modalités comme les expressions faciales, la tension musculaire et la direction du regard. Nous avons aussi remarqué que la plupart de ces confusions sont unidirectionnelles. Par exemple, la peur panique a été confuse avec l'anxiété, mais l'anxiété n'a pas été perçue comme Peur Panique. De même la honte a été perçue comme tristesse mais la tristesse n'a pas été perçue comme honte.

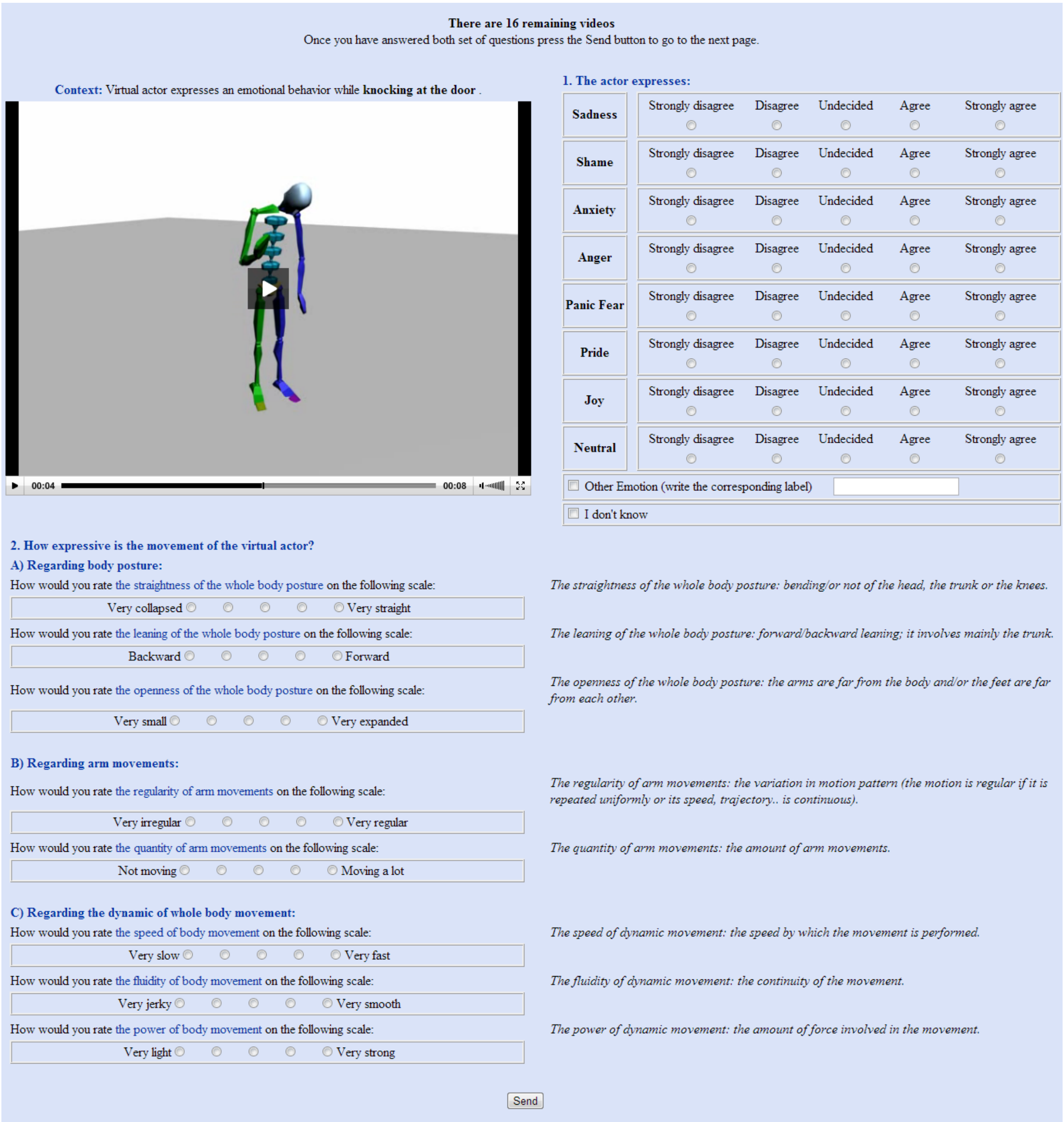


Figure 7: Capture d'écran du protocole d'expérimentation utilisé pour valider la base de données Emilya.

	Human recognition of emotions across all the actions							
	CCR = 41%							
	Anxiety	Pride	Joy	Sadness	Panic Fear	Shame	Anger	Neutral
Anxiety	31	2	5	13	0	2	4	26
Pride	2	17	12	3	0	0	1	48
Joy	8	7	31	7	2	1	4	23
Sadness	2	0	0	73	0	0	0	8
Panic Fear	53	1	1	6	7	1	3	11
Shame	9	0	0	52	0	8	0	14
Anger	10	6	6	4	0	0	43	14
Neutral	0	3	2	13	0	0	1	64

Figure 8: Matrice de confusion de la reconnaissance des émotions obtenu lors du test perceptif.

Cependant, ces résultats sont issus de l'approche de l'émotion la plus fréquemment perçue. En examinant les résultats de la perception des émotions sur l'échelle de Likert, nous avons remarqué que les participants ont aussi perçu la honte dans les vidéos de la tristesse, mais avec un score plus faible. De même pour les vidéos qui montrent l'expression de honte. Ceci accentue l'importance de l'utilisation d'une échelle de Likert à la place d'un paradigme de choix forcé. En effet, ceci montre que l'évaluation multi-niveaux nous permet d'approfondir l'évaluation de la perception d'émotion.

La deuxième tâche de notre étude perceptive consiste à caractériser le mouvement de l'expression d'émotion à l'aide d'un ensemble de 8 paramètres décrivant la posture et le mouvement corporel: la rectitude du corps, le penchement vers l'avant ou vers l'arrière du corps, l'extension des membres du corps, la régularité, la quantité, la vitesse, la fluidité et enfin la force du mouvement (voir Figure 7). Ces paramètres ont été largement étudiés dans la littérature en psychologie pour caractériser les comportements corporels émotionnels. Nous avons proposé une définition de ces paramètres en cohérence avec notre système d'annotation multi-niveau. Par exemple, la rectitude du corps se définit sur le niveau anatomique par la rectitude des parties supérieur et inférieur du corps, sur le niveau directionnel par la direction vertical, et sur le dernier niveau par un paramètre de posture. Ces caractéristiques ont été évaluées sur une échelle sémantique différentielle. Chaque paramètre a été évalué sur une échelle bipolaire de 1 à 5. Par exemple la vitesse du mouvement varie entre très lent (niveau 1) jusqu'à très rapide (niveau 5).

Afin d'étudier les patterns de caractérisation de chaque émotion, nous avons mesuré la valeur moyenne de chaque paramètre. Les résultats sont montrés dans la Figure 9. La honte et la tristesse se caractérisent par un mouvement fluide, lent régulier et une posture effondré et fermée. Nous avons remarqué que la honte et la tristesse se caractérisent de manière très similaire, ce qui peut être cohérent avec la confusion qui a eu lieu au niveau de la perception d'émotion. Figure 9 montre aussi que l'expression corporelle de la peur panique se caractérise de manière similaire

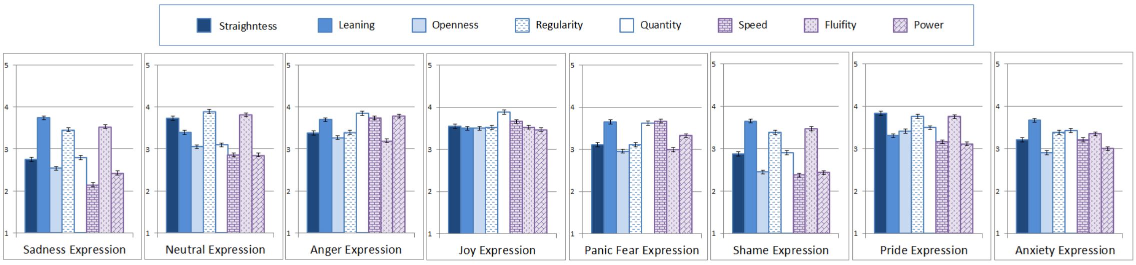


Figure 9: Patterns de caractérisation des expressions des émotions en fonction des valeurs moyennes des paramètres d'expressivités. Les valeurs de moyenne sont graduée de 1 à 5 de la valeur la plus faible jusqu'à la valeur la plus élevée de l'échelle à 5 niveaux.

que l'expression corporelle de la colère. Cependant, ces deux émotions n'ont pas été confuses au niveau de la tâche de perception de l'émotion. Ceci implique la nécessité de caractériser le mouvement corporel expressif avec plus de paramètres afin de mieux distinguer ces expressions corporelles.

Dans cette sections, nous avons présenté les résultats de l'étude perceptive. Nous avons discuté la perception des émotions exprimées dans les mouvements corporels. Nous avons aussi discuté la perception des caractéristiques corporelles expressives. Dans la section suivante, nous présentons l'approche et les résultats de la classification automatique des expressions corporelles en se basant sur l'analyse des données de capture de mouvement.

Classification des expressions corporelles

Dans cette section, nous nous concentrons sur l'analyse des données de capture de mouvement. Nous utilisons les 114 paramètres issus de notre système d'annotation de mouvement pour classifier les expressions corporelles. Différentes approches ont été utilisées dans les études précédentes pour classifier les expressions corporelles comme le perceptron multicouche, Les machines à vecteurs de support, les arbres de décisions ou les forêts aléatoires [Kapur et al., 2005] [Bernhardt and Robinson, 2007] [Kleinsmith et al., 2011]. Dans avons opté pour l'approche des forêts aléatoires. L'utilisation des forêts aléatoires dans notre travail a un double objectif : 1) la fiabilité du modèle de classification en présence d'un nombre large de paramètres et de données, et 2) la possibilité de sélectionner un sous-ensemble de paramètres pertinents en se basant sur la mesure d'importance retournée par le modèle. Nous avons développé une approche itérative permettant de détecter automatiquement le nombre d'arbres suffisant pour construire le modèle des forêts aléatoires. Ceci permet d'avoir un bon compromis entre l'efficacité et la robustesse de la forêt.

Pour comparer la classification automatique avec la reconnaissance humaine, nous avons comparé les taux pour chaque action et chaque émotion. Nous avons

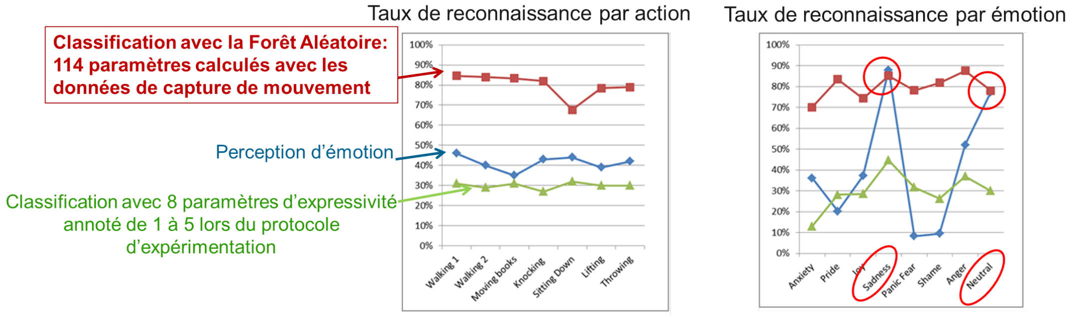


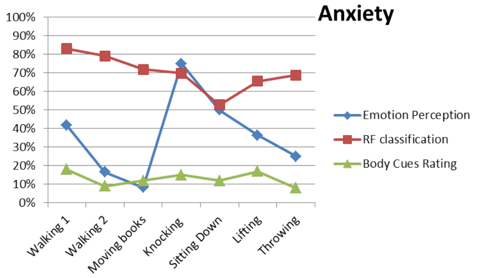
Figure 10: Comparaison des taux de classification humaine et taux de classification automatique des émotions ; par action et par émotion.

construit un modèle de foret aléatoire par action. Nous avons comparé d'abord les taux de classification fournis pour chaque action. Ensuite nous avons comparé les taux de classification de chaque émotion moyennant toutes les actions. Enfin, nous avons comparé les taux de classification de chaque émotion sur chaque action. Figure 10 illustre la comparaison de trois types de mesure: 1) les taux de classification automatique obtenus par l'approche des forêts aléatoires en utilisant les 114 paramètres corporel, 2) les taux de reconnaissance des émotions obtenus lors de l'étude perceptive, et enfin 3) les taux de classification automatique des émotions obtenus par la méthode de regression logisque à l'aide des 8 paramètres annotés lors de l'étude perceptive. Nous avons remarqué que l'utilisation des paramètres de capture de mouvement a donné le meilleur résultat, particulièrement en comparaison avec la classification automatique basé sur des paramètres annotés lors de l'étude perceptive.

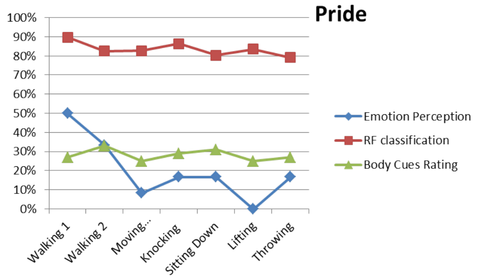
Nous avons trouvé que les classifications automatiques respectives de tristesse, de colère et de neutre s'alignent globalement avec la perception humaine. L'écart de différence entre les taux de classification automatique et de perception est principalement dû aux confusions présentes au niveau de la perception (voir Figure 10).

Nous avons aussi comparé les taux de classification pour chaque expression d'émotion dans chaque action (voir Figure 11). Nous avons trouvé que la tendance des taux de classification automatique s'aligne souvent avec la tendance des taux de reconnaissance humaine. L'expression de certaines émotions est mieux perçue et mieux classifiée dans certaines actions. Par exemple, la colère est mieux perçue et mieux classifiée dans l'action de jeter un objet. De même les expressions de joie et de fierté sont mieux perçues et mieux classifiée dans la marche.

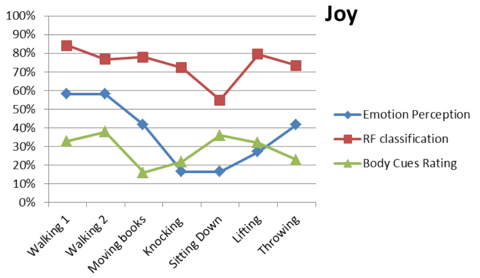
L'étape suivante de notre approche consiste à trouver les paramètres les plus pertinents permettant la caractérisation des expressions d'émotions.



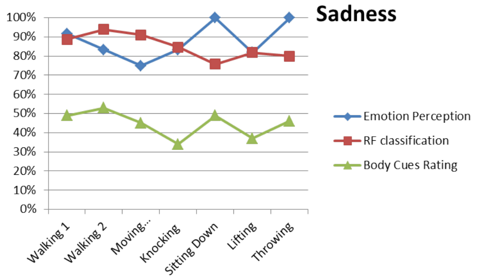
(a) Classification humaine Vs automatique de l'anxiété par action



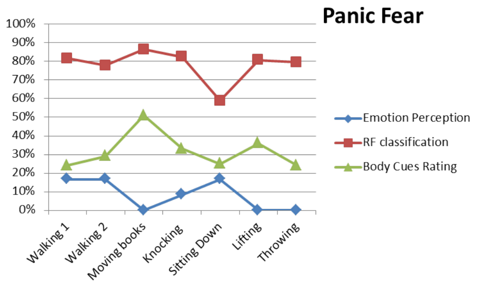
(b) Classification humaine Vs automatique de la fierté par action



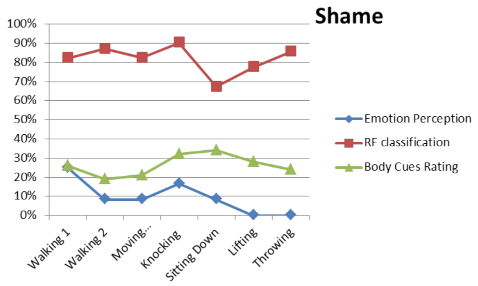
(c) Classification humaine Vs automatique of Joy per action



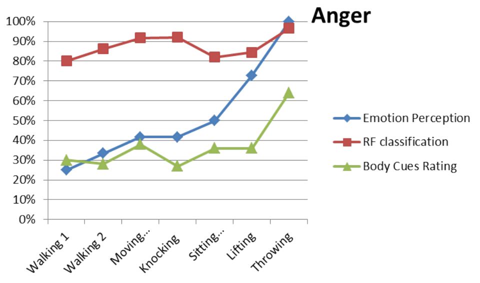
(d) Classification humaine Vs automatique de la tristesse par action



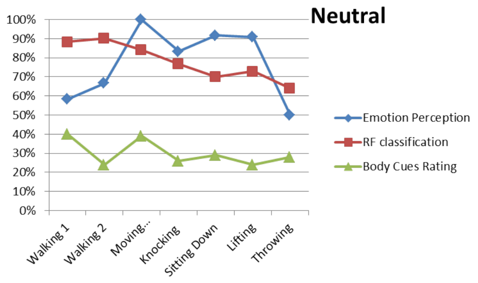
(e) Classification humaine Vs automatique de la peur panique par action



(f) Classification humaine Vs automatique de la honte par action



(g) Classification humaine Vs automatique de la colère par action



(h) Classification humaine Vs automatique de la neuter par action

Figure 11: Classification humaine vs classification automatique des émotions pour chaque action

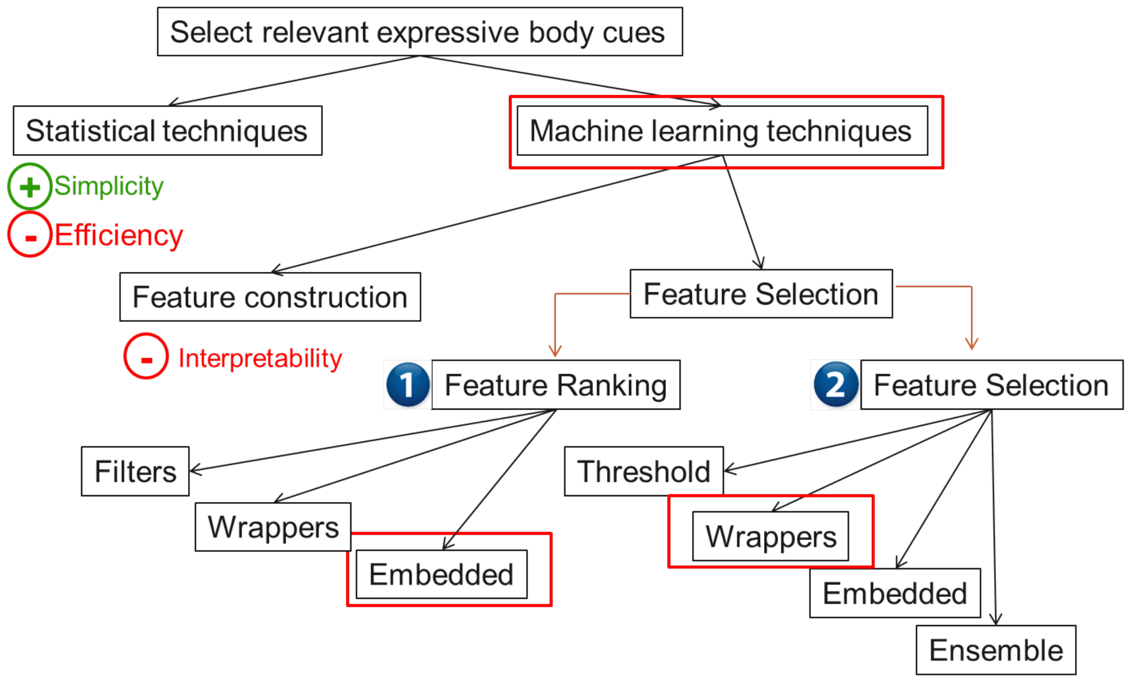


Figure 13: Résumé des méthodes de sélection de paramètres dans proposées dans la littérature.

SÉLECTION DES PARAMÈTRES PERTINENTS

Dans cette section, nous présentons l'approche et les résultats de sélection des paramètres corporels pertinents qui permettent de caractériser les expressions corporelles. Plusieurs approches peuvent être utilisées pour sélectionner les paramètres les plus importants (voir Figure 13). Les techniques statistiques par exemples permettent de trouver de manière rapide les paramètres non pertinents. D'autres méthodes réduisent la liste des paramètres en transformant la liste des paramètres en une liste plus réduite. En revanche, cette méthode rend l'interprétation des paramètres plus compliquée. D'autres méthodes se reposent d'abord sur le classement des paramètres selon leur niveau d'importance avant de sélectionner les paramètres les plus importants. Nous avons utilisé l'approche qui nous permet d'ordonner les paramètres selon leur importance donnée par le classificateur car nous souhaitons extraire les paramètres importants considérés durant la classification. Nous sélectionnons ensuite les paramètres qui permettent d'avoir le meilleur taux de classification. Les étapes de notre approche de sélection de paramètres sont représentées dans la Figure 14. Elles s'expliquent comme suit:

- D'abord, on se base sur la mesure d'importance retournée par la forêt aléatoire. On ordonne les paramètres par ordre décroissant du plus important au moins important.

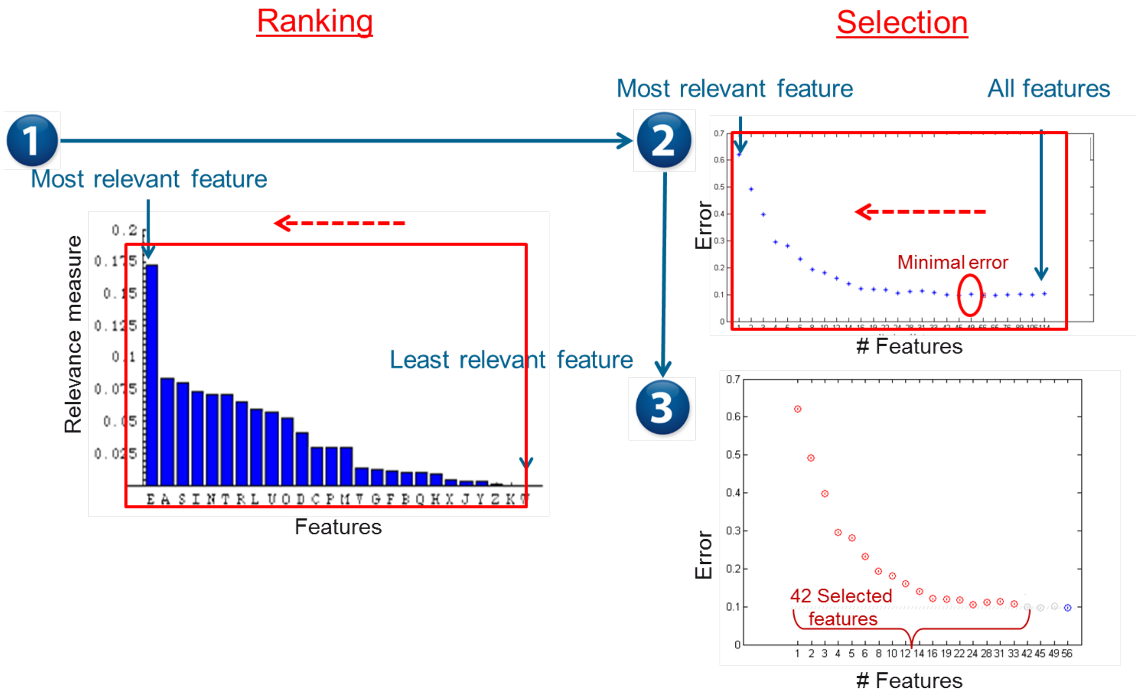


Figure 14: Séquencement de notre approche de sélection des paramètres pertinens.

- Ensuite, on élimine à chaque itération les paramètres les moins importants et on mesure l'erreur du modèle à chaque itération. La deuxième figure représente l'évolution de droite à gauche de l'erreur de classification. On remarque donc que l'erreur se stabilise à partir d'un certain nombre de paramètres.
- Enfin, on choisit les premiers paramètres qui permettent d'avoir une stabilité de l'erreur de classification; l'erreur ne décroit plus de manière significative si on ajoute des paramètres moins importants.

Nous avons appliqué cette approche d'abord sur les données de chaque action, ensuite nous avons étudié les paramètres communs sur toute les actions. Pour chaque action, cette approche a permis de sélectionner une quarantaine de paramètres parmi 114 paramètres initialement considérés. L'intersection des paramètres important sur toutes les actions a conduit à 11 paramètres. Ces paramètres concernent principalement la flexion de la tête, du torse, le penchement du torse vers l'avant et vers l'arrière, la distance entre les pieds, deux paramètres qui concernent l'accélération des mouvements des bras, et trois paramètres qui concernent la vitesse du mouvement des bras.

En utilisant cette approche, nous avons pu sélectionner un ensemble réduit de paramètres pertinents. Ces paramètres vont nous permettre de caractériser les expressions corporelles. Nous avons aussi étudié le rôle de chaque paramètre en se basant sur son classement d'importance fourni par l'approche des forets aléatoires.

Nous avons trouvé principalement 4 catégories de paramètres :

– Comme première catégorie, on trouve les paramètres qui sont considérés comme relativement importants par rapport à la nature de l'action. Par exemple le mouvement du bras droit sur l'axe vertical permet de bien classifier les émotions dans les actions de ramasser et jeter un objet, étant donné que ces actions sont principalement effectuées par le bras droit sur l'axe vertical.
– Comme deuxième catégorie, on trouve les paramètres qui reçoivent un classement important dans quelques émotions et actions. Par exemple la distance latérale entre les pieds a été considérée comme très importante pour classifier la honte dans l'action de jeter un objet. En effet, les jambes se caractérise par une posture assez fermée pendant l'expression de honte dans l'action de jeter un objet, bien que cette action implique principalement le mouvement du bras droit. De ce fait, on remarque qu'on peut trouver des paramètres qui sont classés comme pertinents pour la classification des émotions sans pour autant être intrinsèques à l'action. Nous concluons qu'il est important d'utiliser des paramètres décrivant le mouvement du corps entier afin de capturer l'expressivité corporelle.
– Comme troisième catégorie de paramètres, on trouve ceux qui sont lié à certaines émotions indépendamment de l'action comme le baissement de la tête et le penchement vers l'arrière pour la fierté, et aussi l'accélération du mouvement pour la colère.
– Enfin comme quatrième catégorie, on trouve les paramètres qui sont considérés comme importants pour la classification des 8 émotions dans les 7 actions étudiées

Étudier le classement des paramètres sélectionné nous a permis d'interpréter le rôle de chaque paramètres et par ailleurs de classifier les paramètres en 4 catégories selon leur pertinence par rapport aux émotions et aux actions.

Pour conclure l'étape de sélection de paramètres, nous avons trouvé autour de 40 paramètres pertinents parmi 114 pour chaque action, et 11 paramètres pertinents sur toutes les actions. Ces paramètres vont nous permettre de caractériser les expressions d'émotions.

Caractérisation des expressions corporelles

Dans cette section, nous utilisons les 11 paramètres sélectionnés sur toutes les actions pour examiner les patterns de caractérisation en se basant sur les données de capture de mouvement. Les patterns en couleurs présentent dans la Figure 15 représentent les caractérisations de chaque expression d'émotion sur toutes les actions. Ils sont obtenus avec la moyenne normalisée des 11 paramètres. Cette approche nous permet de visualiser rapidement et facilement la caractérisation de chaque émotion sur toutes les actions.

De manière globale, ces caractérisations peuvent être représentées sur les deux dimensions qui ont été largement utilisées pour représenter les émotions sur deux

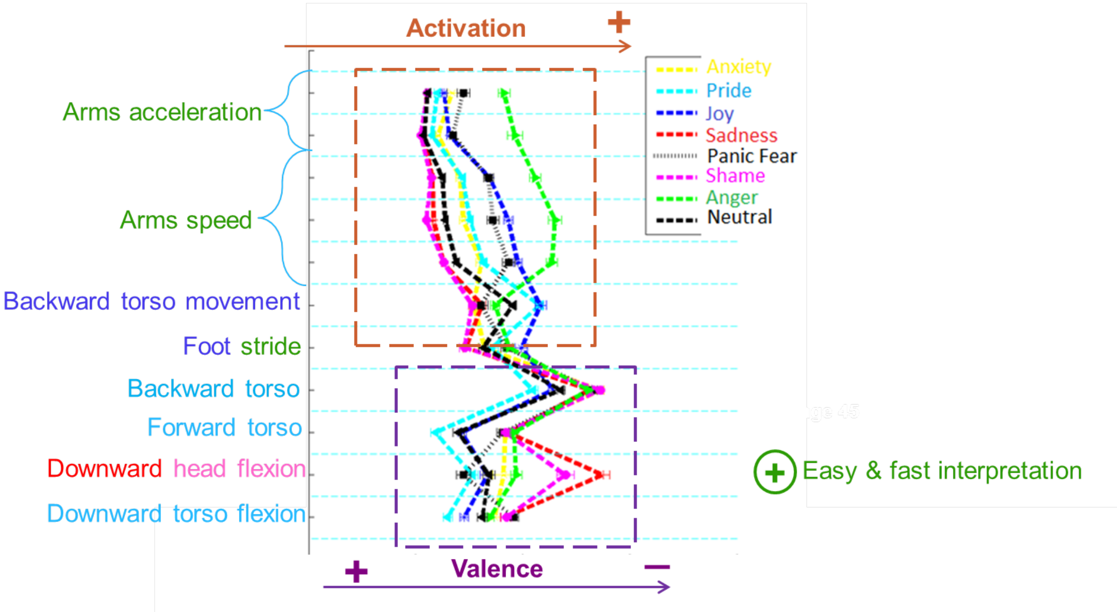


Figure 15: Caractérisation des expressions d'émotion avec les 11 paramètres sélectionnés comme pertinents sur toutes les actions.

axes; l'activation et la valence. L'activation qui est le niveau d'excitation et la valence qui est la polarité de l'émotion. Nous avons remarqué que les paramètres de vitesse et accélération reçoivent les plus grande valeurs avec les émotions qui sont défini avec un niveau d'activation élevé comme la colère, la joie, la peur panique, et les plus faibles valeurs pour les émotions ayant un niveau d'activation faible. De même pour la dimension de valence, nous remarque que la joie et la fierté, qui sont les seuls émotions positives dans notre base de données, se caractérisent par la flexion de torse et de tête les plus faibles, alors que les émotions négatives se caractérisent souvent avec une flexion et un penchement vers l'avant, particulièrement la honte et la tristesse.

Nous avons également utilisé cette approche de caractérisation pour étudier l'effet de l'action sur la caractérisation de chacune des émotions. Par exemple pour l'émotion de colère, nous avons trouvé que le pattern de caractérisation qui concerne la vitesse et accélération du mouvement de "Jeter un objet" est statistiquement différent du pattern de la colère sur les autres actions. Nous avons aussi trouvé que certains paramètres sont influencés par l'action de la même manière sur toutes les émotions. Par exemple la distance entre les pieds est fortement influencée par l'action de marche. De même la flexion du torse est fortement influencée par l'action de s'assoir.

Cette approche nous a permis de caractériser les émotions de manière simple avec les 11 paramètres. En deuxième lieu, nous avons utilisé les paramètres sélec-

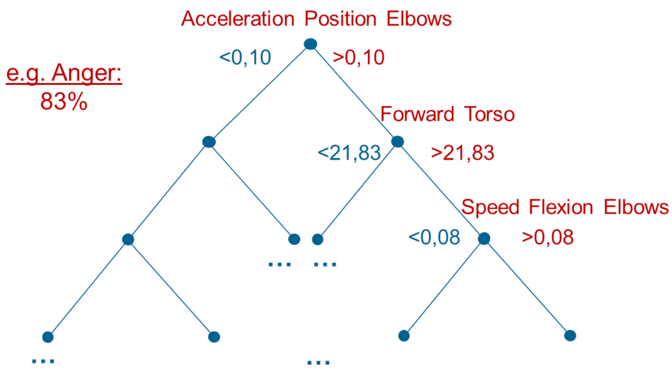


Figure 16: Exemple d'une règle de classification de l'expression de Colère.

tionnés sur chaque action pour générer des arbres de décisions optimisés. L'avantage principal des arbres de décision se manifeste par l'interprétation du modèle de classification. Les inconvénients major sont l'instabilité aux changements des données et l'ajustement taillé du modèle. Pour gérer ces problématiques, nous avons utilisé les paramètres déjà sélectionnés par les forêts aléatoires. De plus, nous avons effectué une étape d'élagage de l'arbre pour éliminer les branches les moins utiles. Nous donnons un exemple d'une règle extraite de l'arbre pour caractériser l'expression corporelle de la colère (voir Figure 16). La colère est classifiée avec une probabilité de 83% avec cet ensemble de règles : l'accélération des mouvements des coudes, le penchement vers l'avant du torse et la vitesse de flexion des coudes soit supérieure à certains seuils.

De manière globale, nous mentionnons quelques résultats déduits à partir des arbres de décisions construites avec nos données:

- Les paramètres sélectionnés sur l'entête de l'arbre sont les paramètres les mieux classé par les forets aléatoires.
- Les paramètres sélectionnés à la racine de l'arbre concernent généralement la vitesse et l'accélération permettant de distinguer les expressions des émotions les plus actives des émotions les moins actives. La rectitude du torse permet souvent de distinguer les émotions négatives des émotions positives .
- Les paramètres sélectionnés à la fin des branches sont particulièrement liés à des émotions spécifiques.
- Les émotions confuses au niveau de la perception humaine partage des règles communs.

Conclusion

Dans cette section nous résumons chacune des étapes de ce travail tout en rappelant les principales contributions. Nous discutons aussi les limites et les perspectives de ce travail.

- Pour la première étape, nous avons proposé un système d'annotation multi-niveaux des mouvements expressifs, à partir duquel nous avons proposé une liste de 114 paramètres corporels. Ce système d'annotation est utile pour caractériser l'expression corporelle dans différentes actions.
- Pour la deuxième étape, nous avons enregistré une nouvelle base de données (Emilya), contenant une large variété d'émotions et d'actions, contenant environs 8000 fichiers de capture de mouvement et plus que 8000 fichiers vidéos. Nous avons effectué une étude perceptif pour étudier la perception des émotions exprimée.
- Pour la troisième étape, nous avons calculé les 114 paramètres à partir des données de capture de mouvement et nous avons utilisé la méthode des forets aléatoires pour étudier la classification des expressions d'émotion sur différentes actions. Nous avons comparé les taux de classification avec la perception humaine des émotions.
- Pour la quatrième étape, nous avons déduit les paramètres importants considérés pendant la classification des émotions, et ceci pour chaque action et sur toutes les actions. Nous avons aussi interprété le rôle d'importance de ces paramètres et nous avons distingué 4 catégories de paramètres selon leur importance par rapport à l'action et à l'émotion.
- Enfin nous avons utilisé ces paramètres sélectionnés pour explorer la caractérisation de l'expression corporelle. Nous avons étudié l'effet de l'action sur la caractérisation et nous avons étudié la caractérisation sur toutes les actions.

Les contributions major de ce travail de thèse porte sur la collection et la validation d'une nouvelle base de données, et la proposition d'un système d'annotation multi-niveaux génétique pour différentes actions. Nous avons étudié l'effet de l'action sur l'expression et la perception des comportements émotionnels corporels et nous avons exploré les paramètres expressifs les plus importants.

Cependant, ce travail n'est pas privé de limites. Nous mentionnons quelques limites identifiés. D'abord, nous n'avons pas étudié la corrélation et la dépendance entre les paramètres sélectionnés. De plus, nous n'avons pas considéré d'autres niveaux de descriptions comme le niveau fonctionnel et spatial qui peuvent enrichir notre système d'annotation. Aussi, nous n'avons pas considéré les paramètres qui dérivent la trajectoire du mouvement. Enfin, bien que nous ayons utilisé une méthode d'induction pour enregistrer les expressions d'émotions, nous n'avons pas considérés l'étude des expressions exprimés de façon naturelle et spontanée.

Dans nos futurs travaux, il serait intéressant d'étudier la corrélation entre les paramètres expressifs identifiés comme pertinents pour la classification et la caractérisation des expressions corporelles. Il serait aussi intéressant d'étudier la reconnaissance de l'expression d'émotion multimodale à travers les données audio-visuelle enregistrés dans la base de données Emilya. Enfin, la synthèse des comportements émotionnels corporels à partir des règles de caractérisation nous sera utile pour vérifier si la caractérisation basée sur les paramètres sélectionnés pourra capturer l'expressivité corporelle. A long terme, nous pourrons enrichir notre système d'annotation du mouvement corporel pour mieux décrire l'aspect dynamique du mouvement (e.g. mesurer la fluidité du mouvement, décrire la trajectoire du mouvement).

Part I : Introduction

1 Introduction

"*Humans, whether in individual or interactive setting, consciously or not, communicate a rich mosaic of nonverbal messages with their bodies, faces and voices*" [J.A. Harrigan, 2005]. Nonverbal communication is achieved using wordless cues through our facial expressions, vocal prosody and body posture and movement. It has been shown in several psychological researches that all of these cues may infer our emotional states. Based on a wide range of human perception studies, it has been also proved that humans can identify, to some extent, emotional states expressed and displayed through facial and bodily expressions of other humans.

The recognition of emotions from facial and bodily expressions has been recently widely explored in affective computing field in order to build "intelligent", "affective" and "interactive" machines in a number of areas such as education, health care, games and entertainment. For instance, intelligent tutorial systems [D'Mello and Graesser, 2009] can react appropriately to adapt the style of learning when the learner is no more interested or motivated. Configurations of upper body posture while being seated (e.g. leaning forward/ backward) has been used as a relevant indicator of the learner's interest [Mota and Picard, 2003]. The interest and the motivation of the learner are strongly affected by the emotional states experienced by the learner such as failure, frustration, fear or anxiety [Kleinsmith and Bianchi-Berthouze, 2013], [Kapoor et al., 2007].

The recognition of facial and bodily expressions is also useful in games scenario to maintain the motivation of the players. For instance, the level of the game can be adapted according to the emotion experienced by the players and displayed through their facial and bodily expressions. The automatic recognition of players' emotions during video game scenario has been explored based on their expressive postures and movements [Savva et al., 2012], [Kleinsmith et al., 2011]. During video games and interactive systems, the credibility of the virtual characters is also required to maintain the motivation of the player/ user. In addition to the rendering of virtual characters, natural body movement and emotions experience can strongly affect the naturalness and the credibility of virtual characters [Ennis et al., 2013]. Another application that makes use of bodily expressiveness is interactive system that provides feedback during dance performance [Alaoui et al., 2012]. It has been shown that providing a visual feedback according to the expressiveness of dancers fosters the explorative and the expressive usage of the interactive system allowing

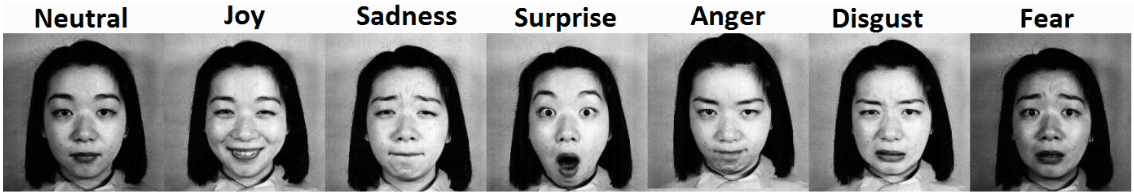


Figure 1.1: Facial expressions of Neutral and the six basic emotions from the Japanese Female Facial Expression (JAFFE) Database.

the dancers to better express themselves. In addition to education, entertainment and games scenarios, the recognition of patients' emotions - in particular from their bodily expression - has been useful in chronic pain rehabilitation to adapt the therapy according to their experienced emotions [Haugstad et al., 2006] [Kleinsmith and Bianchi-Berthouze, 2013].

1.1 From facial to bodily expression of emotion

It has been widely asserted that emotional states can be displayed through our facial expressions (See Figure 1.1). The face has been widely considered as the principal modality to communicate emotions by the means of facial expressions [Ekman and Friesen, 1978]. Ekman [Ekman and Friesen, 1969] proposed the evidence of the recognition of basic emotions from facial expression: Anger, Joy, Sadness, Fear, Disgust, Surprise (See Figure 1.1).

Bodily expression has been widely considered as providing a backdrop for a better recognition of facial expressions [J.A. Harrigan, 2005]. When facial and bodiy expressions of emotions are shown together in an image or a video, they mostly ensure better emotion recognition scores. Gelder [De Gelder, 2006] reported that "*an angry face is more menacing when accompanied by a fist, and a fearful face more worrisome when the person is in flight (that is, running away)*". Gunes and Piccardi [Gunes and Piccardi, 2006] observed that emotion perception leads to better results when subjects observe both facial expressions and gestures than when they observe only facial expressions or gestures. Tracy and Robins [Tracy and Robins, 2008] showed that Pride expression shown through facial expression and upper body posture was reliably recognized and distinguished from Happiness.

Indeed, early researchers have widely assumed that nonverbal affective messages conveyed through body posture and movement can only infer the intensity of emotion expression. However, recent researchers from different disciplines have shown that emotions can be expressed and recognized from solely bodily expressions. Several perceptual studies have been conducted to explore the human ability to recognize emotions from body posture and movement. They have shown that humans are able to identify, to some extent, emotional states from the perception of bodily expressions without including facial expressions [Kleinsmith and Bianchi-Berthouze, 2013]

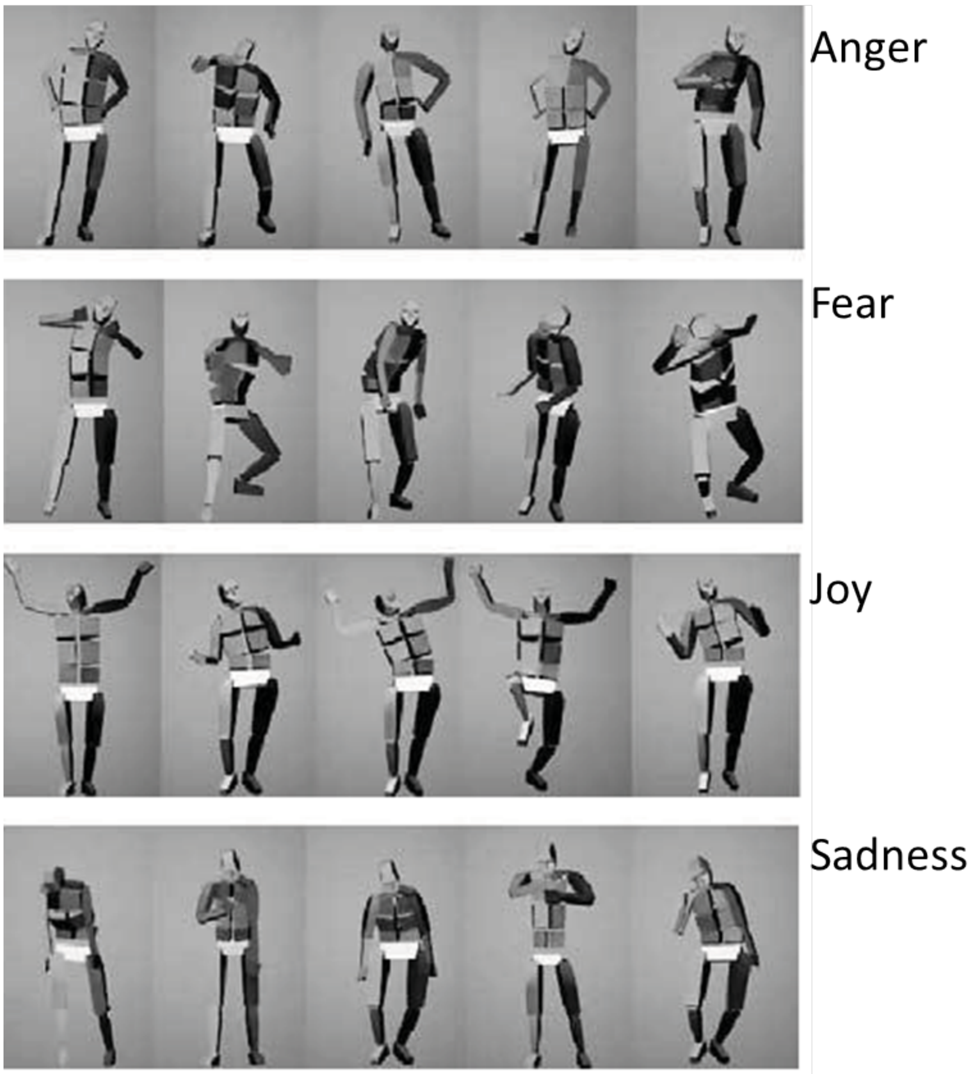


Figure 1.2: Examples of prototypical body posture. From top to bottom, they represent body expressions of Anger, Fear, Joy and Sadness (Screenshots from [Kleinsmith et al., 2009])

[Burgoon et al., 2010] [Wallbott, 1998] [Atkinson et al., 2004] [Dahl and Friberg, 2007] [Gross et al., 2010] [Roether et al., 2009].

1.2 Bodily expression of emotions

In psychological researches [Meijer, 1989] [Wallbott, 1998] [Burgoon et al., 2010] [Atkinson et al., 2004] [Montepare et al., 1999], emotion expression in body movement and body posture has been studied from either explicit or implicit bodily expressions.

- Explicit expression of emotion refers to prototypical expression. The expression of emotion is achieved through specific gestures [Dael et al., 2011] [Kleinsmith et al., 2009]. Figure 1.2 shows some snapshots of prototypical bodily expression of Anger, Fear, Joy and Sadness as studied in [Kleinsmith et al., 2009]. For instance, the expression of Joy can be achieved through raising arms (See Figure 1.2).
  Some emotions seem to be characterized with similar patterns across different studies [Wallbott, 1998], [Dael et al., 2011]. For instance, several researches highlighted that the expression of sadness is conveyed through arms relaxed along the body and downward head flexion [Wallbott, 1998], [Dael et al., 2011], [Coulson, 2004]). However, studies may report very different body postures and movements for the expression of other emotions. For instance, different prototypical expressions of Disgust have been described in previous works [Coulson, 2004] [Wallbott, 1998] [Atkinson et al., 2004].
- Implicit expression of emotion refers to the communication of emotions during a specific movement task (See Figures 1.4 and 1.5). Movement tasks can refer to daily actions (e.g. walking, throwing, knocking), music performance (e.g. piano performance) or choreographed dancing [Montepare et al., 1999] [Pollick et al., 2001] [Gross et al., 2010]. Indeed, emotions are communicated in body movements through the modulation of the movement task such as increasing the amount of arms swing and stride length during Anger expression in walking (See Figure 1.3). Figure 1.3 shows an example of neutral and angry walking as presented in [Hicheur et al., 2013].
  Several perceptual studies have been conducted to explore the characterization of emotions expressed during implicit expression (e.g. during knocking [Gross et al., 2010], walking [Montepare et al., 1999] and music performance [Dahl and Friberg, 2007]). It has been shown that similar characterization of some emotions have been found across different movement tasks. For instance, the expression of Sadness has been mostly characterized by slow movement (when expressed during music performance [Dahl and Friberg, 2007], knocking [Gross et al., 2010] and walking [Montepare et al., 1999]) and contracted posture (when expressed in knocking [Gross et al., 2010] and walking [Montepare et al., 1999]).

1.3 Automatic recognition of bodily expression of emotion

Lately, bodily expression of emotion has received a growing interest in computer sciences and affective computing researches. Several studies have been conducted to automatically recognize emotional states from bodily expression [Camurri et al., 2003] [Alaoui, 2012] [Kapoor et al., 2007] [Bernhardt and Robinson, 2007] [Kleinsmith et al., 2011]. For instance, Camurri et al. [Camurri et al., 2004] proposed a framework called EyesWeb devoted to the analysis of expressive dancing based on image and signal processing techniques. Their framework was also useful to recognize emotions experienced in dancing movement[Camurri et al., 2003]. Recently, Kleinsmith et al. [Kleinsmith et al., 2011] explored the recognition of players' affective states during computer game scenario. In another work, Kleinsmith et al. [Kleinsmith et al., 2009] highlight the need to endow humanoid robot with the ability to express emotions through body posture.

1

As reported in a recent survey of affective body expression perception and recognition [Kleinsmith and Bianchi-Berthouze, 2013], the reasons for the increasing interest on automatic recognition of bodily expression of emotion can be summarized as *scientific*, *social* and *technological* factors.

– Firstly, psychological works highlighted the importance of bodily expression in conveying emotions. In addition to its relevance in multimodal expression when combined with facial expression, bodily expression alone can also convey emotional states [Kleinsmith and Bianchi-Berthouze, 2013] [Burgoon et al., 2010] [Wallbott, 1998]. It has also been shown that some emotions (such as deception [Ekman and Friesen, 1969, Ekman and Friesen, 1974]) are easier conveyed through the body than through the face [Karg and Samadani, 2013] [Aviezer et al., 2012]. Considering bodily expression for the recognition of emotions is also useful when the expressed emotion is recognized at a distance [Karg and Samadani, 2013].
– Secondly, bodily expressions are highly useful to build affect-aware applications such as intelligent tutorial systems for education [Neill and Caswell, 1993] [D'Mello and Graesser, 2009] and reactive rehabilitation in clinical therapy [Haugstad et al., 2006].
– Thirdly, as technologies surround more and more our daily life, it has become essential to incorporate affect recognition within a human-computer interaction [Kleinsmith and Bianchi-Berthouze, 2013]. For instance, recent computer games such as Nintendo Wii (e.g. Just Dance game) are based on whole body movement. Considering the emotion of the player in such a game can increase the motivation of the player and facilitate the human computer interaction. Besides, early studies on emotion expression in body movements mostly relied on video recordings [Wallbott, 1998] [Montepare et al., 1987]. Thanks to the recent development of new motion-capture technologies, it becomes easier to obtain accurate information on 3D body posture and body movement allowing

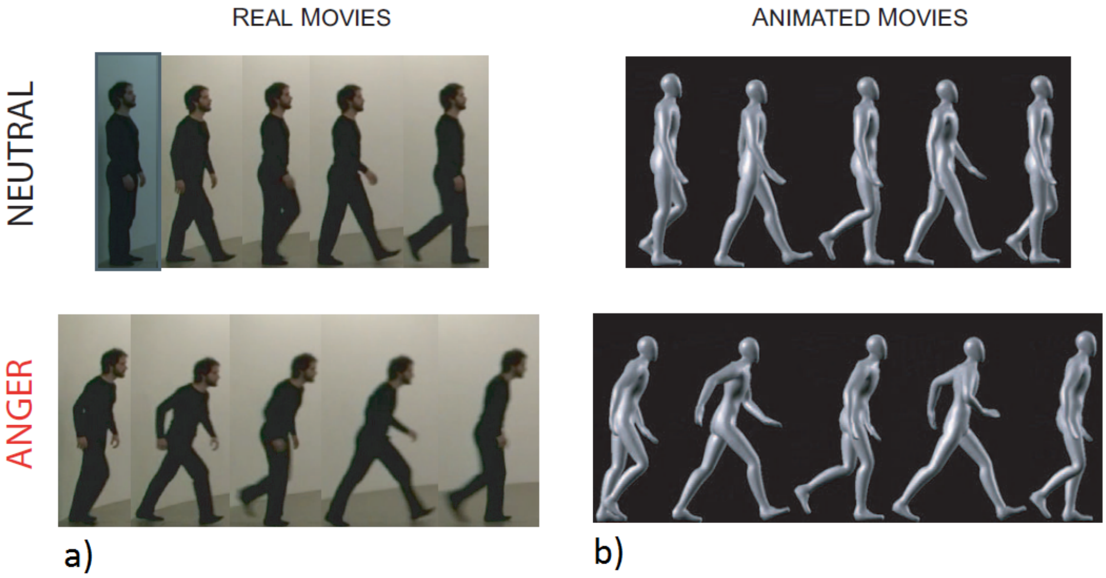


Figure 1.3: Snapshots of video-recording and motion-capture movies illustrating the neutral and angry gaits. (Screenshots from [Hicheur et al., 2013])

a better understanding of bodily expression of emotions.

1.4 Research Challenges

Despite the increasing need to develop computational models enabling the elaboration of affect-aware technologies, there is a clear lack of insights into how these models should be designed [Kleinsmith and Bianchi-Berthouze, 2013]. We present the challenges we encountered in designing such models.

Collecting bodily expression of emotions :
Most of the studies that aim to develop an affect-aware model rely on a database of emotion expression. One of the issues the most agreed upon is dealing with the recognition and the characterization of natural and spontaneous emotion expression. Due to the difficulty of gathering spontaneous expressive body movement, most of the previous studies on bodily expression of emotions have been based on acted database. The collection of emotion expression is mostly followed by a perceptual study to ensure the matching between the emotion being expressed by the actor and the emotion perceived by other subjects. In some acted database, exaggerated and stereotypical expressions lead to artificial behavior that do not represent naturalistic expressions encountered in daily situations. An affective database should offer a good compromise between the reliability of emotion expressions and the precision of movement data. The precision of motion capture data is useful to build accurate computational models in order to capture bodily expression of emotions from daily

situations. Recently, a new trend emerges toward the recognition of naturalistic and accurate expressive behaviors captured during computer game scenarios [Kleinsmith et al., 2011].

Perception of bodily expression of emotions :

The human recognition of emotions expressed in body movement is achieved through perceptual studies relying on participants judgment of bodily expression of emotions. These perceptual studies consist in the perception of stimuli (e.g. pictures or movies) depicting expressive body movement and the recognition of the expressed emotion.

One critical feature of emotion perception studies is the choice of stimuli. Indeed, it has been demonstrated that humans can recognize emotions expressed in the animation replicate where body movement is reproduced on a virtual computer avatar (See Figure 1.2, 1.4 1.5). However, few discussions have been related to the difference of the perception of emotions from real (cf Figure 1.3 a)) and from animated movies (cf Figure 1.3 b)). Moreover, the effect of the virtual avatar model on the perception of animated body movement is still considered as an open research question [Karg and Samadani, 2013].

Another critical feature of emotion perception studies is the labeling approach used to assess the perceived emotion. Different approaches can be adopted to assess the emotion being perceived by the subjects. The approach that has been the most used in previous perceptual studies consists in a forced-choice paradigm [Kleinsmith et al., 2011] [Atkinson et al., 2004], where the subjects are asked to select one or two emotion labels from the forced-choice list (e.g. *select the best representative emotion of an expressive behavior: 1) sadness, 2) joy or 3) anger*). However, it has been shown that offering an open-choice (e.g. *indicate, from your free choice, the best representative emotion of an expressive behavior*) resulted in a rich repository of perceived emotions [Winters, 2005] [Russell, 1994]. Another alternative of labeling is the use of multi-labeling approach relying on the rating of agreement or confidence in perceiving emotions and also other factors such as the level of empathy of participants[Kleinsmith and Bianchi-Berthouze, 2013]. Employing multi-labeling approaches favorites the use of several statistical techniques to deeply explore the perception of emotions.

From the perception to the modeling of bodily expression of emotions :

Psychological researches have shown the human ability to recognize emotions expressed in body posture and movement. Several perceptual studies have revealed that humans are also able to characterize bodily expression of emotions through gross body movement cues (e.g. the overall motion is slow/fast, the body posture is expanded/ contracted). However, reproducing this human ability in computer machine is still a challenging task.

Firstly, human body is characterized with a complex structure involving high number of degrees of freedom. Recent motion capture technologies have been pro-

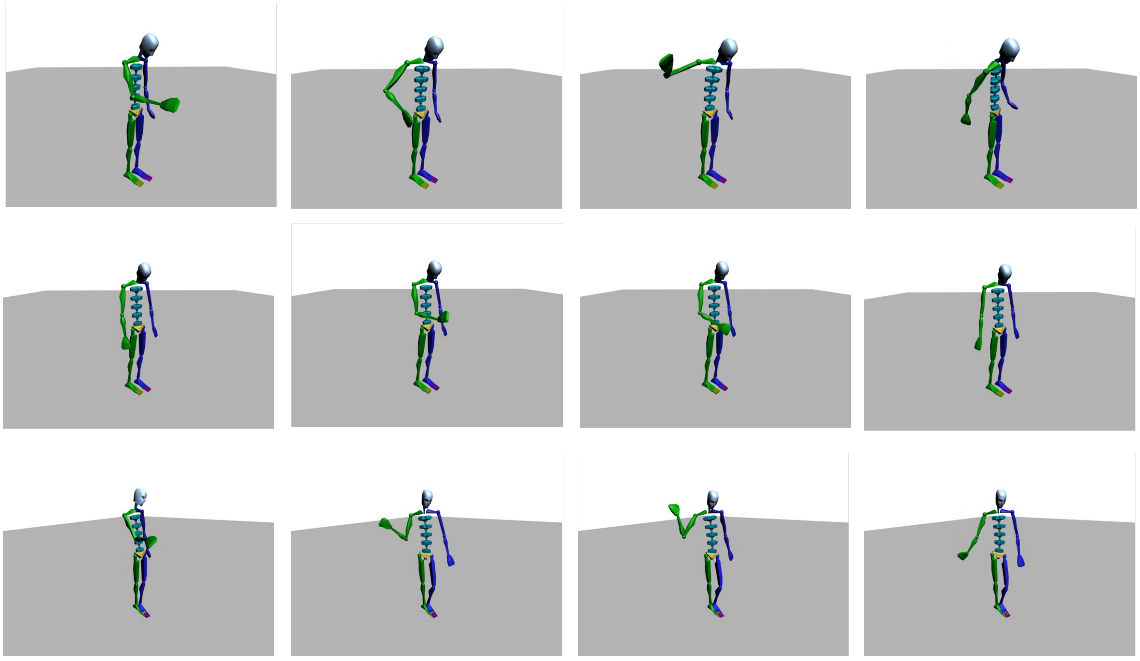


Figure 1.4: Snapshots of Anger (first row), Sadness (second tow) and Pride (third row) expression in Throwing action (Emilya database)

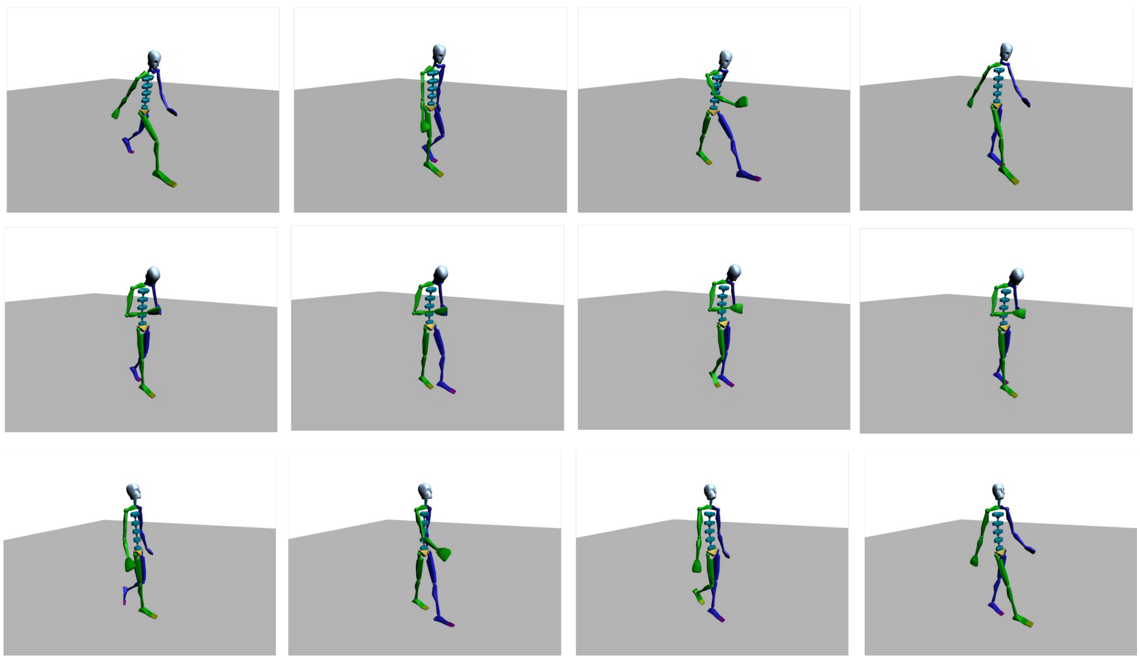


Figure 1.5: Snapshots of Anger (first row), Sadness (second tow) and Pride (third row) expression in Walking action (Emilya database)

posed to accurately acquire 3D body movement in terms of rotations and positions of body segments [Roetenberg et al., 2009]. Human body models used in such techniques are mostly based on a simplified model, ignoring fingers and toes and focusing on few vertebra of the spine (See Figure 1.2). It has been shown that such body models are sufficient to capture the emotional content of body movement. However, it remains unclear whether adding detailed movement of fingers, toes and a complex spine model can contribute to a better recognition and characterization of emotional body expression.

Secondly, a key challenge in building affective computational models is to offer a compromise between the complexity of the modeling of bodily expressions and the interpretability of their characterization. Indeed, modeling expressive body movement based on 3D positions and orientations of body segments achieves high scores of emotion recognition [Kapur et al., 2005]. However, the interpretation of emotion expression characterization is highly constrained by the movement task being studied. Indeed, the configuration of body posture and the motor behavior of movement is highly dependent on the performed movement task (e.g. walking, knocking). Such an approach offers a good solution when the aim is to explore the expression of emotion in a specific movement task. But it is less useful when the aim is to explore the expression of emotion in different movement tasks (e.g. how emotions modulate movements for walking, knocking, lifting...). Thus, there is a need to explore body cues that capture the emotional content regardless of the movement task. Such body cues may enable the generalization of emotion expression characterization across different movement tasks.

Generalization of implicit bodily expression of emotions :

Several perceptual and kinematic studies have been conducted to explore implicit bodily expression of emotion [Gross et al., 2010], [Pollick et al., 2001], [Dahl and Friberg, 2007] [Castellano et al., 2007] [Roether et al., 2009] [Hicheur et al., 2013]. Except few works [Dahl and Friberg, 2007], most of these previous studies focused on a specific movement. Similar patterns across studies have been found regarding the expression of some emotions (e.g. sadness as being expressed with slow motion and collapsed body posture across different actions). However, the effect of the performed movement task on the recognition and characterization of emotions expression remains unclear. There is a need to go beyond action-dependent emotion recognition and characterization to provide better insights into the expression of emotions in body movement.

The recognition and the characterization of bodily expression of emotions across different movement tasks mainly relies on 1) a database that contains a variety of emotions expressions in movement tasks and 2) body cues that allows the characterization and the interpretation of bodily expression of emotions across different movement tasks.

Other issues :

Several other issues have been raised in the development of affect recognition models from body posture and movement [Kleinsmith and Bianchi-Berthouze, 2013] [Karg and Samadani, 2013]. Due to the complexity of emotion quantification and modeling, previous studies mostly relied on a subset of discrete emotion label categories (e.g. sadness, anger) [Kleinsmith and Bianchi-Berthouze, 2013]. Only few studies explored the mapping of bodily expression of emotions to affective space based on dimensional theoretical emotion modeling [Patterson et al., 2001] [Kleinsmith and Berthouze, 2007]. Besides, studies relying on discrete models tend to focus on a reduced subset of emotion categories from the six basic emotions.

Another issue is that of inter-individual differences in emotion experience. There is evidence that bodily expression of emotion is affected by individual movement bias [Bernhardt and Robinson, 2007]. Indeed, affective databases have to rely on the expression of several actors to avoid the personal bias of emotion expression [Scherer et al., 1991]. Computational models have to deal with the problem of individual movement bias [Bernhardt and Robinson, 2007]. Moreover, the effect of gender on the recognition of bodily expression of emotions has been less explored [Karg and Samadani, 2013].

1.5 Our research focus

In this thesis, we focus on implicit bodily expression of emotions. We also aim to explore the effect of action on the recognition and the characterization of expressed emotions. We summarize our research focus in the following points:

1. Exploring the perception and characterization of emotional body expression across different daily actions.
2. Exploring the effect of daily actions on the perception and the characterization of emotional body expression.
3. Performing the automatic classification and characterization of emotional body expression across different daily actions.
4. Studying the effect of daily actions on the classification and the characterization of emotional body expression.
5. Selecting the most relevant body cues that capture the emotional content of expressive body movement across different action.

As discussed in the previous section, a number of challenges are encountered in achieving these goals. In our research work we tackle several of these challenges.

First of all, we need to rely on a database of bodily expression of emotions. As previous databases of implicit body expression mainly focused on a specific movement task and/or a limited range of emotions, there is a need to collect a new database to explore the expression of different emotions in different actions. In our work, we collect a new database of emotional body expression in different daily actions. Eleven actors participated in collecting emotion expression, allowing us to study the effect of inter-individual differences in emotion expression. This database contains the expressions of 8 emotions in 7 daily actions. Due to the complexity of emotion modeling, we focus on a set of discrete emotion categories. In addition to Sadness, Anger, Neutral, Joy and Fear which have been studies in previous databases, we also consider the expressions of Pride, Anxiety and Shame. To avoid exaggerated and stereotypical behaviors, we hired a professional acting director to give the actors 7 training sessions. His goal was to ensure the successful communication of emotion expressed through body movement. We explicitly asked the acting director to avoid exaggerated behaviors while working with the actors during the training sessions. Video as well as motion capture recordings are employed to collect expressive body movements. As such, we achieve a good compromise between the reliability of emotion expressions and the precision of movement data. Based on motion capture data, we are able to explore the classification of emotional body expression across daily actions and to study the effect of daily actions on this classification (see third and fourth research goals)

We also conduct a perceptual experiment to validate this database, that is to ensure the matching between the expression and the perception of emotional body expression through an “Emotion Perception” task. Two critical features are related to the perceptual experiments: the choice of stimuli and the labeling approach. Firstly, the stimuli that we create to conduct this experimental study consist in a replication of motion capture animation reproduced on a virtual puppet. Indeed, using a virtual puppet with no sign of gender, culture and facial expression allows reducing bias in perceptual studies. Besides, we propose an automatic approach to parametrize the display of the avatar movement in a 3D virtual environment in order to avoid any other bias related to the displacement of the avatar (e.g. occlusion of one body part by another, change of the distance at which we perceive the avatar in the 3D environment). Secondly, a multi-labeling approach is used to assess the perception of emotions. Unlike forced-choice framework, multi-labeling approach is useful to deeply explore the statistical differences in emotion perception ratings. We offer to the participants the possibility to assign another emotion label from their free choice. In addition to the “Emotion Perception” task, we also asked the participants to perform a ‘ ‘Body Cues Rating” task. Combining the results of these two tasks allows us exploring the perception and the perceptual characterization of emotional body expression across different daily actions and exploring the effect of actions on this perception and this characterization (see first and second research goals).

We also propose a body movement notation system based on several description levels allowing describing expressive movement across different actions. Relying on this body movement notation system and our new database of bodily expression of emotions, we are able to explore the recognition and the characterization of emotions across various actions.

Since our goal is not reduced to the classification of emotions expressed in body movement but it also includes their interpretation and characterization, there is a need to offer a good compromise between the reliability and the interpretability of emotion expression recognition. Indeed, we make use of Random Forest approach that provide both a high accuracy of classification and a measure of relevance for body cues enabling the interpretation of their predictive power and selecting the most relevant ones. Based on the most relevant body cues, we are able to study the motion capture characterization of emotional body expression in different actions and the effect of actions on this characterization (see third and fourth research goals).

1.6 Our Methodology

We present the methodology of the work conducted in this thesis.

– **First step: proposition of a comprehensive body movement notation system**. First of all, we defined a Multi-Level Body Movement Notation System (MLBNS) based on several description levels of body movement. This notation system is useful to characterize expressive movement in different daily actions. It leads to the definition of a large set of features measured from motion capture data. Based on different description levels, these features describe the whole body movement.
– **Second step: creation of a new database (Emilya database) of EMotional body expression In daiLY Actions**. Our second step consists in the creation of a large database of emotional body expression in different daily actions. This database contains a large variety of emotion expression and movement tasks: the expression of eight emotions (Neutral, Anger, Sadness, Joy, Panic Fear, Shame, Pride and Anxiety) in seven daily actions (Simple Walking, Walking with an object in hand, Knocking at the door, Moving Books on a table, Sitting Down, Lifting and Throwing an piece of papers). Thus, it includes Neutral expression, a subset of basic emotions (Anger, Joy, Sadness, Fear) and other emotional states (Pride, Anxiety and Shame).The movement tasks include whole body movement (walking and sitting down), repetitive arms movement (knocking and moving books) and non-repetitive arm movement (lifting and throwing). Composed of more than 8000 motion capture and video recording files, this database constitutes a rich repository of bodily expression of emotions.
– **Third step: automatic segmentation of walking and turning movement tasks**. Our third step refers to the analysis of walking and turning movement tasks based on the kinematic relationship between hips and shoulders motion. Indeed, we found that the shoulders and hips motion are in opposite phase during straight walking but linearly correlated during turning movement task. Thus, this step was useful for the framework of database post-processing step. Indeed, the segmentation of walking and turning movement tasks was based on the detection of the turn interval time that refers to the most linear correlation between hips and shoulders. A comparative study

was also performed to explore the effect of emotion expression and the effect of turning angle on hips and shoulders correlation during turning movement task.

– **Fourth step: proposition of an automatic approach for dynamic stimuli visualization for experimental studies of body language**. Our Fourth step consists in the proposition of an approach for the automatic generation of animated stimuli from a database of motion capture files. Thee stimuli are useful to conduct perceptual experiments. The position and the motion of the virtual camera are assessed automatically through a parametric approach during the creation of these stimuli. The position of the virtual camera has an impact on the visualization of the virtual puppet (e.g. frontal view). The motion of the virtual camera has an impact on the variation of the viewpoint and the distance at which we perceive the virtual agent in the 3D virtual environment.

– **Fifth step: validation of Emilya database**. Our fifth step consists in the development of a perceptual experiment to validate the expression of emotions recorded in Emilya database. The approach of stimuli creation just presented was used to automatically generate stimuli for our perceptual experiment. Through the same perceptual experiment, we asked participants to accomplish two tasks: 1) Emotion perception and 2) Body cues rating. The first and second tasks are respectively used to explore the human recognition of expressed emotions and the perceptual characterization of emotional body expression.

– **Sixth step: comparison of the classification of expressed emotions in different actions based on multi-level notation system using motion capture data**. Our sixth step concerns the classification of expressed emotions using motion capture data. We explore both 1) the classification of expressed emotions across all the actions and 2) the classification of expressed emotions in each action. We provide insights into the effect of the performed action on the classification of emotions. We also provide insights into the contribution of different description levels of body movement to the classification of emotions (e.g. upper body cues versus lower body cues).

– **Seventh step: identification of relevant expressive body cues across different actions**. Our seventh step concerns the identification of the most relevant body cues that capture the emotional content of expressive body movement. An embedded feature ranking method and a wrapper feature selection technique were combined to explore the subset of the most relevant features. This subset of features obtained from this feature selection approach was useful for the characterization and the interpretation of emotional body expression.

– **Eighth step: perceptual and kinematic characterization of emotional body expression across all the actions.** Our eighth step consists in the generalization of emotional body expression patterns across different daily actions. Kinematic characterization is based on the subset of relevant features obtained from our feature selection approach. The perceptual characterization is based on the body cues rating task achieved during the perceptual experiment. The mean rating based on perceptual and kinematic characterization of expressed emotions are discussed and compared.

– **Ninth step: perceptual and kinematic characterization of emotional body expression across "similar" actions.** Our ninth step consists in the exploration of perceptual and kinematic characterization of emotional body expression across "similar" actions. "Similar" actions refer to the actions that involve similar motor behaviors (e.g. lifting and throwing). This step is achieved through the modeling of emotional body expression using Decision Tree models. Decision Tree models are pruned to discuss the most informative decision rules that characterize emotional body expression across "similar" actions.
– **Tenth step: perceptual and kinematic characterization of emotional body expression for each action.** The last step refers to the exploration of the statistical effect of the performed action on the perceptual and the kinematic characterization of emotional body expression. This step allows us to explore the existence of "emotion associated actions" (i.e. the emotions that are particularly affected y a specific action). This step also allows us to disentangle which features are specifically characteristic of an emotion and which are of an action.

1.7 Our Contributions

The scientific contributions of this thesis can be summarized as follows:
– The proposition of a comprehensive body movement notation system for the description of emotional body expression in different daily actions.
– The collection of a new database of emotional body expression in daily actions and its validation through a perceptual study.
– The elaboration of a Multi-Level classification of emotional body expression based on our body movement notation system
– The identification of the most relevant body cues for the characterization of emotional body expression in daily actions.

1.8 Thesis structure

This thesis is organized into 6 parts: 1) Introduction, 2) Related work, 3) Characterization, collection and perception of expressive body movement, 4) Emilya motion capture data analysis, 5)Conclusion and perspectives and 6) Annex.

In the second part of this thesis, we discuss related works from different disciplines. Chapter 2 is devoted to the discussion of body movement notation system in general and to the discussion of expressive body cues used in previous works.

Chapter 5 is intended to discuss related studies on the perception and the characterization of bodily expression of emotions. This chapter refers to both perceptual (human perception) and kinematic (automatic analysis) studies. Chapter 3 is focused on the collection of bodily expression of emotions. It provides an overview of existing affective databases. Chapter 4 is focused on machine learning approaches for feature selection.

In the third part of this thesis, we present our work on the description, the collection and the perception of expressive body movement. Chapter 6 describes our multi-level body movement notation system. Chapter 7 presents the recording and the content of our Emilya database. Chapter 8 illustrates the methodology and the results of our perceptual experiment. This perceptual experiment includes the recognition and the characterization of emotional body expression recorded in the Emilya database.

In the fourth part of this thesis, we present our work on the analysis of motion captured emotional body expression. Chapter 9 describes the results of expressed emotion classification using motion capture data. Chapter 10 explains our feature selection approach intended to identify the most relevant body cues. This chapter also reports a discussion of features ranking according to their relevance measure. Chapter 11 is devoted to the discussion of emotional body expression characterization.

Last but not least, The fifth part (chapter 12) is dedicated to the conclusion of our work and the discussion of future research directions. Finally, the last part of this thesis provides supplementary and detailed material of our methodologies and results.

Part II : Related work

2 Body movements notation systems

We use our body to accomplish different tasks such as daily actions (e.g. sitting, walking, drinking), artistic movements (e.g. dancing), co-verbal gestures (e.g. gestures to illustrate the speech), but also to communicate social attitudes (e.g. dominance) and emotional states (e.g. sadness, anger). Body movement has been studied for several objectives such as trying to understand or to reproduce complex behaviors. These studies mostly rely on the description of body posture and body movement in terms of a set of body cues. Body movement notation system (also known as body movement coding system) are used to define such body cues.

In this chapter, we discuss the different approaches used in previous works to define body movement cues in general. We also discuss expressive body cues used in previous works to characterize expressive body movements.

This chapter is organized as follows: section 2.1 briefly presents the parallelism between body and vocal cues, which has been widely evoked in previous studies of body movement. Section 2.2 is intended to introduce the body movement description levels that have been commonly used in previous body movement notation systems. In section 2.3, we present examples of body movement notation system that have been proposed in the literature. Section 2.4 is devoted to the expressive body cues used to explore the characterization of expressive body movements. Finally, section 2.5 and 2.6 respectively conclude and summarize this chapter.

2.1 Parallelism between expressive vocal and bodily features

It has been asserted in previous works that body movement can be somehow assimilated to speech [Gross et al., 2010, Tilmanne and Dutoit, 2011]. A number of characteristics can be found in both speech and body movement. For instance, complex motion can be split and composed of isolable elements or motion primitives [Birdwhistell, 1970] [Bernhardt and Robinson, 2007], in the same way as speech can be described in term of words or phonemes [Tilmanne, 2013].

Besides, vocal characteristics changes affect speech as bodily characteristics changes affect body movement. For instance, intrinsic characteristics (e.g. age, gender) and emotional state can influence body movement as they can also influence speech [Tilmanne, 2013]. An interesting concept underlying this assumption is the composition of motion into a “verb” (the task component: the basic motion such as walking) and an “adverb” (the style component: the way the motion is performed) [Rose et al., 1998]. The “verb” corresponds to the linguistic component of the speech, while the “adverb” corresponds to the expressive component of the speech. Similarly, a motion (e.g. walking) can change according to the emotion being expressed (e.g walking angrily). Several studies have shown that humans are able to differentiate emotions expressed in body movement according to the manner by which movements change (e.g. sad walking versus angry walking) [Gross et al., 2010, Hicheur et al., 2013]. However, it is not clear whether emotional states modulate body movements in the same way across different actions (such as knocking, lifting, walking ...). In this thesis, we aim to provide deeper insights into the effect of emotion expression on body movement across different daily actions. To achieve this goal, there is a need to rely on a body movement notation system that can be used to characterize emotional body expression in different daily actions. In the next section, we introduce the description levels that have been commonly used in previous works to characterize body movements. Shortly later in Section 2.3, we discuss body movement notation systems proposed in previous works based on these description levels and we focus on expressive body movement characterization in Section 2.4.

2.2 Description levels of body movement notation system

Body movement notation systems rely on a set of description levels which serve to give better insights on how body features can be defined. In this section, we provide a summary of the description levels that have been commonly used in previous body movement notation systems. In section 2.3, we present examples of body movement notation systems relying on such movement description levels.

2.2.1 Functional level

Functional analysis of gestures are based on the functional level of their description. The categorization of gestures along the functional level is based on underlying states revealed by nonverbal behavior [J.A. Harrigan, 2005] and it is mostly applied in the context of interaction. The functional level allows the classification of gestures according to their meaning. Ekman and Freisen [Ekman and Friesen, 1969] proposed a notation system that categorizes the types of messages conveyed in nonverbal behaviors into five classes of gestures; illustrators, regulators, affects displays, adaptors (or manipulators) and emblems. Three main categories of gestures can be distinguished to group the five classes of gestures proposed by Ekman and Freisen [Ekman and Friesen, 1969]; gestures as support to speech, gestures independent of speech and gestures as an affective means of communication.

– **Gestures as support to speech**:

  - **Illustrators**: Illustrators are defined as "*movement directly tied to speech, serving to illustrate what is being said verbally*" ([Ekman and Friesen, 1969], p. 68). These conversational actions accompanying speech does not have a sense when viewed independently from speech. They just illustrate and accentuate the verbal contents. It was observed in [Ekman and Friesen, 1972] that the quantity of the illustrators increases when the person (who is talking) is excited and it decreases when the person is tired or depressed.
  - **Regulators**: Regulators are defined as "*the actions which maintain and regulate the back-and-forth*" flow of conversation between speakers and listeners ([Ekman and Friesen, 1969], p. 82). They represent nonverbal conversational mediators, such as head nods, eye contact, smiles or utterances such as "mm-hmm".
- **Gestures independent of speech** :
  - **Emblems**: Emblems are symbolic gestures which are dependent of the culture and whose meaning can be understood without the presence of speech. An emblem is a "*specific verbal translation known to most members of a subculture, and is typically intended to send message*" ([Ekman and Friesen, 1977], p. 38). It can have a verbal translation into one word, two words or a complete sentence. Thumbs up (good) and thumbs down (bad) are two examples of emblems known in American culture.
- **Gestures as an affective means of communication**
  - **Affects displays**: They convey messages about affective states. They refer primarily to facial expression of affect, but they can also refer to bodily expression of affect. It has been shown that the facial expressions of the basic emotions are universally recognized [Ekman, 1989].
  - **Manipulators (or adaptors)**: Originally labeled "adaptors", manipulators refer to the actions in which one part of the body does something to another body part (like lip licking, scratching one's hand, hand-to-hand movement) or to an object (like playing with a pencil). Manipulators that refer to the manipulation of body part are called "self-adaptors", "self-manipulator" or "body manipulator". Although self-adaptors are displayed mostly without the intention to communicate a particular message, they tend to convey affective information to observers concerning the emotional state, the pathology, the deceptiveness and also general personality traits of encoder in a nonverbal communication [J.A. Harrigan, 2005]. Ekman and Freisen ([Ekman and Friesen, 1972], p. 362) report that "*self-adaptors occur more frequently when the person is in private rather than a public place, when alone rather than in the presence of others, when not in any way involved with others rather than with others, when listening rather than speaking in conversation*". Ekman and Freisen [Ekman and Friesen, 1972] considered that self-adaptors are mostly performed unconsciously and they are usually associated with negative emotions (such as anxiety). Wallbott [Wallbott, 1998] observed that self-adaptors are more frequent during the expression of

shame and fear rather than the other studied emotions. Nevertheless, Dael et al. [Dael et al., 2011] showed in their study that self-adaptors can occur also when expressing positive emotion like pleasure and relief.

These communicative gestures were used and detailed in other studies. McNeill's approach [McNeill, 2005] characterized only communicative arm gestures accompanying speech into five categories; iconics, metaphorics, deictics, beats and emblems. These gesture models were used later in computational approaches [Cassell, 2000]. Some gestures can be affected to the emblematic category of gesture or a support of speech (like head nod; to say ok, or to follow the speech of the other person).

The computational measurement of the description of body movement presented above according to the functional level is not an easy task. Hager and Ekman [Hager et al., 1995] report that "*The variety of configurations, orientations, and locations that equivalent behaviors at this level might assume would seem to make them a difficult candidate for computational measurement*".

2

2.2.2 Anatomical level

Anatomical description level describes the structure of the body. The human body can be duplicated into several major segments including the head, upper limbs, lower limbs and the torso [Levy and Duke, 2003]. The body structure can also be viewed in a global manner as a whole entity (e.g. "Whole body moves or leans towards a forward position" [Dael et al., 2012], body shape description based on the bounding box surrounding the whole body [Camurri et al., 2004]). Dael et al. [Dael et al., 2012] also provide a detailed description of the body structure considering several segments such as the head, the trunk, right and left shoulders, right and left elbows, right and left wrists,...).

2.2.3 Directional level

The directional level defines the spatial dimensions in which movement is possible. The human anatomy system is based on three planes (and three axes perpendicular to the corresponding planes) defined with respect to the human standard position; sagittal plane, transverse plane and coronal (lateral) plane (See Figure 2.1). The movement change in the sagittal plane means a change of the orientation in the forward direction or the backward direction. The change in movement shape in the coronal plane can be defined as a lean in the side direction to the left or to the right. Finally the movement change on the transverse plane means a rotation around the vertical axis to the left or to the right.

2.2.4 Posture/Movement level

Several previous works differentiate between two fundamental aspects of body moves; body posture and body action/movement [Dael et al., 2012]. The features

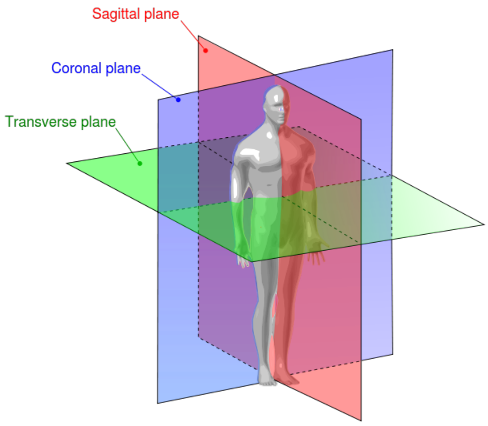


Figure 2.1: The description of the directional level with Human anatomy planes.

used to describe the posture or the movement of the body can be widely different for each notation system depending on the context in which body moves occur (e.g. during interaction [Busso et al., 2008], during acted gesturing for the expression of emotion [Dael et al., 2011], during a daily movement such as walking or knocking[Bernhardt and Robinson, 2007]). For instance, among others, body posture can be described as a particular configuration of each joint angle [Coulson, 2004], as a snapshot of a sequence of movement [Kleinsmith et al., 2011] or as the average of rotation/ position values across the whole sequence of movement [Bernhardt and Robinson, 2007]. Similarly, among others, body movement can be described based on action units (e.g. head node) [Dael et al., 2012] or on movement quality features (e.g. speed) [Dahl and Friberg, 2007].

Dael et al. [Dael et al., 2011] [Dael et al., 2012] focused on the communication of emotions through body gestures. In their body movement notation system called BAP (the Body Action and Posture notation system), they introduced Action units and Posture units. They defined Action (e.g. Forward–backward head shake) and Posture (e.g. Left arm at side) units according to the anatomical and directional description levels. Laban Movement Analysis (LMA) [Laban, 1988] includes the description of the shape (posture) and the quality of expressive movements dynamics in dance performance.

2.2.5 Space level

The notion of space has been firstly introduced by Laban [Laban, 1988]. Space can refer to two categories; personal space known as Kinesphere (See Figure 2.2) and the environment which is the space that surround the Kinesphere (that is the environment space). Kinesphere is defined as "*the sphere around the body whose periphery can be reached by easily extended limbs*" [von Laban and Ullmann, 1966]. The space component that refers to the environment involves the location of the body

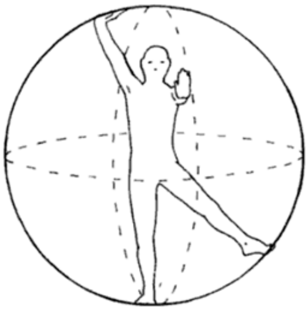


Figure 2.2: The kinesphere [Bartenieff and Lewis, 1980].

in the space and the trajectory of body movement in the space. The relationship between the body and the environment has been mostly considered for dancing movement analyses [Laban, 1988] [Camurri et al., 2004].

2.3 Body movement notation systems

In this section, we discuss examples of body movement notation systems based on the description levels presented in section 2.2. Body movement notation systems are defined in [Karg and Samadani, 2013] as "*an efficient tool for systematic and compact representation of movements that capture both structural characteristics and expressive qualities of the movements*".

According to Burgoon et al. [Burgoon et al., 2010], body movement notation systems can be categorized into *functional* and *structural* approaches. *Functional* notation systems are based on the functional description level discussed in section 2.2. They refer to the categorization of body movement and gestures using verbal labels such as emblems and illustrators (See the work of Ekman and Friesen [Ekman and Friesen, 1969]). *Structural* notation systems describe structural details of posture and movement dynamics to describe "what bodily movements look like" [Karg and Samadani, 2013]. In this section, we focus on *structural* body movement notation systems as they are more appropriate for describing body shape of body movement.

SPAFF: The Specific Affect Coding System :

Advancing researches have been conducted for developing coding system on facial behavior. The Facial Action Coding System (FACS) developed by Ekman and Freisen in 1978 [Ekman and Friesen, 1978] is a common standard to code the movements of individual facial muscles. It was mostly used to categorize the facial expression of emotions and it is also a powerful tool used in psychology and affective computing. However, there is a lack of consensus on a common notation schema for the description of emotional body expression. SPAFF (The Specific Affect Coding

System) [Coan and Gottman, 2007] is a notation system that includes both facial and bodily cues to describe affecive expressions. This notation system is heavily influenced by FACS, but not all FACS codes are in included in SPAFF.

SPAFF relies on verbal content, facial behaviors, voice tones, and other forms of communication. It describes both the code of affects from "Speakers" (those who are observed using the code) and the "Receivers" (those the speakers are speaking to). SPAFF includes the code of Neutral expression, 5 positive affects expressions (Affection, Enthusiasm, Humor, Interest and Validation) and 12 negative affects expressions (Anger, Belligerence, Contempt, Criticism,, Defensiveness, Disgust, Domineering, Fear / Tension, Sadness, Stonewalling, Threats and Whining) [Coan and Gottman, 2007]. Each expression is described through 4 coding cues: Function, Indicators, Physical Cues, and Counterindicators. We give an example of using these cues for the code of Sadness expression as provided in [Coan and Gottman, 2007]:

- *Function*: interpersonal communication.
  Example for Sadness: "Sadness code refers to behaviors that communicate loss, resignation, helplessness, pessimism, hopelessness, or a plaintive or poignant quiescence " [Coan and Gottman, 2007].
- Indicators: provide information about behaviors that probably occur during the expression of an affect.
  Example for Sadness: Sighing, Pouting/Sulking, Resignation, Crying, and Hurt feeling.
- Physical Cues: provide information about physical reactions that probably derive from the presence of an expression.
  Example for Sadness: "AUs 1, 6, 15, 17, 1+6, 1+15, 1+6+15, 1+6+15+17. Shoulders may droop, and individuals may hang their heads or look down. The lips and the chin may tremble. The voice may quaver in terms of pitch and amplitude and may occasionally break" [Coan and Gottman, 2007].
- Counterindicators: behaviors that may not derive from the expression.
  Example for Sadness: No back channel, Relief and Happy tears.

Birdwhistell coding system :

Birdwhistell [Birdwhistell, 1970] developed a notation system of movement behavior based on the anatomical level with respect to the three spatial dimensions. Since he was against the distinction between verbal and nonverbal behavior, he created a coding system that describes the structure of movement based on linguistic principles. In his work, he compares the 'kineme' which describes a body movement, to a phoneme, which is defined as a unit of speech. A kineme can be for example a lateral head sweep or head tilt [Birdwhistell, 1970]. The structural notation system proposed by Birdwhistell includes "motion qualifiers" used to define the degree of muscular tension during movement performance, the duraction and the range of the movement.

Bernese coding system :

The Bernese system is based on body action and posture units to describe body movement of a sitting individual [Frey and Von Cranach, 1973]. It provides a detailed description of different body parts: head, shoulders, trunk, arms, hands, legs and feet. The three Cartesian axes (sagittal, vertical and transverse) are used to code body parts posture and movement. However, the Bernese system is limited to the code of nonverbal behavior of sitting individuals.

The Bernese system can be used to describe the spontaneous movements that occur during interaction. It was used in the field of computer animation to imitate the body movement of two persons during interaction [Bente et al., 2001]. In their perceptual study, it has been showed that the observer's impressions of original and animated characters in the sequence of interaction were similar. Their results prove that the Bernese system is detailed enough to describe body movement when a person is seated an engaged in an interaction with another partner.

Body posture coding systems :

Other works were limited to the study of static posture. In the work on interpersonal attitudes, Mehrabian [Mehrabian, 1972] focused on body posture and orientation to create a more simplified coding system. The orientation of body was coded according to the orientation of head and body toward or away from the interlocutor. The closeness/openness and the symmetry/asymmetry of the different body limbs were also used to code the body posture. Tracy and Robinsons ([Tracy and Robins, 2007] proposed a nonverbal coding system describing the posture of the upper body parts to study the bodily expression of pride. In the work of Coulson [Coulson, 2004], several simplifying criteria were applied to the model of body in order to describe an expressive body posture. The description of postural expression was concentrated on the upper body parts with symmetric criteria on the upper limbs postures. The six joint rotation taking into account in this study were head bent, chest bent, abdomen twist, shoulder abduct, shoulder swing, and elbow bend. Coulson [Coulson, 2004] introduced a new parameter called the weight transfer parameter which defines the position of the mass center. This parameter was coded as forwards, backwards or neutral to simplify the description of the lower body parts.

BAP: Body Action and Posture coding system :

Dael et al. [Dael et al., 2012] have recently proposed a new body action and posture coding system called BAP. Their system is based on the GEMEP corpus (Geneva Multimodal Emotion Portrayals) describing a set of emotion expressions. The corpus contains 17 emotions expressed by 10 actors. Dael et al. [Dael et al., 2012] paid attention on the distinction between postures and actions. Postures are basic units defined as the alignment of one or a set of articulations [Dael et al., 2012]. Action units are local excursion composed of three steps (starting point, duration, and end point) [Dael et al., 2012]. The BAP coding system provides a detailed description of body postures and actions considering the sagittal, vertical and lateral dimensions. With regards to the anatomical level, the code of body postures was defined both in a local manner (considering one limb such as left arm, head, trunk...)

and in a global manner (the posture of the whole body). Each body action unit was coded with regards to a particular body segment. The BAP system includes also a functional classification of hand gestures based on the work of Ekman and Freisen [Ekman and Friesen, 1972] and McNeill [McNeill, 2005]. The upper body parts were more detailed than the lower body parts, but the fingers positions and actions were not taking into account in the BAP system.

Laban notation system :

The Laban notation system is one of the well known notation systems based on the quality of movement [Laban, 1988]. It appeared initially to describe the movements of dancers, but it has been argued that its use has important implications in other fields such as the study of bodily expression of emotion [Gross et al., 2010] or the study of personality [Levy and Duke, 2003]. Laban notation system focuses on the way in which body movement are performed regarding the quality of movement dynamic as well as the form and the shape of the body. Four major concepts were defined to describe movement in the basis of this system; Body, Shape, Effort and Space.

- **Body**: Composed of several modalities, the Body component has an important role to describe the movements relying on the anatomical description level. The Body concept defines the body parts involved in the movements as well as their connection, influence and sequencing. It provides the information about the body organization and the movement order.
- **Shape**: Shape component refers to the way the body changes shape, moves and 'sculpts' itself in space during movement. It is based on both anatomical and directional description levels. Shape component describes not only the changes in the relationship of the body parts to one another, but also the changes in the relationship of the body parts to the environment that surrounds the body [Levy and Duke, 2003]. Laban described Shape change through three distinct qualitative aspects; **Shape Flow**, **Shaping movement** and **Directional Movement**. **Shape Flow** refers to the form of the body itself through the spatial relationship among body parts. Generally speaking, Shape flow movement is described as growing or shrinking [Dell, 1977] based on its relationship to the body center or core. Growing refers to spreading movements that are directed outwards from body center. Shrinking refers to enclosing movements that are directed inward the body center. **Shaping Movement** was defined first by Laban in a simple form as gathering or scattering with regards to the space, which is very similar to the definition of Shape Flow movement. Based on spatial dimensions defined in Space component (horizontal, vertical and sagittal) of Laban system, Shaping Movements were more detailed and categorized. According to these dimensions of spatial orientation, there are specific terms used to define the six shaping possibilities; spreading, enclosing, rising, sinking, advancing and retreating. **Directional Movement** refers to the actions that are motivated by a particular point or

object in the environment. In other words, the directional aspect defines the bridge between the body and a particular object in the environment when the body is directed towards it. The Directional Movements can either be spoke-like (straight) or arc-like (curving) [Levy and Duke, 2003].

– **Effort**: The third component of movement, Effort, describes how the body performs the movement with regards to the dynamic aspect of movement. Laban defined Effort as “a person’s attitude towards movement or the dynamic, qualitative variables of movement” [Levy and Duke, 2003]. Effort is composed of four subcategories; **Space**, **Weight**, **Time** and **Flow**. Each is defined and evaluated in a range between two opposite extremes: ’indulging’ in the quality and fighting against the quality [Chi et al., 2000]. **Space** refers to attitude towards a chosen pathway (direct/indirect) to study the level of attention paid by the mover. **Weight** describes how strong the movement is (Strong/Light) or the amount of force involved on it. Strong weight indicates the increasing of pressure into movement or the ability to overcome the pull of gravity. **Time** defines the speed of movement by the lack or the presence of urgency in the movement (Sustained/Sudden). Finally **Flow** refers to the attitude of the mover toward controlling or releasing his movement (Bound/Free).

– **Space**: The Space component focuses on the connection between body movement and the environment with regards to the location of the body, the paths, the spatial patterns and directions of movement in space.

The Laban Movement Analysis (LMA) [Laban, 1988], considers the body as a three-dimensional structural composed of length, width and depth. These three dimensions are used to describe movement changes; horizontal dimension (width), vertical dimension (length) and sagittal dimension (depth) [Chi et al., 2000]. Laban [Laban, 1988] considered that all body movements executed without changing stance (movement that occurs only by extending limbs) can occur in a person’s own three-dimensional spherical space called ‘ kinesphere” (See Figure 2.2).

The shape change in the horizontal and sagittal in LMA correspond respectively to the shape change in the coronal plane and sagittal plane of the human anatomy system. However, the shape movement change in the vertical dimension in LMA is different from the shape change in the transverse plane of the human anatomy system; the change in the vertical dimension in the LMA is manifested in the Upward/Downward movements while the change in the vertical dimension in the human anatomy system is manifested in Left/Right rotation.

Based on Laban Movement Analysis (LMA) [Laban, 1988] (Shape component), the shape of body movement can be described in a general way into Three dimensional direction (e.g. the spreading movement) or using the three-dimensional structural organization of movement direction (e.g. horizontal, vertical..).

2.4 Expressive body cues for the study of bodily expression of emotions

In this section, we focus on expressive body cues used in previous studies on Affective Computing and psychology to study the expression of emotions in body movement. As proposed in [Karg and Samadani, 2013], three approaches are commonly used to define such body cues: 1) Expressive body cues based on notation systems, 2) Constraint-based expressive body cues (named hand-selected in [Karg and Samadani, 2013]), and 3) Expressive body cues based on perceptual studies in psychology.

2.4.1 Expressive body cues based on notation systems

Only few works rely on body movement notation systems to define expressive body cues. To the best of our knowledge, only one study [Dael et al., 2011] made use of the recently proposed BAP notation system for the study of emotion expression in body movement [Dael et al., 2012]. Applying such a system requires a manual annotation of videos which is a highly time-consuming task. Recently, Velloso et al. [Velloso et al., 2013] proposed a new framework called AutoBAP to enable the automatic annotation of body action and posture units as provided in BAP notation system.

The Laban notation system has received particular interest in the studies that aim to characterize expressive body movement. Laban Movement Analysis (LMA) has been adopted for automatic recognition [Gross et al., 2010] and generation [Chi et al., 2000] of affects expressed in body movement. Effort and Shape components factors were particularly widely used to characterize bodily expression of emotion. The analysis of body movement based on these two components is known as Effort-Shape analysis, derived from Laban Movement Analysis. In their perceptual study, Gross et al. [Gross et al., 2010] showed that Effort and Shape Flow can provide important information in the study of emotion expression in body movement. In the field of computer animation, Chi et al. [Chi et al., 2000] proposed a computational model based on the Effort-Shape analysis called the EMOTE model.

Camurri et al [Camurri et al., 2004] proposed a collection of computational models called EyesWeb expressive gesture processing library for real-time expressive body movement analysis. Based on LMA notation system, the EyesWeb library was designed to describe expressive dance considering that the dance task is a main artistic expression of human movement [Camurri et al., 2004]. The EyesWeb library includes three main sub-libraries: The motion analysis library, the space analysis library and the trajectory analysis library. Castellano et al. [Castellano et al., 2008] proposed an extension of the EyesWeb expressive gesture processing library that focuses on the dynamic of movement expressivity.

The EyesWeb expressive gesture processing library has been proposed to analyze expressive body movement. As it is mainly based on video analysis, using EyesWeb

2

framework for expressive movement analysis is useful for real-time interaction. Several movement dynamics features are proposed within the EyesWeb framework to describe the whole body movement in the environment. However, less postural features are proposed which prevent a fine-grained description of body segments posture.

2.4.2 Constraint-based expressive body cues

Constraint-based expressive body cues are defined according to some constraints such as the nature of motion capture data, the kinematic properties of human motion or both. Constraint-based expressive body cues do not always rely on psychological findings.

Indeed, the definition of the body cues can be constrained by the type of sensors used to capture motion data such as pressure sensors [Mota and Picard, 2003] and Wii remote [Amelynck et al., 2012]. Other studies used statistical measures based on the joint angles or positions obtained from a motion capture system. Kapur et al. [Kapur et al., 2005] calculated 2 statistical measures (mean and standard deviation) based on the Cartesian coordinates of 14 markers' positions in addition to their first and second derivatives. These markers were placed on strategic locations to capture the whole body movement during emotional body expression.

The definition of body cues can be also constrained by the movement tasks. Multon [Multon, 2013] reviewed the main methods used in previous works to characterize walking motion. He focused on four categories of walking motion features: global features (e.g. step length), kinematic features (e.g. joint angles), dynamic features (e.g. ground reaction force) and muscle activity features. Roether et al. [Roether et al., 2009] describe body posture during expressive walking in terms of averaged flexion angles of 11 joints: head, spine, pelvis, left and right shoulder, elbow hip and knee. They also consider the speed of walking in their analyses of motion-captured expressive walking. In [Karg et al., 2010], Karg et al. used a different set of kinematic features to analyze expressive walking recorded using motion capture data. The kinematic features used in their study involve three features describing the stride (velocity, length and duration) and other describing the minimum, mean and maximum of specific joint angles (e.g. head, neck, shoulder, elbow, thorax). While Karg et al. [Karg et al., 2010] focused on the right side of the body based on the symmetric property of walking movement, the coordination of left and right body limbs could be of high interest for the recognition of affects in "symmetric movement" such as walking.

Bernhardt and Robinson [Bernhardt and Robinson, 2007] assumed that "*for the analyzed knocking motions only the right arm exhibits significant movement*". Thus, they focused on 4 features based on the movement of the right arm: maximum distance of hand from body, average hand speed, average hand acceleration and average hand jerk. However, it has been shown in previous works that, in addition to the main body segment involved in the movement task (e.g. right arm for knocking), other body posture and movement are also considered as relevant for the expression of emotion in body movement (e.g. torso movement during knocking [Gross et al., 2010] and head movement during expressive piano performance [Castellano et al., 2008]).

2.4.3 Expressive body cues based on perceptual studies in psychology

Several psychological studies were conducted to explore the characteristics of emotional body expression through perceptual experiments. Later, several studies from affective computing filed have been based on such body cues.

Expressive body cues proposed in psychological studies :
Several psychological researches tend to define general expressive body cues in order to characterize bodily expression of emotions. Meijer [Meijer, 1989] used general dimensions to describe body movement aiming to study their contribution to the attribution of emotion: "(1) trunk movement (stretching, bowing); (2) arm movement (opening, closing); (3) vertical direction (upward, downward); (4) sagittal direction (forward, backward); (5) force (strong, light); (6) velocity (fast, slow); and (7) directness (direct, indirect)". The latter three body cues were inspired from the Laban notation system. Wallbott [Wallbott, 1998] used a set of body cues based both on some specific actions and positions (like downward/backward shoulders, downward/backward head) and three movement qualities: movement activity, spatial extension, and energy.

Montepare et al. [Montepare et al., 1987] explored the effect of emotion expression on walking patterns while focusing on 4 characteristics of walking: 1) the amount of arm swing (no swing arms/ swing arms alot), 2) the stride length (short strides/ long strides), 3) the heavyfootedness (light-footed/ heavyfooted) and 4) the straightness (slouches / stands up straight).

Coulson [Coulson, 2004] proposed a number of expressive body posture configurations to explore the expression of emotions in body posture. The features concerned for the configuration of body posture refer to 6 joint rotations: 1) the head bend, 2) the chest bend, 3) the abdomen twist, 4) the shoulder adduct/ abduct, 5) the shoulder swing, and 6) the elbow bend. An additional feature named "weight transfer" was also proposed to specify the movement of the mass center.

Aronoff [Aronoff, 2006] used in his study two geometrical properties of posture originally appeared in the ballet dance. The first parameter concerns the angular form of the elbows. The second parameter concerns the whole body and it is defined as the angular form of body in the lateral plane. Both of these parameters take one of the following qualitative measures: round, straight and angular.

Expressive body cues adapted from psychological studies :
Dahl et Friberg [Dahl and Friberg, 2007] conducted a perceptual study where

they asked the participants to rate their perception of emotions expressed in musician's performances. The participants were also asked to rate 4 body cues that characterize the expressiveness of the movement: the amount, the speed, the fluency and the regularity of movement. These body cues were based on psychological findings and were defined according to three conditions: 1) ability to describe general motion patterns (not specific to any body part such as the head or the torso), 2) correspondence between motion and musical cues, and 3) ability to capture the expressiveness of body motion regardless the instrument played by the musicians.

Using GEMEP database [Bänziger et al., 2012] and computer vision techniques, Glowinski et al. [Glowinski et al., 2011] propose to compute 5 features inspired from previous psychological researches [Wallbott, 1998]. The features are: 1) energy, 2) spatial extension, 3) smoothness and continuity of movement, 4) forward/ backward leaning of the head and 5) spatial symmetry/ asymmetry of the hands according to the horizontal and the vertical axis.

Alaoui et al. [Alaoui et al., 2012] proposed a computational model for the modeling and the recognition of three movement qualities that describe dancing movements: Breathing, Expanding and Reducing. This approach is mainly used to describe dancing movement. The proposed movement quality features focus on the description of the whole body movement while ignoring the description of some articulations that may convey important information of emotional behaviors (such as head upward/downward movement [Wallbott, 1998]). Introducing "local" body cues related to some particular articulations is helpful to better describe the expression of emotions in body movements.

2.4.4 Summary

In the field of expressive body movement characterization, only few body movement notation systems have been used. Laban notation systems [Laban, 1988] is one of the most well known notation systems that describe expressive body movement. It includes 4 main components: Body, Shape, Effort, and Space. It has been used for the perceptual characterization of emotional body expression [Gross et al., 2010] and for the kinematic analysis of expressive movement [Camurri et al., 2004]. BAP (Body Action and Posture) is another body movement notation system recently proposed by Dael et al. [Dael et al., 2012]. The development of BAP notation system is based on a corpus of prototypical expressions of emotions (GEMEP database [Bänziger et al., 2012]). It provides a detailed description of prototypical expressions. However, it does not include the description of movement dynamics. Besides, BAP units are coded in a binary form (e.g. absence vs presence of a posture/ action units).

We note that most of the previous studies on emotional body expression characterization did not rely on these notation systems [Roether et al., 2009] [Bernhardt and Robinson, 2007] [Kapur et al., 2005]. When the goal is to study the perceptual characterization of emotional body expression, previous studies rely on expressive body cues as proposed in psychological researches. When the aim is to study the recognition and/ or the characterization of bodily expression using kinetic or kinematic data, previous studies rely on a reduced subset of features. In several works, this reduced subset of features is constrained by the movement task (e.g. arms movement features for Knocking action [Bernhardt and Robinson, 2007], flexion angles for Walking action [Roether et al., 2009]). However, there is a need to go beyond action-dependent body cues and to explore different body parts for a deeper analysis of emotional body expression.

Tables 2.1 and 2.2 provide an overview of the body cues used previously in respectively perceptual and automatic analyses of expressive body movement. These tables provide an extension of the list of expressive body cues discussed in [Kleinsmith and Bianchi-Berthouze, 2013].

2.5 Conclusion

We discussed five description levels that are commonly used in previous body movement coding systems; functional, anatomical, directional, posture/movement and space. In this chapter, we focused on structural body movement notation systems based on these description levels. However, we did not focus on functional level.

Different structural body movement notation systems exist. There is a lack of consensus on a common notation system for the description of body movement. This lack is due to the complexity of body structure and the variety of domains in which the body movement notation systems are applied (nonverbal communication of emotions, dance, social interaction, co-verbal behavior ...).

In the study of emotional body expression recognition and characterization, only few works were based on body movement notation systems. Several previous works were focused on constraint based body cues (e.g. arms movement features for expressive Knocking motion) or psychological findings (e.g. collapse/ straightness of the whole body). We note that only few body movement notation systems are intended to describe emotional body expression such as SPAFF and BAP notation systems. Such notation systems provide a fine-grained description of prototypical expression of emotions. However, they are not well adapted to describe how emotion expression modulate our body movement (e.g. how emotion expression affect the way we walk). Laban notation system has been used for the analysis of expressive movement and for the characterization of emotional body expression. Due to the lack of quantification of Laban notation system and the lack of systematic mapping between affective states and its components [Karg and Samadani, 2013], studies relying on Laban movement notation system for the analysis/ synthesis of expressive body movements have been widely limited to the Effort-Shape analysis [Gross et al., 2010].

We conclude that there is a need to propose a body movement notation system that offers a good trade-off between 1) a detailed description of body movement as offered in BAP notation system and 2) a description of movement quality as offered in Laban notation system.

2.6 Summary of chapter

- We present 5 body movement description levels that have been commonly used in previous body movement coding systems: functional, anatomical, directional, posture/ movement and space description level.
- We discuss some examples of body movement notation systems such as SPAFF, Bernese, Body posture coding systems, BAP coding system and Laban notation system.
- We discuss the different approaches adopted in previous works to define expressive body cues for emotional body expression characterization. While some works focus on specific body cues constrained by the motion capture data or the movement task being performed, other studies relied on psychological findings obtained from perceptual experiment. Only few studies relied on previous body movement notation systems, and most of them make use of Laban movement notation system.
- Tables 2.1 and 2.2 provide an overview of the expressive body cues that have been used to conduct perceptual or kinematic analyses of bodily expression of emotions.

Table 2.1: Summary of movement and posture characteristics used in previous perceptual studies for expressive movement characterization.

Body Cues	**Post./ Mvmt.**	**Body Action**	**Short Description**	**Refs**
Movement activity	Mvmt.	Prototypical expression, walking	The amount of movement; a lot of action, almost no action	[Wallbott, 1998], [Montepare et al., 1999]
The amount of arms swing	Mvmt.	Walking	Swinging the arms a lot or not	[Montepare et al., 1987]
The amount of movement (large/none)	Mvmt.	Music performance	The overall measure of the physical magnitude of the movement patterns	[Dahl and Friberg, 2007],
Movement dynamics/energy/power	Mvmt.	Prototypical expression	The energy of the movement	[Wallbott, 1998]
Regularity of movement (regular/irregular)	Mvmt.	Music performance	The variation in movement patterns over the performance.	[Dahl and Friberg, 2007]
Fluency of movement (jerky/smooth)	Mvmt.	Music performance, walking	The smoothness of movement patterns	[Dahl and Friberg, 2007], [Montepare et al., 1999]

Speed of movement (slow/ fast)	Mvmt.	Music performance	the overall number of movement patterns per time unit	[Dahl and Friberg, 2007]
Velocity	Mvmt.	Abstract gestures	Fast - Slow movement	[Meijer, 1989]
Effort-Shape: Limb	Mvmt.	Knocking	Moves close to body, contracted - Moves away from body, expanded	[Gross et al., 2010]
Effort-Shape: Space	Mvmt.	Knocking	Indirect, wandering, diffuse - Direct, focused, channeled	[Gross et al., 2010]
Directness	Mvmt.	Abstract gestures	Direct-Indirect movement	[Meijer, 1989]
Effort-Shape: Energy	Mvmt.	Knocking	Light, delicate, buoyant - Strong, forceful, powerful	[Gross et al., 2010]
Effort-Shape: Time	Mvmt.	Knocking	Sustained, leisurely, slow - Sudden, hurried, fast	[Gross et al., 2010]
Effort-Shape: Flow	Mvmt.	Knocking	Free, relaxed, uncontrolled - Bound, tense, controlled	[Gross et al., 2010]
Force	Mvmt.	Abstract gestures	Strong (muscles tensed) - light (muscles relaxed)	[Meijer, 1989]
The angular form of the body in the lateral plane	Post.	Dancing	The angular form of body in the lateral plane: round/ straight/ angular	[Aronoff, 2006]
The angular flexion of the elbows	Post.	Dancing	The angular form of elbows: round/ straight/ angular	[Aronoff, 2006]
Elbow bend	Post.	Computer generation	Flexion (Positive rotation) of elbow	[Coulson, 2004]
Arms symmetry	Post.	Prototypical expression	The symmetry of arms postures and movement	[Dael et al., 2011]
Straightness.	Post.	Walking	Slouches/ stands up straight	[Montepare et al., 1987]
Heavy-footedness	Post.	Walking	Using heavyfooted/ lightfooted	[Montepare et al., 1987]
Strides length	Post.	Walking	Using short/long strides when walking	[Montepare et al., 1987]
Effort-Shape: Torso	Post.	Knocking	Contracted, bowed, shrinking - Expanded, stretched, growing	[Gross et al., 2010]
Chest bend	Post.	Computer generation	The downward (Positive rotation) /upward (negative rotation) of chest	[Coulson, 2004]
Abdomen twist	Post.	Computer generation	The abdomen was specified as twisting to one side only or not at all.	[Coulson, 2004]
Shoulders	Post.	Prototypical expression	Up, Backward, Forward	[Wallbott, 1998]
Shoulder swing	Post.	Computer generation	The forward (Positive rotation) / backward (negative rotation) of shoulder	[Coulson, 2004]
Shoulder adduct/ abduct	Post.	Computer generation	Negative values relate to arms above shoulder level (abduction), positive values to arms towards the side of the trunk (adduction).	[Coulson, 2004]
Head bend	Post.	Computer generation	Downward (Positive rotation) / Upward (Negative rotation) of the ehad	[Coulson, 2004]

Weight transfer	Post.	Computer generation	Forwards or backwards posture	[Coulson, 2004]
Trunk	Post. / Mvmt.	Abstract gestures	Stretched or bowing movement of the trunk	[Meijer, 1989]
Arm	Post. / Mvmt.	Abstract gestures	Opening symmetrically the arms in front of the body or upward	[Meijer, 1989]
Vertical direction	Post. / Mvmt.	Abstract gestures	Upward or downward movement related to the trunk, head, knees and eyes	[Meijer, 1989]
Sagittal direction	Post. / Mvmt.	Abstract gestures	Forward or backward movement of the whole body	[Meijer, 1989]
Expansiveness/ Spatial extension	Post. / Mvmt.	Prototypical expression	The spatial openness of the body	[Wallbott, 1998]
Upper body	Post. / Mvmt.	Prototypical expression	Away from camera, collapsed	[Wallbott, 1998]
Head	Post. / Mvmt.	Prototypical expression	Downward, Backward, Turned sideways, Bent sideways	[Wallbott, 1998]
Arms	Post. / Mvmt.	Prototypical expression	Lateralized hand/arm movements, Stretched out frontal, Stretched out sideways, Crossed in front of chest, Crossed in front of belly, Before belly, Stemmed to hips	[Wallbott, 1998]
Hands	Post. / Mvmt.	Prototypical expression	Fist(s), Opening/closing, Back of hand(s) sideways, Self-manipulator, Illustrator, Emblem, Pointing (index finger)	[Wallbott, 1998]

Table 2.2: Summary of movement and posture characteristics used in previous automatic analyses for expressive movement characterization.

Body Cues	**Post./ Mvmt.**	**Body Action**	**Short Description**	**Refs**
Whole body motion	Mvmt.	Abstract gestures	The mean values of velocity and acceleration and the standard deviation values of position, velocity and acceleration.	[Kapur et al., 2005]
Hands and elbows motion	Mvmt.	Knocking	Maximum distance of hand/elbow from body, Average hand/ elbow speed, Average hand/ elbow acceleration, Average hand/ elbow jerk	[Bernhardt and Robinson, 2007]
Angular velocity	Mvmt.	Posture	Angular velocity (of right forearm, arm and hand),	[Savva et al., 2012]
Periodicity	Mvmt.	Dancing	The mean of autocorrelation measure of all the end effector positions	[Alaoui, 2012]
Fluidity	Mvmt.	Arms movement	The jerk index of movement	[Pollick et al., 2001]

Flexion-angle trajectories	Mvmt.	Walking	A set of spatiotemporal motor primitives defining the flexion-angle trajectories of eleven major joints including head, spine, pelvis, and left and right shoulder, elbow, hip and knee joints.	[Roether et al., 2009]
Quantity of motion	Mvmt.	Music performance	Movement dynamics features describing the quantity of motion which is defined as an approximation of the amount of detected movement, based on Silhouette Motion Images.	[Castellano et al., 2008]
Head movement velocity	Mvmt.	Music performance	"Movement dynamic features describing the velocity of head movement. The module of velocity is computed based on the coordinates (x, y) of the barycenter (i.e., the centroid) of the pianist's head automatically extracted from the background."	[Castellano et al., 2008]
Head and hands motion	Mvmt.	Prototypical expression of emotion	Movement dynamics features describing the position and velocity of head and hands: Energy, spatial extension, smoothness, symmetry, head leaning	[Glowinski et al., 2011]
Body posture	Post.	Affect-Expressive posture	A set of 24 low-level features that represent the lateral, frontal and vertical extension of body parts (head, shoulders, elbows, hands, heels), and the inclination of the head and shoulders	[Kleinsmith and Berthouze, 2007],[De Silva and Bianchi-Berthouze, 2004]
The mean and peak joint angles	Post.	Knocking	The mean and peak joint angles of some particular articulations	[Gross et al., 2010]
The range of motion	Post.	Knocking	The difference between maximum and minimum values of joint angle related to a particular joint	[Gross et al., 2010]
Verticality	Post.	Dancing	The report between the height and the width of body silhouette	[Alaoui, 2012]
Shoulders angles	Post.	Dancing	the angle formed by the upper arm segment and the vertical axis	[Alaoui, 2012]
Body extension	Post.	Dancing	The maximal distance between CoM and all the end effector of the body	[Alaoui, 2012]
Legs openness	Post.	Dancing	Distance between feet	[Alaoui, 2012]
Weight transfer	Post.	Dancing	Distance between CoM and the middle between the feet	[Alaoui, 2012]
Increase/ decrease	Post.	Dancing	The increase or decrease of a postural feature value and the related variation	[Alaoui, 2012]

Rotation of Head, neck, collar, shoulders, elbows, wrists, torso, hips and knees	Post.	Affect-Expressive posture	"Joint angles describing: stretched outward or backward - stretched inward or forward movement "	[Kleinsmith et al., 2011]
Foot stride	Mvmt.	Walking	Velocity, length and duraction of foot stride during walking	[Karg et al., 2010]
Statistical measures of joint angles	Mvmt.	Walking	The minimum, the mean and the maximum of head, neck, shoulder, elbow, thorax	[Karg et al., 2010]

3 Collection of bodily expression of emotions

The study of emotional body expression often relies on the analysis of a database of body expression of emotions. The recording of emotion expression in body movements can be performed in various contexts such as during human-to-human or human-to-avatar dyadic interaction, games playing, sports playing (e.g. tennis) or daily actions (e.g. walking, knocking,...). Some works focused on constraint-free emotion expression; the actor is asked to freely express emotions through body postures and gestures without any constraint of a specific action.

There is a large consensus on the multimodality of emotion expression[Scherer and Ellgring, 2007]. The different signals tend to complement each other rather than duplicating the same information [Cowie et al., 2011]. Multimodal database for emotion expression are useful not only to have complete information about the expression of emotion, but also to study the interaction between the different multimodal channels. However, previous researches point out that the emotion expression doesn't seem to be distributed equally between the different modalities [Cowie et al., 2011].

The development of databases for emotion-oriented systems is a challenging task [Cowie et al., 2011]. As it is difficult to collect spontaneous emotional behavior, the collection of emotional body expression has been widely based on acted data and induction techniques to elicit emotions.

The recording of databases of body movement is also a challenging task at the level of recording practicalities. Gathering accurate data from all modalities (face, gaze, voice, body...) requires very sophisticated recording equipment. Sensors-based motion capture systems mostly allow obtaining more accurate data than video-based recording. However, the latter ones are considered as non-invasive motion capture systems and they are sometimes preferred to foster the naturalness of the actors' behaviors at the cost of data recording accuracy.

In this chapter, we explore different issues related to the recording of emotional body expression. In section 3.1, we discuss the advantages and the drawbacks of naturalistic and acted expressive behavior. Section 3.2 is devoted to the discussion of data induction techniques to elicit emotions. In section 3.3, we discuss the techniques used to record body movement. Section 3.4 provides an overview of existing databases.

3.1 Naturalistic vs acted expressive behavior

One of the most contentious issues in the recording of emotion expression databases is the difficulty to record as natural and spontaneous expressive behaviors as possible. Many years ago, it was reported in [Wallbott and Scherer, 1986] that "*the ethical and practical difficulties in experimentally inducing strong emotional states, or even recording naturally occurring emotional experiences, has been one of the perennial problems in the scientific study of emotion*".

Spontaneous (or naturalistic) affective behaviors have been obtained in real-life situations such as in video game playing [Kleinsmith et al., 2011], therapy sessions or reality television [Devillers and Martin, 2012]. Data recorded in real-life contexts involve many important information such as physiological arousal which gives rise to changes in muscle tone or reactivity. Acted affective data result from the acted reaction of an "actor" who is aware of the goal of the study. The term "actor" is commonly applied in the context of acted behavior recording.

Spontaneous expression of emotion has been widely considered as harder to collect. Acted data are easier to collect, but they are often considered as "unnatural" or "artificial". One critical aspect of acted data is the exaggeration of the expression. Only few works [Atkinson et al., 2004] tackled the issue of the exaggeration of emotional body expression and its effect on the perception of emotion.

The question that can be raised when recording acted data is: how we can guarantee that we bring out genuinely the required emotional state? A number of solutions have been proposed to deal with this question. Firstly, the use of professional actors working with a director can give more chance to collect emotional behaviors that are the closest to natural behaviors. Secondly, acted data often rely on induction approaches to create and record emotion expression. Several induction techniques were proposed in the literature. They are designed to create an affective state by exposing the subject to some particular conditions or by helping the subject imagining some particular situations. In the following section, we present some of the induction approaches that have been used in the literature. Finally, a post perceptual study is mostly required to ensure the matching between the intended emotion and the perceived one. In the context of emotional body expression, such a perceptual study consists in a perception task where participants are asked to recognize emotions based on their perception of expressive body movement.

Acted data has also potential advantages like ensuring the structure of the data and the quality of recording [Cowie et al., 2011]. Indeed, the recording of acted expressions allows obtaining clearly defined and more controlled emotional states as the actors are mostly asked to express a specific category of emotion or affect. Thus, acted expressions can be very useful given the difficulty to record naturalistic expression of emotions in lab setting [Bänziger and Scherer, 2007]. However, more efforts have to be conducted using acted databases to explore to what extent acted expressions can be useful in real-life situations (e.g. automatic recognition of emotional body expression) [Kleinsmith and Bianchi-Berthouze, 2013].

Due to the difficulty of naturalistic emotion expression recording, most of the experiments on emotional body expression analysis have been carried out on acted data recorded by actors [Devillers and Martin, 2012]. Only few recent studies have taken up this challenge by focusing on spontaneous data recording for instance during video game play [Savva et al., 2012, Kleinsmith et al., 2011].

3.2 Data induction techniques

Acted expressions can be recorded using Verbal labels based approaches. In such approach, the actors are asked to express an emotion defined only with a label such as Anger or Sadness. That means that the only information provided to the actor to express a particular emotional state is a label corresponding to an emotion. This technique is very limited due to the poorness of emotion description through a single word which results in a wide variety of self-interpretation of the label between actors.

Several approaches were proposed in the literature to induce affective states on demand hoping for a resulted acted emotional expression that is the closest to the natural emotional behavior of human. Those approaches aim to achieve acceptable balances between naturalness and the advantage of acting data. They try to record emotions by inducing them. Different categorization of induction techniques have been proposed [Bänziger et al., 2012] [Devillers and Martin, 2012]. In the following, we cite some of the most used techniques.

– **Scenario based approaches:**
  Scenario based approach can be considered as a complementary approach of Verbal labels based induction technique [Bänziger et al., 2012]. Scenario based approach is based on typical scenarios that better describe verbal emotion label in order to contextualize the expression of emotion [Bänziger et al., 2012]. In scenario based approaches, the actor is asked to read a written copy of a scenario and to imagine how he would feel in the corresponding situation. This approach is used in several previous emotional behaviors collection [Dael et al., 2011] [Wallbott, 1998] [Montepare et al., 1987]. Bänziger et al. [Bänziger and Scherer, 2010] used three scenarios that describe the situation in which a given emotion is elicited.
– **Visual or audio influence:**
  This technique is based on the visual influence of some pictures or video sequences to elicit emotions. The AVLaughterCycle Database [Urbain et al., 2010] was recorded based on visual influence induction approach. The subjects were asked to watch funny video to elicit laughter. Personality traits of 35 participants are inferred in [Abadi et al., 2015] based on the analysis of their naturalistic reactions to 16 emotional videos.

– **Autobiographical memories paradigm for emotion elicitation:**
  Autobiographical memories paradigm consists of asking the subjects (actors)

to remember an event that they lived previously and that can be associated with a given emotion. Gross et al. [Gross et al., 2010] used this approach for emotion elicitation to record and analyze expressive knocking motion.

– **Others**:
  Computer or video games [Kleinsmith et al., 2011], physical induction, conversational interactions and other induction techniques were also proposed in previous works to induce emotional states [Cowie et al., 2011] [Devillers and Martin, 2012].

3.3 Technical recording of body movement

The format of the recorded data within each database depends on the recording equipment. For instance specific sensors may be used to measure physiological reactions such as heart beat and skin temperature [Zacharatos et al., 2014]. Pressure sensors have been also used at the level of the seat or the back of the chair to capture different categories of seated posture [Mota and Picard, 2003]. Such sensors are useful to ensure the privacy of the user as his/her identity cannot be retrieved from such data. Physiological sensors are mainly employed in the context of naturalistic emotional behavior recording. In this chapter, we focus on motion capture and audio-visual based recording equipment as they were the most used in the context of acted emotional body expression recording.

3.3.1 Audio-Visual recording techniques

Audio-visual recording refers to the use of cameras providing video clips. The use of several cameras provide different viewpoints of each motion sequence [Bänziger et al., 2012]. In the context of studying emotional body expression, video clips are mostly used for perceptual studies where the participants are asked to recognize the emotion being expressed by the actors. However, video clips can be also used to build computational models for expressive motion analysis based on the extraction of 2D features through computer vision techniques (e.g. background subtraction, silhouette extraction...) [Glowinski et al., 2011].

3.3.2 Motion capture recording techniques

Motion capture techniques mostly refer to sensors based techniques. They produce 3D information about the posture and movement of the subject. Employing high technical quality, one is able to decode and analyze physical properties of expressive body movement [Volkova et al., 2014a]. Classical motion capture systems are based on the use of markers (also called sensors) attached to the body. New markerless motion capture techniques have been recently proposed.

– Optical systems: Optical motion capture systems are based on computer vision techniques to track the position of several markers placed on specific joints of

Figure 3.1: Examples of suit used in sensors-based motion capture systems

(a) Vicon motion capture system [CMU,]

(b) Gypsy mocap system by Animazoo [gyp,]

(c) IGS-180 mocap system from Animazoo [IGS,]

(d) 3DSuit mocap system from Innalabs [3ds,]

(e) Xsens MVN mocap system from Xsens [Xse,]

3

the body (See Figure 3.1). Many cameras are mostly used to track the markers position. The use of more cameras is appreciated to deal with occlusion problems. The Vicon motion capture system is one of the most well known optical motion capture systems [Roether et al., 2009] [Hicheur et al., 2013] [Busso et al., 2008]. While providing accurate motion capture data, advanced computer vision techniques are required to track the markers positions.

- Mechanical systems: Mechanical motion capture systems are based on mechanical device placed around the body to estimate the posture and the movement of the subject. Gypsy system from Animazoo technology is an example of mechanical motion capture system. It is based on a sophisticated exoskeleton (See Figure 3.1). Mechanical techniques mostly provide data with high accuracy. However, the device used to capture the motion of the body can be considered as cumbersome from the subjects viewpoint. The comfort of the subject is an important point to consider in expressive body motion recording as it can affect the expressive content of the movement.
- Magnetic systems: Magnetic motion capture systems use an artificial magnetic field sender and sensors located in key positions of the body. MotionStar from Ascension technology is an example of magnetic motion capture systems. Unlike optical systems, magnetic motion capture systems does not suffer from the issues related to tracking and the occlusion of the markers. However, they are very sensitive to the presence of any other magnetic field and metallic objects in the environment.
- Inertial systems: Inertial motion capture systems use inertial sensors each composed of a three axis accelerometer, a three axis gyroscope and a three axis magnetometer. We can cite three inertial motion capture systems that use around 17 inertial sensors for full body motion capture; Xsens MVN [Xse,], IGS-180 provided by Animazoo [IGS,] and 3DSuit provided by Innalabs [3ds,] (See Figure 3.1). Inertial systems are less sensitive to the environment control than magnetic or optical systems.
- Markerless Motion Capture Technology: Markerless motion capture systems attempt to provide 3D motion capture data using markless sensors such as depth camera. There is no need to attach specific markers to the body of the subject. Kinect from Microsoft is one the most well known and the most used markless techniques for full body motion capture. The advancement of these techniques is still in progress. While they provide a good compromise between the reliability of the data and the natural behavior of the subject, they did not achieve the accuracy of sensors-based motion capture techniques yet.

3.3.3 Summary

We mainly distinguish two types of recording equipment; audio-visual based recording and motion capture based recording. Several motion capture systems exist nowadays. Probably optical motion capture systems are the most expensive and the most complex. They provide accurate motion capture data when sophisticated computer vision algorithms are used. However, the motion capture session has to be performed in very high controlled environment (e.g. the light has to be controlled, the subject can only perform motion while being surrounded by all the cameras). Magnetic systems also require high controlled environment (e.g. no other magnetic filed neither metallic objects should be present in the environment). Unlike optical and magnetic systems, Mechanical systems require less controlled environment but they are more cumbersome from the user viewpoint. Wireless inertial systems can be used in inside or outside setting. They have shown good compromise between the accuracy of the 3D motion capture data, the ease of use, the sensitivity to the environment, the time of calibration, and post-processing of the data. Finally, recent Markerless motion capture technology has been proposed. However, this technology is still an active field of research and Markerless techniques provided nowadays cannot replace accurate sensors-based motion capture systems.

3.4 Databases of expressive body behavior

Due to the recent increase of interest in studying emotion expression in body movement, several databases have been recorded during the last two decades. From the perspective of recording equipment, recent databases can be split into three categories; audio-visual recording based databases, 3D motion capture recording based, and finally multi-media databases that are based on both video and motion capture recordings [Karg and Samadani, 2013] [Kleinsmith and Bianchi-Berthouze, 2013] [Volkova et al., 2014a]. Audio-visual recording provides video clips and/or audio information. 3D motion capture of body movement provides accurate information about the 3D posture and movement of body segments.

Thanks to the recent progress in recording technology, several high quality databases of human body movement are available nowadays. CMU database (from CMU Graphics Lab) is a rich repository of motion capture data and one of the most known motion capture databases (http://mocap.cs.cmu.edu/). It encompasses around 2600 motion capture files of different types of motion such as walking and jumping. The motion capture files of this database are also converted to different formats. Most of the recorded motions cover actions, very few contain emotional content. The KUG (Korea University Gesture) database [Hwang et al., 2006] is another repository of gestures which contains both 2D video data and 3D motion capture data. The database consists of different types of non-emotional gestures including "normal" (e.g. sitting, walking), "abnormal" (e.g. falling) and "command" (e.g. drawing 'X') gestures. However, KUG database was not dedicated to the study of emotional body expression. The HumanEva database [Sigal et al., 2010] is another non-emotional gestures and actions recorded using videos and an optical motion capture system. The actions recorded in this database include walking, dancing and jumping.

Other scholars recorded databases for the study of expressive social interaction between a human and a virtual agent [Demulier et al., 2014] [McKeown et al., 2012]. Recently, in the framework of ILHAIRE project (http://www.ilhaire.eu/), two databases of laughter communication in social interactions were recorded; UCL-ILHAIRE database [Griffin et al., 2013a] and MMLI database [Niewiadomski et al., 2013] (Multimodal multiperson corpus of laughter in interaction). A new multi-modal database, called "EmoPain" is introduced in [Aung et al., 2015] to study naturalistic pain-related affective expressions including facial and vocal expressions, body movement and muscle activity. Other databases are also constructed based on the extraction of sequences of video clips from movies and TV programs [Clavel et al., 2011] [Eroglu Erdem et al., 2014].

In the following of this chapter, we will focus on the databases based on induction techniques to explore the expression of emotion in body posture and movement. Not all the databases that we present are available for research community, but the discussion of previous databases allows providing an overview of the databases that have been recorded so far for the study of emotional body expression.

3.4.1 Examples of audio-visual based databases

As examples of audio-visual based databases we can cite: BEAST [de Gelder and Van den Stock, 2011], GEMEP [Bänziger et al., 2006], and FABO [Gunes and Piccardi, 2006] databases.

- BEAST [de Gelder and Van den Stock, 2011]: The Bodily Expressive Action Stimulus Test (BEAST) database consists of 254 stimuli of expressive postures acted by 46 actors. The database is designed for the study of the perception of whole body expression of 4 emotions; Anger, Fear, Happiness and Sadness. This database is restricted to a set of pictures of expressive body postures. The face is blurred to avoid the influence of facial expressions on the perception of body posture.
- GEMEP [Bänziger et al., 2006] [Bänziger et al., 2012]: GEneva Multimodal Emotion Portrayals is a corpus for the study of multimodal emotional expressions. Three digital cameras were used to record video clips of emotional body expression sequences from three viewpoints. Both facial expressions and upper body movements were provided in each video clip. The actors were 10 professional theater actors. A professional theater acting director was also recruited to work with the actors. 18 emotion categories were considered in this database but not all of them were expressed by the actors. The 18 emotions were split into 6 categories [Bänziger et al., 2012]; 1) Positive valence and High arousal (Joy, Amusement, Pride), 2) Positive valence and Low arousal (Pleasure, Relief, Interest), 3) other emotions of Positive valence (Admiration, Tenderness, Surprise which can be considered as neither a positive nor a negative emotion), 4) Negative valence and High arousal (Hot anger, Panic fear, Despair), 5) Negative valence and Low arousal (Cold anger, Anxiety, Sadness), and 6)

finally other emotions of Negative valence (Disgust, Contempt, Shame). This database was validated by 57 participants. Only 17 emotions were used for validation purpose (Shame was not considered as it did not satisfy the criteria of the stimuli used for the validation study).

- FABO [Gunes and Piccardi, 2006]: FABO is a Bimodal Face and Body Gesture Database. It is designed for automatic analysis of human nonverbal affective behaviors. 23 actors were asked to express emotions through face and gestures. Their behaviors were recorded using two digital cameras. Similarly to GEMEP database [Bänziger et al., 2012], several emotional categories were considered in FABO database: Uncertainty, Anger, Surprise, Fear, Anxiety, Happiness, Disgust, Boredom and Sadness.

Video based recording is mostly considered as a fast and a simple recording device. However, building accurate computational models based on video clips is impaired by the difficulty to extract features that describe the 3D movement of the whole body. Thanks to recent advancement in motion capture technology, several motion capture based databases have been gathered during the last years.

3.4.2 Examples of motion capture based databases

Thanks to the recent advancement of motion capture technology and motivated by the accuracy of motion capture data comparing to audio-visual data, several works turned their attention to the collection of motion capture data of expressive body posture and movement. For instance, such databases were recorded and described in [Ma et al., 2006], [Kleinsmith et al., 2006b], [Roether et al., 2009], [Hicheur et al., 2013], [Hicheur et al., 2013], [Kleinsmith et al., 2011], [Savva et al., 2012] and recently in [Volkova et al., 2014a].

- Ma et al. [Ma et al., 2006] gathered a motion capture database for the study of identity, gender, and emotion perception from biological motion. 30 actors participated in collecting this database. They were asked to express 4 emotions (Neutral, Anger, Happiness and Sadness) while doing 4 daily actions; walking, knocking, lifting and throwing. This database as well as a 3DS max script allowing the creation of biped figure based stimuli have been made available to the research community. Their database contains around 4000 expressive movement sequences, leading to a rich repository of emotional body expression in daily actions. However, the expressed emotion are restricted to a subset of the main basic emotions.
- UCLIC database [Kleinsmith et al., 2006b]: UCLIC database is an expressive acted full body postures. 13 actors were asked to express 4 emotions (anger, fear, happiness, and sadness) through their body posture. The procedure leaded to 108 expressive postures. 70 participants were asked to recognize the emotions expressed in body posture reproduced on a biped figure.
- A database of expressive walking [Roether et al., 2009]: The authors recorded

the expression of anger, happiness, sadness, fear in walking action. 25 actors participated to the recording of their database and 21 other participants were asked to recognize the expressed emotions from walking movement reproduced on body model of a puppet.
- A database of expressive walking [Hicheur et al., 2013]: The authors recorded a smaller database of expressive walking compared to the one recorded in [Roether et al., 2009]. However, they synthesize more stimuli by changing the speed and the orientation of upper body animations.
- Mockey Database [Tilmanne and Dutoit, 2011]: This database was mainly designed for stylistic walk analysis and synthesis. Only one actor was recruited to record expressive walking based on 11 walk styles (e.g. decided, drunk, cool, afraid,...).
- AffectME [Kleinsmith et al., 2011]: Unlike motion capture databases introduced in this section so far, AffectME is based on non-acted emotional body expression. It consists of naturalistic full body postures database recorded during a video game play. A cross-cultural perceptual study was performed to explore the recognition of emotions from expressive body postures reproduced on a computer avatar.
- A database of expressive movement recorded during WII tennis game [Savva et al., 2012]. Similarly to the work of [Kleinsmith et al., 2011], Savva et al. [Savva et al., 2012] employed video game based scenarios to collect spontaneous expressive behaviors. Unlike in [Kleinsmith et al., 2011] which is restricted to expressive postures, expressive body movement were recorded and reproduced on a computer avatar. The clips of computer-animated avatars are then used to conduct a perceptual study where the participants are asked to recognize which emotion was expressed in body motion.
- The MPI Emotional Body Expressions Database [Volkova et al., 2014a]: This database was recorded in the framework of narrative scenarios. The actors were asked to imagine to tell a story to a child. As such, no specific emotion induction techniques were used to collect expressive behaviors. The expressiveness of body movement results from the expressive content of the narrative scenario. Similarly to GEMEP database, Volkova et al. [Volkova et al., 2014a] considered on a larger set of emotions than most of the other databases. 11 emotions were considered in their database; five emotions of Positive valence (amusement, joy, pride, relief, surprise), five emotions of Negative valence (anger, disgust, fear, sadness, shame) and neutral.

Other scholars tend to multi-media databases containing synchronized video and motion capture data. These databases are useful to compare the human recognition of emotions expressed in facial and bodily movement.

3.4.3 Examples of multi-media databases

Multi-media databases refer to the databases that record both video clips and motion capture data. Video and motion capture data are mostly synchronized. As examples of multi-media databases that have been recently collected and described in [Busso et al., 2008], [Metallinou et al., 2010] and [Gross et al., 2010].

– IEMOCAP [Busso et al., 2008]: IEMOCAP is an Interactive Emotional dyadic Motion Capture database. Both scripted and spontaneous spoken communication scenarios were employed to elicit specific emotions; Happiness, Anger, Sadness, Frustration and Neutral. Both video clips and 3D motion capture data were also recorded. However, motion capture data were only provided for facial expressions and hand movements (using few markers around the hands). IEMOCAP was annotated based on video clips. The annotation was performed using categorical labels of emotions (expressed emotions + disgust, fear, surprise and excited) as well as dimensional labels such as valence, activation and dominance
– USC CreativeIT [Metallinou et al., 2010]: USC CreativeIT database contains expressive full body behavior recorded during dyadic interactions based on a two-sentence exercise and its paraphrase. No specific emotions were expressed but a diverse expression of emotions and intentions are obtained during the dialog. Expressive behaviors of the full body were recorded using Vicon motion capture system, HD Sony Video Camcorder and microphones. The data was annotated according to three affect dimensions (valence, activation and dominance) and theater performance properties (interest, naturalness and creativity). Only video clips were used for the annotation process.
– Expressive knocking movement [Gross et al., 2010]: Unlike USC CreativeIT and IEMOCAP databases which focus on dyadic interaction, the database recorded in the work of Gross et al. [Gross et al., 2010] consists of emotion expression during a daily action; knocking at a door. Although knocking movements mainly imply arms movement, full body movements were recorded. Gross et al. [Gross et al., 2010] asked 6 female actors (drama students) to express 6 emotions while knocking at a door; Anger, Anxiety, Sadness, Pride, Contentment, and Joy. 35 participants also participated as observers to recognize the expressed emotions through a perceptual study. Both motion capture data and video clips were obtained for each recording session, but only video clips (from side view) were used for the perceptual study. The face region was blurred to avoid the influence of facial expression on the recognition of emotion from expressive body movement.

3.4.4 Discussion

Different approaches and contexts were adopted to record emotion communication in body movements. Several scholars focus on the communication of emotions through body actions, posture units and communicative gestures (e.g. GEMEP database [Bänziger et al., 2006] [Bänziger et al., 2012]). In these works, the actor is usually asked to freely express an emotion without any constraints [Bänziger et al., 2006] [Wallbott, 1998]. Some researches highlight the importance of studying emotion expression during an interaction (e.g. IEMOCAP [Busso et al., 2008]). Emotion can be communicated implicitly while the subject is performing some movement tasks (e.g. walking) [Roether et al., 2009, Gross et al., 2010]. Some researches focused on collecting bodily expression of emotions in daily actions [Ma et al., 2006, Dahl and Friberg, 2007, Gross et al., 2010, Roether et al., 2009]. These studies investigate the effect of emotions on movement qualities.

In the context of emotional body expression recording, only few databases were based on synchronized audio-visual and motion capture recording. Most of existing databases are based either on audio-visual recording or on motion capture technology. 3D motion capture databases provide accurate information about the 3D information (3D rotation, position or both) related to body movement. Thus, they are more appropriate for low-level analysis of expressive body movements than video clips based databases. However existing databases based on 3D body movement recording so far are limited to a small range of emotional states, mostly reduced to a subset of the basic emotions [Ma et al., 2006] [Kleinsmith et al., 2006b] [Roether et al., 2009] [Hicheur et al., 2013]. Few recent motion capture databases considered more emotions than the basic ones [Volkova et al., 2014a]. Nevertheless, they focused on a specific movement task. There is a need to go beyond a restricted subset of emotional categories and a reduced subset of movement tasks in order to study the emotional body expression in several daily actions.

Multi-media databases recorded in [Metallinou et al., 2010], in [Gross et al., 2010] and in [Busso et al., 2008] were annotated only using video clips. Gathering both video clips and 3D motion capture data is useful for several purposes. For instance, video clips are useful to help the manual identification of the beginning and the end of a motion sequence (e.g. motion phase of repetitive knocking action). Besides, video clips can be used as a complementary resource to compare the perception of emotions from body language using motion capture based stimuli (i.e. 3D motion reproduced on a biped figure) and using audio-visual data.

Inspired from the summary of existing databases provided in [Volkova et al., 2014a], we provide in Table 3.1 an overview of the databases discussed previously. Similarly to [Volkova et al., 2014a], the databases are described in terms of 6 aspects: the Body Part (e.g. Full Body, Upper Body), the action (e.g. walking), acted or naturalistic data, the number of samples, actors and raters, the format of the data and finally the number of expressed emotions.

Table 3.1: Examples of databases used to study emotional body expression. *others** stands for 3 affect dimensions, interaction tendencies and performance ratings. (To be continued in the next page)

Database	**Bod Part**	**Body action**	***Acted* or *Naturalistic***	**Samples, Actors, Raters**	***Mocap* or *Video***	**Nb of emotions**
[Ma et al., 2006]	Full Body	Daily actions	A	4080, 30, 0	M	4
UCLIC [Kleinsmith et al., 2006b]	Full Body	Free expression	A	108, 13, 70	M	4
FABO [Gunes and Piccardi, 2006]	Face + Upper Body	Free expression	A	1900, 23, 0	V	10
IEMOCAP [Busso et al., 2008]	Face + Hands	Dyadic interaction	A+ N	10039, 10, 6	M + V	9 + 3 affect dimensions
[Roether et al., 2009]	Full Body	Walking	A	388, 25, 21	M	4 + Neutral
[Tilmanne and Dutoit, 2010a]	Full Body	Walking	A	540 steps,1, 0	M	11 Styles
[Gross et al., 2010]	Full Body	Knocking	A	42, 6, 35	M + V	6
AffectME [Kleinsmith et al., 2011]	Full Body	Video game play	N	103, 11, 8	M	4 + 2 affect dimensions
USC CreativeIT [Metallinou et al., 2010]	Full Body	Dyadic interaction (Two-sentence exercise and paraphrases)	A	>1200, 19, 5	M+V	8 *others**
BEAST [de Gelder and Van den Stock, 2011]	Full Body	Scenario-based expression	A	254, 46, 19	V	4
[Savva et al., 2012]	Full Body	Wii tennis games	N	161, 9, 7+9	M	8
GEMEP [Bänziger et al., 2012] [Bänziger et al., 2006]	Face + Upper Body	Free expression	A	1260, 10, 57	V	17
[Hicheur et al., 2013]	Full Body	Walking	A	50, 3, 20	M	5
MPI [Volkova et al., 2014a]	Upper Body	Narrative scenario	A+ N	1447, 8, 55	M	11

3.5 Conclusion

In this chapter, we discuss few issues related to the collection of emotional body expression databases. One of the most agreed upon issues in collecting databases of emotional expressions is the need of naturalness and spontaneity of emotional behaviors [Kleinsmith and Bianchi-Berthouze, 2013]. Gathering data with such quality is complex and researchers tend to find compromise between data richness and recording quality. While naturalistic data allows capturing rich information (such as physiological arousal, changes in muscle tone or reactivity), using acted data offers potential advantages like ensuring high quality of recording [Cowie et al., 2011]. Due to the difficulties related to the collection of spontaneous and natural emotional expressions, emotion induction methods are often used. Several approaches were proposed in the literature to induce affective states on demand [Cowie et al., 2011]. They aim to achieve an acceptable compromise between data naturalness and the advantage of acting data. Scenario based approach is one of the most used approach for data induction [Dael et al., 2011], [Wallbott, 1998, Montepare et al., 1987]. In scenario based approaches, the actor is asked to read a scenario and to imagine how he would feel in the corresponding situation.

Most of the databases relying on acted data are followed by a perceptual study to validate the matching between the emotions that were intended by the actors and the emotions recognized by other participants. Acted data are mostly recorded using audio-visual recording equipment or motion capture data equipment. Although the analysis of digital video can be used for producing computational models of multimodal behavior, 3D motion capture of body movement provides more accurate information about 3D posture and movement of some specific body joints of subjects, allowing one to produce more accurate computational models. While Markerless motion capture systems were recently proposed to foster the naturalness of actors, sensors-based motion capture systems still provide more accurate results.

Previous emotional body expression databases were mostly tailored to specific research purposes, which shape and constrain the technical methodologies of the recording. For instance, when recording emotion communication in dyadic interaction while being seated, only upper body movement are often recorded. Apart few examples (e.g. [Wallbott, 1998, Dael et al., 2012, Volkova et al., 2014a]), scholars tend to focus on a subset of the six basic emotions. Moreover, only a limited range of motions, often reduced to one movement task (e.g. walking or knocking at the door) is studied.

The increase development of databases recording is mainly due to two issues; 1) the lack of databases freely available for research community and 2) the need to develop new databases specific for each particular research question.

3.6 Summary of chapter

- We discussed different techniques to record emotional body expression, in particular naturalistic or acted data.
- Different induction techniques were proposed to elicit emotions in the liter-

ature such as visual influence, scenario-based approach and autobiographical memories paradigm [Cowie et al., 2011].

- We report on two types of recording in the context of acted emotional body expression databases: audio-visual recording and motion capture recording. Recent technologies of 3D motion capture of body movement provide more accurate information about 3D body posture, allowing one to explore more efficiently the complex structure of the body and to provide accurate interpretation of the most relevant body cues conveying emotions.
- Different motion capture systems are available nowadays. Optical solutions offer high quality recording at the cost of high computational time and high complexity of data post-processing. Besides, they require a high controlled environment. Magnetic motion capture systems also require high controlled environment. Mechanical motion capture systems can offer good quality data but the cumbersome suit may limit the expressiveness of the actor. Inertial motion capture systems offer good compromise between the ease of use and the ease of post-processing. Besides, they are less sensitive to the control of environment.
- Previous databases of emotional body expression in daily actions were mostly restricted to a smaller subset of emotions and/or a smaller subset of actions (mostly reduced to one specific action).
- Table 3.1 provides an overview of the databases used to study emotional body expression.

4 Features selection approaches for the identification the most relevant expressive body cues

One of the primordial steps in the studies of emotional body expression is the definition of a set of body cues to characterize expressive body movement. Different approaches were adopted in previous works (e.g. relying on a particular coding scheme) as reviewed in Chapter 2. Another important step in the study of emotional body expression is the identification of the most relevant body features that capture the most important characteristics of emotional body expression. Different approaches have been proposed in Affective Computing researches to identify the most prominent expressive features. One approach is to rely on the most important body cues as highlighted by psychological researches. A similar approach is to focus on action-dependent body cues (e.g. elbow flexion feature for the analysis of Knocking motion). As body structure is characterized by a high degree of freedom, it is worth exploring body cues covering different body parts to give a deeper insight on which cues convey emotional expression. Thus, another approach consists in using machine learning techniques to identify the most relevant expressive features starting from a large set of features that describe the whole body movement. This chapter is intended to give an overview of feature selection techniques proposed in machine learning research community.

This chapter is organized as follows: Section 4.1 introduces the general purposes of dimensionality argued with some examples while focusing on the study of emotional body expression. Sections 4.2 and 4.3 summarize respectively two categories of machine learning approaches for feature selection: Feature construction approaches and Feature subset selection approaches. Last but not least, section 4.4 concludes our discussion on feature selection techniques and motivates the relevance of the most suitable approach for our work. Finally, section 4.5 summarizes this chapter.

4.1 Dimensionality reduction

Feature selection is commonly used as a pre-processing step to machine learning algorithms. High-dimensional data (i.e. the existence of a large amount of features) usually affect the training process in classification tasks in term of efficiency (complexity and training time) and performance (classification accuracy), in particular with the presence of several irrelevant and/or redundant features. Several learning methods tend to over-fit in the presence of irrelevant and/or redundant features.

The reduction of data dimensionality is then often motivated by the sensitivity of learning models to high-dimensional data in term of efficiency, over-fitting and accuracy, but it can be also desired for other objectives such as the facilitation of data visualization/interpretation and the reduction of the measurement and storage requirements [Guyon and Elisseeff, 2003]. In the following subsections, we provide two examples of applications that require the reduction of data dimensionality for interpretation purpose. We focus on the study of emotional body expression.

4.1.1 Medical applications

Several recent applications suffer from high-dimensional data, in particular in medical diagnosis (including many bioinformatics applications). In many cases, it is required to identify a subset of relevant features for the prediction of a particular concept. A typical example in bioinformatics is the need to select a subset of genetic markers relevant to the prediction of a certain disease [Strobl et al., 2007]. Motivated by the high-dimensional nature of many tasks in bioinformatics, several techniques of features selection were proposed in specific applications in bioinformatics. In their paper, Saeys et al. [Saeys et al., 2007] provide a review of feature selection techniques in bioinformatics.

4.1.2 Emotional body expression

Recently, a growing interest has been attributed to the expression of emotions in body movement. Recent motion capture techniques are available nowadays allowing to capture the complex structure of the whole body. Hence, it is interesting to explore which particular body cues are the most informative and useful to characterize the expression of emotions. However, the study of emotional body expression is relatively recent and previous studies mostly focused on critical expressive body features for the perception of emotions expressed in body movement. Indeed, several studies were based on perceptual studies to explore the most important expressive features for the perception of emotions expressed in body movement [Meijer, 1989, Montepare et al., 1987, Wallbott, 1998]. Thus, the selection of a subset of expressive body cues in Affective Computing research field has been widely based on psychological findings. In recent years, there has been an increase of interest on kinematic analysis of emotional body expression using motion capture data [Kapur et al., 2005, Castellano et al., 2007, Bernhardt and Robinson, 2007].

The selection of features is mostly motivated by the need to get deeper insights into which body cue is the most relevant to discriminate between emotions. To identify the most relevant features, several studies explored the statistical power of body cues using statistical tests. Other studies focus on the selection of the most prominent features based on machine learning techniques to explore the predictive power of body cues. In the following paragraphs, we provide some of the previous works that attempt to select and identify the most relevant expressive body features in Affecting Computing field.

Using statistical tests :

Based on one-way ANOVA and Tukey's HSD test, Camurri et al. [Camurri et al., 2003] explored the statistical effect of emotions expressed in full-body movement (dancing) on four body cues: overall duration of time, contraction index, quantity of motion and motion fluency. The overall duration of time is defined as the overall duration of motion. The contraction index measure is based on the axes of the ellipse that surrounds the body. It indicates "how the dancer's body uses the space surrounding it" and it is related to Laban's "personal space". The quantity of motion is directly related to the amount of detected movement. Fluent motion in defined as a movement performed in a continuous, " harmonic" way, implying "few and long motion phases". The authors found that the duration of motion is significantly longer for grief expressions and the quantity of motion is significantly higher in anger and joy expressions than in grief expression.

Castellano et al. [Castellano et al., 2008] extracted a number of features and apply statistical techniques to explore the effect of each feature. The features used in their study explain the temporal dynamics of the velocity of head movement and the quantity of upper body movement to automatically analyze expressive body movement in piano performance. The expression of five moods (personal, sad, allegro, serene and over-expressive) were considered in one professional musician's performance. Based on statistical techniques (analysis of variance and pairwise comparisons), they found that head velocity features showed higher impact on the discrimination of expressed emotion compared to the features describing the quantity of motion. The authors explained this result as follows: piano performance does not imply significant difference in movement quantity between different emotions (except Sadness which received significantly lower values of movement quantity features).

Using machine learning techniques :

In another study by Castellano et al. [Castellano et al., 2007], machine learning techniques were used to select the best subset of features related to the temporal profiles of 5 motion cues. The motion cues are; quantity of motion and contraction index of the body, velocity, acceleration and fluidity of the hand's barycentre. Using a wrapper and a correlation based FS approaches, it was found that the quantity of motion and the contraction index of upper body contribute better than the speed and the fluidity of motion in discriminating between different emotions expressed in "Arms rising in coronal plane" action. A decision tree was also built to interpret the prediction power of each body cue (e.g. the maximum of Quantity of motion seems to discriminate between "high arousal" emotions and "low arousal" emotions).

Bernhardt and Robinson [Bernhardt and Robinson, 2007] extracted statistical measures of arms movement qualities measured based on 3D motion capture data in

4

order to classify four emotions expressed in knocking action. Although they obtained successful classification results, they did not provide insight on which body cues contribute better in emotions discrimination.

Roether et al. [Roether et al., 2009] applied sparse regression to find out the most "*critical*" kinematic expressive features that characterize the expression of emotions during walking. The downward flexion of the head was considered as highly important for Sadness expression during walking while the flexion of elbows was relevant to characterize Anger and Fear expressions during walking.

4.1.3 Discussion

Statistical techniques are useful to provide a first interpretation on the discrimination power of each feature. In the studies of bodily expression of emotions, statistical techniques can also be used to quickly identify irrelevant body cues for which no significant main effect of expressed emotions is revealed. However, they do not guarantee to find a small, sufficient and optimal subset of body cues for interpretation or classification purpose. Besides, Feature Selection approach based on statistical measures ignores body cues dependencies.

In machine learning research community, several feature selection techniques are proposed to select an optimal subset of body cues that better contribute to the discrimination between emotions expressed in body movement than statistical methods. We follow this approach. In our work, we aim to start from a large set of features that describe the whole body movement and to identify a comprehensive set of relevant kinematic features for the characterization of emotional body expression across different daily actions. In the next sections, we will focus on the different approaches that one can use to select the most relevant features. Based on the nature of the features resulted from the process of data dimensionality reduction, the approaches that aim to reduce data dimensionality can be categorized into two categories; feature construction approaches and feature selection approaches [Guyon and Elisseeff, 2003].

4.2 Feature construction approaches

Feature construction approaches transform data of a high dimensional space into a lower dimensional space by extracting more compact feature subsets. The mapping of a high dimensional space into a low dimensional space relies on the creation of a new space of features. The new set of features is usually built either using a linear, a non-linear or a manifold transformation of the original set of features. Non-linear and manifold data transformation approaches capture the complexity of data structure but loose the interpretability of the features. The Principal Component Analysis (PCA) is one of the most widely used linear data transformation approach. It results in the creation of a new set of features called the principal components. Each component is a linear combination of the original set of features. Intuitively, one can assume that the most important features present in the high dimensional space correspond to the features that receive high coefficient values (features that contribute to the composition of the first principal components in PCA with high coefficient values). However, PCA is blind to the classification task as it belongs to unsupervised approaches. The focus of the approach is the maximization of variances rather than the maximization of the separation between the classes. Linear Discriminant Analysis (LDA) belongs also to linear data transformation approaches, but differs from PCA as it considers the difference between the classes.

4.3 Feature subset selection approaches

Feature subset selection approaches preserve the original features and highlight the most relevant ones. They facilitate data interpretation as they preserve the meaning of a subset of the original features set.

There is a large literature on feature selection approaches. Recently, Narsky and Porter [Narsky and Porter, 2014b] proposed to summarize the process of subset feature selection in two steps; 1) the ranking of features according to their "relevance" scores, and 2) the selection of the best subset of features. The process of feature ranking consists of assigning a measure of importance to each feature [Narsky and Porter, 2014b]. *This measure is usually nonnegative, with a large value indicating an important variable* [Narsky and Porter, 2014b]. The features are ranked according to their relevance measures to form an ordered list of features. The process of feature selection consists in the identification of a subset composed of the most relevant features [Narsky and Porter, 2014b].

4.3.1 Feature ranking

Based on the interaction between feature ranking scheme and the induction approach adopted in a given classifier, feature ranking approaches have been previously categorized along three categories; 'Filters', 'Wrappers' and 'Embedded' [Guyon and Elisseeff, 2003], [Blum and Langley, 1997] [John et al., 1994] [Kohavi and John, 1997] [Narsky and Porter, 2014b].

4.3.1.1 Filters

Filters rank features based on the intrinsic properties of the data without learning a classification model. Univariate filter methods are the simplest form of filtering scheme [Blum and Langley, 1997]. Multivariate filter methods are another form of filters.

Univariate filters :

Univariate filter approaches are based on individual feature relevance score to rank features. Individual feature relevance is evaluated based only on the intrinsic properties of the data by the calculation of simple statistics and measures, considering its correlation with the class label [Inza et al., 2004]. For instance, Fisher's criterion is a measure for variables ranking based on the ratio between class variance and within-class variance. Other data-based relevance measures are illustrated in [Guyon and Elisseeff, 2003]. Univariate filter approaches consist in measuring relevance measure for each feature individually, independently of the context of other features. Thus, ranking features based on individual feature's contribution ignore features dependencies.

Multivariate filters :

Unlike Univariate ranking methods which consider the individual feature's contribution, multivariate filter methods consider the contribution of a set of features to a class [Guyon and Elisseeff, 2003] [Saeys et al., 2007]. An example of Multivariate filters is Correlation-based feature selection (CFS) [Hall and Smith, 1999]. Multivariate filters are known to be slower than Univariate filters. However, Guyon and Elisseeff [Guyon and Elisseeff, 2003] illustrate through several examples that selecting a subset of features that together have good predictive power- considering the dependencies between features- is more reliable than selecting features according to their individual relevance score.

Both Univariate and Multivariate filters are independent from the classification model. Feature ranking based on Filters ranking can be considered as a preprocessing step to build a learning machine model [Guyon and Elisseeff, 2003]. They are mostly less time-consuming than feature ranking approaches that are more linked to the learning or the prediction of the classification model.

4.3.1.2 Wrappers

Wrapper methods are based on the prediction accuracy of a chosen classifier to evaluate the goodness of each feature [Guyon and Elisseeff, 2003], [Kohavi and John, 1997]; the more the prediction accuracy is increased when considering a given feature, the more this feature is considered as relevant. Thus, the prediction accuracy of the classifier can be used to assess a score of relevance to the features. Wrapper methods take advantage from the predictor only in term of the prediction accuracy, regardless of the algorithm of the classifier. As such, the learning model is used as a black box. Indeed, the knowledge of the algorithm of the classification is not needed [Kohavi and John, 1997]. As the prediction accuracy is used to assess the relevance of each feature, wrapper approaches can be highly time-consuming.

4.3.1.3 Embedded methods

Embedded feature ranking methods search for the optimal subset of features during the construction of the model. They refer to an implicit built-in feature ranking in a data mining algorithm. The relevance of features is estimated during the framework of learning process. Then the ranking of features is achieved using their model-based relevance measures. Several embedded methods have been proposed in previous works [Breiman, 2001],[van der Laan, 2006],[Strobl et al., 2008], [Féraud and Clérot, 2002a].

The quantification of feature relevance score in embedded approaches depends upon the induction algorithm. Recursive partitioning methods for decision trees (including ID3 [Quinlan, 1986], CART [Breiman et al., 1984] and C4.5 [Quinlan, 1993] algorithms) involve an embedded strategy for feature ranking and selection. A decision tree algorithm is built through an iterative process. The data is split depending on the value of a specific feature at each iteration. For instance, the relevance of a given feature can be estimated as the sum of gains in the Gini index over all splits based on this feature [Narsky and Porter, 2014b].

Furthermore, it is possible to measure the relevance of each feature in an ensemble method that combine several trees. Random Forests (RF) approach is a tree-based ensemble method [Breiman, 2001]. The permutation importance measure of Random Forests approach has been widely used to rank features [Genuer et al., 2010] [Pang et al., 2012] [Díaz-Uriarte and Alvarez de Andrés, 2006] and different strategies were applied to select the most relevant features based on their ranking measures [Li et al., 2004] [Svetnik et al., 2004], [Díaz-Uriarte and Alvarez de Andrés, 2006].

Recently, classical statistical models such as Support Vector Machine (SVM) [Guyon et al., 2002, Maldonado et al., 2011, Maldonado and Weber, 2011] [Rakotomamonjy, 2003] and Neural Network [Ball et al., 2002, Féraud and Clérot, 2002b] were used to extract knowledge of the features selected by the model.

The computational cost of embedded techniques vary according to the induction algorithm of the classifier, but they are usually known to be characterized with lower computational cost than wrapper techniques [Guyon and Elisseeff, 2003]. Another interesting advantage that concerns embedded methods is the ability to explore all the original features allowing selecting the most relevant ones discovered during training process. Besides, embedded feature ranking techniques are useful for the interpretation of a classification model. Indeed, “robust” classification models such as Neural Networks and Random Forest approaches prove high performances but the model is usually hard to explicit. Such models are often referred to as “black box” models. The need to explain such a classification/ regression model is sometimes as crucial as the need to obtain “robust” and stable classification/ regression results (e.g. in several scientific domains such as medical decision support). The relevance measure of features returned by such models is highly useful to interpret the predictive power of each feature. Indeed, such feedback is helpful to interpret what goes behind the stores of such “black box” models. For instance, feature relevance measures returned by a Random Forest model have been used in Bioinformatics to study the impact of different amino acid properties for protein evolution or predicting peptide binding [Strobl et al., 2008]. Díaz-Uriarte et al. [Díaz-Uriarte and Alvarez de Andrés, 2006] highlight the need to identify the most relevant genes for gene expression data analysis.

4.3.1.4 Summary

Filters measure features relevance scores based on the intrinsic properties of the data. Using Filters, noisy (irrelevant) features are removed. However, Filters do not guarantee to find a small, sufficient and optimal features subset for a given classifier. As Filters scheme is totally blind to any classifier, the effect of a subset of features selected according to Filters approaches on the performance of the classifier is totally ignored [John et al., 1994]. Filter methods are considered as a pre-processing to the learning models (as they have no interaction with the model). Wrappers or embedded methods consider both the interaction between the features and the model and the dependencies between features. The main drawback of wrapper methods is the computational costs, which make them less efficient than filter methods [Guyon and Elisseeff, 2003] [Saeys et al., 2007]. The computational cost of Embedded techniques depends upon the induction algorithm of the classifier, but it is usually lower than the computational cost of Wrapper techniques. Model-based relevance measure returned by embedded techniques is highly useful to interpret the classification/ regression model.

In practice, Wrappers and Embedded approaches are most often favored for accuracy, while Filters are mostly preferred for speed [Narsky and Porter, 2014b]. When coupled with a "strong" classifier, it has been shown that Wrappers and Embedded methods can identify relevant features more efficiently than Filters [Narsky and Porter, 2014b].

4.3.2 Feature selection

Feature ranking techniques discussed in section 4.3.1 are useful to estimate a measure of relevance for each feature allowing their ranking according to their usefulness. However, the ranking of features does not inform us on the exact subset of features that we should retain. Feature selection approaches discussed in this section aim to identify the subset of features that should be retained (and consequently the subset of features that should be eliminated).

4.3.2.1 Threshold based feature selection

In practice, any ranking feature approach can be turned into a feature selection approach by introducing a Threshold [Narsky and Porter, 2014b]. Threshold based feature selection approach is based on the decreasing ranking of features according to their relevance measures. It consists in selecting a subset of features that receives the highest relevance measures according to a specific threshold [Li et al., 2004] [Narsky and Porter, 2014b]. [Blum and Langley, 1997]. That is, only features with relevance values above this threshold are retained. The choice of this threshold can be subjective or resulted from a statistical test [Narsky and Porter, 2014b]. Threshold based feature selection mostly fail to capture the most relevant features and to eliminate redundant features [Narsky and Porter, 2014b].

4.3.2.2 Wrapper approach for feature selection

A more sophisticated feature selection approach consists in using Wrappers. When used for feature selection purpose, a Wrapper method consists of a search algorithm that is 'wrapped' around the classifier to find an optimal subset of features. The most known recursive greedy search algorithm is the Sequential forward selection (SFS) and the Sequential backward elimination (SBE) [J., 1978]. Such approaches are not necessary based on a preliminary feature ranking step. Usually, greedy search of the optimal subset of features among all the features can be very expensive computationally [Saeys et al., 2007].

4.3.2.3 Embedded method for feature ranking and Wrapper method for feature selection

Recent works proposed to combine Embedded approach for feature ranking with a Wrapper recursive approach to find the optimal subset of features [Svetnik et al., 2004], [Díaz-Uriarte and Alvarez de Andrés, 2006], [Genuer et al., 2010], [Rakotomamonjy, 2003], [Guyon et al., 2002]. Such method is called hybrid feature selection approach . When coupled with Embedded feature ranking method, Wrapper feature selection strategy mostly refers to an iterative search; at each iteration, the error of the model is computed.

Different strategies of iterative feature selection search have been distinguished previously (e.g. recursive feature selection vs non-recursive feature selection strategy) [Hapfelmeier and Ulm, 2013] [Svetnik et al., 2004] [Svetnik et al., 2004], [Díaz-Uriarte and Alvarez de Andrés, 2006] [Pang et al., 2012] [Genuer et al., 2010].

Properties of an iterative search :

- Non-recursive feature selection strategy: In non-recursive feature selection strategy, the relevance measure of each feature is calculated only once during the first training process [Svetnik et al., 2004] [Ghattas and Ben Ishak, 2008].
- Recursive feature selection strategy: In recursive feature selection strategy, the relevance measure of each feature is recalculated; at each iteration, the ranking of features is updated.
- Forward feature selection strategy: In forward feature selection strategy (also known as ascendant strategy) the features are added sequentially at each iteration [Ghattas and Ben Ishak, 2008] [Genuer et al., 2010] .
- Backward feature selection strategy: In backward feature selection strategy, the features are eliminated sequentially at each iteration [Svetnik et al., 2004] [Díaz-Uriarte and Alvarez de Andrés, 2006].

– One-by-One iterative feature selection strategy: add/ remove a single feature at each iteration [Genuer et al., 2010].
– Fraction-by-Fraction iterative feature selection strategy: add/ remove a fraction (a subset) of features at each iteration. The size of this fraction of features can be proportional to the initial set of features (e.g. 20% of the size of the initial set of features [Díaz-Uriarte and Alvarez de Andrés, 2006]) or the updated set of features at each step (e.g. the half of current features size [Svetnik et al., 2004]).

Previous works tend to use either forward feature selection [Genuer et al., 2010] or feature backward elimination [Díaz-Uriarte and Alvarez de Andrés, 2006] for non-recursive feature selection strategy. Recursive feature selection strategy is usually based on backward elimination technique.

Selection of the best subset of features :

After a number of runs of the iterative search algorithm, we obtain an averaged error of the model at each iteration. The features subset that correspond to the smallest averaged error rate is usually chosen. Díaz-Uriarte [Díaz-Uriarte and Alvarez de Andrés, 2006] propose to choose the smallest number of features with an error rate within n standard errors (s.e.) of the minimum error rate across all the iterations. The commonly used rules to choose the best subset of feature are "0 s.e." and "1 s.e.", meaning that n is usually equal to 0 or 1. These rules are used to define how far the selected solution can be from the minimal error solution [Díaz-Uriarte and Alvarez de Andrés, 2006]. Using “0 s.e.” rule is equivalent to select the subset of features corresponding to the smallest error rate across all the iterations. Using “1 s.e.” rule (also called the one standard error rule [Pang et al., 2012]) is equivalent to select the smallest model (the smallest subset of features) with an error less than the minimal OOB error augmented by its empirical standard error [Genuer et al., 2010]. Using the “1 s.e.” rule often tends to result in smaller subset of feature [Díaz-Uriarte and Alvarez de Andrés, 2006] [Genuer et al., 2010] [Pang et al., 2012] and it is known as a classical trick to deal with instability when searching for the smallest averaged error rate [Genuer et al., 2010].

Iterative approaches used in previous works :

Guyon et al. [Guyon et al., 2002] proposed to use the weights of the features in the SVM-formulation to discard features with small weights using Recursive Feature Elimination (RFE) strategy. Rakotomamonjy [Rakotomamonjy, 2003] proposed similar framework as an extension of the work described in [Guyon et al., 2002]. Ball et al. [Ball et al., 2002] used the weights ranking of the input masses in a neural network classifier along with a stepwise procedure in order to identify the most relevant masses that contribute to the prediction of tumor grade.

Díaz-Uriarte et al. [Díaz-Uriarte and Alvarez de Andrés, 2006] used non-recursive backward elimination strategy to select features based on permutation importance measure returned by Random Forest model. At each step, they eliminate 20% of the features that have the smallest permutation importance measure and they retrain the new forest using the remaining features. The set of features that lead to the smaller error rate is chosen as the best optimal feature set. Although they found promising results with RF comparing with other approaches, the impact of highly correlated features on their feature selection approach was not investigated. Pang et al. [Pang et al., 2012] developed a similar iterative feature elimination method based on a random survival forest to identify a set of prognostic genes.

Guner et al. [Genuer et al., 2010] used RF to built a feature selection approach based on four successive steps; 1) feature Ranking: Rank the features in a descending order, 2) feature elimination: Filter the features (eliminate irrelevant features) based on the average of their relevance scores and the corresponding standard deviation across 50 runs of the model, 3) feature selection for interpretation: Select a ranked list of features based non-recursive forward feature selection algorithm, 4) feature selection for prediction: Apply an ascending stepwise algorithm to select an optimal subset of features in term of the predictor. During the fourth step (ascending stepwise algorithm), the algorithm starts with the most important features based on the list of features obtained from step 3), but unlike the third step (non-recursive feature selection algorithm), the features are added only if the resulted error rate decrease (that is, if the error rate gain exceeds a threshold).

Svetnik et al. [Svetnik et al., 2004] compared the performance of recursive and non-recursive wrapper feature reduction algorithm used to select a subset of features in order to study the evolution of the error rate in terms of the number of features. The wrapper algorithms were both based on a backward elimination of features based on their permutation measure, where half of the features that receive the lowest permutation measure is eliminated at each iteration. The authors showed that the error rate increases with the reduction of features from a particular number of features. But most importantly, they showed that the non-recursive wrapper algorithm performs better than the recursive one, saying that recursive approach is more prone to over-fitting.

Pseudo code of the iterative search of the optimal subset of features :

Despite the different criteria that can be chosen to build an iterative feature selection approach, the resulted approaches are actually similar in such a way that a pseudo code can abstract the principal methodology to summarize the successive steps [Hapfelmeier and Ulm, 2013]. The following pseudo-code adopted from [Hapfelmeier and Ulm, 2013] can be used to summarize the main steps of an iterative feature selection approach:

1. Build the model using a) the whole set of features or b) the most relevant features and assess the prediction error (usually cross-validation error).

2. Compute the importance measures of features.

3. (a) Reject a subset of least important features (backward approach) or b) add a subset of less important features (forward approach), and refit the model. A subset may contain one or more features.

4. Assess the prediction error (usually cross-validation error).
5. Return to (a) step 2 (for recursive approach) or (b) step 3 (for a non-recursive approach) until no further features can be a) rejected / b) added.
6. Choose the model with (a) the lowest error or (b) the sparsest model with an error within n standard errors to the lowest error (e.g. according to the 1 s.e. rule [Díaz-Uriarte and Alvarez de Andrés, 2006]) .
7. Often the preceding steps are based on averaged findings to achieve higher stability. Therefore, steps 1–5 can optionally be repeated separately, in conjunction and within cross-validation runs. Another approach consists of repeating separately step 1-2 and step 4 to get respectively the average of relevance scores across several runs and the average of error rate across several runs [Genuer et al., 2010].

When using Random Forest as an embedded feature ranking approach, an important criterion is the error rate used for evaluation purpose: OOB-error (Out-Of-Bag) [Díaz-Uriarte and Alvarez de Andrés, 2006] [Genuer et al., 2010] or cross-validation error [Svetnik et al., 2004] can be used for this purpose.

4.3.2.4 Ensemble feature selection approaches

Another growing interest has been shown for ensemble feature selection methods which are based on the same idea of ensemble learning for classification [Saeys et al., 2008]. The reason for this interest is the need for a stable subset of features (that is not dependent on the change of the database neither on the feature selection technique). The authors in [Saeys et al., 2008] argued that a feature subset selected from a single feature selection technique may be not stable to the change in the database, and may not correspond to the same feature subset selected by another feature selection technique. Saeys et al. [Saeys et al., 2008] showed that the selection of a subset feature by aggregating the results of several versions of the same feature selection method leads to a more stable feature subset than the one created using a single feature selection method. However, only subtle changes were found for Random forests method, which is considered as an ensemble method by itself. The latter finding highlights the stability of Random Forest based feature selection.

4.3.2.5 Summary

Feature selection methods have been categorized into 4 approaches; 1) Threshold based feature selection, 2) Wrapper approach 3) Hybrid approach combining Embedded feature ranking and Wrapper feature selection, and 4) ensemble feature selection.

The simplest form of feature selection approaches can be reduced to the application of a threshold on features relevance measures. However, this approach does not result in an optimal subset of features. Wrapper approaches are more efficient but they suffer from a high computational time. Recently, several works proposed to combine embedded feature ranking approach with wrapper feature selection approach. This combination refer to an hybrid approach based on an iterative search of the optimal subset of features. Although different criteria were adopted to define the iterative search, they can all be summarized into a common pseudo-code. Hybrid approaches have been known to be highly efficient to capture an optimal subset of the most relevant features. They have been applied on several bioinformatics applications, in particular using Random Forest approach. Hybrid approaches can also suffer from high computational time as they imply a wrapper approach. However, they rely on the classification model (through embedded feature ranking) more closely than wrapper approaches applied on the initial set of features. Finally, other works highlighted the need to apply ensemble feature selection approach to obtain a stable subset of features. However, an ensemble of Random Forests has shown only subtle improvements as a single Random Forest method is categorized as an ensemble method.

4.4 Conclusion

The reduction of data dimensionality can be carried out to achieve several goals such as to reduce CPU time and memory requirements, to improve the prediction accuracy for a classifier sensitive to noisy features and to facilitate the interpretation or the visualization of the analysis. In our work, the reduction of data dimentionality is useful to identify body features that capture the most important characteristics of emotional body expression.

We have distinguished feature construction from feature subset selection approaches for the purpose of data dimensionality reduction. Feature construction methods extract new set of features, whereas feature subset selection techniques select a subset of the original features. Feature construction approaches are used to reduce data dimensionality when the hold of feature's meaning is not of concern, as they consists of constructing a new (and less meaningful) subset of features. In our work, we aim to classify expressive body movements based on explicit body cues. Feature subset selection approaches facilitate data interpretation as they preserve the meaning of a subset of the original features set. Thus, they are more suitable for our work.

Starting from a large set of features, we aim to identify the most relevant features that contribute the best to the classification of expressed emotions in each action. Our goal is not restricted to selecting the best optimal subset of features for classification purpose. Instead, our goal consists in exploring the most relevant features to consider during the learning process of a given classifier. As such, hybrid approaches combining Embedded feature ranking and Wrapper feature selection are the most suitable for our work. Besides, it has been shown in several works that Random Forest approach provide an optimal and a stable subset of features. Thus, we adopt Random Forest approach in our work to identify the subset of the most relevant expressive body cues. Based on the relevance measure returned by the model, we are also able to discuss the usefulness of each selected feature.

In this chapter, we provided an overview of Feature selection techniques in machine learning community. The reader is invited to read [Narsky and Porter, 2014b] and [Guyon and Elisseeff, 2003] for more detailed review of feature selection approaches.

4.5 Summary of chapter

- **Feature construction vs feature selection:** Feature construction approaches are used to reduce data dimensionality when the hold of feature's meaning is not of concern. As we prefer to preserve the feature meaning, we opt for feature selection approach.
- **Feature subset selection:** Feature subset selection can be summarized into two steps; 1) Ranking of features according to their relevance measure, and 2) Selection of the most relevant features.
- **Feature Ranking:** According to the dependency between the features and the learning models, feature ranking methods have been categorized into three approaches; 1) Filter, 2) Wrapper and 3) Embedded. Their measure of relevance is respectively based on 1) intrinsic properties of the data, 2) prediction accuracy and finally 3) model-based criteria. Filters are favored for speed, but they are totally blind to any classifier. Wrapper or Embedded feature selection approaches allows identifying a good subset of feature for a given classifier. Unlike filter methods, wrapper techniques integrate the prediction of the classifier in the process of feature selection scheme through a specific search algorithm. Embedded approaches connect feature subset selection process more closely to the learning algorithms.
- **Feature Selection:** Feature selection methods have been categorized into 4 approaches; 1) Threshold-based feature selection, 2) Wrapper approaches, 3) Hybrid approach combining Embedded feature ranking and Wrapper feature selection, and 4) ensemble feature selection. Wrapper approach is more efficient than Threshold based approach. Hybrid approach combines more closely the feature ranking to the classification model and feature selection to the classification performance. Ensemble feature selection approach is preferred to obtain stable results of feature selection, except for ensemble methods (such as Random Forest method).

5 Perception and characterization of bodily expression of emotion

The work of Darwin [Darwin, 1872] is one of the most prominent works on the expression of emotion in man and animals. In his work [Darwin, 1872], Darwin described the body postures and movements that occur in response to some specific emotional states like Joy, Sadness, Pride and Shame. However, early studies on emotion expression were more focused on vocal and facial expression than on bodily expression of emotions. It has been widely assumed that the face is the principal channel to express emotion [Ekman and Friesen, 1978] and that body movements and postures can only provide information about the affect dimension or the intensity of the experienced emotion [J.A. Harrigan, 2005]. An exception has been made for hand gestures assuming that " *hand cues, much like facial cues, might permit perception of the nature of the emotion*" in [Ekman and Friesen, 1967]. In the last two decades, there has been a growing interest in studying the expression of emotion through body movement and posture. Several studies in neuroscience and psychology have shown the importance of body expressions in nonverbal communication of emotion and affect.

Preliminary works on emotional body expression were focused on perceptual experiments to explore the human ability to recognize emotions from body movement. Recently, several works have been focused on automatic analysis to recognize emotions. Researches have been also widely turned toward one critical question; which body cues may be related to the recognition and the characterization of emotion expression? To be able to answer this question, previous researches were mainly based on explicit body cues (predefined body cues).

The structure of this chapter is represented in Figure 5.1. In section 5.1, we discuss the different contexts used in previous works to study the communication of emotions in body movements. In section 5.2 we highlight the differences between the expression of emotions in body posture and in body movement. Then, section 5.3 focuses on previous perceptual studies of emotional body expression recognition and characterization. The following section 5.4 is devoted to the discussion of previous works on automatic analyses of emotional body expression recognition and characterization. In both perceptual and automatic analyses studies, we distinguish studies that focus on affective posture and those which focus on affective body movement. Finally, sections 5.5 and 5.6 respectively conclude and summarize this chapter.

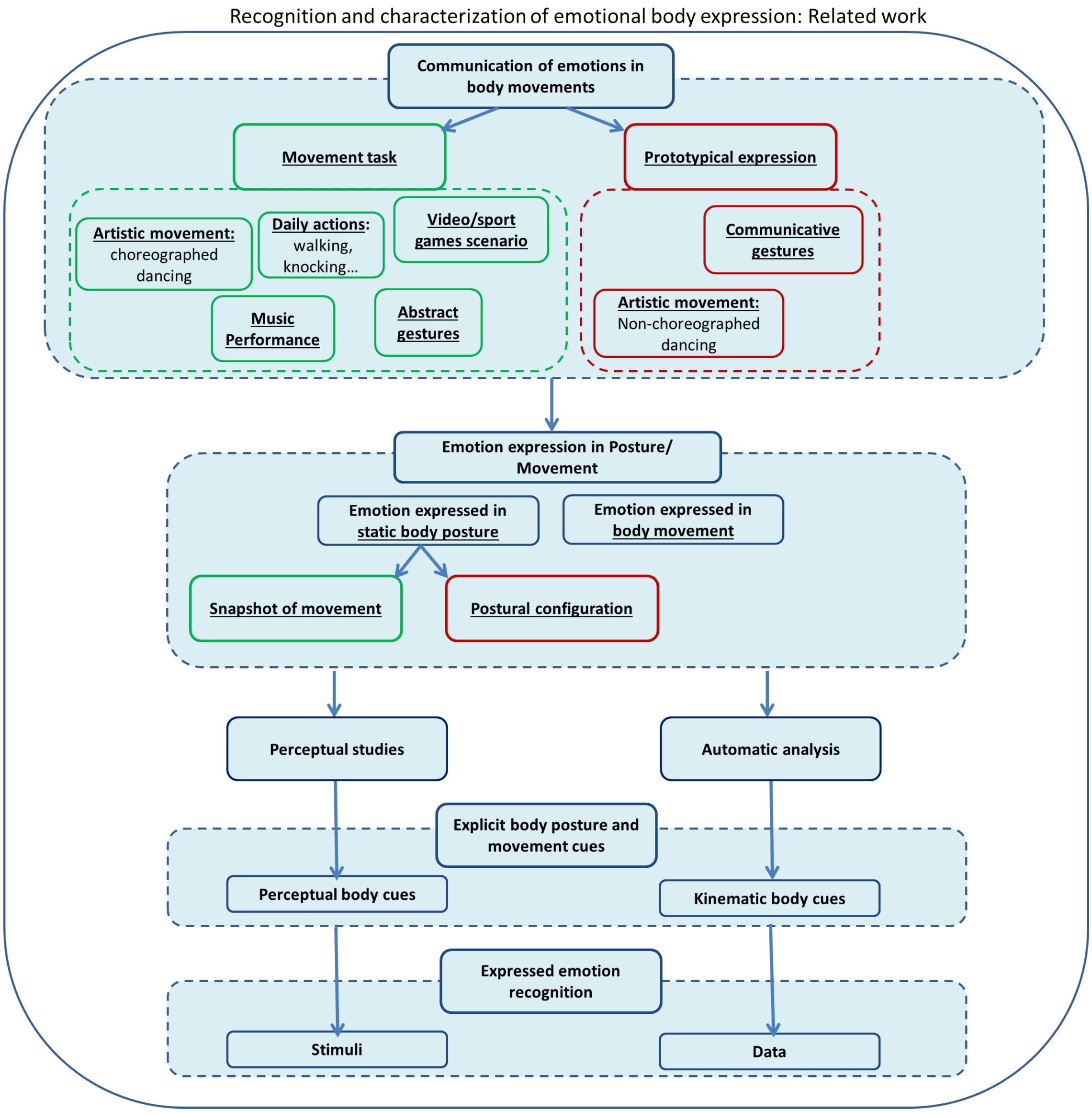


Figure 5.1: Related works on the recognition and the characterization of emotional body expression.

5.1 Communication of emotions in body movements

In this section, we provide an overview of previous studies that explore the communication of emotions in body movements.

The communication of emotions through body movement has been mainly studied from two perspectives: 1) the "**explicit**" prototypical emotion expression and 2) the "**implicit**" expression of emotion during a specific movement task.

"**Explicit**" expression of emotions is directly communicated trough gestures dedicated to convey affects and emotions such as non-choreographed dancing [Park et al., 2004] or communicative gestures that infer affective states (e.g. rising arms to express joy) [Bänziger et al., 2012]. Such communicative gestures can be selected from "a library of movement types" [Karg and Samadani, 2013]. Unlike choreographed dancing, the change of non-choreographed dancing is obtained by the dancers using artistic communicative gestures.

"**Implicit**" expression of emotions is indirectly communicated in body movement during a movement task that do not convey emotions by itself (e.g. walking). The implicit expression of emotion has been studied in different contexts:

– Artistic movement such as choreographed dancing [Camurri et al., 2003], video/ sport game scenario [Kleinsmith et al., 2011] [Savva et al., 2012].
– Music performance such as piano, bassoon and saxophone performances [Dahl and Friberg, 2007, Castellano et al., 2008].
– Abstract gestures such raising and lowering the arms in the coronal plane [Castellano et al., 2007]. Such gestures are not used to accomplish a specific task as daily actions. They also do not communicate any explicit meaning. They are only used to explore the effect of expressiveness on body motion [Karg and Samadani, 2013].
– Daily actions such as walking [Crane and Gross, 2007] and knocking [Gross et al., 2010].
– Video/sport games scenario [Savva et al., 2012] [Kleinsmith et al., 2011]. The study of bodily expression of emotions during a video/sport games scenario is a successful alternative to capture spontaneous and natural emotion expression.

In Implicit expression of emotions studies, the main focus is about the effect of the expression of emotion on body movement; that is, how the expression of emotion modulates body movements [Karg and Samadani, 2013]. For instance, choreographed dancing are based on the same type of movement defined by the choreography. Thus, the modifications of the movement is achieved by the change in the expressivity of the movement (e.g. Joy is expressed with large and fast movements).

5.2 Emotion expression in static body posture

The study of emotion expression in static body posture can rely on different definitions of body posture:

- A particular configuration of body posture [Tracy and Robins, 2007, Coulson, 2004]: body posture may refer to a particular configuration of body structure according to specific measures of joint angles as in [Coulson, 2004] or according to a prototypical affect expressive posture (e.g. postural expression of Pride [Tracy and Robins, 2007]).
- Snapshot of prototypical expression of emotions: such as the "peak" of emotion expression during prototypical expression [Atkinson et al., 2004].
- Snapshot of emotion expression during a specific movement task: such as the instant of an animation sequence for which the body posture is "the most expressive" [Kleinsmith et al., 2011].

5.3 Perceptual recognition and characterization of emotions

The study of explicit and implicit bodily expression of emotions was primarily based on the perception of human observers. Several perceptual studies have shown that human observers can readily recognize emotions from the perception of expressive bodily behavior [Atkinson et al., 2004], [Pollick et al., 2001], [Wallbott, 1998], [Meijer, 1989]. The confusions of emotions recognition (e.g. perceiving Shame expression as Sadness) in perceptual studies are sometimes explained by a similarity of their movement characteristics (e.g. both are characterized with slow movement). Indeed, the human ability to perceive emotions expressed in body movement and postures encouraged the researchers to look for the body cues supporting these perceptual capabilities.

Perceptual experiments consist in asking, individually, a group of participants (ie. observers) their judgment with regard to particular stimuli. In the context of perceptual recognition of emotion, the participants are mostly asked to recognize the emotion expressed in a number of stimuli. A stimulus may refer to a snapshot/ movie of a video recording (See Figure 5.5) or a snapshot/ movie of a motion capture recording (See Figure 5.2). The perceptual recognition of emotions can be achieved through different procedures such as a forced-choice paradigm (selecting a verbal label of emotion from a set of labels) and an agreement scale paradigm (e.g. rating the presence of an emotion on a 5-point scale graduated from "strongly disagree" to "strongly agree"). In the following subsections, we discuss previous works on the recognition and the characterization of emotion expression achieved through perceptual experiments. In section 5.4, we will tackle the recognition and the characterization of emotion expression achieved through automatic analyses.

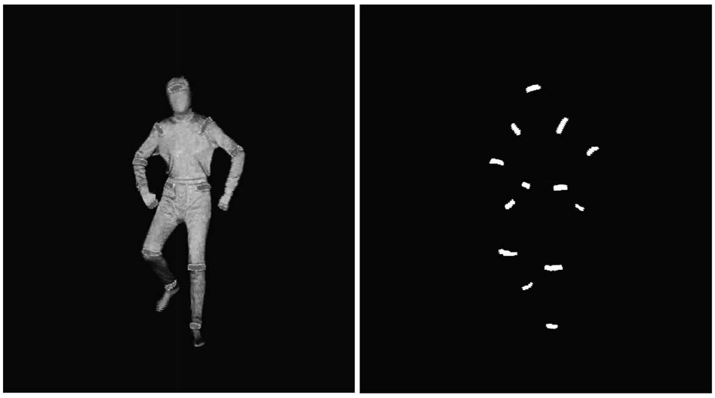


Figure 5.2: Examples of (a) Full Light and (b) Point Light expression of anger corresponding to the peak of the expression (screenshot from[Atkinson et al., 2004])

5.3.1 Emotion expression in body posture

In this section, we discuss the studies that explored the expression of emotions in body posture. Expressive body posture refers to a snapshot of a continuous movement (See Figure 5.2 and 5.5) or to a particular configuration of body segments (See Figure 5.4).

Snapshots of movement :

Atkinson et al. [Atkinson et al., 2004] conducted a perceptual study for emotion recognition from postural body expression. Participants were asked to recognize the emotion expresed in Full Light and Point Light stills stimuli (See Figure 5.2). These stimuli were selected from Full Light and Point Light movies of explicit prototypical expression of emotions. Each stimulus corresponds to the "*peak*" of emotional expression that occur during a sequence of prototypical expression. These stimuli display the expression of 5 emotions (disgust, fear, anger, happiness and sadness) portrayed by 10 actors. The participants were asked to select, among these 4 emotions, the one that is best represented in the stimuli. Based on the human perception of expressive postures, the sadness and happiness were the most easily recognized emotion. It has been shown that postural body expression of disgust was the least recognized. Sadness postural expression was mainly characterized with downward head flexion, crossing/bringing arms in front of the body and bring hands to the face.

Postural configuration :

Stock et al. [Van Den Stock et al., 2007] collected a set of emotional body posture explicitly acted by actors (See Figure 5.3). A perceptual study based on "body expression-matching task" to recognize emotions from the whole body postural expression was presented in their work. Unlike [Atkinson et al., 2004], the aim of

Figure 5.3: An example of stimuli used in [Van Den Stock et al., 2007]: Angry expression on top (reference), Angry expression on the bottom right, and a Sad expression on the bottom left (screenshot from [Van Den Stock et al., 2007])

"body expression-matching task' was to explore the recognition of emotion on the basis of similarity of expressive body posture without making use of verbal labels of emotions (e.g. sadness) (See Figure 5.3). In front of three pictures, the participants were asked to choose the picture that contains the most similar expressive posture picture comparing to a reference picture (See Figure 5.3). The emotions expressed in body posture were anger, fear, happiness and sadness. Similarly to what was found in [Atkinson et al., 2004], Sadness and happiness were the most correctly recognized. It was also observed that fearful postural expression was the least recognized.

Coulson [Coulson, 2004] adopted a different perceptual experiment of emotion recognition from body posture. Based on the psychological descriptions of postural expression of the 6 basic emotions (anger, disgust, fear, happiness, sadness, and surprise), Coulson [Coulson, 2004] generated a number of computer mannequin postures. Each computer mannequin posture was generated according to a particular configuration of 13 joint angles (e.g. 25 ° of downward head flexion...): head/neck, chest, abdomen, two shoulders/upper arms and two forearms, two thighs, two shins and two feet) (See Figure 5.4). A perceptual study based on a forced choice methodology was conducted to explore the recognition of emotions expressed in static body postures. Although the generation of stimuli is different from [Van Den Stock et al., 2007] and [Atkinson et al., 2004], Coulson found similar recognition results: Sadness and happiness, were the most recognized emotions. Only few postures were correctly identified as expressive postures of the disgust emotion.

Unlike [Van Den Stock et al., 2007] and [Atkinson et al., 2004], Coulson [Coulson, 2004] also explored the effect of three dimensional stimuli view on the perception of emotion from body posture. Thus, each configuration was rendered according to three viewpoints: front, side and rear (See Figure 5.4). This work is based on the assumption that our perception of emotions expressed in body postures depends

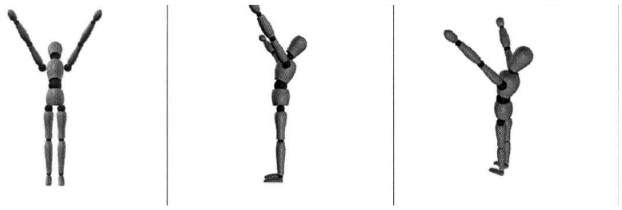


Figure 5.4: An example of expressive posture configuration generated in [Coulson, 2004]: front viewpoint on the left, side viewpoint in the middle, and rear viewpoint on the right (screenshot from [Coulson, 2004])

upon the viewpoint of the body posture [Daems and Verfaillie, 1999]. On one hand, it was found that the viewpoint of expressive postures mostly did not has significant affect on the recognition task. On the other hand, Sadness was the best recognized from the side viewpoint due to the bending configuration of chest and head.

The expressive postural configurations proposed by Coulson [Coulson, 2004] were based on symmetrical arms posture. This symmetry criterion was introduced to simplify the configuration of expressive postures [Coulson, 2004]. However, other researches [Dael et al., 2011] have shown that the symmetry of the arms is an important postural feature in the study of bodily expression of emotion.

Tracy and Robins [Tracy and Robins, 2007] focused on the expression of Pride in body posture. They showed that the emotion of pride can be readily recognized and distinguished from the emotion of joy in bodily expression rather than in facial expression. Their study is based on a set of explicit prototypical postural expression of Pride. They suggest that among the set of postural features that suits the expression of pride through body posture, the best one corresponds to head tilted slightly backward, arms akimbo with hands on hips, and expanded arms posture with a small smile. Tracy and Robins [Tracy and Robins, 2007] point out that at least one of these postural features must occur (with a small smile) to recognize pride in body posture.

As such, these studies report that sadness, happiness and pride are the most recognized from body posture based on postural cues that explicitly communicate emotion. Other studies focused on the perceptual recognition of emotions in body movements. These studies will be discussed in the following subsections.

5.3.2 Emotion expression in body movement

Recent studies clearly demonstrate that variations in body movement, manifested by body actions and postural changes, can convey further information about the affective state [Atkinson et al., 2004], [Dael et al., 2011]. In a comparative study by Atkinson et al. [Atkinson et al., 2004], it was found that participants achieve better scores of emotion recognition when judging video clips than when judging static images.

Prototypical expressions :

Several studies explored the perceptual recognition and characterization of emotions expressed in body movements. In addition to the recognition and characterization of expressive posture discussed in Section 5.3.1, Atkinson et al. also explored the recognition and characterization of emotions expressed in body movement [Atkinson et al., 2004]. It was observed that body expressions of Fear were characterized by backward whole body movement. Angry expression was characterized by forward whole body movement with expansive and aggressive movement. Bodily expression of Fear and Disgust shared some common actions: rising hands towards the head. While such an action involved covering and holding the head with hands when expressing fear, it involved touching the face when expressing disgust. Anger and Happiness expressions in body movement were confused by observers in recognition task [Atkinson et al., 2004]. The authors claimed that this is due to the fact that both Anger and Happiness are characterized by expansive and energetic movement. Two specific types of movement including raising arms and shaking the fists were also presented in Happy and Angry bodily expression.

Dael et al. [Dael et al., 2011] also examined the expression of emotion in prototypical expression of emotions. Their work was based on GEMEP database and on a set of body action and posture units defined in BAP coding system (Body Action and Posture) [Dael et al., 2012]. Figure 5.5 shows an example of Elated Joy portraying taken from GEMEP database. The authors studied 12 different emotions: Elated Joy, Amusement, Pride, Pleasure, Relief, Interest, Hot Anger, Panic ear, Despair, Cold Anger, Anxiety and Sadness. Sixteen principal behavior factors were judged to be the most relevant physical cues in the expression of emotion through the body. They include specifics patterns of actions and postures related to the arms, the head, the neck, knee and the whole body. The authors [Dael et al., 2011] highlighted several findings that we summarize as follows:

- The actors tend to express hot anger by forward body lean, amusement by touching or manipulation, discontinuous movements, head laterally straight and oriented toward the interlocutor.
- Pleasure portrayals were characterized with head tilted up and averted and asymmetrical arm actions.
- Sadness and Relief were regrouped into one cluster despite their differences along the valence dimension; Sadness is considered as a negative emotion whereas Relief is considered as a positive one. The common point between the two emotions is the absence of goal and thus the absence of significant activities. Sadness and Relief were both characterized with arms at rest and hands in pocket. The authors proposed to label the group “Sadness+Relief” as disengagement or resignation [Dael et al., 2011].
- No distinctive pattern was concluded for despair and Anxiety.
- Backward body lean or movement combined with upward gaze and lateral

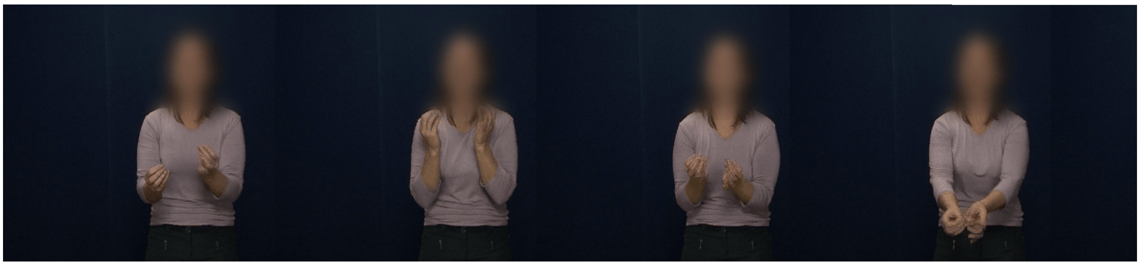


Figure 5.5: Snapshots from Elated Joy expression from GEMEP database [Bänziger et al., 2012], [Dael et al., 2012]

trunk lean characterize Anxiety expression as well as Panic Fear.

- When encoding Elated Joy (and sometimes Pride), symmetrical up-down repetitive arm actions were frequently used (See Figure 5.5). This is aligned with the assumption of Frijda [Frijda, 1986] saying that "*Movements orthogonal to the direction of locomotion are conspicuous in joyful behavior as in excited behavior*" (page 37).
- Confusions were observed in the portrayals used to encode Elated Joy and Panic Fear (symmetrical arm action and knee movement). This was observed also in the work of Walbott [Wallbott, 1998].
- Interest was encoded several times by forward trunk lean and facing head orientation.
- The forward lean that characterizes Hot Anger was also aligned with the definition of angry actions mentioned in the work of Frijda [Frijda, 1986] (page 31) as being *directed toward the offender or toward objects that are likely as offenders*.

Wallbott and Scherer [Wallbott and Scherer, 1986] defined a set of body cues to characterize prototypical bodily expression of emotions. The body cues are the number of hand movements, head orientation and mean frequency some qualitative measures (fast, expansive, energetic, active or pleasant movement behavior). The emotions were: joy, sadness, anger and surprise. 12 students were asked to watch the videotapes of recorded portrayals and to evaluate the emotional body expression using the proposed body cues. Sadness was characterized highly different from the other emotions in term of downward/averted head orientation and frequent hand movement such as the shrugs (indicating helplessness). The lack of energy, expansiveness and dynamism were also typical for sadness. The feature of pleasantness was observed as a critical feature that discriminate between positive and negative emotions. Downward head orientation received the highest values for sadness. Most expansive movement was judged for surprise, followed by anger and joy.

In another study by Wallbott [Wallbott, 1998], the perceptual characterization of prototypical bodily expression of 14 emotions was explored: elated joy, happiness, sadness, despair, fear, terror, cold anger, hot anger, disgust, contempt, shame, guilt,

5

pride, and boredom. The actors enacted the expression of these emotions in body movement through communicate gestures while pronouncing “standard sentences” [Wallbott, 1998]. Trained participants were asked to indicate perceived body cues of prototypical expressions in a free format. Then a subset of body cues that received high agreement is kept.

On one hand, it was observed in [Wallbott, 1998] that, some features were either typical or mostly frequent for some emotions like crossing arms for pride, self-manipulators for shame,upward arm movement for elated joy, pointing movement for hot anger and shoulders backward for disgust. On the other hand, some features seemed to characterize some groups of emotions rather than one single category; the collapsed upper body posture was present mainly when experiencing negative emotions like shame, sadness and boredom, backward head movement was observed for pride, elated joy and disgust, lifting shoulders for elated joy and hot anger, self-manipulators for shame and fear, and finally illustrators, movement activity, dynamics of movement and high expansiveness occurred mainly when experiencing elated joy and hot anger. A discriminant analysis on the set of proposed features showed that some patterns of movement features can be assigned to some emotions despite the confusions that occur between some emotions. The confusions didn’t occur only between emotions of same valence (e.g. between happiness and pride), but also between distinctive emotional states along the valence dimension (e.g. terror and elated joy). Wallbott [Wallbott, 1998] highlighted the importance of the upper body parts, particularly hand and arm postures and movement in the distinction between emotions.

Emotion expression during walking :

In addition to the work that focused on prototypical expression of emotions as discussed above, several studies explored the recognition and the characterization of bodily expression of emotions in walking action. Walking movement is a constrained motor task that involves a repetitive and a symmetric pattern. The expression of emotion in walking action is mostly manifested through the modulation of walking pattern according to the expressiveness of the motion. It has been shown in several studies that humans are able to recognize emotions expressed in walking pattern [Montepare et al., 1999] [Hicheur et al., 2013] [Montepare et al., 1987] [Karg et al., 2010] [Crane and Gross, 2007]. Studies also looked at body cues that characterize the expressiveness of walking pattern (e.g. foot stride, arms swing...).

Montepare [Montepare et al., 1999] conducted two perceptual studies based on the videotaped displays in order to explore the recognition and the characterization of emotion expression during walking pattern. The emotions were joy, sadness, anger and neutrality. The first group of observers was asked to identify the category of emotion from the perceived videotaped displays. The second group of observers was asked to rate six body movement cues when watching the different video clips using 7-point scale: 1) very smooth-very jerky, 2) very stiff-very loose, 3) very soft-very hard, 4) very slow-very fast, 5) very expanded-very contracted, 6) almost no action-a lot of action. The participants recognized the portrayed emotions significantly above chance levels. Neutral and Angry expressions were better recognized than Sad and Happy expressions. To some extent, Angry expressions were confused with Happy expressions and Sadness expressions were confused with Neutral expressions. Happy and Angry movements were both characterized as being full of actions (high quantity of movement), jerky (with high intensity for angry displays), fast, hard and expanded. Happy movements were however perceived as being looser than angry movement. Displays of sadness were observed to be very smooth, loose, slow, soft, contracted, and lacking in action. Neutral displays were characterized similarly to Sad displays. Indeed, it turns out that the emotions that were confused during the recognition task are characterized with similar patterns of body cues.

In another study by Montepare et al. [Montepare et al., 1987], the authors explored the perception and the characterization of four emotions expressed in gait (sadness, happiness, anger and proud). They asked 10 participants to identify the emotion from several video clips acted by five actors and to rate 4 body cues using a 7-point bipolar scales: 1) amount of arms swing, 2) strides length, 3) heavy-footedness and 4) posture straightness. The authors observed that most of the misidentification of proud and happy expressions was related to the performance of two actors. However, sadness was identified by the participants in most of the trials. It was also observed that angry gaits were characterized by a high level of heavy-footedness. The use of long strides was rated as higher in angry and proud gaits than in happy gaits. The sad displays have received the lowest rate of the use of long strides and arms swinging. The straightness of the body was rated higher in proud gaits.

Emotion expression during knocking :

Knocking action is another constrained motor task. It has been shown that emotion expression can be conveyed in knocking motion [Gross et al., 2010] [Pollick et al., 2001].

In order to study the effect of emotions on the form and the dynamic of movement during knocking action, Gross et al. [Gross et al., 2010] proposed to use a simplified version of the Effort-Shape analysis (derived from the Laban movement analysis) to better understand expressive body movement. The objective of their experiment was to assess the effects of 6 emotions on body movement: Anger, Anxiety, Sadness, Joy, Proud and Contentment. Their principal aim was to investigate how much the Effort-Shape analysis could characterize the different emotional states. 31 participants were asked to watch video clips performed by six actors displaying emotion expression in knocking action. They were also asked to rate the movement quality according to 6 factors, each composed of two opposite qualities. Two factors (one for the torso and one for the limb) were used to describe the shape of the body. Four factors were used to describe the Effort component: flow (bound/free), weight/energy (forceful/light), time (sustained/ quick) and space (indirect/direct).

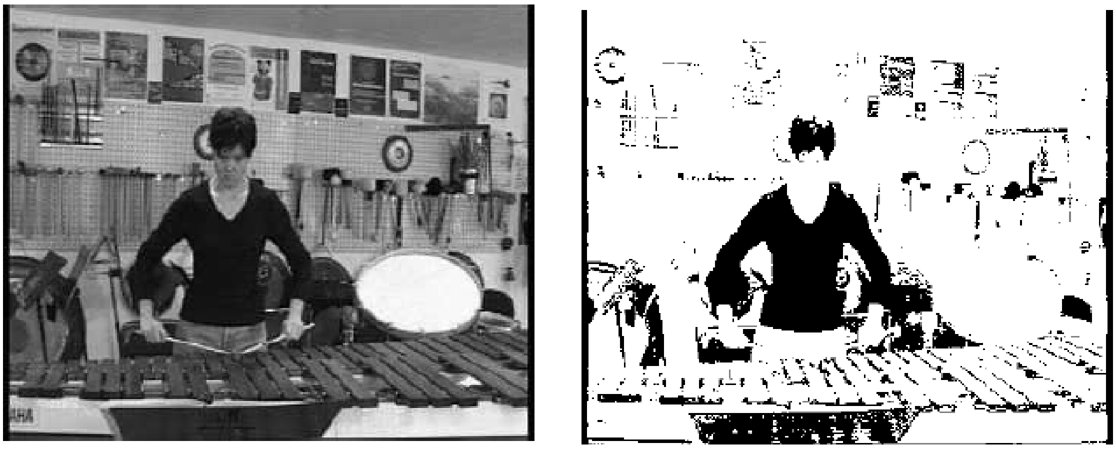


Figure 5.6: Original (on the left) and filtered (on the right) videos images [Dahl and Friberg, 2007] (screenshot adapted from [Dahl and Friberg, 2007])

The authors observed that each of the emotional trials group can be associated with a unique set of factors. While angry movements were characterized by forceful movement and expanded limbs movement, joyful movements were expanded and free. Sad movements were observed to be sustained and bound, and anxious movements were similarly characterized. Proud and content, which are low activation emotions, were characterized by opposite patterns on the energy and space factors with stronger and indirect movement for proud.

Gross et al. [Gross et al., 2010] also asked the participants to recognize the emotion expressed in knocking movement. They found that negative emotions (anger, anxiety, sadness) are better recognized than positive emotions (pride, joy and contentment). Overall, Anger expression in knocking action was the best recognized.

Emotion expression during music performance :

In addition to the perceptual studies that focus on the expression of emotions in communicative gestures, and in daily actions (e.g. walking, knocking), other perceptual studies turned their attention to the expression of emotions during music performance. It has been shown in several studies that expressing emotions in body movement can affect the way by which we play on musical instruments (e.g. piano performance) [Dahl and Friberg, 2007] [Castellano et al., 2008].

Based on a serial of perceptual studies, Dahl and Friberg [Dahl and Friberg, 2007] tried to characterize the expression of 4 emotions in music performance: Happiness, Sadness, Anger and Fear. For this purpose, they proposed four movement cues: the amount (none/large), the speed (slow/fast), the fluency (jerky/smooth) and the regularity (irregular/regular) of body movement. In each perceptual study, the stimuli consist in video clips for which the content was modified through digital modifications to avoid the bias of the facial expressions and the environment (See

Figure 5.6.

In their first perceptual study [Dahl and Friberg, 2007], 20 participants were asked to observe video clips of a professional percussionist playing on her instrument and to rate both the emotional content and movement cues on a scale from 0 to 6. Fear was always hardly recognized and characterized. Observers tended to attribute similar level of movement cues for Happiness and Anger: large, fast and jerky description of the movement with a particular high level of the latter cues related to angry performances. The body movements of Sad performances were, however, described as small, smooth, slow and regular.

Their second perceptual study [Dahl and Friberg, 2007] was conducted with two professional woodwind players: a bassoon player and a saxophonist. Fear was not successfully communicated to the observers. Dahl and Friberg [Dahl and Friberg, 2007] observed that the participants tend to correlate the Angry intentions with a jerky movement.

Emotion expression in abstract gestures :

The perceptual studies of emotion expression in body movement were not limited to the prototypical expression and specific movement tasks. Indeed, other researchers explored the human ability to recognize emotions in abstract gestures (abstract gestures were introduced in section 5.1).

Meijer [Meijer, 1989] supposed that general movement features might describe the expressive value of body movements in abstract gestures. In order to study the contribution of general features to the attribution of emotion, participants were asked to rate 7 features based on the observation of video clips performed by three actors in 12 different emotional states [Meijer, 1989]. The features used in his study were: the trunk movement (stretching/bowing), arm movement (opening/ closing), vertical direction (upward/ downward), sagittal direction (forward/backward), force (strong/light), velocity (fast/slow), and directness (direct/indirect). It was concluded that the trunk movement plays a critical role to discriminate between positive and negative emotions except for Anger. Displays of anger were always perceived as strong and displays of grief were always observed as slow. Surprise was mostly characterized by a backward movement. The expression of joy was mainly characterized with straight body posture, upward movement with raised arms and fast movement. Grief was characterized with slow, light, downward directed body movement and arms closed around the body.

5.3.3 Summary

Several perceptual studies were realized to explore human recognition and characterization of bodily expression of emotions. While it has been reported that sadness and happiness can be successfully communicated in static body postures, most studies were oriented toward the study of emotion expression in body movement in several contexts: prototypical expression of emotion as well as emotion expression during movement tasks such as daily actions (e.g. walking) and music performance (e.g. piano performance). Besides, it has been shown that the dynamic of emotion expression can improve the human recognition of bodily expression of emotions.

Overall, sadness and anger expressions mostly received the best recognition scores. Fear expression in body movement was the least recognized. The difficulty of the recognition of fear in bodily expression was also reported in the study of facial expression [Milders et al., 2003]. The expression of fear is strongly depending on the kind of threat or the antecedent of fear like the darkness, the fear of getting hit, of being rejected [Van Den Stock et al., 2007].

The perceptual characterization of emotional expressive body movement based on the ratings of few body cues is a challenging task. Previous perceptual studies tend to use a small subset of body cues for the characterization of emotional body expression. Across several studies, sadness expression was mostly characterized by slow motion, low energy, and collapsed body posture. Anger expression was mostly characterized by fast motion, high energy and expanded body posture.

Table 5.1 provides a summary of the results found in previous perceptual studies related to the human recognition and characterization of emotional body expression. The content of this table represents an adaptation of the recent review provided in [Kleinsmith and Bianchi-Berthouze, 2013]. In Table 5.1, we focus on the emotions that have been the most studied in previous studies. Detailed description of less frequently used emotions (e.g. sympathy, antipathy, warmth...) can be found in [Kleinsmith and Bianchi-Berthouze, 2013]. Unlike in [Kleinsmith and Bianchi-Berthouze, 2013], we also provide the recognition rates for emotions and we additionally highlight the nature of feature (describing body posture, body movement or both) and the nature of the analysis in recognition and characterization tasks: perceptual (human perception) or kinematic (automatic analysis). The results of recognition and the characterization of emotions in kinematic analysis will be discussed in the next section.

5

5.4 Automatic analysis of emotional body expression

The previous section was devoted to discuss the perception and the perceptual characterization of bodily expression of emotions. In the current section, we discuss related works on the automatic recognition and the characterization of bodily expression of emotions. Automatic analysis are based on “kinematic” body cues to recognize and characterize bodily expression of emotions. The kinematic description is required to build computational models for the automatic analysis of expressive movement.

5.4.1 Automatic analyses of emotion expression in body posture

Previous works on automatic analysis of emotion expression using body posture cues were mainly based on 3D motion capture data. A collection of 3D motion

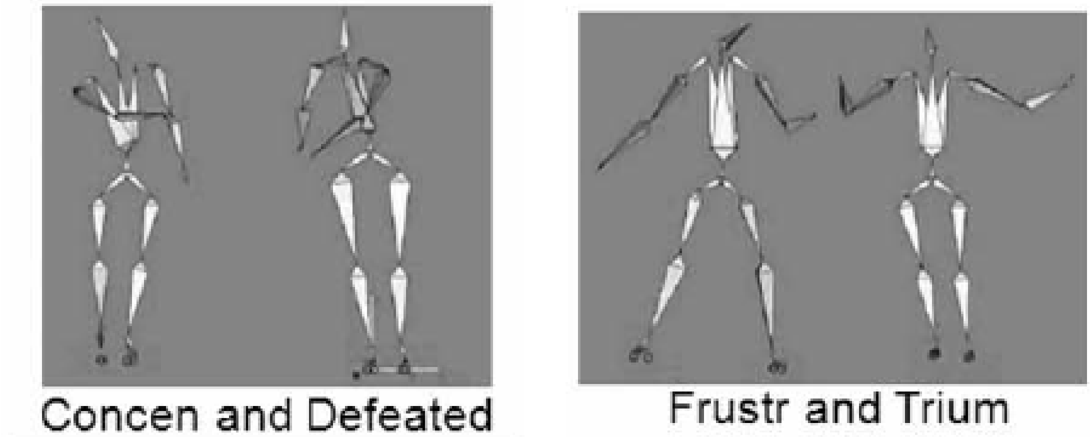


Figure 5.7: Expressive body postures expressing Frustrated & triumphant and concentrating & defeated groups of affective states (screenshot adapted from [Kleinsmith et al., 2011])

capture affective postures is obtained by selecting the snapshot representing "most expressive instant of the gestures" in [De Silva and Bianchi-Berthouze, 2004]. Based on this collection of expressive body postures, De Silva and Bianchi-Berthouze [De Silva and Bianchi-Berthouze, 2004] used a set of 24 low-level features to investigate their relevance in discriminating four basic emotions: angry, fear, happy, and sad. This set of features is mainly based on the projection of the 3D distances between different body joints (e.g. shoulder and elbow) along the lateral, frontal and vertical planes, and further information concerning the orientation of the head and shoulders. Around 100 affective postures were classified. It has been found that vertical and lateral direction based features are important to discriminate between Sadness and Joy; rising arms/lateral extension for Joy expression and bending arms in Sadness expression.

Similarly to [De Silva and Bianchi-Berthouze, 2004] (based on the same database and similar set of body posture cues), Kleinsmith and Berthouze [Kleinsmith and Berthouze, 2007] studied the relationship between expressive body postures and affective dimensions. The affective dimensions used in their study were valence (pleasure), arousal (alertness), potency (control), and avoidance (avoid/attend to). Statistical analyses showed that some postural features are more important than others in the characterization of affective dimensions level [Kleinsmith and Berthouze, 2007]. The openness of the body, in particular on the lateral plane, showed an important effect on the activation dimension. The vertical extension of right arm was a discriminate feature for the potency.

Unlike the snapshots of acted affective postures collected in [De Silva and Bianchi-Berthouze, 2004], Kleinsmith et al. [Kleinsmith et al., 2011] collected snapshots of non-acted affective postures of unprofessional actors when playing a video game.

The 3D motion capture data was collected using the Gypsy motion capture system from Animazoo [Kleinsmith et al., 2011]. Mostly the subjects are not aware of the true purpose of the study in order to ensure the natural and non-acted aspect of affective bodily expression. The difficulty is to define the exact emotion or affective state that was felt at a particular moment to study the corresponding affective posture. For this purpose, Kleinsmith et al. [Kleinsmith et al., 2011] were based on the label judgments of three external students on the expressive postures. After analyzing all the proposed labels, the affective states were categorized into four classes; concentrating, defeated, triumphant and frustrated. Then, an automatic recognition of emotion categories as well as of affective dimensions was conducted. The low-level features used to describe postures were based on the Euler angles of following joints; head, neck, collar, shoulders, elbows, wrists, torso, hips and knees. A statistical analysis showed that the upper body and arms are the most important features describing the posture. In particular, it was observed that the leaning of the torso in the sagittal plane and sometimes the lateral and vertical extension of arms and shoulders take different values on two different groups of emotions; frustrated and triumphant versus concentrating and defeated (See Figure 5.7). Frustrated and triumphant affective states, considered as more active states, were characterized with upright or slightly backward bending of the torso. While concentrating and defeated affective states, considered as less active states, were characterized by forward bending of the torso.

The works discussed above are based on a snapshot of expressive body movement. The characterization of expressive body postures was based on motion capture data (See Figure 5.7). In the following subsection, we will focus on the automatic recognition and characterization of emotion expression in body movement.

5.4.2 Automatic analysis of emotion expression in body movement

Several recent works have been conducted to explore the emotional body expression in different movement tasks such as arms movement based actions (lifting, knocking), walking, music performance and abstract gestures. These works are discussed in the following paragraphs.

Emotion expression during arms movement based actions :

Several works focused on bodily expression of emotions in arms movement based actions. They mostly focus on a reduced subset of kineamtic features based on arms movement cues.

Bernhardt and Robinson [Bernhardt and Robinson, 2007] extracted statistical measures of arms movement qualities measured based on 3D motion capture data in order to classify four emotions expressed in knocking action. Although they proved successful classification results, they did not provide insight on which body cues contribute better in emotions discrimination.

Pollick et al. [Pollick et al., 2001] used dynamic movement features to characterize drinking and knocking movement performed by two actors with 10 different affective states: afraid, angry, excited, happy, neutral, relaxed, sad, strong, tired and weak. They construct a psychological space dimensions based on the results obtained in a perceptual study when the observers were asked to categorize an affective point-light display of human arm movement. While the first dimension was strongly similar to the activation dimension of the two dimensional emotional model [Russell, 1980], the second dimension was similar to the valence dimension. The movement kinematics used in this study were the duration, average velocity, peak velocity, peak acceleration, peak deceleration and jerk index of movement. It was shown that a high level of energy (e.g. high peak of velocity, high average of velocity...) was correlated with active affects (e.g. angry, excited and happy) and a low level of energy was correlated with passive affects (e.g. sad, tired and weak).

Patterson et al. [Patterson et al., 2001] showed that the human classification accuracy for angry and sad expression in arm movement is affected by the variation of the speed during the movement. This result is congruent with previous findings in perceptual studies. Several perceptual studies have shown that, with regards to the speed of movement, angry bodily expression is perceptually characterized by fast and sudden movement while sad bodily expression is perceptuall characterized by slow movement [Montepare et al., 1999]. In this study [Patterson et al., 2001], the artificial rise of the speed of movement in the sequence corresponding to the sad experience leaded to categorize it as angry. The variation of the movement duration also affected the intensity of Anger expression. Thus, raising the speed of angry movements leaded to categorize them as more intense and decreasing the speed of angry movements leaded to categorize them as less intense.

Gross et al. [Gross et al., 2010] investigated the ability of a kinematic analysis of body movement to discriminate six emotions expressed in kocking (Angry, Anxious, Content, Joyful, Proud and sad). The motion capture data of knocking movement were used to extract movement features. The angular data of the main articulations involved in the movement were calculated and only those which concern the sagittal plane were kept (i.e. flexion angles). The average and the peaks of joint angles, the range of motion (difference between maximum and minimum values), the angular velocity of elbow movement and their maximum values, the average knocking duration and the knocking duration for every movement cycle were measured. Sadness trials of the knocking were characterized by slow movement (the longest movement time) in the first and last movement cycles. Shoulder flexion posture and elbow flexion variation received the highest value in angry trials. The frequency of knocking was most important for the joyful trials than the other emotional trials. Gross et al [Gross et al., 2010] showed that the perceptual rating of movement quality factors (based on Effort-Shape analysis Analysis) was aligned with the kinematic analysis.

Emotion expression during walking :

In addition to the studies that focus on arms movement, other studies explored kinematic features of the whole body movement to characterize the expression of

5

emotion during walking action. Roether et al. [Roether et al., 2009] studied the expressive gaits of four emotions (anger, happiness, sadness and fear) performed by 25 individuals. Their analysis was based on a set of spatio-temporal motor primitives defining the flexion-angle trajectories of eleven major joints including head, spine, pelvis, and left and right shoulder, elbow, hip and knee joints. The proposed parameters were modeled using an unsupervised learning techniques and statistical methods. The automatic analysis of the trajectories has revealed interesting postural and movement features that were, to some degree, critical to differentiate between some emotions. The head inclination or downward head rotation was a prominent postural feature for sad gaits. The expression of sadness was also characterized by low elbow flexion-angle whereas this features received increasing values in angry and fearful gaits. High values of the variation in the amplitudes of joint flexion-angles, and high values of the average of gait velocity were correlated with happiness and anger while low values were correlated with sadness and fear. The authors observed that knee and hip movement are considered as important features to characterize the emotion of fear in walking while shoulder and elbow movements were important when expressing emotions with high arousal anger and happiness. It has been also shown that human recognition of emotions expressed in gait were inline with their automatic recognition.

Emotion expression during music performance :

Other studies focused on the characterization of emotions expressed during music performance. Castellano et al. [Castellano et al., 2008] extracted a number of features that explain the temporal dynamics of the velocity of head movement and the quantity of upper body movement to automatically analyze expressive body movement in piano performance. The expression of five moods (personal, sad, allegro, serene and over-expressive) were considered in one professional musician's performance. Based on statistical techniques (analysis of variance and pairwise comparisons), they found that head velocity features showed higher impact on the discrimination of expressed emotion compared to the features describing the quantity of motion. The authors explained this result as piano performance does not imply significant difference in movement quantity between different emotions (except Sadness which received significantly lower values of movement quantity features).

Emotion expression in abstract gestures :

It has been shown that, in addition to daily actions and music performance movements, emotion expression can be also conveyed in abstract gestures. Castellano et al. [Castellano et al., 2007] explored the kinematic characterization of emotion expression in abstract gestures. In [Castellano et al., 2007], machine learning techniques were used to select the best subset of features related to the temporal profiles of 5 motion cues. The motion cues are; quantity of motion and contraction index of the body, velocity, acceleration and fluidity of the hand's barycentre. Using a wrapper and a correlation based feature selection approaches, it was found that the quantity of motion and the contraction index of upper body contribute better than the speed and the fluidity of motion in discriminating between different emotions expressed in "Arms rising in coronal plane" action. A decision tree was also built to interpret the prediction power of each body cue. It has been shown that the maximum of Quantity of motion seems to discriminate between "high arousal" emotions and "low arousal" emotions.

5.4.3 Summary

In recent years, there as been an increase of interest in automatic recognition of emotions from body movement using posture and movement quality features [De Silva and Bianchi-Berthouze, 2004, Kapur et al., 2005, Castellano et al., 2007, Bernhardt and Robinson, 2007]. Only few scholars attempted to discriminate affective postures using features that describe the form of body posture. Most of the previous works focused on the recognition of emotions expressed in body movement using features that describe the form of body posture, the dynamics of body movement and spatio-temporal features. Body posture features describe the form of body posture such as the amount of openness/closeness. Movement dynamic features refer to the properties of the dynamic aspect of movement such as the speed or the acceleration of movement. Movement dynamic features can also refer to temporal characteristics of the motion such as the number of peaks that occur in a knocking motion.

Several studies explored the statistical or predictive power of body cues either for interpretation (to get deeper insights into which body cue is the most relevant to discriminate between emotions) or classification (to enhance classifier efficiency and performance) purpose. Anger and Sadness expressions were widely explored in automatic analysis studies. High recognition scores were mostly achieved for Sadness and Anger expressions.

However, most of the previous studies that analyzed bodily expression of emotions in daily actions focused on a limited range of emotional labels, actions and body cues [Bernhardt and Robinson, 2007, Hicheur et al., 2013, Gross et al., 2010, Kleinsmith and Bianchi-Berthouze, 2013]. In a recent survey on affective body expression perception and recognition [Kleinsmith and Bianchi-Berthouze, 2013], the authors conclude that there is a need to go beyond the focus on a limited set of movement tasks in the analysis of bodily expression of emotions and that *this requires a more systematic investigation of the types of features (e.g., local versus global) that change the affective message carried out by the body independently from the semantic meaning (action) that it conveys.*

Previous automatic analyses of emotional body expression were mostly focused on the communication of emotions in a movement task rather than emotion communication through prototypical expressions (e.g. using communicative gestures). Indeed, the automatic recognition of emotions expressed in communicative gestures requires the automatic detection of specific communicative gestures conveying emotions (e.g. making a fist to express Anger). There is a large repository of movement

5

types and gestures that can be used to explicitly express emotions in body movement. However, gesture recognition is still a challenging task in our days, which may explain the lack of studies related to the automatic analysis of emotion expression based on body action and posture units. The automatic recognition of emotions expressed in a specific movement task with a well known pattern (such as walking) requires less effort; it has been shown that a good recognition score can be achieved using simple statistical measures of body cues such as the average of arms movement speed [Bernhardt and Robinson, 2007].

Table 5.1 also includes a summary of previous works conducted on the automatic recognition and kinematic characterization of emotions expressed in body movement. Similarly to the summary of perceptual studies, we focus on the emotions that have been the most studied.

5.5 Conclusion

It has been widely shown that facial expressions convey emotions. When combining facial and bodily expressions, it has been shown that the perception of emotions is biased toward body expression. Recently, there has been a growing interest on the study of emotion recognition solely from body posture and movement.

Early studies on emotional body expression reported the great influence of body movement in the perception of emotion. It has been also shown that humans are able to recognize emotions from solely body movements. Several works from psychology attempted to find out what qualitative body cues is guiding human perception of emotions expressed in daily actions. However, the rating of expressive body cues in perceptual studies is mostly limited to a reduced set of cues. Besides, perceptual rating of body cues allows a qualitative description of some predefined cues (e.g. the collapse of body posture or the speed of movement) while emotional body expression is better and more accurately described in terms of kinematic body cues measured using 2D or 3D motion capture data.

Recent studies explored techniques enabling automatic recognition of emotions based on video or 3D motion capture data analysis. A high amount of similarity can be found between results from perceptual studies and from automatic analysis studies of emotional body expression. For instance, several perceptual studies and automatic analysis reported that Sadness and Anger expressions were the best recognized. Besides, Sadness expression has been characterized by slow movement and bending posture in both perceptual and automatic analyses, in the same way as Anger expression has been characterized by fast, energetic and expanded movement.

However, previous analyses were mostly restricted to the analysis of expressive movement in a single daily action. Thus, they mainly used an action-dependent body coding schema, which makes the generalization of the results difficult due to the lack of common coding schema for movement [Kleinsmith and Bianchi-Berthouze, 2013, Castellano et al., 2008]. Besides, a reduced set of body cues measured according to specific body parts (mainly the ones involved in the movement task) have been mostly considered. As body structure is characterized by a high degree of freedom, it is worth exploring body cues covering different body parts to give a deeper insight on which cues convey emotional expression.

As spontaneous expression of emotions is hard to collect, most of previous works were based on dataset of actors portraying emotions through their body. Usually, the actors are asked either to portray prototypical expression of emotion in individual or interaction settings, or to express emotions while performing daily actions.

We can distinguish two categories of explicit body cues defined in previous works as an attempt to understand the process of emotion recognition from body movement; movement-type based body cues and movement-quality based body cues.

Movement-type based cues are sometimes referred to as " propositional" cues which can be defined as "*established signs to transmit meaning*" [Camurri et al., 2003] (e.g. rising arms to express Joy). They are also referred to as gestures and body actions units (e.g. Up-down head shake [Dael et al., 2011]). As movement-type based body cues describe body movement and gestures in a fine-grained fashion, they are mostly used to characterize prototypical expression of emotion.

Movement-quality based cues refer to the manner by which a body movement is performed using spatio-temporal or movement dynamics features(e.g. the speed of head movement). They are sometimes referred to as non-propositional cues [Camurri et al., 2003] as they describe the way by which a movement is carried through. Movement-quality based cues do not rely on a specific kind of action, but they refer to the characteristics of movement tasks (e.g. walking, dancing or knocking at the door). Unlike perceptual studies, most of automatic analyses aim at recognizing emotions from body posture and movement focused on movement-quality based body cues. As movement-quality based body cues abstract the intrinsic properties of the action itself to focus on the manner by which a movement is performed, they were mostly used to characterize the expression of emotions in specific movement tasks. However, few studies attempted to characterize prototypical expression of emotion using movement-quality based features [Glowinski et al., 2011].

5.6 Summary of chapter

- **Communication of emotions in body movements:** It has been shown in previous works that affective states can be conveyed in body movement in different ways. Some studies focused on prototypical expression of emotions (e.g. communicative gestures conveying emotions such as making a fist to express anger). Other studies focused on the expression of emotion during a movement task such as artistic movements, daily actions, music performances, video games scenario and other abstract gestures.
- **Emotion expression in Posture/ Movement:** Perceptual (based on human perception) and kinematic (e.g. automatic analysis using motion capture data) studies on bodily expression of emotions can be split into two groups: those which focus on static postures and those which focus on body move-

ments. Static postures are obtained from a particular configuration of body segments or from a movie of expressive movement (i.e. a snapshot of the movie).

- **Perceptual studies:** We discuss previous perceptual studies that explore the human recognition of emotions expressed in body posture and movement. Some works explored both the recognition and the characterization of bodily expression others were focused on the perceptual rating of few body cues to characterize bodily expression of emotions.
- **Automatic analyses:** We discuss previous works on automatic analysis of expressive body movement. These works aimed to automatically discriminate between emotions expressed in body posture/ movement through a set of kinematic features (e.g. using motion capture data).
- **Perceptual and automatic recognition of emotion:** Both Sadness and Anger expressions in body movement have been widely considered as the best recognized. In several perceptual and automatic recognition studies, Fear expression has been considered as the least recognized in both expressive body posture and body movement.
- **Perceptual and automatic characterization of emotion:** Both 1) Sadness and 2) Anger expressions in body movement have been widely characterized respectively with 1) slow, less energetic movement and bending posture and 2) fast, jerky, more energetic movement and expanded posture.
- **Summary of previous findings:** A summary of previous findings on the recognition and the characterization of bodily expression of emotions is provided in Table 5.1. We mainly highlight the results of the most frequently used emotions: Sadness/Grief, Shame, Happiness/Joy/Elated Joy/ Anger/Hot Anger/ Cold Anger, Fear, Pride, Contentment, Disgust, Surprise and Anxiety.

Table 5.1: Summary of movement and posture characteristics used in previous automatic analyses for expressive movement characterization. Post. and Mvmt. stand respectively for posture and movement. (*) stands for significant recognition (See [Gross et al., 2010]), A/H stands for Automatic/Human recognition.

Emotion	Characterization	A/H	Recognition	Reference
Sadness	Prototypical expression	H	66.67%(Post.)	[Atkinson et al., 2004]
Sadness	Prototypical expression	H	86.94% (Mvmt.)	[Atkinson et al., 2004]
	Perceptual characterization Mvmt.: Dropping the head , Mvmt.: Crossing/bringing arms in front of the body , Mvmt.: Bringing hands to the face			
Sadness	Music performance	H	78.4%	[Dahl and Friberg, 2007]
	Perceptual characterization Mvmt.: Low Speed , Mvmt.: High Regularity , Mvmt.: High Fluency **Perceptual characterization**: Basson:			

	Mvmt.: Fairly regular **Perceptual characterization**: Saxophone Mvmt.: Very low Quantity of motion			
Sadness	Computer generation **Perceptual characterization** Post.: Forwards head bend , Post.: Forwards chest bend , Post.: No abdominal twisting , Post.: Arms at the side of the trunk	A	27.4%	[Coulson, 2004]
Sadness	Knocking **Perceptual characterization** Mvmt.: Long overall movement time but proportionately low time spent actually Knocking , Mvmt.: Small amplitude of elbow motion , Mvmt.: Low elbow extensor velocity	H H	56% (*) 17%	[Gross et al., 2010]
Sadness	Prototypical expression **Perceptual characterization** Post.: Collapsed body posture , Mvmt.: Low movement dynamicss	-	-	[Wallbott, 1998]
Sadness	Prototypical expression **Perceptual characterization** Post: Arms at rest and hands in pocket	-	-	[Dael et al., 2011]
Sadness	Walking **Perceptual characterization** Mvmt.: Very smooth movement , Mvmt.: Very loose and soft movement , Mvmt.: Very slow movement , Post./ Mvmt.: Contracted posture/ movement , Mvmt.: Low quantity of movement	H	76%	[Montepare et al., 1999]
Sadness	Walking **Perceptual characterization** Mvmt.: Low arms swing , Mvmt.: Small strides , Mvmt.: Low heavyfooted , Post.: Fairly straight body posture	H	94%	[Montepare et al., 1987]
Sadness	Walking **Kinematic characterization** Post.: Downward head flexion , Post.: Downward torso flexion , Post.: Straight elbow flexion , Mvmt.: Decrease in knee movement	H	89.8%	[Roether et al., 2009]
Sadness	Walking	H	73%	[Karg et al., 2010]
Sadness	Prototypical expression	H	95%	[Kapur et al., 2005]
Sadness	Abstract gestures **Kinematic characterization** Mvmt.: Low quantity of movement , Post./ Mvmt.: Contracted posture/ movement	A	48%	[Castellano et al., 2007]
Sadness	Music performance **Kinematic characterization** Mvmt.: Lowe level of upper body movement , Mvmt.: Slow head movement	-	-	[Castellano et al., 2008]
Sadness	Knocking	A	92.4%	[Bernhardt and Robinson, 2007]
Sadness	Walking **Kinematic characterization**	H	76%	[Crane and Gross, 2007]

	Mvmt.: Low speed of movement , Post.: Downward head flexion , Post./ Mvmt.: Contracted posture/ movement , Mvmt.: Low shoulders and elbow ranges of motion , Mvmt.: Low hip range of motion			
Grief	Abstract gestures	-	-	[Meijer, 1989]
	Perceptual characterization Post.: Bowed body posture , Mvmt.: Downward body movement , Mvmt.: Slow speed of movement , Mvmt.: Light movement (muscles relaxed)			
Shame	Abstract gestures	-	-	[Meijer, 1989]
	Perceptual characterization Post.: Bowed body posture , Mvmt.: Downward body movement , Mvmt.: Low speed of movement , Mvmt.: Light movement (muscles relaxed)			
Shame	Prototypical expression	-	-	[Wallbott, 1998]
	Perceptual characterization Post.: Collapsed body posture , Post.: Downward head posture , Mvmt.: Self-manipulators			
Happiness	Prototypical expression	H	61.67% (Post.)	[Atkinson et al., 2004]
Happiness	Prototypical expression	H	86.67% (Mvmt.)	[Atkinson et al., 2004]
	Perceptual characterization Mvmt.: Raising the arms and shaking of the fists, either above the head or in front of the body , Mvmt.: Jumping up and down , Mvmt.: Skipping			
Happiness	Music performance	H	64.4%	[Dahl and Friberg, 2007]
	Perceptual characterization Mvmt.: Fairly High Quantity of motion , Mvmt.: Fairly High Regularity **Perceptual characterization**: Bassoon Mvmt.: Fairly High speed , Mvmt.: Fairly Low Fluency **Perceptual characterization**: Saxophone Mvmt.: Fairly High speed , Mvmt.: Fairly Low speed			
Happiness	Computer generation	A	78.6%	[Coulson, 2004]
	Perceptual characterization Post.: Head backwards , Post.: No forwards movement of the chest , Post.: Arms raised above shoulder level , Post.: Straight elbows flexion			
Happiness	Walking	H	73%	[Montepare et al., 1999]
	Perceptual characterization Mvmt.: Jerky movement , Mvmt.: Fast movement , Mvmt.: Loose and hard movement , Mvmt.: High quantity of motion , Post.: Expanded limbs			
Happiness	Walking	H	74%	[Montepare et al., 1987]
	Perceptual characterization Mvmt.: High arms swing , Mvmt.: Fairly large strides , Mvmt.: Fairly low heavyfooted , Post.: Straight body posture			
Happiness	Walking	H	75.1%	[Roether et al., 2009]
	Kinematic characterization Post.: Upward head flexion , Post.: Straight toro , Mvmt.: Increase of arms movement			
Happiness	Knocking	A	65.3%	[Bernhardt and Robinson, 2007]
Happiness	Walking	H	53%	[Karg et al., 2010]
Joy	Prototypical expression	H	99%	[Kapur et al., 2005]

Joy	Knocking **Perceptual characterization** Mvmt.: Sudden, hurried, fast time , Mvmt.: Free, relaxed, uncontrolled flow , Post.: Expanded limbs **Kinematic characterization** Mvmt.: “Chin up” position of the head , Mvmt.: High Knocking rate , Mvmt.: Similar peak flexor and extensor elbow velocities	H H	39% (*) 0%	[Gross et al., 2010]
Joy	Abstract gestures **Perceptual characterization** Post.: Straight body posture , Post./ Mvmt.: Backward posture , Post./ Mvmt.: Expanded limbs (arms raised) , Mvmt.: Upward direct movement , Mvmt.: Fast movement	-	-	[Meijer, 1989]
Joy	Abstract gestures **Kinematic characterization** Mvmt.: High quantity of movement , Post./ Mvmt.: Expanded posture/ movement	A	44%	[Castellano et al., 2007]
Joy	Walking **Kinematic characterization** Mvmt.: High speed of movement , Mvmt.: High shoulder and elbow range of motion , Mvmt.: High hip range of motion	H	67%	[Crane and Gross, 2007]
Elated Joy	Prototypical expression **Perceptual characterization** Post.: Shoulders lifted upward , Post.: Backward head posture , Post. /Mvmt: Arms stretched out frontal or upward , Mvmt.: Opening/ closing hands movement , Mvmt.: Illustrators , Mvmt.: High movement activity , Post. /Mvmt: Expansive movement , Mvmt.: High movement dynamics	-	-	[Wallbott, 1998]
Elated Joy	Prototypical expression **Perceptual characterization** Mvmt.: Symmetrical up-down repetitive arm actions	-	-	[Dael et al., 2011]
Anger	Walking	H	57%	[Karg et al., 2010]
Anger	Prototypical expression	H	87%	[Kapur et al., 2005]
Anger	Prototypical expression Prototypical expression **Perceptual characterization** Mvmt.: Expansive movement towards the camera , Mvmt.: Shaking of the fists or stamping of the feet, or both	H H	46.95% (Post.) 85.55% (Mvmt.)	[Atkinson et al., 2004] [Atkinson et al., 2004]
Anger	Music performance **Perceptual characterization** Mvmt.: Fiarly High Quantity of motion , Mvmt.: HighSpeed , Mvmt.: Low Fluency , Mvmt.: Medium Regularity	H	61.9%	[Dahl and Friberg, 2007]
Anger	Computer generation **Perceptual characterization** Post.: Backwards head bend , Post.: No backwards chest bend , Post.: No abdominal twist , Post.: Arms raised forwards and upwards.	A	34.9%	[Coulson, 2004]
Anger	Knocking **Perceptual characterization** Mvmt.: Strong, powerful, forceful energy , Mvmt.: Moves away from the body: expanded limbs	H	78% (*) 67%	[Gross et al., 2010]

5

	Kinematic characterization Mvmt.: Large amplitude of elbow motion , Mvmt.: The largest elbow extensor velocity , Mvmt.: High raised arm (i.e., largest shoulder flexion), , Mvmt.: Long time spent in Knocking , Mvmt.: Large elbow amplitudes			
Anger	Abstract gestures **Perceptual characterization** Post.: Bowed body posture , Post./ Mvmt.: Expanded limbs , Mvmt.: Strong force	-	-	[Meijer, 1989]
Anger	Walking **Perceptual characterization** Mvmt.: Hard and stiff movement , Mvmt.: Fast movement , Mvmt.: Expanded limbs , Mvmt.: Very jerky movement , Mvmt.: High quantity of motion	H	85%	[Montepare et al., 1999]
Anger	Walking **Perceptual characterization** Mvmt.: High arms swing , Mvmt.: Large strides , Mvmt.: High heavyfooted , Post.: Straight body posture	H	90%	[Montepare et al., 1987]
Anger	Walking **Kinematic characterization** Post.: Angular elbow flexion , Post.: Head downward flexion , Mvmt.: Increase of arms movement	H	70.3%	[Roether et al., 2009]
Anger	Abstract gestures **Kinematic characterization** Mvmt.: High quantity of movement , Post./ Mvmt.: Expanded posture/ movement	A	90%	[Castellano et al., 2007]
Anger	Knocking **Kinematic characterization** Mvmt.: Energetic and forceful movement	A	92.4%	[Bernhardt and Robinson, 2007]
Anger	Walking **Kinematic characterization** Mvmt.: High speed of movement , Mvmt.: Increase of step frequency , Mvmt.: High shoulder and elbow range of motion , Mvmt.: High hip range of motion	H	62%	[Crane and Gross, 2007]
Hot Anger	Prototypical expression **Perceptual characterization** Post.: Lateralized hand/ arm movement , Post.: Shoulders lifted upward , Post.: Arms stretched out frontal , Mvmt.: High movement activity , Mvmt.: Expansive movements , Mvmt.: High movement dynamics	-	-	[Wallbott, 1998]
Hot Anger	Prototypical expression **Perceptual characterization**	-	-	[Dael et al., 2011]
Cold Anger	Prototypical expression **Perceptual characterization** Post.: Lateralized hand/ arm movement , Post.: Arms stretched out frontal	-	-	[Wallbott, 1998]
Fear	Prototypical expression	H	91%	[Kapur et al., 2005]
Fear	Prototypical expression	H	64.17% (Post.)	[Atkinson et al., 2004]
Fear	Prototypical expression	H	91.11% (Mvmt.)	[Atkinson et al., 2004]

	Perceptual characterization Mvmt.: Movement away from the camera: contracting or cowering movements , Mvmt.: Raised hands, especially in front of the face			
Fear	Music performance **Perceptual characterization** Mvmt.: Low Quantity of motion **Perceptual characterization**: Bassoon Mvmt.: Low Fluency , Mvmt.: Fairly High speed , Mvmt.: Medium Regularity **Perceptual characterization**: Saxophone Mvmt.: High Fluency , Mvmt.: Low Speed , Mvmt.: High Regularity	H	19.1%	[Dahl and Friberg, 2007]
Fear	Computer generation **Perceptual characterization** Post.: Head backwards , Post.: No abdominal twist , Post.: Forearms raised , Post.: Weight transfer is either backwards or forwards	A	76.4%	[Coulson, 2004]
Fear	Abstract gestures **Perceptual characterization** Post.: Bowed body posture , Mvmt.: Backward body movement , Mvmt.: High speed of movement	-	-	[Meijer, 1989]
Fear	Prototypical expression **Perceptual characterization** Mvmt.: Opening/closing hands movement , Mvmt.: Many self- manipulators	-	-	[Wallbott, 1998]
Fear	Prototypical expression **Perceptual characterization** Post.: Backward body lean , Post.: Upward gaze , Post.: Lateral trunk lean	-	-	[Dael et al., 2011]
Fear	Walking **Kinematic characterization** Post.: Upward head flexion , Post.: Forward/ bent torso , Post.: Upper-arm retraction , Post.: High knee flexion , Post.: High angular elbows flexion , Mvmt.: Decrease in knee movement	H	77.1%	[Roether et al., 2009]
Pride	Prototypical expression **Perceptual characterization** Post.: Head titled backward and a small smile , Post.: Arms akimbo with hands on hips , Post.: Expanded arms posture	H	high	[Tracy and Robins, 2007]
Pride	Knocking **Perceptual characterization** Mvmt.: High energy **Kinematic characterization** Mvmt.: “chin up” position of the head , Mvmt.: Large elbow amplitudes	H H	56% (*) 0%	[Gross et al., 2010]
Pride	Prototypical expression **Perceptual characterization** Post.: Backward head posture , Post.: Arms crossed in front of the chest	-	-	[Wallbott, 1998]
Pride	Walking **Perceptual characterization** Mvmt.: High arms swing , Mvmt.: Large strides , Post.: Straight body posture	H	56%	[Montepare et al., 1987]
Content-ment	Knocking **Perceptual characterization**	H	33% (*) 17%	[Gross et al., 2010]

	Mvmt.: Light, delicate, buoyant , Mvmt.: Moves away from the body, expanded limb shape			
Contentment	Walking	H	74%	[Crane and Gross, 2007]
Disgust	Prototypical expression	H	34.17% (Post.)	[Atkinson et al., 2004]
Disgust	Prototypical expression **Perceptual characterization** Mvmt.: Bringing a hand (occasionally two hands) to the face, especially in the region of the mouth and nose , Mvmt.: Turning the face away from the camera , Mvmt.: Dropping the head , Mvmt.: Swiping a hand in front of the face, often repeatedly, as if dispersing a bad smell	H	75.28% (Mvmt)	[Atkinson et al., 2004]
Disgust	Abstract gestures **Perceptual characterization**	-	-	[Meijer, 1989]
Disgust	Computer generation	A	18.9%	[Coulson, 2004]
Disgust	Prototypical expression **Perceptual characterization** Post.: Collapsed body posture , Post./ Mvmt.: Backward or forward shoulders , Post.: Downward head posture , Post.: Arms crossed in front of chest , Mvmt.: Inexpansive movements	-	-	[Wallbott, 1998]
Surprise	Computer generation **Perceptual characterization** Post.: Backwards head and chest bends , Post.: Abdominal twisting , Post.: Arms raised with forearms straight	A	0.9%	[Coulson, 2004]
Surprise	Abstract gestures **Perceptual characterization** Post.: Straight body posture , Post./ Mvmt.: Backward stepping	-	-	[Meijer, 1989]
Anxiety	Knocking **Perceptual characterization** Mvmt.: Bound, tense, controlled flow **Kinematic characterization** Mvmt.: Short movement times , Mvmt.: Constrained torso range of motion (small torso amplitude)	H H	65% (*) 33%	[Gross et al., 2010]
Anxiety	Prototypical expression **Perceptual characterization** Post.: Backward body lean , Post.: Upward gaze , Post.: Lateral trunk lean	-	-	[Dael et al., 2011]

Part III : Description, collection and perception of expressive body movement

TELECOM
ParisTech

6 Expressive body movement notation system

In chapter 2, we discussed body movement notation systems and expressive body movement cues used in previous studies. We concluded that there is a need to propose a body movement notation system that offers a good trade-off between 1) a detailed description of body movement as offered in BAP notation system and 2) a description of movement quality as offered in LABAN notation system. This chapter presents our body movement notation system intended to characterize emotional body expression in different daily actions by including a detailed description of posture and movement as well as of movement quality.

We propose a coding system based on the representation of movement quality and body shape through several description levels. This Multi-Level notation system focuses on the manner by which a daily action is performed while abstracting its semantic meaning. Considering the combination of different description levels allows an accurate and detailed description of expressive body movement. Our Multi-Level notation system can be used for the perceptual and the motion capture characterization of expressive body movement. In this Chapter, we focus on the description of motion capture characterization. The perceptual characterization adapted from our Multi-Level notation system is presented in Chapter 8.

In this chapter, we firstly describe the description levels used to define our body movement coding system. Secondly, we describe the set of motion capture features derived from our body movement notation system. Finally, we present an example of expressive walking analysis using our set of motion capture features.

6.1 Multi-Level body movement notation system (MLBNS)

Our aim is to explore the classification and the characterization of emotional body expression across different actions. To achieve these goals, we rely on a set of body features allowing the description of expressive movement across different daily actions. That is, we need to go beyond action-dependent expressive body features. We propose a movement quality based coding system that allows us to describe how the expression of emotions modulates body movement across different daily actions. Based on different body description levels, our Multi-Level Body Movement Notation System (MLBNS) allows defining a large set of body movement characteristics. Using a Multi-Level notation system allow us exploring the different

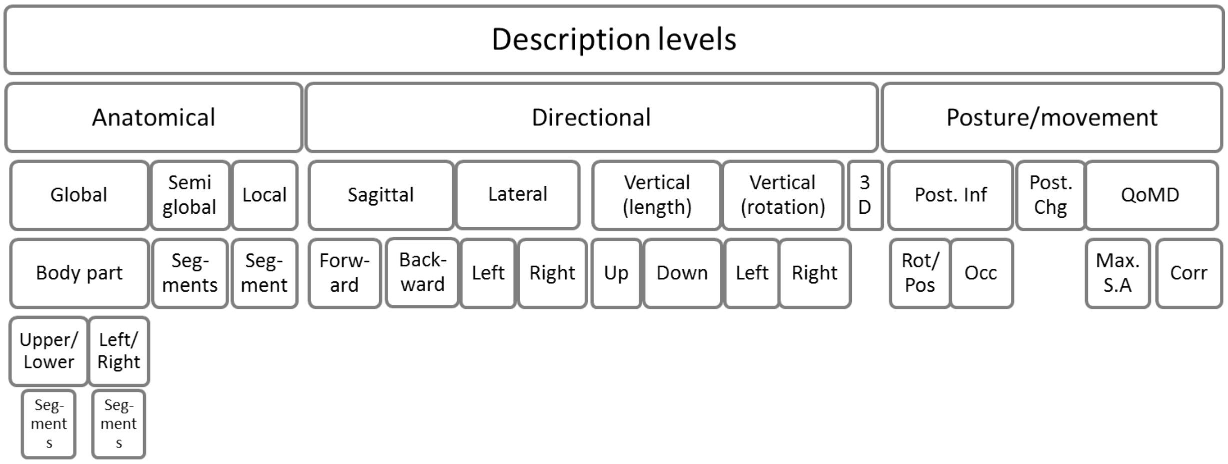


Figure 6.1: Body movement features description levels (Post. Inf stands for Postural Information, Post. Chg stands for Postural Changes, QoMD stands for Quality of Movement Dynamics, Max.S.A stands for Maximum value of the Speed and the Acceleration and Corr stands for the Correlation)

6

aspects of emotional body expression characterization. Unlike several previous works that focus on some particular aspects of body movements (e.g. elbow flexion features to study emotion expression in Knocking action [Bernhardt and Robinson, 2007]), we consider several components of body movement description levels in order to provide deeper insights into the characterization of emotional body expressions.

Three main dimensions are considered in this notation system to describe the characteristics of body movement: Anatomical, Directional and Posture/Movement dimensions. The whole set of description levels that we consider in our body movement coding schema is illustrated in Figure 6.1.

In the next subsection, we describe Directional, Anatomical and Posture/Movement description levels of our Multi-Level notation system. In these subsections, we present each description level separately. In section 6.2, we will present the whole set of body features while combining these description levels.

6.1.1 The Directional dimension

The Directional dimension describes movement directions. Movement direction has been traditionally defined with respect to three directions, which refers to the orthogonal plans of the body: the Sagittal, Lateral, and Vertical plans [Dael et al., 2012], where Sagittal plan refers to forward and backward directions, Lateral plan refers to left and right sides, and Vertical direction refers to left and right rotations around the Vertical axis. In LABAN Movement Analysis (LMA) system, the Vertical change of movement shape is manifested in upward and downward directions [Laban, 1988]. Inspired from the accuracy of BAP notation system [Dael et al., 2012] and the reliability of LABAN notation system in describing movement expressivity, we consider both the length and rotation dimensions in Vertical direction; Vertical (length: Upward/Downward) and Vertical (rotation: Left/Right).

The Directional dimension is then defined through five possible directions: Sagittal direction, Lateral direction, Vertical direction defining Vertical length of body shape, Vertical direction defining left or right rotation and finally 3D information (See Figures 6.1 and 6.2).

- Sagittal: Sagittal direction refers to sagittal extension of body limbs in terms of two directions: Forward and Backward (See Figure 6.2). In order to avoid the bias of individual size, we measure the relative Forward and Backward extensions of body limbs according to their maximal extension as shown in Figure 6.7c.
- Lateral: Lateral direction refers to lateral orientation or extension (See Figure 6.2).
  Lateral orientation captures Left and Right leaning of the head or the torso. In the definition of body cues used to characterize expressive movement (See section 6.2), we omit the difference between left and right Lateral orientation.
  Lateral extension corresponds to the relative extension of body limbs according to the Neutral body shape: Openness (limbs extended Outside the Neutral lateral bounding box) or Closeness (limbs remaining Inside the Neutral lateral bounding box) (See Fig. 6.6). As such, we also omit the difference between left and right lateral extension as mentioned for the lateral orientation.
  We also refer to the relative Lateral Openness/ Closeness of body limbs regarding the maximal body shape to avoid the bias of individual size (See Fig. 6.6). Thus, the Lateral extension is relative to 1) the Neutral lateral body shape and to 2) the Maximal lateral body shape.
- Vertical (Length): This direction refers to the flexion of body joint or the vertical position of body limbs (See Figure 6.2). Both the flexion and the vertical position are described in terms of two directions: Upward and Downward.
  Downward flexion involves the bending of the head, the torso, the knees and the angular flexion of the elbows. Upward flexion involves the straightness of the torso, the head and the round flexion of elbows. We proposed to measure angular and round elbow flexion as it was introduced in [Aronoff, 2006]. In our set of motion capture body features, we do not consider the measure of torso straightness. We define the straightness of the torso as the absence of torso bending.
  Downward and Upward positions of arms and lower body limbs are measured according to the position of the hips. Similarly to the Sagittal and Lateral extension measures, we compute the relative Upward and Downward extensions of body limbs (arms and legs) according to the their maximal extension (See Figure 6.7a).
- Vertical (Rotation): This direction refers to the left/right rotation of the head and the torso. Similarly to the Lateral orientation, we omit the difference

6

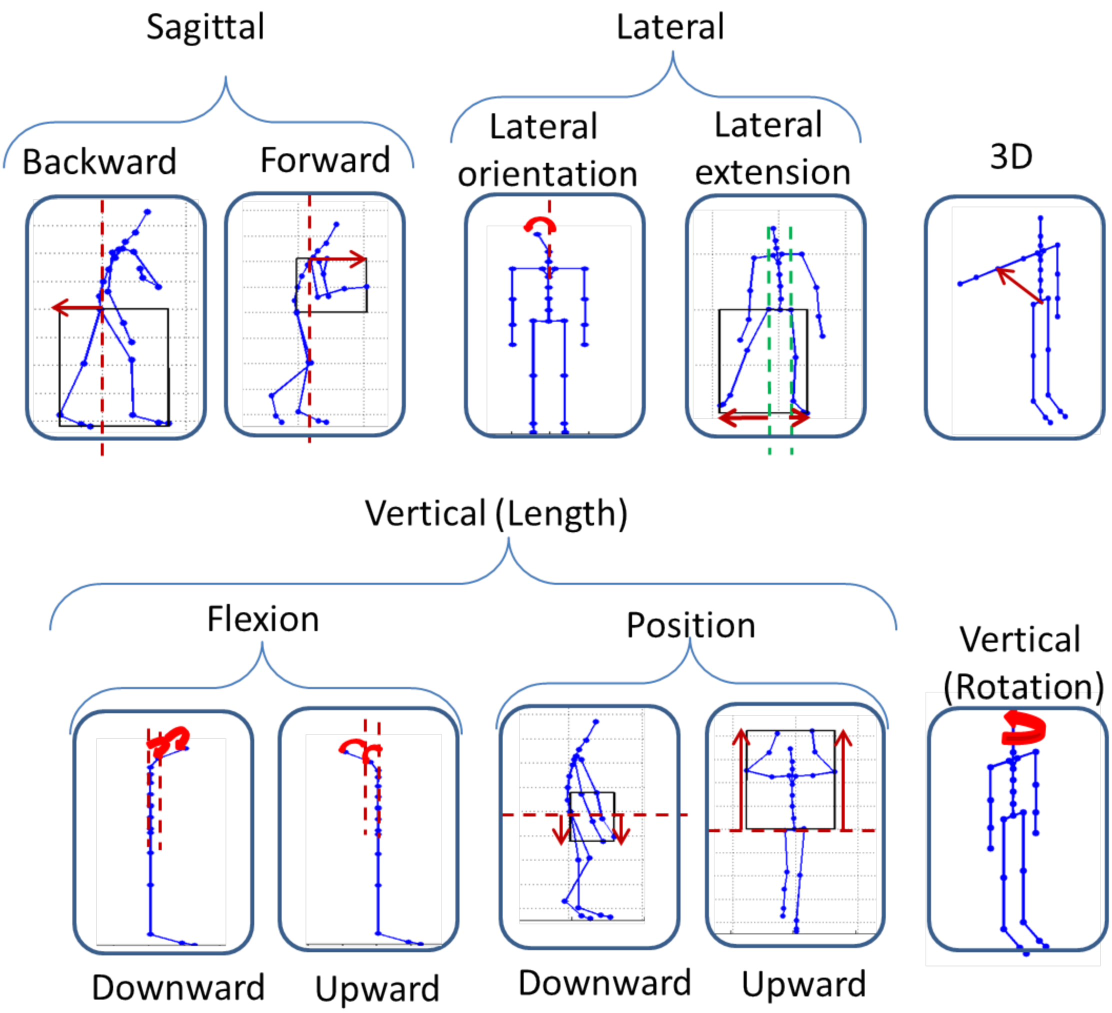


Figure 6.2: Examples of body features measured according to Sagittal, Lateral, Vertical (Length), Vertical (Rotation) and Three Dimensional (3D)

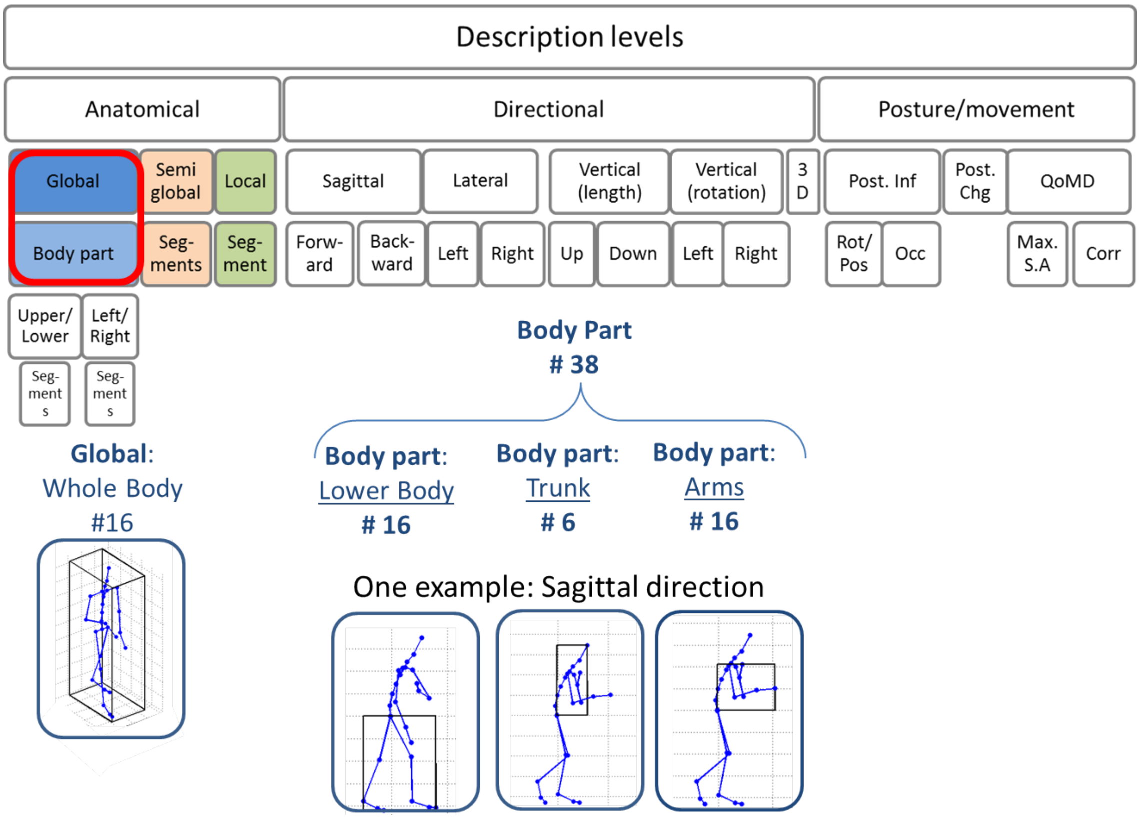


Figure 6.3: Anatomical dimension: Description of Global and Body Part description levels. # stands for the number of features that describe each body cue.

between left and right rotation in the definition of body cues.

– 3D: Three Dimensional direction refers to the 3D extension of hands, elbows and feet as well as to the 3D relative position of one body segment regarding another (e.g. 3D distance between right hand and head).

6.1.2 Anatomical dimension

Anatomical dimension describes the body segments used to characterize a body posture or movement. We distinguish three different description levels that describe the Anatomical dimension: Global, Semi-Global and Local.

In order to describe these Anatomical description levels (Global, Semi-Global and Local), we distinguish two terms: body cues and body features. Body cues are defined by the Anatomical and Directional description levels. Features are used to describe body cues based on Posture/ Movement description levels. For instance, head flexion is a body cue defined by the Local Anatomical description level and the Vertical (Length) Directional description level. It can be described through 4 features from Posture/ Movement level: the maximum rotation value, the maximal

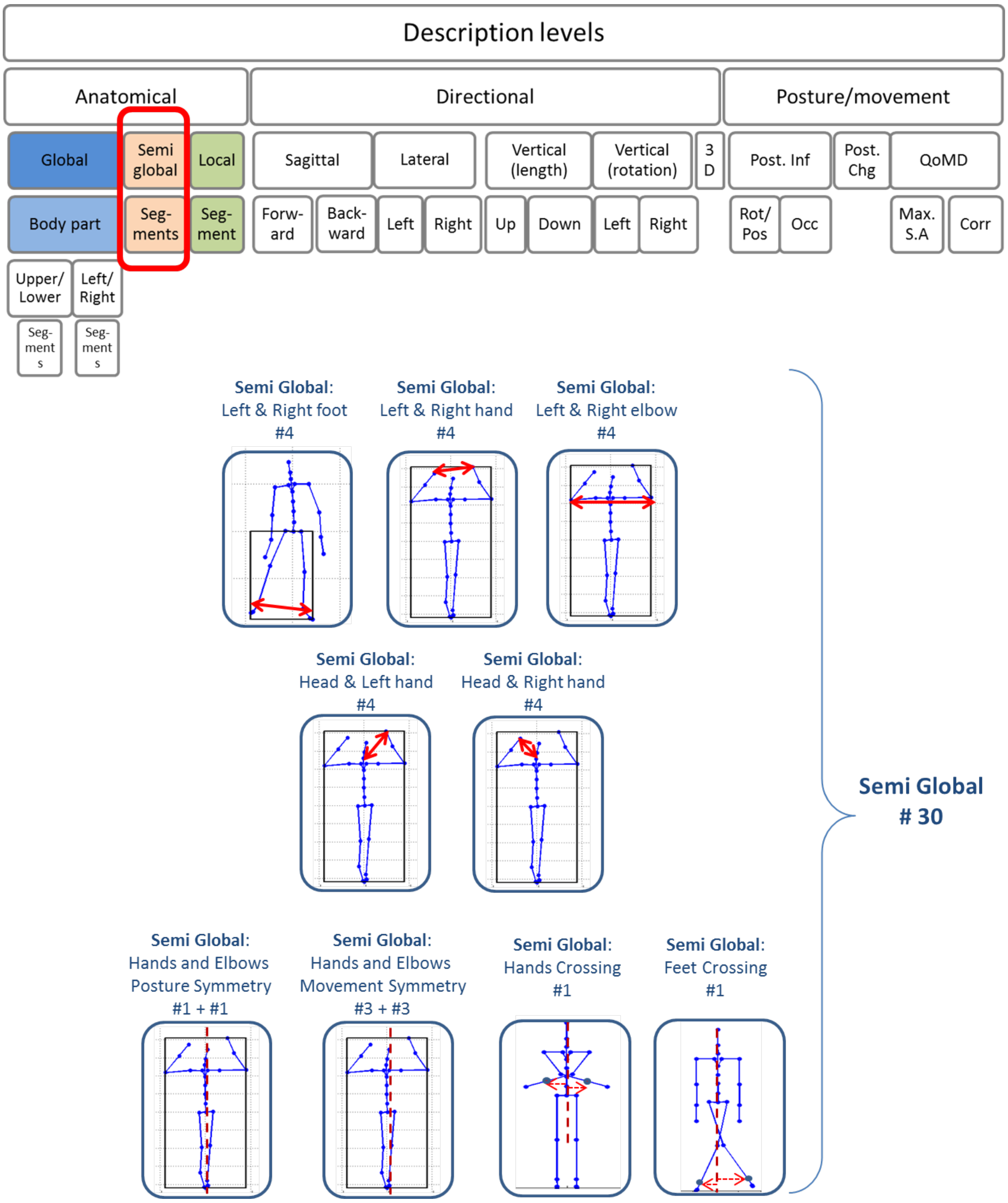


Figure 6.4: Anatomical dimension: Description of Semi Global description level. # stands for the number of features that describe each body cue.

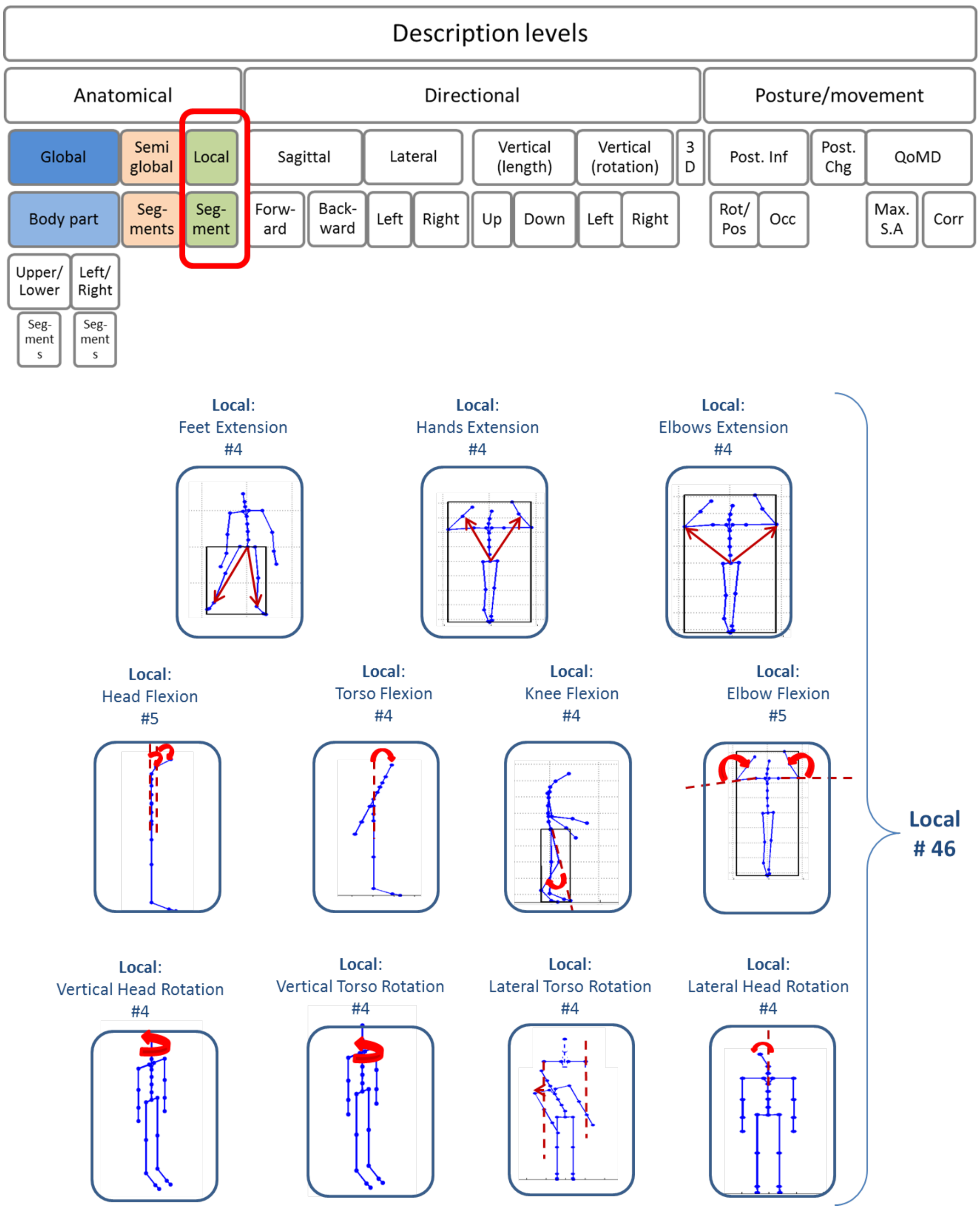


Figure 6.5: Anatomical dimension: Description of Local description level. # stands for the number of features that describe the bounding box.

speed of head flexion, the maximal acceleration of head flexion, and finally the standard deviation of head flexion motion. The notation of body cues and features will be used in the followed discussions.

– Global: Global Anatomical level involves the whole body to refer to the 3D bounding box surrounding body shape based on the Sagittal, Vertical and Lateral direction. But it can also refer to the bounding box surrounding some particular body parts. We define three body parts:
  – Lower body limbs
  – Arms
  – Trunk: involves the torso and the head (it is referred to as TorsoHead).

  Figure 6.3 shows the bounding boxes surrounding the whole body, Lower Body limbs, Arms and Trunk. In Figure 6.3, the bounding boxes surrounding Body Parts are shown from a side view, but they actually refer to the 3D bounding boxes. We note that the bounding box that surrounds the whole body is described through 16 features. Similarly, the bounding boxes surrounding Lower body limbs, Arms and the Trunk are described respectively through 16, 16 and 6 features. These features result from Directional and Posture/ Movement description levels. They will be illustrated later on in section 6.2. In total, we define 38 features for Body Part description level and 16 features for Global description level.
– Semi-Global: Semi-Global level involves some specific body segments to refer to the coordination, the relationship or the relative positions of multiple body segments. For instance, the distance between the head and the right hand refers to the relative position of the right hand according to the head position. The correlation between right and left hand movement refers to the coordination between hands motion.

  Figure 6.4 shows the set of 11 body cues that we consider in Semi Global description level. These body cues are:
  – Feet Relationship
  – Hands Relationship
  – Elbows Relationship
  – Hands Posture Symmetry
  – Elbows Posture Symmetry
  – Hands Motion Symmetry
  – Elbows Motion Symmetry
  – Right Hand - Head Relationship
  – Left Hand - Head Relationship
  – Arms Crossing
  – Feet Crossing

  Similarly to what we report for Global description level, each body cue in Semi Global description level is described through a number of features (# refers to the cardinal of features) (See Figure 6.4). A visual representation of these body cues is provided in Figure 6.4. They are discussed and described

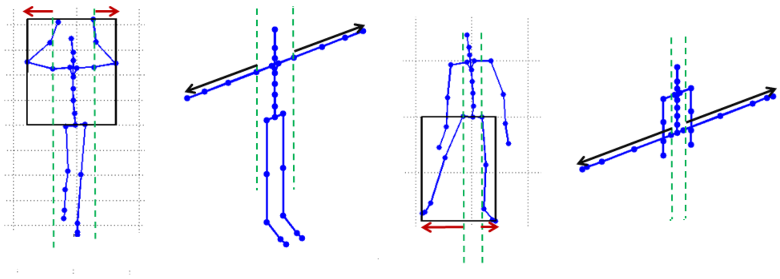


Figure 6.6: Lateral extension of arms and feet (red arrows) measured as a percentage of extension according to the corresponding neutral (green dashed lines) and maximal Lateral body shape (black arrows).

in details in Section 6.2 and also in Appendix chap:TablesBodyCues.

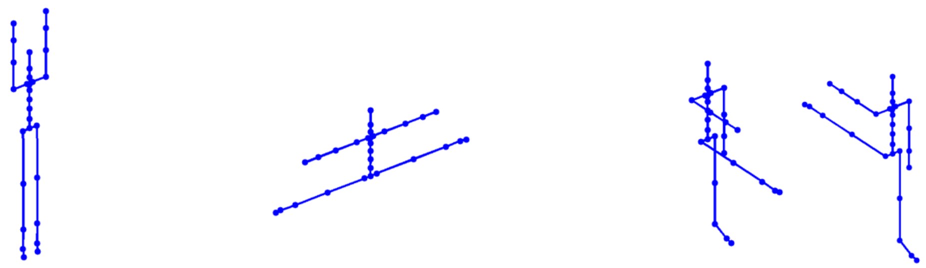
- Local: Local Anatomical level involves one particular body joint to describe its rotation (e.g. downward head flexion) or its position regarding the origin of the coordinate system.
  11 body cues are defined according to the Local description level:
  - Hands Extension
  - Elbows Extension
  - Feet Extension
  - Lateral Torso Rotation
  - Lateral Head Rotation
  - Elbows Flexion
  - Head Flexion
  - Torso Flexion
  - Knee Flexion
  - Vertical Head Rotation
  - Vertical Torso Rotation

  Similarly to what we report for Global and Semi-Global description levels, each body cue in Local description level is described through a number of features (# refers to the cardinal of features) (See Figure 6.5). A visual representation of these body cues is provided in Figure 6.5. They are discussed and described in details in Section 6.2 and Appendix E.

6.1.3 The Posture/Movement dimension

The Posture/Movement dimension aims to distinguish three categories of movement characteristics; features describing the postural information (body shape), features describing the postural change and variation, and finally features describing the dynamics of movement (See Figure 6.1).

- Postural Information (Post): Postural information refers to 1) the maximum/minimum value of a body cue throughout a motion sequence defined either with rotational or positional cue or to 2) the occurrence of specific postural



(a) Maximal Vertical Extension (b) Maximal Lateral Extension (c) Maximal Sagittal Extension

Figure 6.7: Maximal limbs extension

configurations (e.g. crossing or symmetry of left and right limbs).
Using motion capture data, we measure the absolute maximum/ minimum value of a given body cue if only one peak is detected during the whole motion sequence. If more than one peak is detected (e.g. several peaks of elbow flexion in knocking action), we measure the average of all local maximum/ minimum values of a given body cue.
As discussed in subsection 6.1.1, we propose to measure the extension of body limbs as a relative measure according to their maximal extension; so as to take into account the individual skeleton size of each actor (See Fig. 6.6). As such, the maximal/ minimal position of body limbs is proportional to the maximal extension of body limbs.

– Postural Changes (STD): Postural changes refer to the quantity of movement. It is defined in our notation system as the standard deviation of movement.
– Quality of Movement Dynamics (Speed/ Acc/ Corr): The quality of movement dynamics refers to the dynamic properties of a continuous motion. In our work, we focus on the maximal speed and acceleration of a single motion and on the correlation between two time-series. The speed and acceleration of the motion correspond to the first and second motion derivative. The correlation between two motions is obtained through Pearson correlation measure.

Posture/ Movement dimension is used to assess a number of features to describe body cues. As shown in Figure 6.5, 6.3 and 6.4, 4 features are measured for each body cue. These features correspond to 1) a postural feature, 2) a speed feature, 3) an acceleration feature and finally 4) a standard deviation feature.

6.1.4 Summary

The purpose behind the proposed description levels is to give a more comprehensive definition of body movement features. It is useful to characterize emotional body expression in different daily actions. Hence, it allows comparing the classification and the characterization of emotions expressed in different actions. In our Multi-Level notation system, we abstract the meaning of the movement task and we focus on the shape of movement.

Regarding Anatomical description level, we differentiate Global, Semi-Global and Local levels. Regarding the Directional description level, we consider both Vertical (Length) and Vertical (Rotation) components. We also consider the three-dimensional direction in addition to the Sagittal, Lateral, Vertical (length) and Vertical (rotation) directions. This component (3D direction) is useful to describe 3D body limbs extension (e.g. 3D hand extension from the body center) and coordination (e.g. 3D distance between left and right hand).

Our body movement coding system allows the description of expressive body movement within the interpersonal space of the body, referred to in the literature as the LABAN's " personal space" or "Kinesphere" [Laban and Ullmann, 1966]. "Kinesphere" is defined as '*the sphere around the body whose periphery can be reached by easily extended limbs without stepping away from that place which is the point of support when standing on one foot*" [Laban and Ullmann, 1966]. Thus, we do not consider the relationship between the body and the space. In our work, the interpersonal space is defined as the bounding box that surround the body when the maximal Anatomical limbs extensions are reached as shown in Figure 6.7. The maximal limbs extensions that we define does not takes into account the personal flexibility of the body. For instance, maximal backward extension of legs is hard to achieve (See Figure 6.7). However, the maximal limbs extensions are useful to measure the proportional openness of body limbs according to the body size. We take into account the personal size of body limbs to define the maximal bounding box of the body.

In the following section, we present our set of motion capture body features based on our body movement notation system. This set of motion capture features is obtained from the combination of the description levels presented in this section. It will be used for the classification of motion captured bodily expression of emotions (See Chapter 9).

6.2 Motion Capture body features

Based on the proposed body description levels presented in section 6.1, 114 motion capture features have been defined; they are used to represent an expressive movement with 3D motion capture data. The detailed definition and illustration of these 114 features is provided in Appendix E. In this section, we summarize this set of 114 features firstly according to the Anatomical dimension (See Appendix E) and secondly according to the Posture/ Movement dimension of our notation system (See Figure 6.8). Finally, we summarize the quantification of our 114 features according to the Anatomical, Directional and Posture/ Movement description levels. Features name are presented in three levels notation: Posture/ Movement level - Directional level - Anatomical level (e.g. Post-Outside-LowerBody). Appendix E illustrates in details the body cues introduced in Figures 6.3, 6.4 and 6.5.

Description of motion capture features according to Anatomical dimension:

Table E.1 lists the set of features defined in Global Anatomical level (16 features). These 16 features are not included in our set of 114 features. They are only used to compare the classification of expressive movements based on features describing the whole body movement (Global level) and their classification based on features from more detailed level (e.g. Local). This comparative study is illustrated in Chapter 9.

Table E.2 lists the set of features defined in Body Part Anatomical level (38 features). These features describes the bounding boxes that surround the Arms, Lower body limbs and the trunk according to the Sagittal, Lateral and Vertical (Length) directions as well as the Posture, the Speed, the Acceleration and the standard deviation of bounding box movement.

In Table E.3 E.4, we summarize the features described respectively in Semi-Global and Local Anatomical levels. For each body cue, we give the corresponding number of features (# refers to the cardinal of features set). In total, 30 and 46 motion capture features are measured respectively in Semi Global and Local description levels (See Appendix E). Four features are computed over a motion sequence and used to define each body cue: 1) the maximum value of posture (the average of local maximum values or the absolute maximum value), 2) the standard deviation of the motion, 3) the maximum value of speed, and 4) the maximum value of acceleration. Considering for instance *Feet Relationship* body cues, 4 features are measured; 1) the postural information (the maximal 3D distance between the feet), 2) the postural changes defined as the standard variation of the motion (the standard variation of 3D distance between the feet) and 3) the maximum value of the velocity and acceleration of motion (the maximal velocity and acceleration measures of the 3D distance between the feet). In few cases, two postural features are defined; they correspond to the maximum and the minimum values of the posture (e.g. Downward and Upward Head Flexion). Thus, 5 features are used to describe such a body cue.

For the body cues *Hands Posture Symmetry*, *Elbows Posture Symmetry*, and the occurrence of *Arms Crossing* and *Feet Crossing*, the features correspond to the percentage of frames for which the corresponding event occurs.

Description and quantification of motion capture features according to all description levels:

Figure 6.8 shows the distribution of our set of 114 features according to Posture/ Movement description levels. The results are split into 6 blocks of features corresponding to 1) Posture features, 2) Speed features, 3) Acceleration features, 4) Postural Changes (standard deviation) features, 5) Correlation features and finally 6) Occurrence of symmetry and crossing limbs features. For each block of features, we present the direction, the sub-direction (e.g. Forward/ backward for Sagittal direction), the body segments according to Anatomical description levels. Besides, a color-based graphical representation is adopted to distinguish the features according to the Anatomical description levels: Body Part features (in blue), Semi Global features (in orange) and Local features (in green).

Finally, Figure 6.9 summarizes the quantification of the proposed set of 114 features according to each dimension of our body movement notation system: Anatomical, Directional and Posture/ Movement dimension.

6.3 One example: Expressive walks characterization and analysis

Let us illustrate our body movement notation system with an example. In this example, we use motion capture data of expressive walks. Our aim is to evaluate if our body movement notation system allows clustering expressive movements.

For this study, we use the Mockey database proposed by Tilmanne et al. [Tilmanne and Dutoit, 2011]. This database contains different expressive gaits, among which: Proud, Cool, Decided, Sad, and Afraid walks. We manually segment each walk sequence to keep only the straight walk cycles by removing the turn phase.

We use the Principal Component Analysis (PCA) to reduce the data dimensionality and to extract the most important movement characteristics. Three principal components (PC), that have an eigenvalue greater than 1, are retained. The first, second and third PC explain respectively 41.50%, 17.55% and 14.06% of the total variance. Together, they explain 73.12% of the total variance. Figure 6.10 shows the projection of expressive walking samples in the PCA space according to the first and second PC (See Figure 6.10a) and to the first and third PC (See Figure 6.10b). We observe that the styles of walking are highly separated, in particular according to the first and third PC (See Figure 6.10b). In order to find the prominent features that contribute to the linear combination of the resulted PC, we explore the five features that receive the highest coefficients for the first, second and third PC.

According to the first PC, the first five relevant features refer to the speed of hands, elbows and torso movement, elbows flexion and the distance between feet. In both Figures 6.10a and 6.10b, we observe that the first PC allows discriminating Decided, Cool and Afraid styles, while Sad and Proud styles are somehow mixed. Decided style, followed by Cool and Afraid Styles, appear to be characterized with the highest speed of movement, the highest flexion of elbows and the largest steps during walking. The cluster Proud and Sad receive the lowest values.

The features that contribute with high coefficients to the second PC regroup the posture and standard deviation of the flexion of lower body limbs, Lateral arms openness, upward head flexion and the 3D openness of hands. In Figure 6.10a, we observe that the Afraid style is highly discriminated from the other styles. Based on the visualization of the different walking behaviors, we deduce that Afraid style is particularly characterized with high openness of hands, in particular in Lateral direction.

Finally, according to the third PC, the first five relevant features refer the downward flexion of head, forward and backward torso leaning, elbows flexion and forward arms movement. Figure 6.10b shows that we can discriminate between Afraid style, Proud+Cool and Sad+Decided styles based on the third PC. Taking a close look

6

6

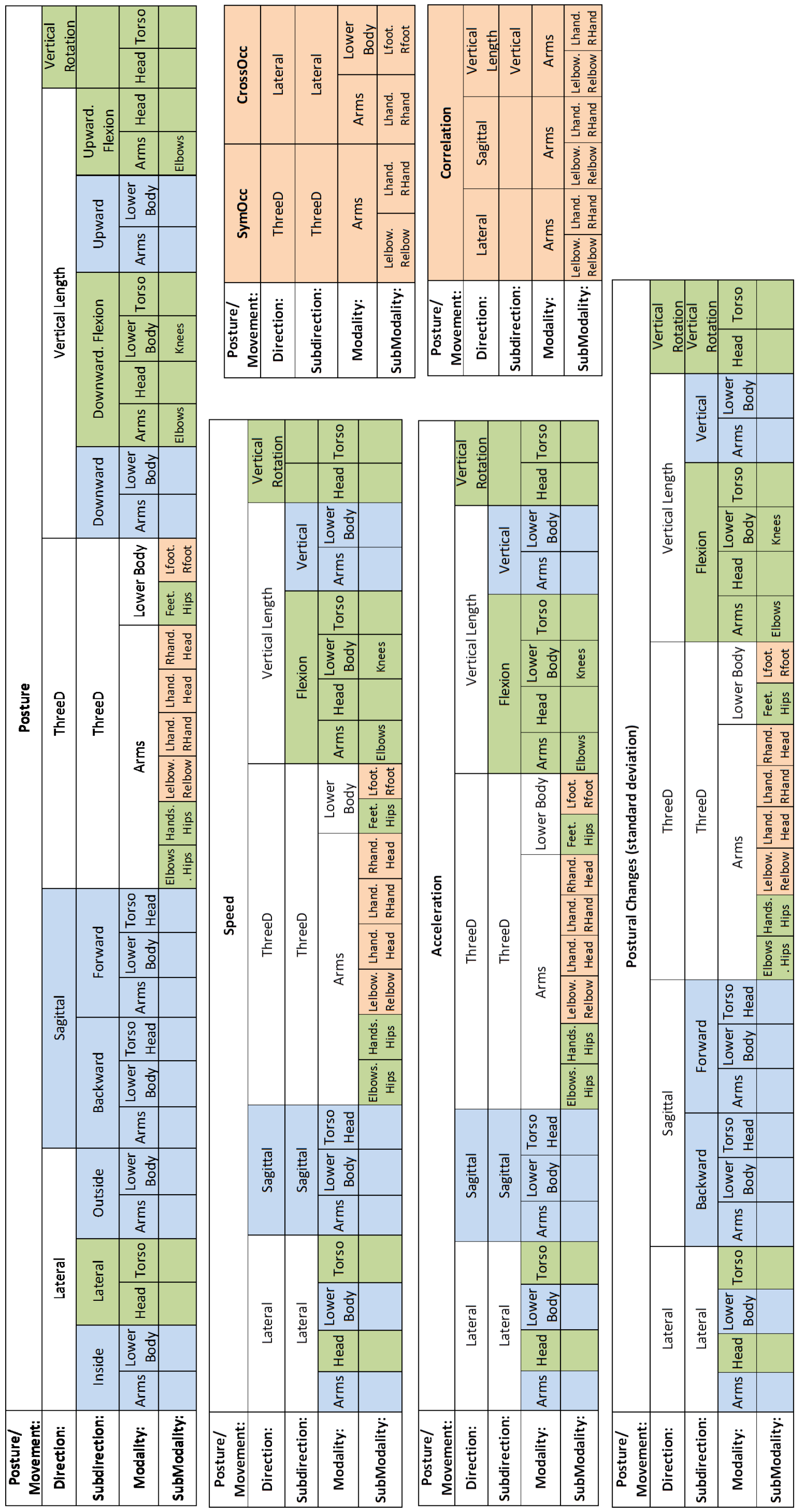


Figure 6.8: Presentation of our set of 114 motion capture features according to all the description levels of our Multi-Level notation system.

Anatomical	Body Part (#38)	Arms	#16
		TorsoHead	#6
		Lower Body	#16
	Local (#46)	Arms	#13
		Head	#13
		Torso	#12
		Lower Body	#8
	Semi Global (#30)	Arms	#25
		Lower Body	#5
	Total:		#114
Directional	Lateral		#22
	Sagittal		#20
	ThreeD		#34
	VerticalLength		#30
	VerticalRotation		#8
	Total:		#114
Posture/Movement	CrossOcc		#2
	Corr		#6
	SymOcc		#2
	STD		#26
	Posture		#32
	Speed		#23
	Acceleration		#23
	Total:		#114

Figure 6.9: Quantification of the set of 114 features according to all the description levels of our Multi-Level notation system.

to the walking behaviors, we deduce that Sad and Decided styles are characterized with high downward head flexion and torso forward leaning, while Proud and Cool are characterized with the lowest downward head flexion, the least torso forward leaning and the highest torso backward leaning.

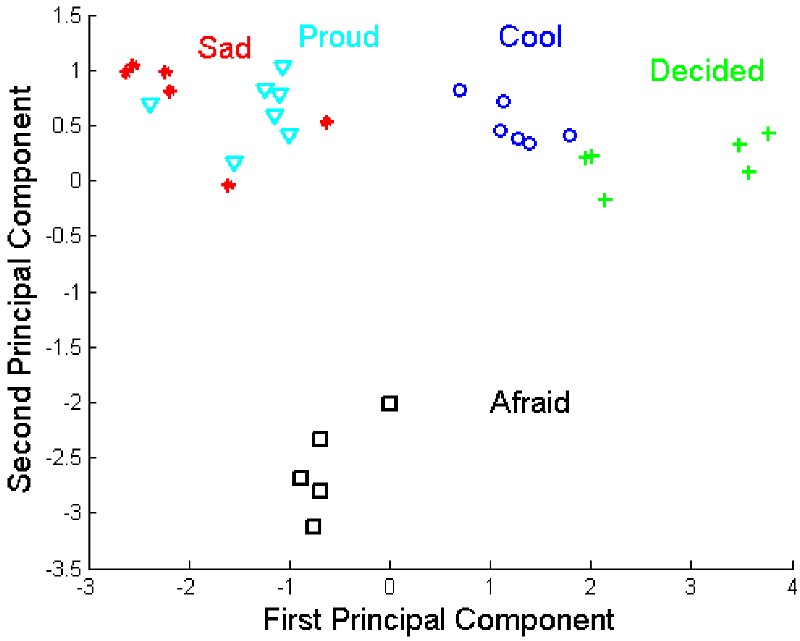
Thus, we conclude that our Multi-Level body movement notation system can be used to discriminate different styles of expressive body movement. In Chapter 9, we will use our Multi-Level notation system to classify emotions expressed in different daily actions.

6.4 Discussion: MLBNS vs related works

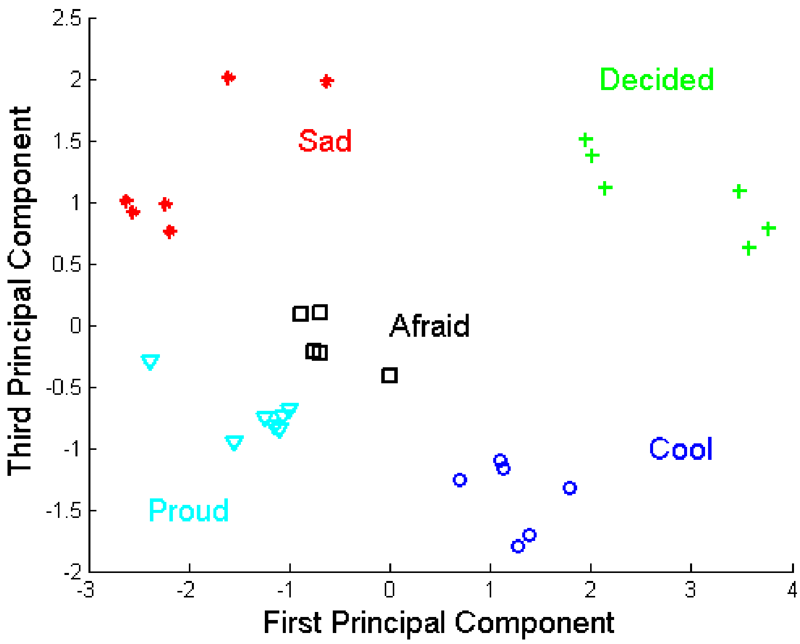
In this section, we briefly compare out Multi-Level body movement notation system with two previous notation systems that have been widely used to describe expressive body movements: BAP [Dael et al., 2012] and LABAN [Laban, 1988] notation systems. These notation systems have been introduced and described in Chapter 2.

BAB notation system has been introduced to describe prototypical emotional body expressions, while MLBNS has been introduced to describe implicit expression of emotions in body movements (i.e. how the expression of emotions modulates body movements). Unlike BAB which focuses on posture and action units, in MLBNS we abstract the semantic meaning of the gesture (e.g. crossing arms) to focus on the body shape (e.g. contracted vs expanded body posture) and the movement quality (described in our work through the velocity and acceleration of each body cue as well as the correlation between the movements of two body segments). Besides, our coding schema differs from BAB notation system at the anatomical description level as we do not focus on the local rotational aspect of body joints (e.g. forward/backward torso movement), but we also consider the description of the bounding boxes surrounding body parts or the whole body, and the relationship between body segments (e.g. the distance between hands).

Figure 6.11 provides a brief comparison between our MLBNS and the LABAN notation system components. For instance, MLBNS does not consider the displacement of the body in space, while it is represented by the Space component of LABAN notation system. Indeed, MLBNS describes the movement of body segments occurring inside the "Kinesphere" [Laban and Ullmann, 1966] (see section 6.1.4). Regarding the Effort component of LABAN notation system, MLBNS only considers the discretization form of some movement dynamics features: the speed, the acceleration and the correlation of joints movement. In LABAN notation system, the Body component defines the body parts involved in the movements as well as their connection, influence and sequencing. In MLBNS, the body is described along the Anatomical description level, spread over the Global, Semi-Global and Local levels. There are three spatial dimensions defined in Space component (horizontal, vertical and sagittal) of LABAN system. In LABAN, shaping movements can be categorized into the following shaping possibilities; spreading/enclosing, rising/sinking, advancing/retreating. Spreading/enclosing corresponds to the extension along the Lateral direction in MLBNS, rising/sinking to the Vertical (length) direction and ad-



(a) Projection of expressive walking samples according to the first and second PC



(b) Projection of expressive walking samples according to the first and third PC

Figure 6.10: Projection of expressive walking samples from Mockey database according to the first, second and third principal component.

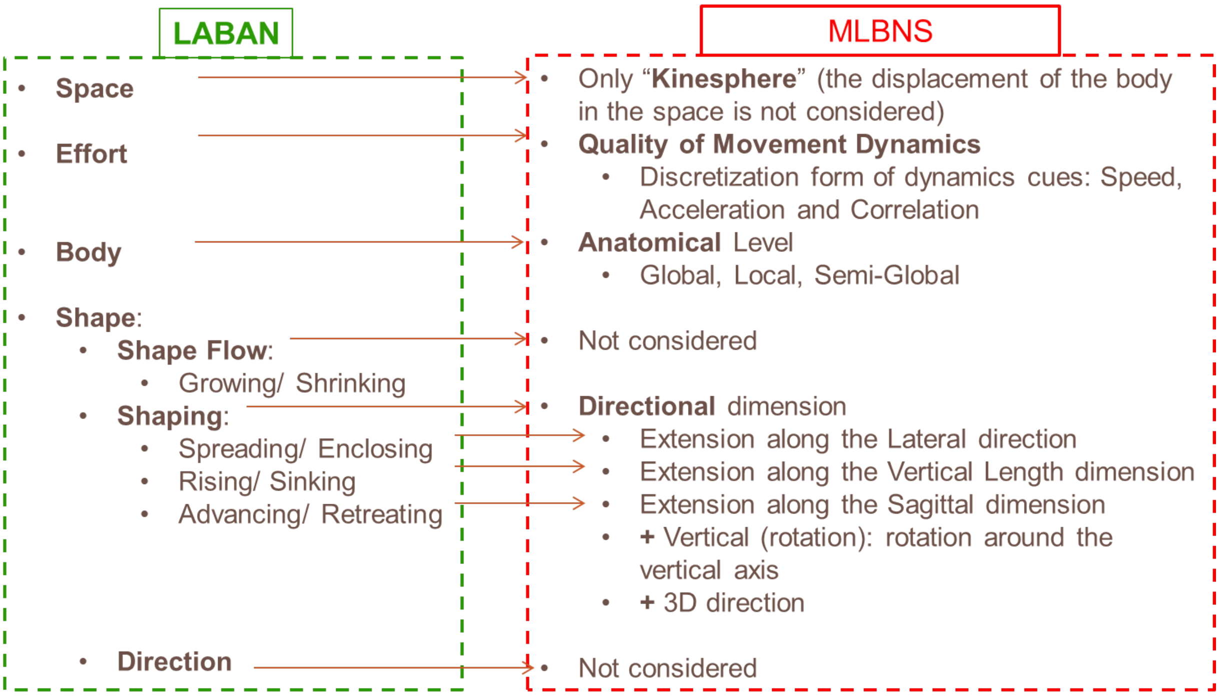


Figure 6.11: Multi-Level Body Movement Notation System (MLBNS) vs LABAN notation system

vancing/retreating to the Sagittal direction. Besides, MLBNS includes the Vertical (rotation) which describes the vertical rotation of body joints around the vertical axis and the three dimensional direction which describes the overall extension of body limbs. Finally, MLBNS does not include Shaping Flow and Directional qualitative aspects of Shape component that are present in the LABAN notation system.

6.5 Conclusions

In this chapter, we described our proposed Multi-Level body movement notation system based on several description levels. Considering different levels altogether provides a more complete illustration of body movement; it also allows the description of different movement tasks in both qualitative and kinematic analysis.

We also presented our set of motion capture features derived from the proposed body movement notation system. We aim to use this set of motion capture features to characterize emotional behaviors in different movement tasks (as walking, lifting and throwing an object).

Our preliminary analysis shows that this set of Mutli-Level motion capture features is able to discriminate between five expressive walking styles. A principal component analysis was performed to study the clustering of expressive walks characterized through a subset of our proposed body cues. Based on the projection of the data on the principal components, we observed that the walk styles are highly differentiated mainly based on the first and the third principal components.

The proposed coding schema can be used for both perceptual (See Chapter 8) and kinematic (See Chapter 9) methods. Considering for example body straightness feature, which refers to postural information (in Posture/Movement level). This body cue can be defined in perceptual experiment according to the length of body shape in upward and downward directions (in Directional level) while considering the global shape of the body (in Anatomical level). The straightness of the body shape involves mainly the trunk, the head and the knees posture [Meijer, 1989]. Hence, if 3D rotation data is provided for each body joint, this body cue can be defined in kinematic analysis by mean of upward/downward rotation of head, trunk and knees (three "local" body segments in Anatomical dimension).

6.6 Summary of Chapter

- We proposed a movement quality coding system encompassing several descriptions levels, namely "Anatomical" (Global, Semi-Global, Local), "Directional" (Sagittal, Lateral, Vertical:Length, Vertical:Rotation and Three dimensional) and "Posture/Movement" (Posture, Postural changes and Movement dynamics) description levels.
- We proposed a set of 114 motion capture features derived from the proposed body movement notation system.

– Our set of motion capture features was firstly summarized in terms of body cues in Figure 6.3, 6.4 and 6.5. It is described in details in Appendix E.
– Based on Mockey database, we showed that the proposed set of 114 motion capture features allows discriminating 5 walking styles: Decided, Cool, Proud, Sad and Afraid.

7 Emilya database collection

In chapter 3, we discussed the different databases that have been created to study emotional body behaviors. In our work, we aim to explore the classification and the characterization of emotional body movements across different daily actions. Thus, we propose a new database of emotion expression called Emilya for EMotional body expression In daiLY Actions.

Emilya database is a new repository of expressive and emotional body movements. It encompasses a larger set of emotions and movement tasks (actions) than the ones considered in previous databases. Besides, Emilya is a multi-media database containing synchronized audio-visual and motion capture data recorded using an inertial motion capture system. In this chapter, we describe the collection and the content of Emilya database.

This chapter is organized as follows: firstly, we describe the procedure that we adopt to record emotional body expression. Secondly, we describe the recording of motion capture data and its synchronization with video recording. Then, we describe the content of Emilya database. Finally, we conclude this chapter and we provide a summary of its content.

7.1 Emotional expression recording

In this section, we describe the procedure used to record emotional body expression and the training sessions given to the actors with the help of a professional director.

7.1.1 Daily Actions

In our work, we focus on emotional body expression in daily actions. To build our database, we consider a wide range of daily actions that involve the whole body, in particular the upper body and arms for their involvement in communicating emotional states [Kleinsmith et al., 2011]. The actions are walking, Sitting Down (SD), Knocking at a Door (KD), Lifting (Lf) and Throwing (Th) an object (a ball made of paper) with one hand, and Moving objects (Books) (MB) on a table with two hands. For the walking actions, we asked the actors to walk back and forth along the long side of the recording room. Two types of walks were considered in our database to capture two types of arms behavior during walking action: a Simple Walk (SW) and a Walk with an object in Hand (WH). Some of those actions were already used in past studies and considered as relevant to discriminate between different styles of the same movement [Tilmanne and Dutoit, 2011, Pollick et al., 2001, Gross et al., 2010]. We asked the actors to perform each action four times in a row to capture a large set of data. Each actor was free to repeat the actions without any constraints on how exactly it should be; the repeated performances could be different in term of movement quality or form. As such, the Emilya database contains an intra-variation of emotional behaviors in addition to the inter-variation obtained from different actors.

A continuous sequence consisting of the series of all the actions with just one trial per action was also recorded.

7.1.2 Scenarios

One of the most agreed upon issues of emotion oriented databases is the natural aspect of the expressive behavior. Due to the complexity of recording natural expression of emotions in daily actions with motion capture data, we turn our attention to the recording of acted data. A scenario-based approach is adopted for data induction as it provides a good compromise between its reliability to induce affects on demand and its simplicity from the perspective of the actor. Each scenario includes the description of a situation, which is assumed to elicit a given emotional state [Bänziger et al., 2006]. During the recording of each motion sequence recording, the actor was asked to read the scenario which is written on a paper, to imagine that he/she is living this situation and to perform the proposed actions.

Most of the scenarios that we used to elicit emotions on demand correspond to the scenarios used to collect emotional behaviors in GEMEP database [Bänziger et al., 2012] and others proposed in the work of Scherer et al. [Scherer et al., 1991]. However, as more scenarios were required, we proposed other scenarios following the same approach as in [Scherer et al., 1991] and [Bänziger et al., 2012]. We asked 10 participants to select the emotion label that corresponds the most to each of these scenarios. The results showed that the participants mostly recognize the emotion that corresponds to the scenario. During each recording session, the actors were provided with the scenarios, the emotion that is attached to each scenario as well as the emotion definition. In Table F.1, we precise the source of each scenario that we used. Combining the scenarios proposed in [Bänziger et al., 2012] and [Scherer et al., 1991] and the scenarios that we proposed, we obtained four different scenarios for each emotional state. During the recording sessions, only three scenarios were used. The fourth one was used to replace a scenario when the actor felt unable to imagine the specific situation.

Two scenarios used respectively for Anger and Sadness emotions are provided here as examples, but the whole list of scenarios is provided in French in Table F.1:

"I hoped to sleep late on Sunday morning, but my neighbor started very noisy work in his house at 7am. I felt so angry that I decided to go and scold him."

"I got a call to tell me that my favorite aunt suddenly died."

The emotions considered in our study are Joy (Jy), Anger (Ag), Panic Fear (PF), Anxiety (Ax), Sadness (Sd), Shame (Sh), Pride (Pr) and Neutral (Nt). Those emotions were selected to cover the arousal and valence dimensions. It has been shown in previous works that the expression of those emotional states can last a period of time [Dael et al., 2011], which makes their expression through body movement more or less extensible unlike reactive emotions such as surprise.

7.1.3 Actors

The actors were eleven (6 females and 5 males) graduate students. The mean age was 26 ranging from 23 to 28. They were motivated to participate to the construction of our database and they gave informed consent that their motion capture data as well as their video could be used and published for research purpose. Besides, we hired a professional acting director to give the actors 7 training sessions, each lasting three hours. The acting director was also invited for one recording session and was paid for his services. Each training session lasted three hours in which the acting director tried to keep the actors at ease as much as possible.

7.1.4 Training sessions

During the training sessions, the acting director made the actors aware on how to use their body to express affects through daily body actions. At no time, did the acting director show how an action should be performed. Rather, the training sessions consisted of improvisation tasks; each actor was asked to create and play a scenario and to communicate an intended emotion through some daily actions. S/he had to do so through her/his body movement; s/he could not talk. The other actors had to recognize the emotion s/he tried to express based on her/his expressive body movement (not from her/his facial expressions). Each actor was free to express the proposed emotions without any constraints on how exactly it should be. The acting director did not aim to align the different styles of actors' expressivities. His goal was to ensure the successful communication of emotion expressed through body movement. We were also aware about the risk of obtaining exaggerated expressive behaviors. Thus, we explicitly asked the acting director to avoid exaggerated behaviors while working with the actors during the training sessions.

7.1.5 Recording sessions

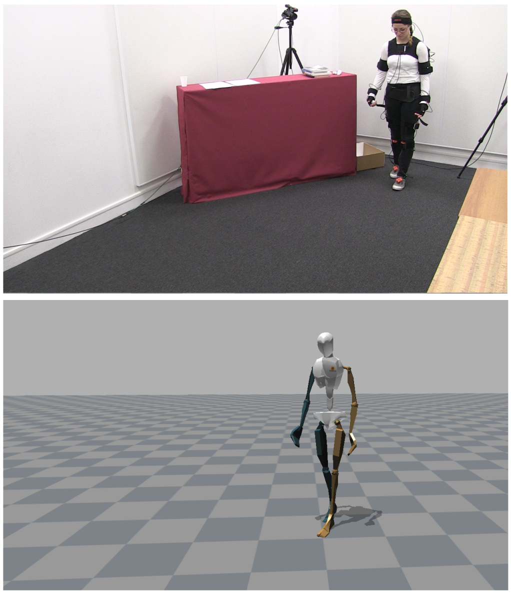


Figure 7.1: The extraction of the same posture from one video sequences (related to the second camera) and 3D motion capture data

Recording sessions lasted approximately four hours including the break times and the calibration steps. Due to the actors' schedules and sometimes to some delay related to technical issues of recording, the recording sessions were sometimes split into two recording sessions.

During each recording session, the actors were provided with the scenarios, the emotion that is attached to each scenario as well as the emotion definition. The order of emotions, scenarios as well as actions was randomized for each actor. Continuous sequences were recorded only for two scenarios (and not for three scenarios as it is the case for individual actions) because we wanted to reduce the recording, as the database is quite large with individual actions, and it takes time to record all the actions for all the scenarios.

In the following of this thesis, CS, SW, WH, MB, KD, SD, Lf and Th stand respectively for the actions Continuous Sequence, Simple walk, Walk with an object in hand, Move books, Knock at the door, Sit down, Lift and Throw an object. We also refer to Anxiety, Pride, Joy, Sadness, Panic Fear, Shame, Anger and Neutral respectively as Ax, Pr, Jy, Sd, PF, Sh, Ag and Nt.

7.2 Body movement recording

Bodily behaviors can be recorded using audiovisual recording and/or 3D motion capture recording (See Chapter 3). Audio-visual recording provides 2D digital video that contains the visual content of body movement and audio information. Although the analysis of digital video can be used for producing computational models of multimodal behavior [Camurri et al., 2004], 3D motion capture of body movement provides more accurate information about 3D posture and movement of some specific body joints of subjects, allowing one to produce more accurate computational models [Roether et al., 2009].

In our database, we record both audio-visual and motion capture data. We use specific hardware to perform the synchronization between those recording types (See Section 7.2.3). The motion capture took place in a professional audio-visual recording studio in our institute, Telecom ParisTech in France. For each sequence of body movements, we recorded 3D motion capture file as well as video file.

7.2.1 3D motion capture:

We used the inertial motion capture system Xsens [Roetenberg et al., 2009] to record the 3D motion data of the whole body. The MVN Mounting straps were used as the straps allows interchanging setups between different subjects. Indeed, the straps are adaptable for different subject sizes. 17 Motion Trackers (MTx) were used to capture the movements of 23 body segments. Each Motion Tracker contains 3D linear acceleromaters, 3D rate gyroscopes and 3D magnetometers. The 17 Motion Trackers are attached through the straps to the body segments to measure their motions. They are placed around the pelvis, the sternum, body extremities (hands, feet and head) and finally to the lower and upper arms and legs as well as the shoulders. The orientation and the position information are obtained for each body joint. Figure 7.1 shows the visualization of motion-captured data reproduced on a computer avatar.

7.2.2 Video acquisition:

Besides the 3D motion capture data obtained for each sequence of movement, we recorded four MXF (Material eXchange Format) video files of full HD resolution (1280*720) from four cameras placed in the four corners of the studio. Two cameras were dedicated to capture a general view of the room while the other two cameras were placed carefully to capture the face and the upper body when the actor performed actions involving mainly upper body. Canon XF105 cameras were used for this purpose. We recorded audio even though we did not explicitly ask actors to express emotions through the voice.

7.2.3 The synchronization of video and 3D motion capture data

The video files were synchronized with the motion capture files through some specific hardware. The Rosendahl nanosyncs HD, a professional video and audio sync reference generator, was used to generate a common Time Code (TC). The generated TC is read from the cameras through the outlet Genlock/TC. The Alpermann card PLC PCIe is used to read the TC in the computer. Using the MVN

Figure 7.2: Still frames depicting three viewpoints (corresponding to three cameras) relevant for the visualization of the Moving books action

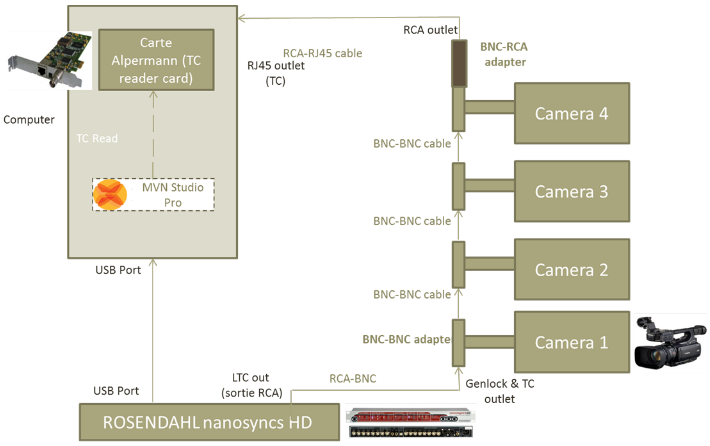


Figure 7.3: Synchronization of video and 3D motion capture data

time code plug-in of the MVN Studio software (Xsens software), we were able to read the TC from the Alpermann card. Thus, for each sequence of movement, the corresponding motion capture file contains the same TC as the video files recorded from the cameras (See Figure 7.3). The TC integrated in motion capture and video files is useful to extract the same sequence of movement in both audio-visual and motion capture data (See Figure 7.1).

7.3 Data post-processing and quantification

As explained in section 7.1, the actors were asked to express emotions in different actions, where each action is repeated four times. The segmentation of motion sequence aims to separate the different types of action (walking, knocking, sitting down...) and the different trials of the same action.

As the order of actions sequencing was different from one actor to another (randomly chosen), and given the fact that the actors were not asked to perform a particular “key” posture between the different actions, it was difficult to develop an automatic approach for the segmentation of those different actions. Thus, a manual segmentation was preferred to separate the different types of action.

While the segmentation of the motion sequence into different actions was performed manually, the segmentation of the same action into single trials was performed manually, semi-automatically or automatically according to the performed action as explained in the next paragraphs.

7.3.1 Data capture files post-processing

Three types of segmentation process were used in order to separate multiple repetitions of the same action; Manual, Semi-Automatic and Automatic.

Manual segmentation: :

A manual segmentation was performed for Lifting, Throwing and Sitting down actions as the start and the end of these actions was difficult to detect automatically. A detailed description of the beginning and the end of each action was provided as a reference for manual segmentation.

Semi-Automatic segmentation: :

A semi-automatic segmentation was performed for Knocking at the door and Moving Books actions. Although we always give the instructions to move to a neutral pose (called N−Pose) between each successive action repetitions, actors sometimes forget to perform this step. Therefore, a manual segmentation was performed for the actions of Knocking at the door and Moving books when the actor did not return to N−Pose between successive trials. The automatic segmentation of these actions (in which the actor returned to N−Pose) was performed automatically based on kinematic measures related to the position of the hands along the vertical axis.

Automatic segmentation: :

7

An automatic segmentation was performed for Walking actions (Simple Walking and Walking with an object in the hand) to separate when the actors walked along the long side of the recording room and when they turned. The automatic segmentation of the walking action was based on the work conducted on the analysis of walking and turning task (See Chapter A). In Chapter A, we investigate the relationship between upper and lower body parts through the relationship between shoulders and hips movement. We used our database to investigate this relationship during turning when expressing different emotions and the eNTERFACE08 3D [Tilmanne et al., 2009] database to study neutral turning behavior with different angles. Like hips movement, we found that shoulders movement is characterized with a non linear (sinusoidal) behavior during straight walk. However, shoulders and hips are in opposite of phase during walking, mainly due to arm swings. Based on the relationship of shoulders and hips turning angles, we found that - unlike straight walk - shoulders and hips movement follow a strong linear relationship during turning task for different turn angles (45 $^\circ$, 90 $^\circ$, 135 $^\circ$ and 180 $^\circ$) (See Chapter A). We found that the linear relationship is stronger for higher turn angles. In Emilya database, the actors are asked to walk back and forth along the long side of the recording room. Thus, they performed 180 $^\circ$ turning between two successive straight walks.

Therefore, the separation between walking and turning tasks was based on the detection of the Turn Interval Time in which the relationship between hips and shoulders is the most linear. The Turn Interval Time is defined by the start (onset) and the end (offset) of a given turning behavior. The approach used to detect the start and the end of each turning behavior is described in Chapter B.

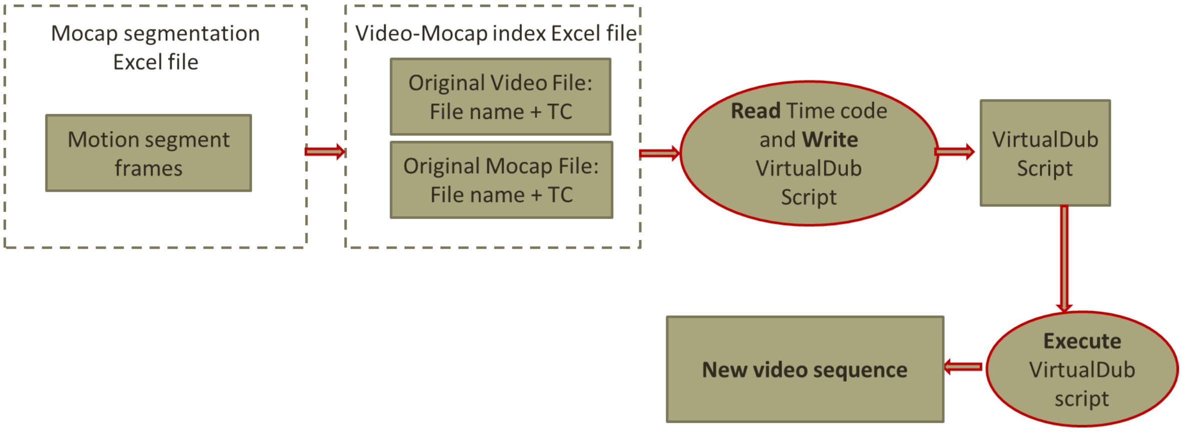


Figure 7.4: Automatic video files segmentation based on TC and motion-capture files segmentation frames

7.3.2 Video files post-processing

Since video and motion capture data were synchronized, we can find the start and the end of the video sequences that correspond to motion capture data using the TC information. Consequently, the process of video files segmentation (MXF files) was based on the results of motion capture files segmentation using MediaInfo JAVA library, VirualDub tool and the start and end TC of the segmented motion capture files. The videos resulted from the video segmentation process were compressed and converted to AVI format (Audio Video Interleave).

The process of video files post-processing is described in Figure 7.4. The process mainly consists of two steps: 1) write and 2) execute a VirtualDub script. First of all, we dispose of two Excel files. The first Excel file (called *Mocap segmentation Excel file*) contains the motion segment frames (resulted from manual, semi-automatic or automatic segmentation of mocap files). The second file (called *Video-Mocap index Excel file*) contains the file name and the TC of each mocap file and the corresponding video files (each corresponding to one viewpoint). Based on the frames of mocap file segmentation, the first and the last TC of original mocap file, the first and the last TC of the original video file, we are able to identify the frames of the new video segment. Using a JAVA program, we automatically create a VirtualDub script that contains the necessary commands to read each video file and to create a new sequence of video segment. A single VirtualDub script may contain several commands allowing the processing of several video files.

7.3.3 Data quantification

The segmentation process led to the duplication of the motion sequence according to the different actions and the individual repetition of each action. Each actor

Table 7.1: Motion capture data quantification per action and emotion. NbV stands for the number of viewpoint of the camera relevant to visualize bodily behaviors in the videos.

	Ax	Pr	Jy	Sd	PF	Sh	Ag	Nt	Total	NbV
CS	22	27	23	26	24	27	25	22	196	x4
SW	135	135	135	128	137	135	127	93	1025	x4
MB	141	136	129	131	145	130	123	96	1031	x3
KD	135	135	134	126	141	140	123	93	1027	x2
WH	132	136	134	126	136	136	128	94	1022	x4
Th	140	136	132	120	136	124	126	92	1006	x3
Lf	133	137	133	126	134	136	132	88	1019	x3
BS	133	136	133	136	132	137	135	96	1038	x2
SD	133	136	133	136	132	137	135	96	1038	x2
Total	1104	1114	1086	1055	1117	1102	1054	770	8402	

was asked to express 8 emotional states in 7 actions. Each described through 3 scenarios (except for Neutral which was described through 2 scenarios only). Only one scenario per emotion was used with the professional acting director. For each emotion scenario, the actors performed each action 4 times. As such, for a particular emotion expressed in a particular action, we obtained around 136 motion sequences: 3 emotion scenarios * 4 action repetitions * 11 actors + 1 emotion scenario * 4 action repetitions * 1 professional actors. Besides, for a particular emotion expressed in a continuous sequence (CS), we obtained around 22 motion sequences: 2 repetitions of CS * 11 actors.

Table 7.1 describes the quantification of motion capture data per action and per emotion. The number of motion sequences correspond to the "theoretical" numbers presented above minus some segments for which the recording session did not work properly or plus some extra repetitions. The quantification of video files is based on the number of viewpoint of the camera according to each action (See Table 7.1). In total, we obtained 8402 motion capture sequences composed of 8206 single actions sequences plus 196 continuous sequences.

7.3.4 Files representation

The final video sequences resulted from the automatic segmentation of the original video files are saved as .avi files. Motion captures files are represented through three types of files;

- MVN files; which are binary files that can be read only with MVN Studio software,
- Bvh files (biovision hierarchical data); which are standard animation files.
- Excel and Matlab files; which contain all the information related to the 3D

motion. Each file contains root-related positions, world-related positions or orientation in Angle-Axis representation.

Root-related positions are obtained from world-related positions through a change of the Cartesian Coordinate System. The process of changing the Cartesian Coordinate System consists of rotating the world-related 3D positions according to the inverse of body orientation (See Figure 7.5). The body orientation is estimated through the turning angle of hips. As explained in Chapter A, the projection of right hip position and left hip position along the X and Z axes informs us about the turning angle of the hips vector (Y being the vertical axis).

In our analyses based on motion-capture data (Chapter 9, 10 and 11), we use Matlab files that correspond to root-related positions and orientation in Angle-Axis representation. We do not use world-related positions as we do not consider the displacement of the body in the space in our body movement coding scheme (See Chapter 6).

Besides, two types of Excel files are used in our database in order to save all the information related to the segmentation of motion sequences. The first type of Excel file (referred to as *Mocap segmentation Excel Files*) is related to the segmentation process results. They contain the frames defining the start and the end of each created motion segment. The second type of Excel file (referred to as *Video-Mocap index Excel files*) is related to the matching between the motion capture file and the video sequences obtained after each recording session. As explained in section 7.3.2 and Figure 7.4, these Excel files were useful to automatically segment video files.

7.4 Conclusion

In this paper, we introduce a new database of emotional body expression in daily actions. This database constitutes a rich repository of emotional expression in body movements. Eleven actors expressed 8 emotions while performing 7 actions. Each action was repeated four times to capture a wide range of data. During the recording session, we tried to keep the actors at ease as much as possible at the cost of extensive manual post-processing. Indeed, we did not ask the actors to return to a particular posture between different actions. Such a restriction could be helpful to ease the automatic segmentation task, but it could also affect the naturalness and the continuous emotional behavior. Besides, we did not impose restrictions to perform the actions such as the number of knocks at the door in order to foster inter-variability in emotion expression.

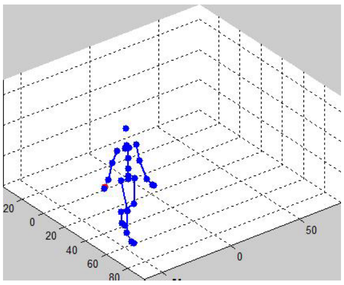
Our database includes synchronized audio-visual and motion capture recording. While the video files include facial and bodily emotional expression visualized from different camera viewpoints, motion capture files include three-dimensional data of the whole body movement.

Composed of more than 8000 motion capture and video files of the expression of basic emotions and other in different daily actions, Emilya database aimed to characterize emotional body expression across different daily actions.

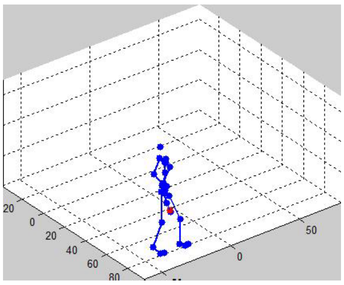
Our next step is to validate this database. That is we want to examine if the emotions that were intended to be expressed by the actors are perceived as such by naive participants. We also want to understand where confusion in perceiving emotions arises and between which emotions. This validation step is described in the next chapter.

7.5 Summary of chapter

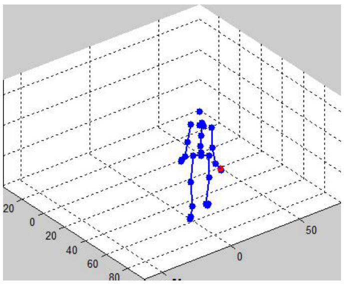
- To collect Emilya database, we follow induction technique where scenario-based induction technique was adopted to induce emotions as it provides a good compromise between its reliability to induce affects on demand and its simplicity from the perspective of the actor.
- For each action, the actor was asked to express each emotion by mean of three scenarios (except for Neutral for which we use only two scenarios) and to perform 4 repetitions for each action.
- We record both digital videos and 3D motion capture data in Emilya database providing a rich multi-media dataset of emotional and expressive body movement.
- We obtained more than 8000 motion capture and video files of emotional body expression in daily actions.



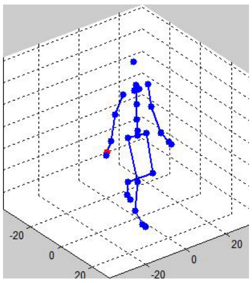
(a) World-related positions in frame i



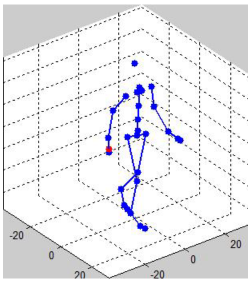
(b) World-related positions in frame j



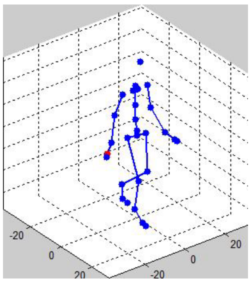
(c) World-related positions in frame k



(d) Root-related positions in frame i



(e) Root-related positions in frame j



(f) Root-related positions in frame k

Figure 7.5: World-related positions versus root-related positions in three frames (named i, j,k) of a motion sequence involving walking and turning

8 Emilya database validation

In this chapter, we use the Emilya database to explore three issues based on a perceptual study: 1) how intended emotions are perceived, 2) how intended emotions are characterized through expressive body cues, and 3) how intended and perceived emotions are automatically classified based on the human rating of body cues. We address these three issues through two main studies which are "Emotion Perception" and "Body Cues Rating". Henceforth, we refer to the emotion we asked the actors to portray as expressed or intended emotion and to the emotion that was perceived by the participants in a given stimulus as perceived emotion.

In this chapter, we firstly provide a brief overview of previous works on the labeling and the perception of emotional body behaviors in section 8.1. In section 8.2, we describe the design of our perceptual study intended to validate the Emilya database. Before discussing the results of this perceptual study, we present the inter-rater consistency in section 8.3. Section 8.4 provides the results of the emotion perception study. Section 8.5 discusses the characterization of emotional body expression. Section 8.6 reports the results of automatic classification of emotions based on perceptual body cues ratings. In section 8.7, we compare the results of emotion perception provided from two different pool of participants. Finally, we conclude and summarize this chapter.

8.1 Labeling and perception of emotional body behaviors in related work

Labeling emotional body behaviors of a database can be used to investigate the multimodal expression of emotional states, to train and test automatic recognition model [Kleinsmith et al., 2011], and to establish lexicon associating multimodal behaviors and emotional states [Niewiadomski et al., 2011]. Several labeling methods have been proposed.

Self-reported feeling and third observers' judgment: Self-reported feeling (user's self-labeling) [Kapoor et al., 2007, Gross et al., 2010] or third observers' judgment (through perceptual experiments) [Kleinsmith and Bianchi-Berthouze, 2013] have been used for emotion expression labeling. In the first approach, the goal is to ask the subjects (here, the actor) what they really felt when their emotion expressions were recorded [Kapoor et al., 2007, Gross et al., 2010]. However, it is very difficult to discern what one really felt when portraying an emotion [Kleinsmith et al., 2011]. Kapoor et al. [Kapoor et al., 2007] assert that "*Self-reported feelings at the end of a task are notoriously unreliable*". Besides, as self-reported feeling only indicates to what extent the emotion expressed by the subject was really felt, this approach does not provide insight on which cues (such as gaze, facial or bodily expressions) participate to the perception of the emotion expression. In the second approach, observers (expert or naive observers) are asked to give their perception of the emotional behaviors of actors. This latter approach, referred to as observers labeling approach, is the most used to label the expressed emotions in databases of emotion expression [Kleinsmith et al., 2011][Gross et al., 2010].

Response format of observers' ratings: The response format of observers' ratings has often been based on forced-choice option, where the attribution of a single label to an expressed bodily behavior is required and the most frequent label is used. However, as reported in [Kleinsmith and Bianchi-Berthouze, 2013], it is important to consider the variability of observers' ratings and to go beyond the forced-choice option using multi-labeling techniques. Many other factors are important and need to be considered with an observer labeling approach such as the observer's level of empathy, the observer's culture and the observer's mood during the perceptual experiment [Kleinsmith and Bianchi-Berthouze, 2013]. However it is very complex to measure their impact on the labeling task. In our work, we do not consider the effect of such contextual factors on emotion labeling.

In the following sections, SW, WH, MB, KD, SD, Lf and Th stand respectively for the actions Simple walk, Walk with an object in hand, Move books, Knock at the door, Sit down, Lift and Throw an object. We refer also to Anxiety, Pride, Joy, Sadness, Panic Fear, Shame, Anger and Neutral respectively as Ax, Pr, Jy, Sd, PF, Sh, Ag and Nt.

8.2 Perceptual experiment design

We have designed an online perceptual study where participants were asked to perform two tasks for each stimulus: 1) rate the emotion they perceive and 2) rate some expressive body cues. In the following sections, we respectively use "Emotion Perception" and "Body Cues Rating" terms to refer to the first and the second tasks.

8.2.1 Stimuli creation

When designing our protocol, the question of the content of the stimuli to be shown to participants was raised: should it be real videos of an actor performance or videos of a virtual actor that reproduces the expressive movement recorded through motion capture data? Using a computer actor that has the shape of a puppet with no sign of gender, culture and facial expression allows reducing bias of these factors in perception studies [Kleinsmith et al., 2011, Volkova et al., 2014b]. In our perceptual study we follow the same approach; the movement of the actors is reproduced on a virtual puppet (See Fig. 8.1). No facial expressions are displayed in the face of the virtual actor but the direction of gaze (given by the nose direction) is visible (See Fig. 8.1). For a given motion capture segment we created a video (stimulus) of the virtual actor reproducing the recorded movement. This process was done automatically; the creation of stimuli is implemented using MAXScript scripting language of Autodesk 3DS Max software (See Appendix C). To ensure that the distance between the virtual camera and the virtual character remains constant (so that it appears with the same size in each stimulus), once again with the aim to reduce any bias when viewing the stimuli, the virtual camera motion follows a non-uniform linear style (See Appendix C), and the position of the virtual camera is automatically set according to a specific viewpoint (defined by an angle of 30 ° at the right of the virtual actor, see Fig. 8.1) (See Appendix C).

8.2.2 Stimuli selection

Since the Emilya database consists of more than 7000 emotional behaviors sequences, we conducted this perceptual study using a subset of the Emilya database (this subset is selected from the set of 8206 expressive body movement sequences, See Chapter 7). It is made of 664 segments. As we aimed to evaluate the perception and the characterization of the expression of all the emotions in all the movement tasks, we selected a subset of the Emilya database by considering randomly one scenario per emotion and one trial per action for all the actors. For instance, for each set of 12 expressive motions performed by one actor (3 scenarios of an emotion * 4 trials of an action), we randomly selected one sample. Most of the actions recorded in the Emilya database consist of repetitive actions (walking, knocking, moving books). For instance, walking is a repetitive pattern involving steps. For those actions, the corresponding stimuli depict a single trial of the action (e.g. moving all the books from left to right). Sitting Down is not a repetitive action, but the corresponding stimuli depict a single trial of Sitting Down action. However, a single trial of Lifting or Throwing can be of too short duration (e.g. less than two seconds) and it may involve only fingers motion (which is not present in the Emilya motion capture data, thus not visible in stimuli). Thus, for Throwing and Lifting actions, the corresponding stimuli depict the four trials of the action (Lifting or Throwing) in order to ease the perception task.

During recording sessions, one actor (Actor1) forgot to perform “Lift” action. Thus, only 6 actions are considered for this actor. Consequently, 664 stimuli were automatically generated for the perceptual study (1 actor (Actor1) * 6 actions * 8 emotions + 10 actors * 7 actions * 8 emotions + 1 (professional) actor * 7 actions * 8 emotions).

8

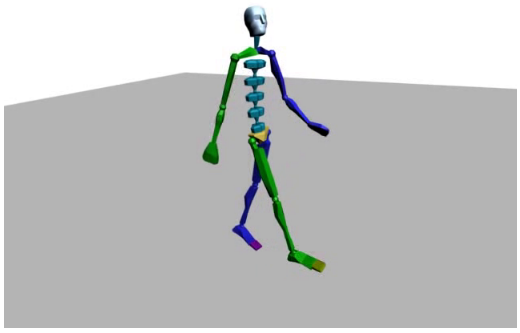


Figure 8.1: One snapshot of one created stimulus using the 3D Studio MAX biped model.

8.2.3 Protocol

8

We found the participants through the Mechanical Turk crowd-sourcing web site. As it is very time-consuming to rate all the 664 segments (without mentioning the 8206 ones, See Chapter 7), each participant had to view only a subset of the stimuli. After conducting some pre-tests on few participants through Mechanical Turk, we have decided to ask each participant to view 16 stimuli (corresponding to the expressive behaviors of 2 actors * 4 emotional states * 2 movement tasks). To cover the set of 664 stimuli, 42 subsets of stimuli (each subset contains 16 stimuli) were created, leading to a set of 672 stimuli (42 subsets of 16 stimuli = the original 664 stimuli + a duplication of certain stimuli). For each new participant, a subset of stimuli is randomly selected and assigned to. As we aimed to obtain 24 answers for each subset of stimuli, we asked for 1008 participants (42 subsets of stimuli * 24 answers).

Our perceptual experiment consists in an online survey, where naive participants were asked to evaluate emotional body expressions recorded in Emilya database based on two tasks: 1) "**Emotion Perception**" task (rate the perceived emotions) and 2) "**Body Cues Rating**" task (rate some expressive body cues that characterize emotional body expressions) (See Figure 8.2). For the first task, the participants were asked to rate their perception of emotion through a multi-labeling approach (using a Likert scale), but they also had the possibility to give another emotion label based on an open-ended option. In "Body Cues Rating" task, the participants were asked to rate 8 expressive body cues.

For each stimulus participants viewed, they were asked to rate their perception of emotion and to rate 8 expressive body cues. The action reproduced on a virtual actor was indicated for each stimulus (e.g. Virtual actor expresses an emotion while Moving books on a table). The order of stimuli in each subset was randomized for each participant.

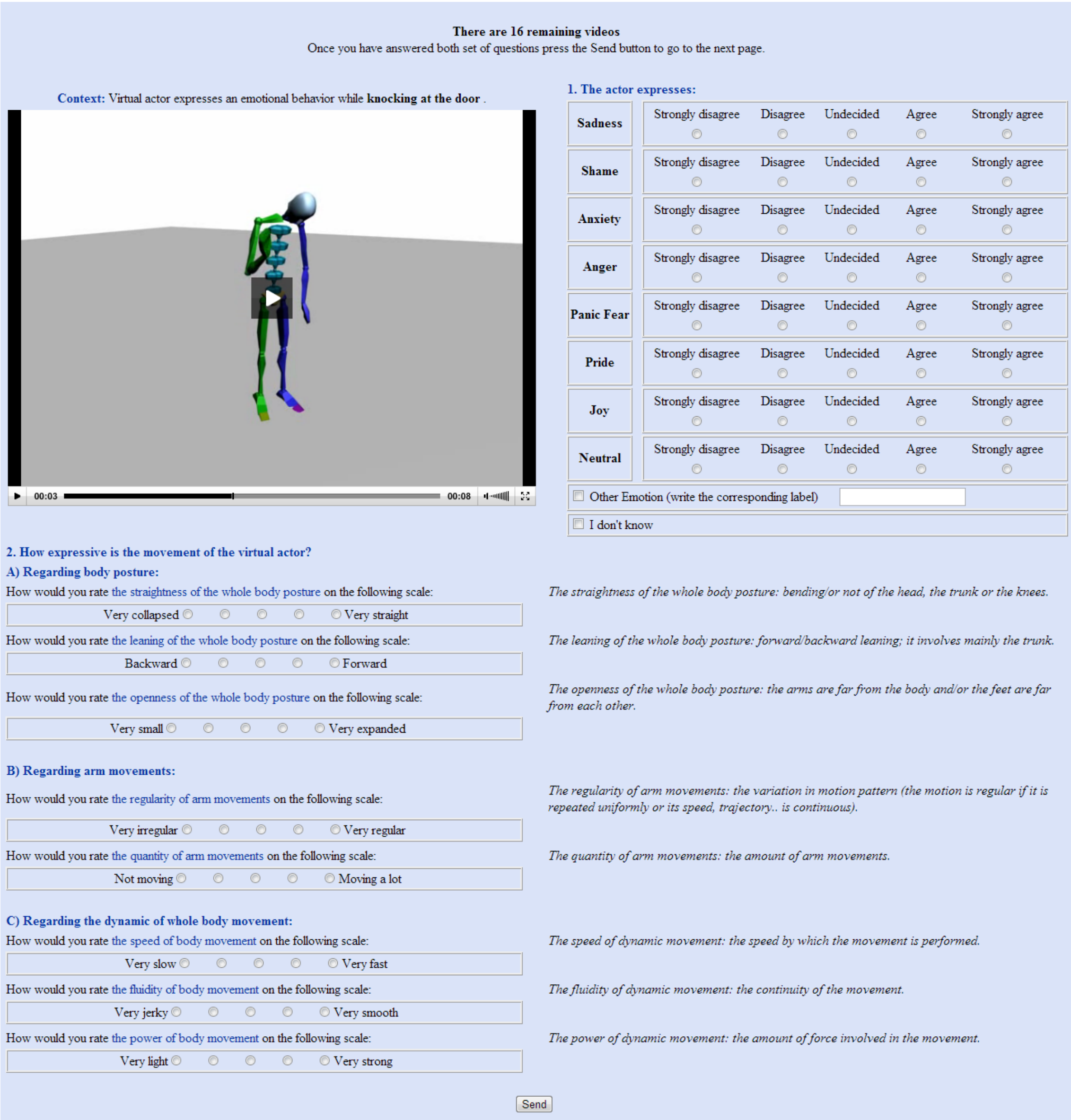


Figure 8.2: Screenshot of our online perceptual study.

8

	Simple Walk	Walk with an object in hands	Move Books	Knock	Sit Down	Lift	Throw
%	2,53%	2,74%	4,08%	3,35%	2,84%	3,13%	2,91%

	Anxiety	Pride	Joy	Sadness	Panic Fear	Shame	Anger	Neutral
%	3,22%	3,27%	3,52%	2,38%	3,32%	2,78%	2,88%	3,22%

	Actor1	Actor2	Actor3	Actor4	Actor5	Actor6	Actor7	Actor8	Actor9	Actor10	Actor 11	Actor12
%	3,30%	3,20%	2,53%	2,90%	5,36%	4,32%	0,97%	1,93%	3,20%	3,45%	2,53%	3,20%

Figure 8.3: Percentages of choosing the "*I don't know*" button per actor, emotion and action.

8.2.4 Participants

The popular crowd-sourcing website Amazon Mechanical Turk (AMT) was used to collect the results of our perceptual study as it provides an easy access to a large, stable and diverse subjects pool. Thus, our perceptual study consisted of an online survey. TurkGate tools (Grouping and Access Tools for External surveys) [Darlow and Goldin, 2013] were used to better control the use of Mechanical Turk with external HITS. We set the qualifications of the participants to the following conditions: HIT Approval Rate (%) for all Requesters' HITs greater than 95, and Number of HITs Approved greater than 1000. It has been shown that using those qualifications will direct the work to the highest quality participants as determined by their approval rate by other requesters in the Mechanical Turk Marketplace.

1008 participants took part in our study (56.01% of females and 43.98% of males, mean age of 36 years old ranging from 15 to 76 years old). The percentage of participants who spent the majority of their life in U.S was 98.21%.

8.2.5 "Emotion Perception" task: Multi-labeling approach

When rating their perception of emotions, participants indicated if they perceived the presence of the 8 emotions considered in the Emilya database on a 5-level Likert scale. The rating refers to the presence of an emotion and not to its intensity. The sentence "the actor expresses" is used for all the Likert items. The scale levels are "Strongly disagree", "Disagree", "Undecided", "Agree" and "Strongly Agree".

We consider that they perceive an emotion if they gave a rating score above the level "3" (Undecided level), i.e. if they agree or they strongly agree on the perception of this emotion. As such, each stimulus used in our perceptual study can be described in term of percentages that depict the presence of each perceived emotion.

	Anxiety	Pride	Joy	Sadness	Panic Fear	Shame	Anger	Neutral	Total
Percentages	6,60%	8,73%	7,14%	7,09%	5,85%	7,44%	6,60%	6,94%	7,05%
Most Frequent Labels	Boredom, Nervous	Boredom Relaxed Confidence	Happiness Relaxed	Boredom Tired	Nervous	Boredom Nervous Shyness	Frustration	Boredom Calm	Boredom

Figure 8.4: Percentages and results of choosing the open-ended option per emotion.

For instance, we found that the multi-labeling perception of Neutral expression in Moving Books action of Actor 3 is constituted of 4.16% of Anxiety, 0% of Pride, 0% of Joy, 0.12% of Sadness, 0% of Panic Fear, 4.16% of Shame, 4.16% of Anger and finally 75% of Neutral, where the percentage of each emotion represents the percentage of participants who perceived it with a score above "3". That is, as 24 participants rated this stimulus, the Neutral expression of Moving Books by Actor 3 was perceived as Anxiety by 1 participant, as Pride by 0, as Joy by 0, as Sadness by 3, as Panic Fear by 0, as Shame by 1, as Anger by 1 and as Neutral by 18 participants.

Participants had also the possibility to choose the "*I don't know*" button if they were not able to answer [Winters, 2005]. We found that the button " I don't know" was barely chosen (3%). Choosing the button " I don't know" can be also considered as short way off from choosing the "Undecided" response option to all the emotion items. We found that the "Undecided" response option was assigned to all the emotion items in 0.7% of the stimuli. Totally, participants were not able to perceive any emotion in 3.7% of the stimuli. Besides, the participants had the possibility to choose the button "*another emotion*" and propose another label of emotion of their free choice [Winters, 2005]. In the next subsection 8.2.6, we briefly summarize the results obtained from choosing another emotion label in the "Emotion Perception" task.

8.2.6 "Emotion Perception" task: Perception of another emotion

The percentage of proposing another emotion label is small (7.05%): it happens at a similar rate for all the stimuli, ranging from 5.85% in stimuli showing Panic Fear expression to 8.73% in stimuli showing Pride expression. A linguistic analysis was performed before analyzing the frequency of each suggested emotion label. Across all the actions and all the expressed emotions, the most frequent label that was suggested by the participants, is "Boredom" (13.36% among the attributed labels). It is also the most frequent label that was proposed for the stimuli corresponding to Sadness, Shame, Pride, Anxiety and Neutral across all the actions. Boredom label has also been used, with high frequency, by participants when they were asked to answer open-ended questions format in previous studies [Russell, 1994]. Looking more closely, we can also note that Boredom label has been the most used in stimuli showing Throwing, Lifting and Moving Books actions. As if those actions are more likely to infer Boredom across all the expressed emotions. Boredom was not necessary the most frequent label used for Sadness, Shame, Pride, Anxiety and Neutral in the other actions. For instance, the most frequent label attributed to Neutral in Simple Walking action is "Confidence" . We note that the second most frequent label attributed to Anxiety is "Nervous", which can be considered as a synonym of Anxiety; "Nervous" and "Anxious" belong to the same synset [Bentivogli et al., 2004], where a synset is defined as a set of one or more synonyms. "Nervous" label is also the most frequent label attributed to Panic Fear. This result is congruent with what can be found in WordNet database: panic is derived from Anxiety [Bentivogli et al., 2004] and that Anxiety and Nervous belong to the same synset [Bentivogli et al., 2004]. The most frequent label assigned to Anger is "Frustration", which has also been signaled in previous studies [Savva et al., 2012, Russell, 1994]. Finally the most frequent label assigned to Joy is "Happiness". According to the WordNet Affect (a hierarchy of "affective domain labels") [Strapparava and Valitutti, 2004], Happiness is considered as an affective category label derived from Joy.

Overall, the emotion labels suggested by the participants when they made used of the open-question "another emotion" are in line with what previous studies reported. In the remaining of our analyses on emotion perception, we will focus on the Multi-Labeling perception of emotions performed in the "Emotion Perception" task (See section 8.4).

8.2.7 "Body Cues Rating" task

We selected a set of 8 body cues that describe body shape, postural changes and the quality of movement dynamics from our multi-level body movement coding schema described in [Fourati and Pelachaud, 2014]. We selected a subset of variables from our coding schema to simplify the rating task. We chose those that are also used in previous works [Dahl and Friberg, 2007, Wallbott, 1998, Gross et al., 2010, Meijer, 1989] to be able to compare our results with these ones. A semantic differential scale is used to rate these 8 body cues [Coolican, 2004]. Each body cue is rated on a bipolar 5-point scale. The definition of each body cue was provided in the interface of the perceptual experiment:

- Power of movement (Very light—Very strong): the amount of force involved in the movement.
- Fluidity of movement (Very jerky—Very smooth): the continuity of the movement.
- Speed of movement (Very slow—Very fast): the speed by which the movement is performed.
- Quantity of arms movement (Not moving—Moving a lot): the amount of arms movement.
- Regularity of arms movement (Very irregular—Very regular): the variation in

motion pattern (the motion is regular if it is repeated uniformly or its speed, trajectory.. is continuous).
– Body openness (Very contracted—Very expanded): the arms are far from the body and/or the feet are far from each other.
– Sagittal Body leaning (Backward—Forward): forward/backward leaning; it involves mainly the trunk.
– Body straightens (Very collapsed—Very straight): bending/or not of the head/ trunk/ knees.

8.3 INTER-RATER CONSISTENCY

The study of emotion labeling and body cues rating based on participants judgment requires investigating the degree of homogeneity in the rating given by different participants. Depending on the goal of the study and the rating strategy, two types of measurements can be used to explore the degree of homogeneity in the rating: the inter-rater agreement measurement and the inter-rater reliability measurement. Inter-rater reliability measures the degree to which different participants make consistent ratings of the same phenomenon. Inter-rater agreement measures the extent at which participants make the exact same rating about the same phenomenon [Salkind, 2010]. While several measures were proposed and used as an agreement index of nominal data between participants (e.g. Cohen's kappa, Fleiss kappa), agreement measures between multiple participants for ordinal data remain less popular [Banerjee, 1999]. The judgment of participants in our perceptual experiment is made on a numerical scale (Likert scale for emotion perception and semantic differential scale for body cues rating). In our study, we are interested to measure the similarity between participants's rating rather than their exact agreement of rating; that is we are interested in understanding if participants are self-consistent in perceiving each emotion and in rating each body cue rather than finding out if participants share the same rating score. Thus, we make use of the measure of inter-rater reliability through Cronbach's alpha coefficient.

The Cronbach's alpha coefficient is used to measure the inter-rater reliability of both, emotions perception (based on Likert scale data) and body cues ratings (based on Semantic differential scale data). Figure 8.5 shows the mean of Cronbach's alpha coefficient that measures the inter-rater reliability of emotion perception and of body cues rating. We discuss the mean of inter-rater reliability per emotion across all the actions and all the actors of the Emilya database (see Figure 8.5).

Figure 8.5 shows that, across all the emotions, it seems that the participants were mostly more self-consistent when rating the perception of emotions than when rating the expressive body cues (significant differences are observed for Anxiety ($p<.05$), Pride ($p<.001$), Joy ($p<.001$), Sadness ($p<.01$) and Anger ($p<.01$)). In both "Emotion Perception" and "Body Cues Rating" tasks, the participants were the most self-consistent when rating stimuli showing Sadness expression (see Figure 8.5). This result suggests that bodily expression of Sadness in the Emilya database is

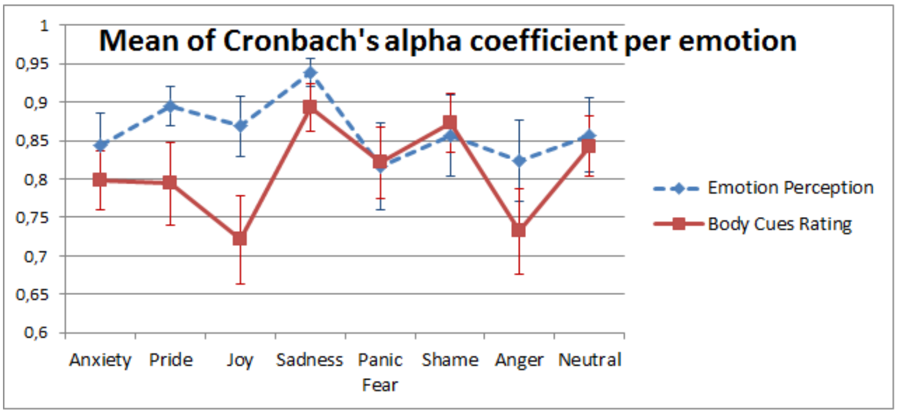


Figure 8.5: The mean of Cronbach's alpha coefficient for inter-rater reliability in "Emotion Perception" and "Body Cues Rating" tasks per emotion. The error bars indicate 95% confidence interval.

more likely to receive similar ratings across different participants in both "Emotion Perception" and "Body Cues Rating" tasks.

Overall, Figure 8.5 shows that the mean of Cronbach's alpha coefficient is always higher than 0.7. Previous studies reported that the acceptable level of Cronbach's alpha is 0.7 [Salkind, 2010]. Thus, this result highlights a strong agreement between the participants for the perception of emotions and for the rating of body cues.

We also quantify the stimuli for which the Cronbach's alpha coefficient received a value above 0.7 level (considered in previous studies as an acceptable level of Cronbach's alpha [Salkind, 2010]). We find that the Cronbach's alpha coefficient that measures the reliability of participants ratings in the "Emotion Perception" task is higher than 0.7 in 90.36% of the stimuli. Thus, high inter-rater reliability of emotion perception is observed for the large majority of the stimuli. Looking in more details on the ratings of "Emotion Perception" task in the remaining 9.64% of the stimuli, we found that low agreements are mostly observed when the expressed emotion is not correctly recognized. We also find that the Cronbach's alpha coefficient that measures the reliability of participants ratings in "Body Cues Rating" task is higher than 0.7 in 84.48% of the stimuli. In the following sections, we will discard the samples that received low agreement ratings in the "Emotion Perception" task and in the "Body Cues Rating" task to explore respectively the recognition rates of emotion perception (sections 8.4.2 and 8.4.3) and the classification rates based on body cues rating (section 8.6).

8.4 Perception of expressed emotions

In this section, we turn our attention on the analysis of the results obtained from the multi-labeling ratings of emotion perception in the "Emotion Perception" task

(See section 8.2.5). In this section, our goal is to explore how expressed emotions are perceived. Three successive steps are performed to achieve a deep understanding of expressed emotions perception.

As a first step, we explore how the expressed emotions are perceived compared to the perception of Neutral. The hypothesis is, given a unit increase in $emotion_i$ perception rating, it is significantly more likely to obtain the expression of $emotion_i$ than the Neutral expression (in other words, the perception rating of $emotion_i$ is significantly positively associated with the expression of $emotion_i$ comparing to the Neutral expression). This analysis is presented in section 8.4.1.

As a second step, we discuss the mean ratings, the recognition rates and the confusions that occur at the level of emotion perception across all the actions. The most frequent label approach is used to measure the recognition rates as performed in similar previous works [Kleinsmith et al., 2006b] [Kleinsmith and Bianchi-Berthouze, 2013]. This approach allows us studying the confusions in the perception of emotions. The results are presented in section 8.4.2.

As a third step, we explore the effect of each daily action on the recognition rates of emotional body expressions (See section 8.4.3). In the following subsections, we present and we discuss the results of each of these three studies.

8.4.1 Perception of Neutral expression Vs perception of other expressed emotions

First of all, we conduct a Multinomial Logistic Regression (MLR) analysis to examine, across all the actions, the statistical effect of a unit increase in each emotion perception rating on the probability of obtaining each expressed emotion compared to the probability of obtaining the Neutral expression (i.e. the Neutral expression is considered as a reference category). Table 8.1 reports the coefficients (i.e. the slopes) of the MLR model and the corresponding p-value levels: 'ns' stands for non-significant and '***', '**', '*' stand respectively for a significant difference with p<.001, p<.01 and p<.05. The coefficients express the effects of emotion ratings on the relative risk of being in one expressed emotion versus the Neutral expression. For instance, the first cell of the Table 8.1 indicates that the probability of being the Anxiety expression compared to the probability of being the Neutral expression increases significantly (p<.01) exp(0.44) times for each unit increase in the Anxiety rating, given all other emotions ratings equal. In other words, it indicates that the rating of the perception of Anxiety is significantly more positively associated with the expression of Anxiety than with the Neutral expression. The same result (in bold, on the diagonal of Table 8.1) is found for the other ratings; the rating of the perception of $emotion_i$ is significantly more positively associated with the expression of $emotion_i$ than with the Neutral expression. Another interesting finding is that the rating of Neutral is significantly more negatively associated with all the other emotion ratings (See the last column in Table 8.1). In other words, that means that the rating of the perception of Neutral is significantly more positively associated

Table 8.1: Statistical results of Multinomial Logistic Regression (MLR): Emotion perception ratings included in MLR model predicting emotion expression. The columns represent emotion perception rating. 'ns' stands for non-significant and '***', '**', '*' stand respectively for a significant difference with p<.001, p<.01 and p<.05.

Ax	Pr	Jy	Sd	PF	Sh	Ag	Nt
Anxiety expression							
+0,44 ***	-0,22 ***	+0,05 ns	-0,28 ***	+0,13 **	+0,02 ns	+0,12 **	-0,33 ***
Pride expression							
+0,04 ns	**+0,1** *	+0,18 ***	-0,26 ***	-0,04 ns	-0,18 ***	+0,17 ***	-0,23 ***
Joy expression							
+0,27 ***	-0,08 *	**+0,42** ***	-0,48 ***	+0,05 ns	-0,18 ***	+0,26 ***	-0,45 ***
Sadness expression							
+0,1 *	-0,18 **	-0,1 ns	**+0,62** ***	+0,02 ns	+0,24 ***	-0,14 **	-0,25 ***
Panic Fear expression							
+0,37 ***	-0,23 ***	+0,02 ns	-0,55 ***	**+0,68** ***	-0,08 ns	+0,03 ns	-0,47 ***
Shame expression							
+0,19 ***	-0,17 **	-0,17 **	+0,17 ***	+0,19 ***	**+0,35** ***	-0,24 ***	-0,26 ***
Anger expression							
+0,13 **	-0,13 **	+0,09 ns	-0,4 ***	-0,06 ns	-0,13 **	**+0,95** ***	-0,48 ***

with the expression of Neutral than with the other expressed emotions (i.e. Anxiety, Pride, Joy, Sadness, Panic Fear, Shame and Anger).

Overall, Table 8.1 shows the increase/ decrease of the probabilities of obtaining each emotion expression (compared to the probability of being the Neural expression) given the unit increase in each emotion perception rating. For instance, according to column 6 of Table 8.1, we can observe that, given a unit increase in Shame perception rating, it is significantly more likely to obtain the expression of Shame or Sadness rather than the Neutral expression; while it is significantly less likely to obtain the expression of Pride, Joy or of Anger rather than the Neutral expression given a unit increase in Shame perception rating.

Table 8.1 can also be interpreted as indicating somehow the statistical patterns of each emotion expression in terms of emotion perception ratings. For instance the pattern of the expression of Shame can be described as follow (see line 6 of Table 8.1): it is significantly more likely to obtain the expression of Shame rather than Neutral expression with a unit increase in Anxiety, Sadness, Panic Fear or Shame perception rating; while it is significantly less likely to obtain the expression of Shame rather than the Neutral expression with a unit increase in Pride, Joy, Anger or Neutral perception rating.

8.4.2 Mean ratings, recognition rates and confusions in emotion perception

In section 8.4.1 we compared the perception of $emotion_i$ in stimuli showing the expression of the $emotion_i$ and in stimuli showing the Neutral expression. In this section, we explore the interaction between the expressed emotions based on emotion perception ratings. We discuss the mean ratings, the recognition rates and the confusions obtained from the multi-labeling rating of "Emotion Perception" task across all the actions.

Mean ratings:

Figure 8.6 shows the different mean ratings of perceived emotion according to each expressed emotion. For instance, we can observe that the expression of Sadness is described by the highest mean rating of the perception of Sadness, followed by the perceptions of Shame and of Anxiety and the lowest mean rating of the perception of Pride and of Joy (See Figure 8.6). Based on the mean rating of emotion perception, Figure 8.6 indicates that Anger and Neural are mainly perceived as such. However, Figure 8.6 shows that Panic Fear, Shame and Pride are mainly perceived respectively as Anxiety, Sadness and Neutral. In order to better quantify the recognition rate of expressed emotions and to give more insights about the confusions occurring at the perception level, we now present the recognition rates and the confusion matrix of emotion perception.

Recognition rates and confusions:

In order to explore the recognition rate of emotions, we need to establish a one-labeling rater-based ground truth. Thus, we use "the most frequent label" approach

[Kleinsmith and Bianchi-Berthouze, 2013]. Table 8.2 shows the confusion matrix of the recognition of emotions based on the most frequent label approach as well as the Recall and the Precision measures for each emotion. The recognition rates are measured using the samples that received high agreements in the "Emotion Perception" task (that is 90.36% of the stimuli, see Section 8.3). Recall measure refers to the fraction of the stimuli showing the expression of $emotion_i$ that are perceived as $emotion_i$. Precision measure refers to the fraction of the stimuli perceived as $emotion_i$ that show the expression of $emotion_i$. The overall emotion recognition rate is 43%, which is above the chance level (12.5%). Low emotion recognition rates were also reported in previous perceptual studies (30% in [Patterson et al., 2001], 42% in [Gross et al., 2010] and 56.45% in [Graham and Jackson, 1993], 35.2% and 18% in [Volkova et al., 2014b]). The expression of Sadness is the best recognized (it receives the highest Recall measure), followed by the expressions of Neutral and of Anger with respectively 88%, 78% and 58% of recognition rates (Recall measures) over all the actions. In [Volkova et al., 2014b], it was also found that the expressions of Neutral, Anger and of Sadness were the best recognized from stimuli showing expressive upper body movements reproduced on a stick-figure. Table 8.2 shows that the recognition rates of the expressions of Anxiety, Pride and Joy over all the actions are above the chance level and correspond respectively to 37%, 20% and 38%. Across all the actions, the expressions of Panic Fear and of Shame are the least recognized. Their recognition rate over all the actions are below the chance level (10% for Panic Fear and 11% Shame).

Anxiety receives the lowest Precision measure (30%) followed by Neutral (33%) and Sadness (44%) (See Table 8.2). Anxiety and Sadness low Precision measures are mainly due to the confusion between Anxiety and Panic Fear and between Sadness and Shame. However, low Precision measure of Neutral is due to the fact that Neutral was highly perceived in different stimuli (e.g. stimuli showing Pride, Anxiety or Joy expression, See Table 8.2). This result was also reported in [Volkova et al., 2014b]; in many cases, participants perceived Neutral in expressive upper body movements when an emotion category other than Neutral was intended by the actor (e.g. amusement, joy, pride, relief, shame...).

The overall emotion recognition rate is relatively low (43%). This is mainly due to the presence of large confusions; Shame is highly confused with Sadness, Pride with Neutral, and Panic Fear with Anxiety.

- Shame is found to be confused with Sadness based on the most frequent label approach (See Table 8.2). However, we observe that this confusion is unidirectional; Sadness is not confused with Shame based on the most frequent label approach (See Table 8.2). Looking at the characterization of expressed emotions based on the mean ratings of emotion perception (See Figure 8.6), we can observe that the expression of Sadness (respectively of Shame) was highly perceived as the emotions Sadness and Shame (respectively Shame and Sadness). So, the unidirectional confusion is only observed based on the most frequent label approach. The confusion occurring at the level of Shame and

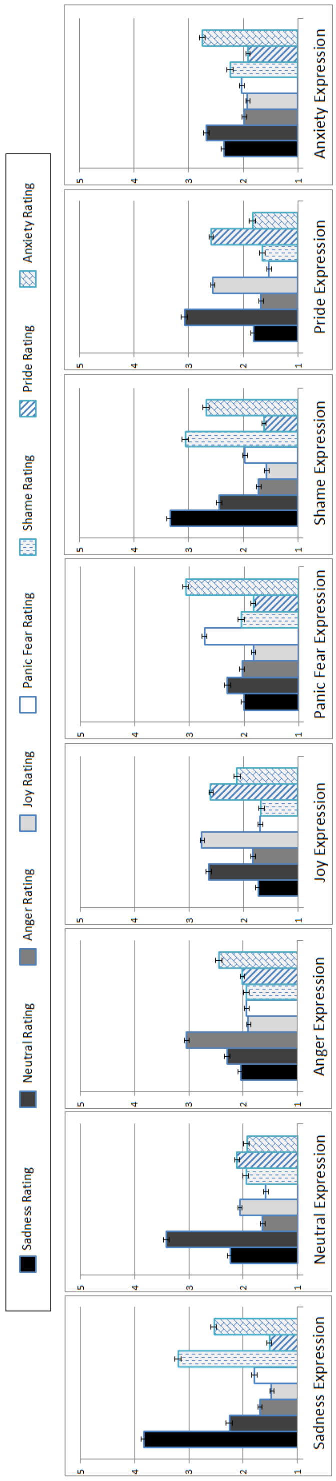


Figure 8.6: Emotion expression description in terms of emotion rating. The mean rating is graduated from 1 to 5 which stands respectively for: "Strongly disagree", "Disagree", "Undecided", "Agree", "Strongly Agree". The error bars indicate 95% confidence interval.

Table 8.2: Confusion matrix for emotion labels: rows depict the expressed emotion and columns depict the perceived emotion. RC stands for Recall (%), PS stands for Precision (%).

	Ax	Pr	Jy	Sd	PF	Sh	Ag	Nt	**RC**
Ax	**28**	1	5	13	0	2	4	22	37%
Pr	2	**16**	12	2	0	0	1	46	20%
Jy	7	7	**30**	6	2	1	4	21	38%
Sd	2	0	0	**72**	0	0	0	8	88%
PF	43	1	1	6	**7**	1	2	9	10%
Sh	7	0	0	48	0	**8**	0	11	11%
Ag	7	4	5	3	0	0	**40**	10	58%
Nt	0	1	2	12	0	0	1	**57**	78%
PS	30%	53%	55%	44%	78%	67%	77%	33%	43%

Sadness expressions may be explained as a similarity at the level of the characterization of their bodily expressions (see section 8.5 for further investigation on this topic). The confusion in Shame and Sadness perception has been also reported in previous works [Volkova et al., 2014b] [Gross et al., 2010]. Gross et al. [Gross et al., 2010] reported in their work that observers perceived Shame in Sadness expression in knocking action based on the mean rating of emotion perception. Volkova et al. [Volkova et al., 2014b] found that Shame was confused with Neutral and Sadness when participants viewed stimuli depicting actors telling a narration while being seated. However, [Wallbott, 1998] and [Volkova et al., 2014b] reported that participants were able to recognize Shame expression in body movements from stimuli showing actors communicating emotions without any constraints of performing a specific action/task (i.e. being an “explicit” expression). Wallbott [Wallbott, 1998] highlighted the importance of self-manipulators in the expression of Shame. Besides, unlike Sadness, Shame has been mostly studied in the context of social settings [Gilbert and Andrews, 1998]. [Gilbert and Andrews, 1998] argues that “*Shame is not only related to internal experiences but also conveys socially shared information about one's status and standing in the community*”. However, the recording of emotional behaviors in the Emilya database is based on individual settings (the actors are alone in a room) and does not include a social context. Besides, the expression of Shame is mostly associated with facial expressions that are characteristics of submissiveness such as eye-gaze avoidance and turning away [Gilbert and Andrews, 1998]. As we do not display facial expressions on the virtual character during the perceptual study, gaze-avoidance behavior is not visible; including gaze direction and facial expressions may be critical for the discrimination between the expression of Sadness and the expression of Shame.

– We find that the expression of Pride in daily actions is mostly perceived as Neutral across all the actions. This result is found in both emotion perception mean rating (See Fig. 8.6) and the most frequent label approach (See Table 8.2). A similar finding was reported in [Volkova et al., 2014b]; Pride expression was mostly perceived as Neutral. We explain the confusion between Pride and Neutral as facial expressions are not displayed during the perceptual task, while they have been considered as a relevant cue for characterizing the expression of Pride [Tracy and Robins, 2007].
– The expression of Panic Fear is mostly perceived as Anxiety (See Table 8.2). The confusion between Anxiety and Panic Fear was also reported in previous works; Dael et al. [Dael et al., 2011] proposed to group Anxiety and Panic Fear expressions under a single cluster that represents the expressive profile of Anxiety. However, we note that a unidirectional confusion is observed in our work; Panic Fear is mostly perceived as Anxiety but Anxiety is not perceived as Panic Fear (See Table 8.2). This unidirectional confusion can be explained as a lexical confusion in emotion labels attribution. According to WordNet [Bentivogli et al., 2004], a synset involving "panic" and "scare" (*sudden mass fear and anxiety over anticipated events*) seems to be derived from a synset involving "Anxiety" and "Anxiousness" [Bentivogli et al., 2004].

Overall, based on the mean rating of emotion perception (See Fig. 8.6) and the most frequent label approach (See Table 8.2), the confusions occurring at the level of emotion perception are mainly between Anxiety and Panic Fear, Sadness and Shame, and between Pride and Neutral. We reported different possible explanations to such confusions such as the lack of contextual factors, the lack of other modalities (e.g. facial expression) or a similarity in the characterization patterns of bodily expressions. Later, in Section 8.5, we examine whether the expressed emotions that are confused in the "Emotion Perception" task share similar characteristics in the "Body Cues Rating" task.

In this section, we discussed the results of emotion perception across all the actions. In the following subsection, we explore the effect of daily actions on the perception of emotion.

8.4.3 Effect of daily actions on the recognition rates

In Table 8.3, we present the results of the recognition rates of the expressed emotion in each action. Each cell of Table 8.3 (except the last row) reports Recall measures for each expressed emotion. Sadness is well recognized in the actions Sitting Down and Throwing and in most of the other actions. Neutral is very well recognized in the action Moving books; it has also been well recognized in the actions Knocking, Sitting Down and Lifting. Anger is highly recognized in the Throwing action. Positive emotions (Joy and Pride) are better recognized in both Walking actions (64% and 45%), where the whole body is involved. This result may suggest that Pride and Joy are better recognized in actions involving the whole

Table 8.3: Percentage of recognition scores for each emotion in each action according to the most frequent emotion in multi-labeling.

	SW	WH	MB	KD	SD	Lf	Th
Ax	42%	18%	10%	**75%**	64%	25%	18%
Pr	**45%**	36%	9%	17%	17%	0%	18%
Jy	58%	**64%**	42%	17%	11%	27%	45%
Sd	92%	83%	75%	83%	**100%**	80%	**100%**
PF	18%	17%	0%	11%	22%	0%	0%
Sh	27%	10%	10%	17%	10%	0%	0%
Ag	25%	40%	44%	45%	55%	89%	**100%**
Nt	60%	80%	**100%**	83%	91%	90%	40%
CCR	47%	44%	37%	45%	48%	38%	43%

body. Anxiety is better recognized in the action Knocking at the door (75%). In the study conducted by Gross et al. [Gross et al., 2010], Anxiety is also perceived with a similar recognition rate when actors expressed emotions while knocking at the door (65%). We also note that Anxiety is perceived with high recognition rate (64%) in the Sitting Down action.

8

The actors worked to express emotions through their body movement. We asked them not to associate the different actions (Knocking, Throwing...) to each emotion. Some emotions were well recognized for specific actions as they can be associated to them (e.g. Throwing an object with Anger). However, some emotions can be well recognized in actions that are not intuitively associated to them (e.g. Throwing an object with Sadness, Knocking with Anxiety, Walking with Joy). In section 8.6.1, we will compare the effect of the action on the recognition of expressed emotions (whose results were just presented) with its effect on the classification of expressed emotions based on body cues ratings.

The last row of Table 8.3 (CCR) reports the Recall measures across all the emotions for each action. The overall percentage of emotions recognition rate (across all the emotions) is similar for each action, ranging from 37% and 38% (in Moving Books and Lifting actions) to 47% and 48% (in Simple Walking and Sitting Down actions).

8.5 Characterization of emotions based on body cues rating

In this section, we focus on the **characterization** of expressed emotions based on body cues rating and in the next section 8.6, we will explore the **classification rates** of each expressed emotion based on body cues rating.

In addition to the "Emotion Perception" task, participants were also asked to perform a "Body Cues Rating" task (See section 8.2.7). The goal of this task is

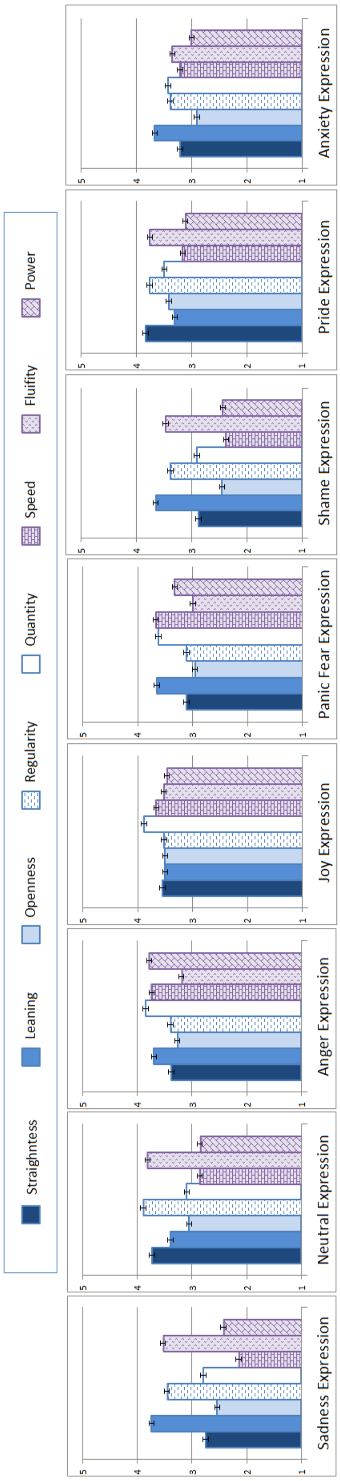


Figure 8.7: Emotion expression description in terms of body cues rating. The mean rating is graduated from 1 to 5 which stands for a bipolar scale for each body cue (see Section 8.2.7). The error bars indicate 95% confidence interval.

Table 8.4: Statistical results of Multinomial Logistic Regression (MLR): Body Cues ratings included in MLR model predicting emotion expression. 'ns' stands for non-significant and '***', '**', '*' stand respectively for a significant difference with p<.001, p<.01 and p<.05.

Strai.	Lean	Open.	Regu.	Quant.	Speed	Flui.	Power
Anxiety expression							
-0,37 ***	+0,37 ***	-0,12 ***	-0,31 ***	+0,24 ***	+0,38 ***	-0,25 ***	-0,01 ns
Pride expression							
+0,08 *	-0,18 ***	+0,24 ***	-0,20 ***	+0,28 ***	+0,15 ***	-0,05 ns	+0,05 ns
Joy expression							
-0,24 ***	-0,01 ns	+0,27 ***	-0,34 ***	+0,49 ***	+0,61 ***	-0,19 ***	+0,19 ***
Sadness expression							
-0,55 ***	+0,54 ***	-0,14 ***	-0,16 ***	+0,05 ns	-0,62 ***	-0,09 *	-0,03 ns
Panic Fear expression							
-0,45 ***	+0,28 ***	-0,11 **	-0,39 ***	+0,22 ***	+0,83 ***	-0,49 ***	+0,17 ***
Shame expression							
-0,45 ***	+0,44 ***	-0,32 ***	-0,24 ***	+0,12 ***	-0,29 ***	-0,11 **	-0,15 ***
Anger expression							
-0,35 ***	+0,22 ***	+0,02 ns	-0,29 ***	+0,35 ***	+0,55 ***	-0,48 ***	+0,69 ***

to study the characterization and the classification of the expressed emotions based on the human rating of some body cues: Straightness (Collapsed-Straight), Sagittal Torso Leaning (Backward-Forward), Body Openness (Contracted-Expanded), Regularity of arms movement (Irregular-Regular), Quantity of arms movement (Not Moving-Moving a lot), Speed (Slow-Fast), Fluidity (Jerky-Smooth), Power (Light-Strong). We study the characterization of expressed emotions based on the mean rating of body cues across all the actions (Figure 8.7) and we also explore which body cues contribute significantly to the characterization of each expressed emotion through a Multinomial Logistic Regression Analysis (MLR). Figure 8.7 depicts the mean ratings of body cues for each expressed emotion across all the actions. Table 8.4 shows the coefficients of the MLR model and the corresponding p-value that explains the effects of the body cues rating on the relative risk of being in one expressed emotion versus the Neutral expression (Neutral expression is chosen as the reference category).

The characterization of each expressed emotion is discussed based on the mean ratings showed in Figure 8.7 and on the MLR model as described in Table 8.4.

Sadness:
Sadness expression is characterized across all the actions by slow (low Speed), more or less smooth and regular (mean rating of Fluidity and Regularity are 3.53 and 3.46; See Figure 8.7). However, we can note that Fluidity and Regularity are significantly more associated with Neutral than with Sadness (see Table 8.4). Sadness expression is also characterized by contracted body shape (low Openness), forward torso leaning (high Leaning) and collapsed body posture (low Straightness) (see mean ratings in Figure 8.7 and MLR coefficients in Table 8.4). In [Gross et al., 2010], Sadness expression in Knocking action is similarly characterized through Effort-Shape analysis. Gross et al. [Gross et al., 2010] reported that Sad trials are characterized with “sustained, leisurely, slow” movement (low Speed). Wallbott [Wallbott, 1998] found that bodily expression of Sadness is characterized by collapsed posture (low Straightness) and low movement dynamics (low Power). In [Montepare et al., 1999], bodily expression of Sadness is perceived as very smooth (high Fluidity), slow (low Speed), and lacking in action (low Movement Quantity). Sadness expression in musicians’ body movements [Dahl and Friberg, 2007] was also characterized by slow (low Speed), smooth (high Fluidity), regular movements and more or less lacking in action (low Movement Quantity).

Shame:
In our study, Shame expression is characterized with a very similar pattern of body cues as Sadness across all the actions. This similarity holds for all the parameters but one; unlike Sadness expression, the rating of arms movement quantity is significantly more positively associated with Shame expression than with Neutral expression (See Table 8.4). The similarity of Shame and Sadness patterns across all the actions can explain the confusion between Sadness and Shame that occurred in the “Emotion Perception” task where Shame is often perceived as Sadness (See Table 8.2, Figure 8.6). Wallbott [Wallbott, 1998] found that Shame expression is mainly characterized

with collapsed posture (low Straightness). Similarly, Meijer [Meijer, 1989] found that Shame expression is characterized with bowed trunk (high Torso Leaning), knees slightly bent and downward body movement (low Straightness).

Anger:
Based on the mean rating of body cues, Figure 8.7 shows that Anger expression is characterized with strong (high Power) and fast (high Speed) movement, high Quantity of arms movement and high forward body Leaning. Table 8.4 shows also similar results: Anger is characterized by high values of Leaning, Speed, Power or Quantity of movement. It is also defined by low values of Straightness, Regularity or Fluidity. Similar characterizations of body movement in Anger expression are found in previous works: High movement activity (high Movement Quantity), high movement dynamics (High Speed and Power) in [Wallbott, 1998], strong, powerful, forceful (high Power) knocking movement in [Gross et al., 2010], and very fast (high Speed), expanded (high Openness) and full of action (high Movement Quantity) in [Montepare et al., 1999]. Unlike results reported in [Wallbott, 1998] and [Montepare et al., 1999] that indicate a positive correlation between Openness and Anger expression, we did not find such a correlation.

Panic Fear:
Panic Fear expression is characterized with a similar pattern as Anger expression (See Figure 8.7). As also reported in other studies [Dahl and Friberg, 2007], body movement Fluidity receives similar mean rating in Panic Fear and Anger expressions. Based on the MLR analysis (See Table 8.4), we found that Anger and Panic Fear were perceived as jerky compared to the perception of Neutral expression. This result has been also highlighted in [Dahl and Friberg, 2007]. Based on the results shown in Figure 8.7 and Table 8.4, we can observe that the patterns of body cues of Panic Fear expression is quite similar to the pattern of body cues of Anger expression. However, these emotions were not confused with one another in the “Emotion Perception” task (see Section 8.4.2). It could mean that more body cues are needed to discriminate these two emotions.

Anxiety:
Except the effect of the Power and the Speed ratings, Table 8.4 shows that the Anxiety expression receives similar ratings as Panic Fear and Shame expressions, except for the Power and Speed ratings. Unlike Panic Fear, no significant effect is found for Power rating (See Table 8.4). While the Speed rating is significantly negatively associated with Shame expression, it is significantly positively associated with Anxiety expression (See Table 8.4).

Joy:
Across all the actions, Joy expression is characterized with high mean rating of arms movement Quantity (See Figure 8.7). Table 8.4 shows that the Openness, the Power, the arms movement Quantity and particularly the Speed are positively associated with the Joy expression while the Regularity, the Straightness and the Fluidity are negatively correlated. Joy expression in musicians’ body movements [Dahl and Friberg, 2007] was also characterized with high Speed and high Quantity

of movements, but with lower rating of Fluidity and Regularity. Similarly to what was reported in [Montepare et al., 1999], we find that Joy expression was rated as smoother than Angry expression, although jerkier than Neutral.

Pride:

Based on body cues rating, we can observe that Pride and Neutral expressions share some similarities in their patterns of characterization (See Figure 8.7). The similarity between Pride and Neutral expressions characterization is congruent with the confusion found in the "Emotion Perception" task; Pride is often perceived as Neutral (See Table 8.2). However, the MLR model indicates that the increase of the Straightness, the Openness, the Quantity or the Speed of movement as well as the decrease of the Leaning or the Regularity cues contributes significantly on the probability of being a Pride expression versus the probability of being a Neutral expression (See Table 8.4).

Summary:

Overall, we can deduce from Table 8.4 that the maximal significant increase of the probability of being a particular emotion compared to the probability of being a Neutral expression is given respectively by the rating of the Openness for Pride (+0.24, $p<.001$), the Straightness for Shame (-0.45 $p<.001$), the Speed for Anxiety, Sadness, Joy and Panic Fear (+0.38, $p<.01$ for Anxiety, -0.62 for Sadness $p<.001$, +0.61 for Joy and +0.83 for Panic Fear $p<.001$), and finally the Power for Anger (+0.69, $p<.001$). This result highlights the presence of "emotion related" body cues such as the Straightness of body posture for Shame and the Power of the movement for Anger expression. It also highlights the presence of relevant body cues across different emotions and actions such as the Speed of movement which has been already considered in previous researches on bodily expression of emotion as a highly relevant expressive body cue [Hicheur et al., 2013] [Roether et al., 2009].

8.6 Classification of emotions based on body cues rating

In this section, we explore the automatic classification of emotions, where the predictor variables correspond to body cues rating. We conduct two analyses. In the first one, we examine how the expressed emotions are automatically classified based on body cues rating. We also compare the results of this classification with the recognition rates of emotion perception presented in Table 8.2. In the second analysis, we explore how the perceived emotions are automatically classified based on body cues rating.

We use the Multinomial Logistic Regression (MLR) method to perform the multiclass classification in these two analyses. MRL method is widely considered as a robust approach for multiclass problem in a reduced feature space [Prinzie and Van den Poel, 2008]. We use the samples that received high agreement rates in the "Body Cues Rating" task (84,48% of the stimuli, see Section 8.3). The three-fold cross-validation approach is used to set the training and the test datasets. The

Table 8.5: The percentage of recognition scores for each expressed emotion in each action according to MLR method.

	SW	WH	MB	KD	SD	Lf	Th	mean
Ax	28%	4%	12%	12%	11%	6%	4%	11%
Pr	15%	53%	20%	37%	27%	26%	28%	29%
Jy	49%	23%	24%	11%	40%	30%	25%	29%
Sd	52%	58%	49%	28%	51%	37%	40%	45%
PF	29%	36%	56%	47%	32%	43%	26%	38%
Sh	25%	18%	24%	40%	35%	33%	22%	28%
Ag	17%	22%	25%	14%	19%	35%	**67**%	28%
Nt	53%	21%	40%	33%	31%	33%	27%	34%
CCR	35%	33%	32%	28%	32%	31%	31%	32%

Table 8.6: The percentage of correct recognition scores for each perceived emotion in each action according to MLR method.

	SW	WH	MB	KD	SD	Lf	Th	mean
Ax	28%	8%	28%	54%	40%	40%	8%	29%
Pr	2%	11%	0%	3%	0%	0%	13%	4%
Jy	48%	51%	5%	0%	12%	1%	28%	21%
Sd	80%	77%	29%	62%	71%	44%	81%	64%
PF	9%	16%	-	0%	0%	-	-	6%
Sh	6%	2%	0%	0%	0%	-	-	4%
Ag	21%	18%	17%	0%	10%	33%	64%	23%
Nt	44%	43%	89%	71%	65%	76%	21%	58%
CCR	49%	41%	57%	54%	52%	52%	51%	51%

process is repeated 20 times to smooth the results. The predictor variables are the ratings of the 8 body cues. The value of each predictor variable ranges between 1 and 5 and represents one level on a 5-point scale (see Section 8.2.7). In the following, we refer to the accuracy of a MLR model as the percentage of Correct Classification Rate (CCR). We also refer to Recall measure of a particular class (emotion) as the emotion-CCR.

8.6.1 Classification of expressed emotions using body cues rating

For each action, a MLR classifier is built to classify expressed emotions. The last row of Table 8.5 depicts the Correct Classification Rate (CCR) of each MLR model. The CCR of expressed emotions classification ranges between 28% (in KD action) and 35% (in SW action). Each cell of Table 8.5 (except the last row and the last column) represents a given emotion-CCR (Recall measure) in a particular action. All emotions-CCR are above chance level (12.5%) in all the actions except for Anxiety classification in Th and WH action (see Table 8.5).

We find that Sadness is the best classified across all the actions in both “Emotion Perception” task (see Table 8.3) and “Body Cues Rating” task (see Table 8.5). Anger is the best classified in Throwing action in both “Emotion Perception” (see Table 8.3) and “Body Cues Rating” tasks (see Table 8.5). This result suggests that Throwing could be the best action (among the actions that we study) to convey Anger in both “Emotion Perception” and “Body Cues Rating” tasks. Joy and Pride expressions are also best classified in walking action in both “Emotion Perception” and “Body Cues Rating” tasks. The classifications of Panic Fear and Shame across all the actions are above chance level in the “Body Cues Rating” task (see Table 8.5) while they were not in the “Emotion Perception” task (see Table 8.3). While Panic Fear expression has never been perceived as Panic Fear in Moving Books action (0% recognition rate, see Table 8.3), its CCR is higher than 50% in automatic classification based on body cues rating (See Table 8.5). Panic Fear and Shame are the least recognized across all the actions in the “Emotion Perception” task (see Table 8.3), but Anxiety is the least recognized across all the actions in the “Body Cues Rating” task (see Table 8.5).

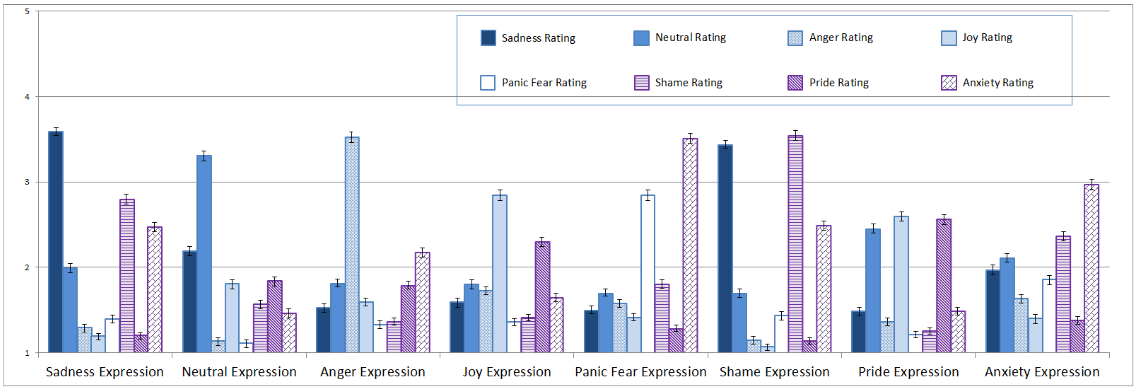
8.6.2 Classification of perceived emotions using body cues rating

As explained in Section 8.4.2, we use “the most frequent label" approach to rate perceived emotions in each stimulus. Seven MLR models are built to classify perceived emotions in each action based on body cues rating. The classification rates are shown in Table 8.6. The CCR of each classifier is shown in the last row of Table 8.6, ranging from 41% for the action Walking with an object in the hands to 57% in the Moving books action. Across all the actions, the classification of perceived emotions (51%, see Table 8.6) is better than the classification of expressed emotions (32%, see Table 8.5) based on body cues rating.

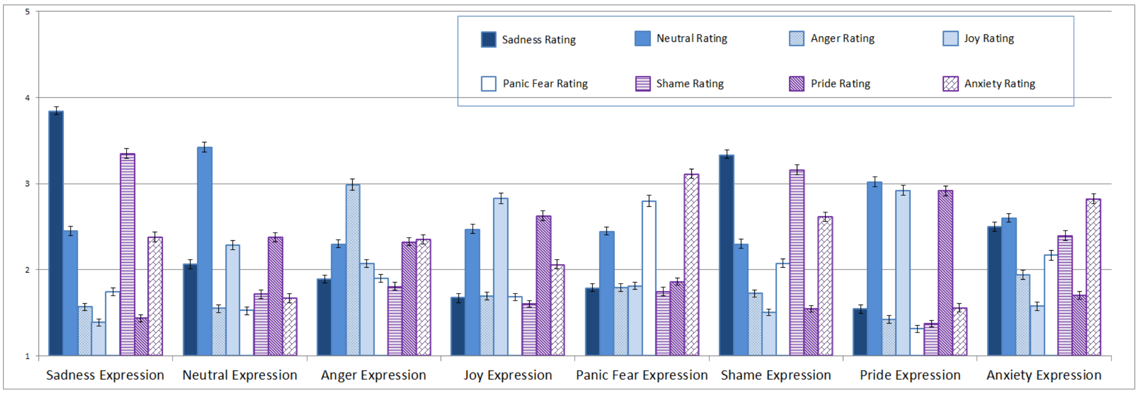
Each cell in Table 8.6 (except the last row and the last column) depicts an emotion-CCR in each action. The hyphen ‘-’ is used when no participant perceived that emotion with a score above “3". Compared to the results obtained in the expressed emotions classification based on body cues rating (see Table 8.5), the classification of perceived emotions across all the actions is better for Anxiety, Sadness and Neutral (see Table 8.6).

We observe a high amount of low emotions-CCR in perceived emotions classification based on body cues rating (see Table 8.6). This result is mainly due to two issues; 1) the confusion existing between the classifications of perceived emotions based on body cues, and 2) the unbalanced problem in multiclass problem.

As we aim to understand to which extent we can classify perceived emotions based on body cues rating, we propose to cluster the emotions (as the most frequent perceived emotions) that are the most confused by participants (see Table 8.2). Thus, we grouped Anxiety&Panic Fear, Pride&Neutral, and Sadness&Shame; it



(a) Emotion expression description in terms of emotion rating in RISC survey for a subset of stimuli.



(b) Emotion expression description in terms of emotion rating in Mturk survey for a subset of stimuli

Figure 8.8: Emotion expression description in terms of emotion rating in RISC survey vs Mturk survey for a subset of stimuli

gave 3 clusters. Joy and Anger expressions data are not considered as we aim to maintain an equal distribution between classes. We performed 7 MLR classifiers (for each action) to classify those three clusters. We found that the overall CCR (across all the actions) is 61%. Across all the actions, the cluster Sadness&Shame received the best recognition rate with 65%, followed by Pride&Neutral with 60% and finally Anxiety&Panic Fear with 56%.

8.7 Emotion perception task results obtained with another pool of participants

Our perceptual study was conducted with a large pool of participants (1008 participants). All these participants were found via Mechanical Turk crowd sourcing website. Most of them spent the majority of their life in US (89.21%). In order to explore the effect of these two factors (US culture and Mechanical Turk crowd sourcing website) on the recognition of emotions, we compare the results of the "Emotion Perception" task with another pool of participants. Similarly to the survey conducted on Amazon Mechanical Turk crowdsourcing website, we carried out an online survey with researchers from different research areas and culture; we have contacted participants though a French mailing list (called RISC) that reaches out researchers in cognitive sciences. Henceforth, we called this survey "RISC survey". Our aim is to compare the results of emotion recognition obtained from "RISC survey" with the results of emotion recognition provided from the Mechanical Turk crowd sourcing website, called "Mturk survey".

RISC survey:
Unlike Amazon Mechanical Turk crowdsourcing website, we did not have access to a large pool of participants on RISC mailing list. Consequently, we conducted the RISC survey using a small amount of stimuli. This set of stimuli was restricted to 64 sequences of emotional behaviors. They refer to the expression of the 8 emotions recorded in the Emilya database and expressed in 2 actions: Moving Books (32 stimuli= 4 stimuli * 8 emotions) and Simple Walking actions (32 stimuli= 4 stimuli * 8 emotions). 287 participants from RISC mailing list took part in RISC survey (66.55% of females and 33.45% of males, mean age of 37.31 years old ranging from 18 to 78 years old). Most of the participants reached through RISC mailing list spent the majority of their life in France (81.53% of them).

Mturk survey:
RISC survey is limited to 64 stimuli (out of 664 stimuli used for the first survey). In order to compare the results provided from RISC and Mturk surveys, we focus on the participants of Mturk survey who evaluate a subset of these 64 stimuli. Out of the 1008 participants who took part in Mturk survey, only 96 participants evaluated the recognition of emotions from this set of 64 stimuli; 61.45% of females and 38.54% of male, mean age of 38.43 years old ranging from 18 to 76 years old. Most of these 96 participants, reached through Mechanical Turk crowd sourcing website, spent the majority of their life in US (96.88% of them).

RISC vs Mturk surveys:
In order to compare the results of emotion perception obtained in Mturk and RISC surveys, we firstly discuss the characterization of the expressed emotions in terms of emotions rating in Figure 8.8. Secondly, we compare the statistical results of two Multinomial Logistic Regression (MLR) models built using the results of each survey.

Figure 8.8 shows that the tendency of emotions perception mean ratings for each expressed emotion in RISC surveys is highly similar to the results found in Mturk survey. For instance, in both RISC and Mturk surveys, the participants perceived Sadness with the highest mean rating in the stimuli showing Sadness expression. Besides, in both RISC and Mturk surveys, the participants perceived Anxiety with

8

the highest mean rating in the stimuli showing Panic Fear expression. Overall, Sadness, Neutral, Joy, Anger and Anxiety were "correctly" perceived in both Mturk and RISC surveys. Panic Fear was "not correctly" perceived in both Mturk and RISC surveys as it was mainly perceived as Anxiety based on the mean ratings of emotion perception. The mean ratings of emotions perception were only slightly different between RISC and Mturk surveys in the stimuli showing Shame and Pride expressions.

We also built two MLR models to compare the statistical results of emotions ratings in Mturk and RISC surveys. The Neutral expression was considered as the reference category. We found that the ratings of Anxiety, Joy, Panic Fear, Shame and Anger are significantly more positively associated with respectively Anxiety, Joy, Panic Fear, Shame and Anger expressions than with Neutral expression. For instance, we found that the probability of being Anger expression increases significantly ($p<.001$) exp(1.75) times for RISC survey and exp(0.93) times for Mturk survey for each unit increase in Anger rating given all other emotions ratings equal. In both RISC and Mturk surveys, we did not found a significant difference between the probabilities of being Pride or Neutral for Pride rating. Based on the data obtained from RISC survey, we did not found a significant difference between the probabilities of being Sadness or Neutral for Sadness rating. This result is not observed based on the data obtained from Mturk survey; the rating of Sadness was significantly ($p<.01$) more positively associated with Sadness expression than with Neutral expression. However, based on the data obtained from Mturk survey, we did not found a significant difference between the probability of being Neutral or Sadness expressions for Neutral rating.

We can conclude that we observed a huge amount of similarity in emotion recognition ratings in both RISC and Mturk surveys. This result suggests that the results of the "Emotion Perception" task using Amazon Mechanical Turk crowdsourcing website seem not to be restricted to the participants found through Amazon Mechanical Turk crowd sourcing website. Similarly, these results seem also not to be restricted to the participants that spent the majority of their life in US as similar results were obtained with participants that spent the majority of their life in France.

8.8 Conclusion

In this Chapter, we described our perceptual experiment on Emilya database. During this perceptual study, we asked participants to perform two tasks: "Emotion Perception" task and "Body Cues Rating" task. The first task is intended to evaluate the perception of emotions. The second task is intended to obtain the perceptual ratings of expressive body cues. Both tasks are based on the perception of emotional behaviors recorded in the Emilya database and reproduced on a virtual actor. Using a computer avatar that has the shape of a puppet with no sign of gender, culture and facial expression allows reducing bias of these factors. As the Emilya database contains a large amount of expressive sequences (more than 8000 sequences), the perceptual study was conducted for a subset of the whole database using Mechanical Turk crowd-sourcing web site. The samples were selected randomly for each actor, action and emotion.

A multi-labeling approach was adopted to label the perceived emotion by participants for all the stimuli used in the perceptual study. We found that the perception of each emotion is statistically positively associated with the corresponding expressed emotion. While Sadness, Anger, Neutral, Joy and Anxiety were perceived as such, Shame was often perceived as Sadness, Pride as Neutral and Panic Fear as Anxiety. We reported several factors that may contribute to these confusions such as the lack of further body features (e.g. facial expressions) coupled with a similarity at the level of bodily characterizations, but also the lack of including contextual factors. Indeed, the participants were not provided with contextual factors regarding the identity of actors (their gender, cultural background) neither the scenario used to elicit emotions. Such factors can have a considerable impact on the recognition of emotions [Kleinsmith and Bianchi-Berthouze, 2013].

Based on a subset of samples, we also compared the results of the “Emotion Perception” task using Amazon Mechanical Turk crowdsourcing website with the one obtained using a French mainling list that reaches out researches in cognitive sciences. As we found similar results, we concluded that the results of the “Emotion Perception” task seem not to be restricted to the participants found through Amazon Mechanical Turk crowd sourcing website. Similarly, these results seem also not to be restricted to the participants that spent the majority of their life in US as similar results were obtained with participants that spent the majority of their life in France.

Moreover, we studied the effect of actions on the recognition rates of the “Emotion Perception” task. Anger was the best recognized in Throwing action, Neutral in Moving Books action, Pride and Joy in Walking action and Anxiety in Knocking action. Overall, Sadness was the best recognized across all the actions, in particular in Sitting Down and Throwing actions. In related works, Sadness expression in body movement has been widely considered as the best recognized emotion [Kleinsmith and Bianchi-Berthouze, 2013, Roether et al., 2009, Gross et al., 2010].

Another aspect we tackled in this paper is the characterization of emotions through body cues across all the actions. We found similar results as those reported in previous studies. A statistical analysis revealed interesting findings about the presence of “emotion related” relevant body cues such as the power of movement for Anger expression.

Finally, we explored the classification of expressed emotions based on body cues rating. The overall classification rate of emotion recognition in “Body Cues Rating” task was lower than the recognition rates observed in “Emotion Perception” task. This result suggests that more body cues are needed to better discriminate between the expressed emotions. Interestingly, the similarity of the effect of an action on the perception of emotion and on the classification rates based on body cues rating reveals interesting findings and suggests the presence of “associated” actions to some expressed emotions, such as Throwing for the Anger expression. However, the effect of an action on the classification of expressed emotions should be studied through further analyses based on motion capture data in order to establish a clear relationship between the expressed emotions and the performed actions. We also compared the classification of expressed and perceived emotions based on body cues rating. We reported an increase of the classification rates of perceived emotions compared to the classification rates of expressed emotions. Clustering the emotions that were confused at the perception level leaded also to an increase of the classification rates.

8.9 Summary of chapter

- The validation of Emilya database was achieved through a perceptual study (an online survey) using Mechanical Turk crowd-sourcing web site. The perceptual study is conducted using stimuli that display Emilya motion capture data reproduced on a computer avatar.
- During the perceptual study, the participants were asked to perform two tasks: 1) "Emotion Perception" task and "Body Cues Rating" task. The results of the first task are used to explore the recognition of expressed emotion. The results of the second task are used to explore the characterization and the classification of expressed emotions based on the perceptual rating of 8 body cues.
- Across all the actions, we find that the rating of emotion perception is significantly positively correlated with the corresponding emotion expression. Based on the most frequent perceived emotion, we find that Sadness, Anger, Neutral, Joy and Anxiety expressions are "correctly" perceived, while Panic Fear, Shame and Pride were respectively confused with Anxiety, Sadness and Neutral. We conclude that the confusions occurred at the level of emotion perception may be due to a lack of contextual factors, to a similarity of bodily expressions (e.g. similar characterization of Sadness and Shame bodily expressions), but also to the lack of other modalities that may contribute to a better recognition of bodily expression of these emotions (e.g. facial expressions).
- The results of the "Emotion Perception" task using Amazon Mechanical Turk crowdsourcing website seem not to be restricted to the participants found through Amazon Mechanical Turk crowd sourcing website as similar results were found using a French mainling list that reaches out researches in cognitive sciences. Similarly, these results seem also not to be restricted to the participants that spent the majority of their life in US as similar results were obtained with participants that spent the majority of their life in France.
- We also discuss and compare the profile of each expressed emotion based on the rating of 8 body cues: the Straightness, the Leaning, the Openness of body posture, the Regularity, the Quantity of arms movement, and the Speed, the Fluidity and the Power of body movements. The profiles of expressed emotions characterization are mostly in line with previous findings in psycho-

8

logical researches (e.g. bodily expression of Sadness is perceived as slow and collapsed).

- We explored the effect of actions on the recognition rates of emotion perception and on the classification of emotion based on body cues ratings. The effect of actions on the recognition rates of emotion perception seems to be similar to the one reported in automatic classification of expressed emotions based on body cues ratings. This result suggests the presence of "associated" actions to some expressed emotions, such as Throwing for the Anger expression.
- Across all the actions, the classification of expressed emotions considering body cues ratings as predictor variables leads to lower classification rate that the perceptual recognition of emotions in "Emotion Perception" task. This result suggests that more body cues are needed to better discriminate between the expressed emotions. However, the classification of perceived emotions leads to better results than the classification of expressed emotions.

Part IV : Emilya motion capture data analysis

9 Multi-level classification of emotional body expression in daily actions

In chapter 8, we explored the human recognition of emotions expressed in the Emilya database based on the perception of animated stimuli. In this chapter, we explore their automatic classification based on a kinematic analysis that makes use of motion capture data. Random Forest approach is used for the classification purpose.

Thanks to the variability of emotions expressions and actions in the Emilya database, we are able to compare the expression of emotion in different daily actions. We also compare the automatic classification and the human recognition of expressed emotions. Moreover, based on our multi-level body movement notation system, we compare the contribution of different types of body cues to the automatic classification of expressed emotions (e.g. Lower body cues vs. Upper body cues). We also explore the contribution of temporal features to the classification of expressed emotions.

This chapter is organized as follows: section 9.1 briefly discusses previous works on the classification of emotions in body movement and the motivation of using Random Forest approach. Section 9.2 briefly introduces Random Forest approach and its parameterization. Section 9.3 is devoted to the results of emotion classification using motion capture data. Our body movement notation system described in Chapter 6 is used for this purpose. Section 9.4 illustrates the comparison between perceptual emotion recognition in "Emotion Perception" task, "Body Cues Rating" task and the automatic classification achieved using RF approach and motion capture data. In section 9.5, we provide deeper insights into the contribution of different description levels of our body movement notation system to the classification of emotions using RF approach and motion capture data. Section 9.6 discusses the contribution of Temporal Profile features. Finally, sections 9.7 and 9.8 conclude and summarize this chapter.

9.1 Classification of emotions in body movement: Related work

Recently, several works used machine learning algorithms to automatically classify emotions expressed in body movement using explicit body cues and motion capture data. Kapur et al. [Kapur et al., 2005] consider the mean of position, velocity and acceleration of 14 body joints to classify four prototypical expression of emotions through the whole body movement. They compare the performance of five different classifiers for the classification of emotions; 1) a logistic regression, 2) a naive bayes with a single multidimensional Gaussian distribution modeling each class, 3) a decision tree classifier based on the C4.5 algorithm, 4) a multi-layer perceptron backpropogation artificial neural network (MLP), and 5) a support vector machine (SVM) trained using the Sequential Minimal Optimization. They found that SVM and MLP provide the best performances. These classifiers also tend to exhibit good generalisation performance. Based on the result provided in [Kapur et al., 2005], Bernhardt and Robinson [Bernhardt and Robinson, 2007] used support vector machines with a polynomial kernel to detect emotions from knocking motion. They used four statistical measures applied to hand and elbow motion; extension of posture, average of speed, average of acceleration and average of jerk. Kleinsmith et al. [Kleinsmith et al., 2011] used multi-layer perceptron to automatically recognize non-acted affective postures using rotation of body joints (e.g. knee, shoulder...) in x,y and z directions.

The use of machine learning techniques in above studies is mostly reduced to the purpose of classification. When the goal is to evaluate the discriminative power of each feature for distinguishing between emotions, statistical tests have mostly been used. In [Kleinsmith et al., 2011], each postural feature was subjected to one-way ANOVAs to evaluate its discriminative power. Based on one-way ANOVA and Tukey's HSD test, Camurri et al. [Camurri et al., 2003] explored the statistical effect of emotions expressed in full-body movement (dancing) on four body cues (overall duration of time, Contraction index, quantity of motion and motion fluency).

In our work, we aim to explore both the prediction accuracy and the discriminative power of each body cue for the classification of emotions. That is a powerful classifier that incorporates an embedded feature ranking strategy. Unlike statistical tests as one-way ANOVAs, Embedded feature ranking methods allow exploring all the original features and automatically selecting the most relevant ones discovered during the training process. A typical model is a decision tree algorithm such as CART, ID3 and C4.5. However, in addition to their instability to changes in learning dataset, the performance of decision tree algorithms is sensitive to the presence of several "weakly relevant" features [Breiman, 2001]. They tend to overfit the dataset. It has been shown that an ensemble method that combines the prediction of several decision trees is better than a single decision tree [Breiman, 2001]. Random Forests approach (RF), introduced a couple of years ago by Breiman [Breiman, 2001], has been known as an efficient non-parametric ensemble method that provides both high performance and an embedded approach for feature selection. Besides, RF approach returns measures of relevance (importance) for each input feature. Consequently, the advantage of using RF approach in our work is double-fold: 1) reliability of the classification model and 2) possibility to select a subset of relevant features based on their relevance measures.

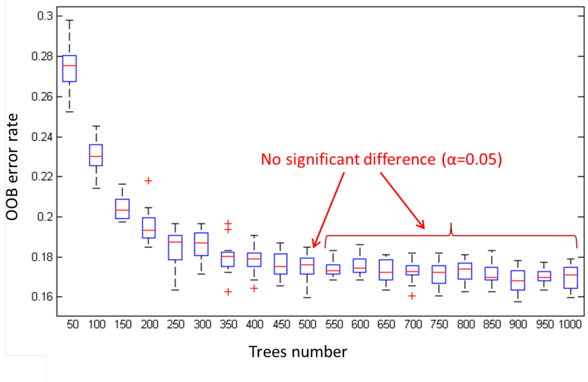


Figure 9.1: The boxplot of the OOB error rates for each RF model run 50 times for each trees number. In this example, the Tukey-Kramer test indicates that there is no significant improvement (α =0.05) of the OOB error rate when using more than 500 trees.

9.2 Random Forests (RF)

In this section, we briefly describe the Random Forest approach and the technique that we use to assess the number of trees. More details of Random Forest approach can be found in Appendix D.

Random Forest (RF) model builds several decision tree predictors which are combined at the decision level. During the training process, each tree in the forest is grown on a bootstrap sample of the data. The samples that were not considered in the bootstrap related to a particular tree are called “out-of-bag” (OOB) data, and they can be useful to obtain the prediction performance of that tree as well as the overall error rate of the model. As such, there is no need for a test set to get the classification error. The OOB error is estimated internally during the learning process (See section D.4, Appendix D).

The number of trees in the forest (ntrees) has to be sufficiently large so that the prediction of the ensemble method can be stabilized [Svetnik et al., 2004]. So far, except few recent works [Oshiro et al., 2012, Díaz-Uriarte and Alvarez de Andrés, 2006], the literature associated with the RF approach gave few details about the optimal number of trees that has to be considered to achieve a good tradeoff between the computational time and the model prediction accuracy. We propose to estimate the optimal number of trees based on statistical tests. As shown in Fig. 9.1, our approach consists of running RF model several times for different values of ntrees parameter. For each run, we measure the OOB error rate of the model. Starting from 50 trees, we increase ntrees parameter with 50 trees until 1000 trees (See Fig. 9.1).

We first apply a One-way Anova test on the sets of OOB error rates to know whether the change in trees number has a significant effect on the OOB error rate.

If the difference is significant, we apply a post-hoc test (we use the Tukey-Kramer test) for a pairwise comparison of the sets of OOB error rates. Finally, the optimal trees number is chosen in such a way that there is no significant improvement in term of OOB error rates (considering a significance level of 0.05) if we increase the number of trees (See Fig. 9.1). In the following sections, the number of trees that we use to build each RF model is determined as just explained. We attribute the default values to the other parameters of RF (the number of variables used for the split of each node in a single tree and the minimum node size) as recommended in previous works [Díaz-Uriarte and Alvarez de Andrés, 2006, Svetnik et al., 2004].

9.3 Classification based on multi-level features

In this section, we describe the motion capture data used for the classification of expressed emotion using RF approach. We recall that the emotions that we asked the actors to express are Joy (Jy), Anger (Ag), Panic Fear (PF), Anxiety (Ax), Sadness (Sd), Shame (Sh), Pride (Pr) and Neutral (Nt). The actions are Simple Walk (SW), Walk with an object in hands (WH), Sitting down (SD), Being Seated (BS) Knock at the door (KD), Move books on a table with two hands (MB), Lift (Lf) and Throw (Th) an object (a piece of paper) with one hand. A manual segmentation of the action Sitting Down (SD) and the action Being Seated (BS) is performed for automatic analysis as Sitting Down action involves the change of whole body posture while Being Seated action involves small changes in body parts posture. However, these actions were not separated in the stimuli created for the perceptual experiment described in the Chapter 8.

9

8206 motion capture files are used for this study; 1038 for BS, 1027 for KD, 1019 for Lf, 1031 for MB, 1025 for SW, 1038 for SD, 1006 for Th and 1022 for WH. Indeed, around 130 motion capture files (considering the variability of actors, emotion scenarios and action repetitions) are available for each emotion expression in a single action.

The features considered for the classification task consists in the set of multi-Level features described in Chapter 6. This set of multi-level features refers to 114 motion capture based features used to describe the expressive movement of a given motion capture sequence. We build 8 RF models for the classification of the expressed emotions using this set of 114 motion capture features. Each RF model corresponds to one action and it is run 50 times to smooth the results as recommended in previous works [Svetnik et al., 2004]. The number of trees used in each model is chosen as explained in section 9.2.

Table 9.1 shows (in the firs row) the results of expressed emotions classification in each action (that is for each RF model). The percentages provided in the first row correspond to the average of classification rates across 50 runs. They are measured based on the OOB error rates (See section D.4 in Appendix D). We refer to this classification rate as CCR_{OOB}. When a Cross-Validation framework is adopted, we refer to the classification rate as CCR_{CV}.

In order to get deeper insight into the classification of each emotion, we also provide in Table 9.1 the individual Recall measure for each emotion expressed in each action. Recall measure is defined as the percentage of observation for a given emotion that are correctly classified regarding the total amount of observation of this emotion. We refer to this recall measure as $CCR_{Emotion}$.

9.3.1 Results of RF classification per action

In this subsection, we discuss the CCR_{OOB} of expressed emotions classification in each action. These results are shown in the first row of Table 9.1.

Walking: SW and WH :

As we can observe from Table 9.1, the classification of emotions using the whole set of Multi-Level features receives the best results in walking actions (the corresponding CCR_{OOB} are 85% for SW and 84% for WH).

Walking has been well used in previous works to explore the human recognition and the the characterization of implicit bodily expression of emotion [Roether et al., 2009] [Hicheur et al., 2013] [Crane and Gross, 2007] [Karg et al., 2010]. These studies showed that humans are able to recognize emotions expressed in walking pattern. They also revealed the effect of emotion expression on the characteristics changes of walking pattern. However, only few works studied the automatic classification of emotions expressed in walking using machine learning techniques [Janssen et al., 2008] [Karg et al., 2010].

Janssen et al. [Janssen et al., 2008] explored the classification of 4 emotions ("normal", happy, sad and angry) expressed in walking using artificial neutral networks. Their analyses are based on Kinetic data (3D ground reaction forces during gait obtained from a force plate in the ground). When using time discrete parameters (e.g. minima, maxima, positions of minima, positions of maxima ...), the classification rate achieves 80.8%. Using time continuous data (whole time courses) allowed the increase of the classification rate to 83.7%. Although the classification using Kinetic data ([Janssen et al., 2008]) and Kinematic data (in our analyses) are not directly compared, our experiments show slightly better classification scores (85% in SW and 84% in WH) than the ones found in [Janssen et al., 2008]. In other words, we showed that the classification of a larger set of emotions (8 emotions) expressed in walking action using our set of multi-level features shows slightly better results than the classification of a smaller set of emotions ("normal", happy, sad and angry [Janssen et al., 2008]) expressed in walking action using Kinetic features [Janssen et al., 2008]. This result highlights the robustness of our multi-level set of features in the discrimination of emotions expressed in walking.

Karg et al. [Karg et al., 2010] compared the performance of several machine learning techniques for the classification of 4 emotions (Neural, Happy, Angry, Sad) expressed in walking. Their analyses are based on kinematic parameters that describe the joint angle trajectories. Across different combinations of feature selection techniques and classification models, the classification rate reached a maximum of

95%, which is highly better than our classification rate (85% in SW and 84% in WH). While their classification model is restricted to the discrimination between 4 expressed emotions, our classification model deals with the recognition of 8 expressed emotions. In Chapter 8 we showed that some expressed emotions (e.g. shame and sadness) were confused at the level of human perception. This result may suggest the similarity of their characterization, which makes the classification task more difficult than the one achieved in [Karg et al., 2010].

Moving Books and Knocking: MB and KD :

Expressed emotions classification receives slightly lower CCR_{OOB} in Moving Books (CCR_{OOB} is 83%) and Knocking (CCR_{OOB} is 82%) actions than in walking actions (85% in SW and 84% in WH).

The result that we obtain for the classification of 8 emotions expressed in Knocking using our set of Multi-Level features (82%) turns out to be slightly better than the one reported in [Bernhardt and Robinson, 2007] (81.1%). In their experiment, Bernhardt et al. [Bernhardt and Robinson, 2007] explored the classification of 4 emotions (Neutral, Anger, Sadness and Joy) using SVM classifier and kinematic data. Few body cues were measured for each motion primitive (maximum distance between right hand/elbow and body, average of right hand and right elbow speed, acceleration and jerk).

Lifting and Throwing: Lf and Th :

The classification of the expressed emotions in Throwing (Th) action (CCR_{OOB} is 79%) is slightly higher than their classification in Lifting action (CCR_{OOB} is 78%). The classification of expressed emotions turns out to be more difficult in Lifting and Throwing actions compared to their classification in Moving Books (MB) and Knocking (KD) actions.

Sitting Down and Being Seated: SD and BS :

Expressed emotions classification is similar in Sitting Down (SD) and Being Seated (BS) actions (CCR_{OOB} are 68%). SD and BS actions seem to be the least expressive actions in Emilya database as they receive the lowest classification rates in emotions classification.

In previous works, the detection of affects and emotions in Sit Down action has been widely based on the seated posture [Mota and Picard, 2003] [D'Mello and Graesser, 2009](that is the body posture during being seated and not the action of sitting down). Unlike our analyses based on Kinematic data, previous studies mainly focus on Kinetic data through pressure sensors mounted on strategic places on the seat and on the back of a chair [Mota and Picard, 2003] [D'Mello and Graesser, 2009]. Besides, these studies show high recognition rate for the discrimination between two or three emotional states (typically the presence or not of an emotional state) but low recognition rate with a large number of emotions.

The experiments presented in [Mota and Picard, 2003] are dedicated to detect the level of interest of the user during human computer interaction. In [Mota and Picard, 2003], a HMM model was applied to classify three categories of interest based on Kinetic data: high interest, Low interest and Taking a break. Their

Table 9.1: The overall CCR_{OOB} (%) for each RF model built for one action and the individual CCR for each emotion for each RF model.

(%)	SW	MB	WH	KD	BS	SD	Lf	Th
CCR_{OOB}	**85%**	83%	84%	82%	68%	68%	78%	79%
CCR_{Ax}	**83%**	72%	79%	70%	52%	53%	66%	69%
CCR_{Pr}	**90%**	83%	83%	86%	78%	82%	83%	79%
CCR_{Jy}	**84%**	78%	77%	72%	54%	56%	80%	74%
CCR_{Sd}	89%	91%	**94%**	85%	76%	76%	82%	80%
CCR_{PF}	82%	**86%**	78%	82%	56%	61%	81%	80%
CCR_{Sh}	82%	82%	87%	**90%**	69%	66%	78%	86%
CCR_{Ag}	80%	92%	86%	92%	84%	80%	84%	**97%**
CCR_{Nt}	88%	84%	**90%**	77%	72%	69%	73%	64%

8-fold cross-validation setup achieves a recognition accuracy of 82.25%. The experiments presented in [D'Mello and Graesser, 2009] is devoted to the recognition of the emotional states experienced by a learner during a tutorial task. The recognition accuracy referring to the discrimination between between two, three, four, and five affective states were respectively 71%, 55%, 46%, and 40% (with chance rates being respectively 50%, 33%, 25%, and 20%). Thus, their classification accuracy decrease with the increase of the number of classes.

9.3.2 Results of RF classification per Emotion

In this subsection, we discuss the effect of actions on the classification of each emotion. This discussion is based on the comparison of the recall measures across different actions. The recall measure of each emotion classification in each action as presented in Table 9.1. We discuss the effect of actions on the classification of Anxiety, Pride, Joy, Sadness, Panic Fear, Shame, Anger and Neutral.

Anxiety: Anxiety expression is best classified in walking actions (Anxiety CCR_{Ax} is 83% in SW and 80.1% in WH), but it is worst classified in Sitting Down and Being Seated actions (Anxiety CCR_{Ax} is 53% in SD and 52% in BS). Most of the actors tend to fidget with the hands when expressing Anxiety. However, fingers motion is not captured by our motion capture system, thus we could not consider such cues in our analyses.

Pride: Pride expression is also best classified in walking actions (90% in SW and 83% in WH) and worst in Being Seated action (78%). Interestingly, Pride expression classification also receives high CCR_{Pr} in actions involving repetitive arm movements; Knocking and Moving Books actions (86% in KD and 83% in MB). Unlike the results obtained for most of the expressed emotions, Pride expression is better classified in Sitting Down than in Being Seated, and in Lifting than in Throwing.

Joy: Similarly to Pride expression classification result, Joy expression is best classified in walking actions (84%) and it is worst classified in Being Seated action (54%). Similarly to the result reported in [Bernhardt and Robinson, 2007] resulted from SVM classifier, the CCR_{Jy} of Joy expression in Knocking motion (Happiness in their study) is lower than the CCR_{Jy} of Neutral, Anger and Sadness. However, the CCR_{Jy} of Joy expression in Knocking reported in our work (72%) is better than the one found in the study of Bernhardt and Robinson [Bernhardt and Robinson, 2007] (65.3%).

Sadness: Sadness expression classification receives the highest classification rate in walking and Moving Books actions (94% in WH, 91% in MB and 89% in SW). The classification error rate of Sadness is very similar in Sitting Down and in Being Seated actions, as well as in Lift and in Throw object actions.

Panic Fear: Panic Fear expression is best classified in Moving Books action (CCR_{OOB} is 86%). We also note that the CCR_{PF} of Panic Fear classification in Walking and Knocking actions receive also high values (82% in SW and 82% in KD). Thus it seems that Panic Fear is the best conveyed in actions involving repetitive motor activity.

Shame: Shame expression classification receives the best classification rate in Knocking action (90%) but the lowest classification rates in Sitting Down and Being Seated actions (66% in SD and 69% in BS).

Anger: The results reported in Table 9.1 indicate that Anger expression is best classified in the actions that mainly involve a particular body part (Knocking, Throwing, Moving Books) rather than the actions that are described through the motion or posture of the whole body (Simple Walking, Sitting Down, Being Seated). These results indicate that Anger expression is better communicated in actions that mainly involve a particular body part than in actions that involve the whole body movement. Previous studies highlighted the presence of forceful movement and high movement dynamics in bodily expression of Anger [Dahl and Friberg, 2007, Wallbott, 1998]. Thus, our results can be explained as all the energy and the power of Anger expression is accumulated in a single body segment rather than dispensed over the whole body. Besides, we find that Throwing action is the best action that conveys Anger expression as Anger classification receives a significantly higher CCR_{Ag} in Throwing action (97%).

Neutral: Neutral expression classification receives the highest CCR_{Nt} in the action of Walking with an object in the hand (90%). The CCR_{Nt} of Neutral classification in Knocking motion (77%) is slightly better than the one reported in [Bernhardt and Robinson, 2007] (74.2%) although they focused on a few set of body cues and they consider less emotions in their classification algorithm (only 4 emotions classified using SVM approach).

9.3.3 Inter-individual classification

In the above subsections, we discussed the scores of RF emotions classification based on the classification accuracy that is estimated internally within the RF model (CCR_{OOB}). In this subsection, we tackle the issue of inter-individual classification as it has been widely reported in the study of emotion recognition in body movements [Karg et al., 2010] [Janssen et al., 2008]. This problem refers to the effect of individuality on the expression of emotion. In the following paragraphs, 1) we explain the difference between Inter-individual classification and OOB based classification (the one presented in previous subsections), 2) we refer to the related works on Inter-individual classification and 3) we present the results of Inter-individual classification.

<u>Inter-individual Vs OOB based classification :</u>
Inter-individual classification corresponds to a cross-validation framework where one actor is left out in the training set [Karg et al., 2010]. To obtain the classification accuracy, the *unknown* samples of that *unknown actor* are classified. This framework is different from the one based on OOB error rate (i.e. CCR_{OOB} discussed in previous subsections). The estimation of CCR_{OOB} can be described as the classification of *unknown* samples of probably *known* actor. Indeed, each tree in the forest is grown on a bootstrap sample (randomly selected) of the data. About 2/3 of this bootstrap are used to construct that tree. As such, the samples left out to estimate the classification accuracy can correspond to *unknown* samples of *known* actors.

<u>Inter-individual classification in related work :</u>
Janssen et al. [Janssen et al., 2008] compared Intra-individual and Inter-individual classification of emotions expressed in walking. Intra-individual classification refers to the recognition of emotions expressed by a particular actor (i.e. training and testing datasets are derived from the samples of the same actor). Across all the actors, Intra-individual classification rate reached 83.7% on average. However, Inter-individual classification remains around the chance level, which means that the actors expressed emotions differently.

Karg et al. [Karg et al., 2010] conducted a similar experiment to explore the classification of emotions expressed in walking action. However, they found that Inter-individual classification reaches above chance level score (69%) and it is comparable with human recognition of emotions (63%). Their Intra-individual classification reaches on average (across actors) 95% of accuracy.

<u>Results of Inter-individual classification :</u>
In order to explore the Inter-individual classification in our database, we build 12 RF models (for 12 actors) for each action (8 actions), that is 96 (12*8) RF models. A cross-validation framework is applied where only one actor is left of the training

dataset for each model and the CCR_{CV} is measured for each model. We measure the average of Inter-individual classification for each action across all the actors. The averaged CCR_{CV} for respectively SW, MB, WH, KD, BS, SD, Lf, and Th actions are 52,54%, 45,28%, 49,68%, 42,57%, 41,96%, 40,32%, 42,54%, 42,56% and 44,68%.

Similarly to [Karg et al., 2010], we found that Inter-individual classification achieves an accuracy that is lower than a framework including individual bias: CCR_{CV} ranges from 40.32% to 52.54% < CCR_{OOB} ranges from 68% to 85%. Thus, considering the individuality allows the increase of the classification of emotions. We also found that the Inter-individual classification scores are above the chance level (40.32% to 52.54% > 25%). As such, despite the classification of unknown actors' emotions, the classification performance outperforms the random level. This result suggests that our classification approach can be generalized to unknown observations while achieving above chance level classification scores.

9.3.4 Summary

We compared the RF classification rates across the daily actions considered in Emilya database. Firstly we compare the classification of emotions for each action (See section 9.3.1). Secondly we compare the classification rates of each emotion (See section 9.3.2).

We found that walking actions are the most expressive action as SW and WH received the best classification rates. The classification of expressed emotions has also shown good classification rates in repetitive arms movement based actions (Knocking and Moving Books). The classification scores that we obtained in these actions are comparable with the results of previous studies. We also found that Sit Down actions (SD and BS) are the least expressive actions as they received the lowest classification rates. However, previous studies on emotion recognition in sitting down action are mostly based on Kinetic data obtained from pressure sensors based data [Mota and Picard, 2003]. As Kinetic data has shown low recognition accuracy for the discrimination between several emotions, combining Kinematic and Kinetic data may significantly increase the recognition rates of emotions expressed in seated postures. In this thesis, we only focus on Kinematic data analysis.

We also found that the classification of each emotion shows mostly the best score in walking actions (SW for Anxiety, Pride and Joy, and WH for Sadness and Neutral). However, Panic Fear, Shame and Anger were respectively the best classified in Moving Books, Knocking and Throwing. Throwing action, in particular, seems to be related to Anger expression. On one hand, Anger is the best classified in Throwing with a high classification rate: 97%. On the other hand, the classification rates of other emotions in Throwing action are not as high as in other actions.

Finally, we explore the problem of Inter-individual classification, that refers to the effect of individuality on the classification of expressed emotions. We found that Inter-individual classification is achieved above chance level but considering individuality allows increasing classification score. Thus, we conclude that our classification model is robust enough to deal with the classification of unknown observations above the chance level. We also conclude that considering a prior knowledge of individuality allows the increase of expressed emotions recognition.

9.4 Human recognition Vs Automatic classification of expressed emotions

In this section, we compare the results of automatic classification of emotions just presented and the results of human recognition of emotions obtained in our perceptual study (See Chapter 8). We compare three measures; 1) Random Forest (RF) classification of emotions using motion capture based features, 2) Human recognition of emotions in "Emotion Perception" task (See Chapter 8) and 3) Multinomial Logistic Regression (MLR) classification of emotions based on their perceptual characterization performed in "Body Cues Rating" task (See Chapter 8). MLR classification refers to the classification of the expressed emotions based on the human rating of 8 body cues along 5-point scale: Straightness, Leaning and Openness of body posture, Regularity, Quantity, Speed, Fluidity and Power of body movement (See Chapter 8).

Both "Emotion Perception" and "Body Cues Rating" tasks were performed during the perceptual study detailed in Chapter 8. In chapter 8, we focused on the stimuli that received high agreement ratings to discuss the results of emotion recognition in the perceptual study. In this chapter, we present the results of emotion recognition in the perceptual study while considering the whole set of stimuli. We compare them with the results of automatic classification of emotions based on motion capture data analysis (introduced in section 9.3).

Firstly, we compare the recognition rates and confusion matrices of emotions recognition across all the actions (See section 9.4.1). Then we compare the recognition rates per emotion and per action. Finally, we compare the recognition rate of each emotion for each action (see section 9.4.2).

9.4.1 Recognition across all the actions

In section 9.3, we discussed the results of RF classification of emotions for each action. In this subsection, we explore the results of RF classification of emotions across all the actions and we compare them with the results of human recognition of emotions.

A RF model is built based on the set of 114 motion capture based features and the data of all the actions. The matrix used for this purpose is composed of 8206 rows (corresponding to 8206 motion capture files) and 114 features. The number of trees is fixed to 500 as no more significant changes of the error are observed with more trees (See section 9.2). RF model is built 50 times to smooth the results. Similarly to section 9.3, the classification rates refer to CCR_{OOB}.

A MLR model is also built based on the perceptual body cues rating performed in "Body Cues Rating" task and the data of all the actions. The matrix used for this purpose is composed of 16128 rows (corresponding to 672 stimuli each annotated 24 times, that is by 24 observers) and 8 features (the 8 body cues used for "Body Cues Rating" task, See Chapter 8). A three-fold cross validation is used to split the data into training and test dataset. The model is built for each fold and the average of Correct Classification Rates and confusion matrix is estimated. This process is repeated 50 times to smooth the results.

Figure 9.2 represents the Correct Classification Rates (CCR) and confusion matrices of human recognition of emotions ("Emotion Perception" task), MLR classification of emotions characterized in "Body Cues Rating" task and RF classification of emotions using motion capture based features.

CCR: RF Vs "Emotion Perception" :

While human recognition of emotions achieves 41%, the automatic classification of emotions performed by RF reaches 75% (See Figure 9.2). Thus, RF outperforms the human recognition of emotions in "Emotion Perception" task. We also note that the relatively low score of correct emotion perception (41% vs 75%) is strongly affected by different factors. We discuss these factors in the following paragraphs.

Firstly, the higher performances seen in automatic classification regarding the human recognition can be due to the fact that people have been "trained" with a larger set of emotions and actions, while the RF model has been trained with a reduced set of emotion labels. Generally speaking, providing a larger set of categories in the training dataset may engender more confusion in the process of classification. Thus, the process of human recognition of emotions can be considered as much more complex than the process of automatic classification task.

Secondly, we note that the relatively low score of human recognition of emotions (41%) is strongly affected by the confusions occurred at the level of some particular emotions (e.g. Sadness and Shame, Pride and Neutral) as discussed in Chapter 8. Such confusions did not occur with the same level with the automatic classification of emotions. As explained in Chapter 8, the recognition of emotions was based on the perception of a puppet animated with the motion capture data. No context was provided to the participants, which makes the recognition of emotions a difficult task. It has been reported in a recent survey on emotion recognition from bodily expression that contextual factors may have a considerable effect on the expression as well as the perception of emotional states [Kleinsmith and Bianchi-Berthouze, 2013].

Thirdly, the automatic classification of emotions achieved with RF model was affected by the identity of actors, which may foster obtaining high classification rates (See section 9.3.3). Indeed, the classification of emotions within RF classifier can be described as the classification of *unknown* samples of *known* actor. As such, the framework of RF classification of emotions includes contextual factors related to the identity of actors. We have showed in section 9.3.3 that preventing RF classifier from such contextual factors (i.e. classifying emotions of an *unknown* actor) leads

	Human recognition of emotions across all the actions							
	CCR = 41%							
	Ax	Pr	Jy	Sd	PF	Sh	Ag	Nt
Ax	**37%**	2%	6%	16%	0%	2%	5%	**31%**
Pr	2%	**20%**	14%	4%	0%	0%	1%	**58%**
Jy	10%	8%	**37%**	8%	2%	1%	5%	**28%**
Sd	2%	0%	0%	**88%**	0%	0%	0%	10%
PF	**64%**	1%	1%	7%	**8%**	1%	4%	13%
Sh	11%	0%	0%	**63%**	0%	**10%**	0%	17%
Ag	12%	7%	7%	5%	0%	0%	**52%**	17%
Nt	0%	4%	2%	16%	0%	0%	1%	**77%**

	Multinomial Logistic Regression (MLR) classification of emotions							
	CCR = 29%							
	Ax	Pr	Jy	Sd	PF	Sh	Ag	Nt
Ax	**10%**	11%	13%	13%	**18%**	12%	14%	9%
Pr	5%	**27%**	**18%**	8%	5%	6%	11%	**19%**
Jy	8%	**17%**	**27%**	5%	13%	2%	**21%**	7%
Sd	5%	7%	3%	**47%**	3%	**19%**	3%	13%
PF	10%	9%	13%	7%	**30%**	5%	**21%**	4%
Sh	8%	7%	4%	**37%**	8%	**20%**	3%	13%
Ag	6%	10%	**18%**	6%	15%	4%	**35%**	6%
Nt	9%	**20%**	7%	13%	6%	8%	6%	**31%**

	Random Forest (RF) classification of emotions across all the actions							
	CCR = 75%							
	Ax	Pr	Jy	Sd	PF	Sh	Ag	Nt
Ax	**64%**	3%	6%	5%	6%	7%	**8%**	2%
Pr	1%	**82%**	**9%**	1%	1%	2%	2%	2%
Jy	3%	**13%**	**67%**	1%	5%	2%	7%	1%
Sd	2%	2%	2%	**83%**	0%	**8%**	1%	2%
PF	4%	3%	4%	2%	**74%**	6%	6%	2%
Sh	4%	2%	2%	**10%**	1%	**78%**	0%	1%
Ag	3%	4%	**7%**	1%	4%	0%	**81%**	1%
Nt	1%	**13%**	2%	4%	0%	4%	1%	**76%**

Figure 9.2: Correct Classification Rate (CCR_{OOB}) and confusion matrices of human recognition of emotions (‘Emotion Perception” task), MLR classification of emotions characterized in “Body Cues Rating” task and RF classification of emotions .

to a classification rate comparable to that of human perception: 44.68% of Inter-individual classification across all the actions vs 41% in human perception. A similar result was observed in [Karg et al., 2010] where the Inter-Individual recognition of walking expressions achieved a similar score to human perception (69% vs 63%) but Person-Dependent recognition score outperforms both of them and reached 95% accuracy.

CCR: RF Vs “Body Cues Rating”:

We also observe from Figure 9.2 that RF outperforms the classification of emotions characterized in “Body Cues Rating” (75% vs 29%). This result highlights the importance of kinematic analysis over the qualitative analysis. The classification of emotions using Multi-Level features is much more successful and reliable than using perceptual measures of few body cues rating.

Indeed, the set of body cues used in the perceptual study consists of 8 body cues (Straightness, Leaning and Openness of body posture, Regularity, Quantity, Speed, Fluidity and Power of body movement). The set of motion capture based cues consists of 114 features. This set of 114 features include further descriptions of expressive body movement that are not considered in the perceptual study.

Besides, the set of body cues in the perceptual study describes the whole body movement (e.g. straightness of the whole body), whereas the set of body cues describes in details the posture and the movement of individual body segments.

Moreover, the perceptual rating of body cues is based on a 5-point scale measurement. As such, body cues refer to discrete measurement ranging from 1 to 5. Such a measurement does allows to capture significant changes in the expressiveness of motion as motion capture features do.

CCR: “Emotion Perception” Vs “Body Cues Rating”:

We observe from Figure 9.2 that the recognition of emotions based on the perceptual rating of body cues (“Body Cues Rating” task) achieves lower score (29%) than in “Emotion Perception” task (41%). As such, the classification of emotions from their perceptual characterization is more prone to the error than the recognition of emotions from the multilabeling rating of emotion perception.

Confusion Matrix: RF Vs “Emotion Perception” Vs “Body Cues Rating”:

Figure 9.2 also shows the confusion matrices of emotions classification achieved by RF and the confusion matrices of emotion recognition in “Emotion Perception” task and “Body Cues Rating” task.

For the sake of simplification, we distinguish different kinds of confusions in emotion recognition and classification; “Weak” vs “Strong” confusion and “Bidirectional” vs “Unidirectional” confusion.

– Strong confusion means that $Emotion_1$ is more recognized/classified as $Emotion_2$ than as $Emotion_1$ (e.g. 58% of Pride are misclassified as Neutral, only 20% of

Pride are correctly classified).

- Weak confusion means that $Emotion_1$ is correctly recognized/classified as $Emotion_1$ but it is confused to some extent with $Emotion_2$ (e.g. 37% of Anxiety are correctly classified, but 31% of Anxiety are misclassified as Neutral).
- Bidirectional confusion means that both $Emotion_1$ and $Emotion_2$ are (weakly or strongly) confused with each other (e.g. 27% of Pride are correctly classified but 18% are misclassified as Joy. Similarly, 27% of Joy are correctly classified, but 17% are misclassified as Pride).
- Unidirectional means that one $Emotion_1$ is (weakly or strongly) confused with $Emotion_2$ but $Emotion_2$ is not confused with $Emotion_1$ (e.g. 31% of Anxiety are misclassified as Neutral but 0% of Neutral are misclassified as Anxiety).

Sadness and Shame confusion:

Sadness and Shame recognition shows a strong unidirectional confusion in "Emotion Perception" task; only Shame is mostly perceived as Sadness (64% of Shame are perceived as Sadness but 0% of Sadness are perceived as Shame, See Figure 9.2). In RF and MLR classifications, a bidirectional confusion occurs to some extent in both Sadness and Shame classification (See Figure 9.2). This bidirectional confusion can be explained as Sadness and Shame expressions could share common characteristics that leads to their bidirectional confusion at the classification level (perceptual and motion capture classifications).

The strong unidirectional confusion of Shame perception in "Emotion Perception" task may be due to the lack of contextual factors. In fact, unlike Sadness, Shame is mostly studied in the context of social behaviors [Gilbert and Andrews, 1998]. It has been argued in [Gilbert and Andrews, 1998] that "Shame is not only related to internal experiences but also conveys socially shared information about one's status and standing in the community". However, the recording of emotional behaviors in the Emilya database is based on individual settings. Besides, Shame expression is mostly associated with facial expressions that are characteristics of submissiveness such as eye-gaze avoidance and turning away [Gilbert and Andrews, 1998]. As we do not display facial expressions on the virtual character during the perceptual study, gaze-avoidance behavior is not visible. Including gaze direction and facial expressions may be critical for the discrimination between Sadness and Shame expression.

Pride and Neutral confusion:

A strong unidirectional confusion occurs in Pride perception; Pride is mostly perceived as Neutral in "Emotion Perception" task (57% of Pride are perceived as Neutral but 4% of Neutral are perceived as Pride, See Figure 9.2). In RF classification, the opposite unidirectional confusion is found; Neutral is mostly confused with Pride but Pride is rarely confused with Neutral. However, Pride and Neutral recognition show a bidirectional confusion based on their perceptual characterization (MLR classification). As only unidirectional confusion is found in RF classification but bidirectional confusion is found in perceptual characterization, this result suggests that the discrimination between Pride and Neutral is more successful based on the analysis of motion capture features than on perceptual characterization.

Pride and Joy confusion:

We observe a weak bidirectional confusion between Pride and Joy in "Emotion Perception" task, perceptual and kinematic classification. We also note that, during "Emotion Perception" task, the participants rate the perception of Pride similarly for Pride and Joy expressions (See Chapter 8). As such, it seems that Pride and Joy share common characteristics at the perceptual and kinematic levels.

Anxiety, Neutral and Panic Fear confusions:

Anxiety and Neutral show a weak unidirectional confusion in "Emotion Perception" task; Anxiety is mainly confused with Neutral in "Emotion Perception" task (37% of Anxiety are correctly perceived, 31% are misclassified as Neutral but 0% of Neutral are misclassified as Anxiety, See Figure 9.2). However, the same confusion did not occur neither in RF classification nor in MLR classification. As such, Anxiety and Neutral confusion is not due to the similarity in their characterization neither in perceptual rating nor in automatic analysis. It can be due to the lack of contextual factors during the task of "Emotion Perception".

In RF classification, a weak unidirectional confusion is observed in Anxiety classification; 64% of Anxiety are correctly classified, 8% are misclassified as Anger, and 7% are misclassified as Shame. Based on its perceptual characterization (in MLR classification), Anxiety seems to be more confused with Panic Fear, which may be explained as Panic Fear is mainly perceived as Anxiety in "Emotion Perception" task.

Panic Fear is weakly confused with Shame and Anger in RF classification (74% of Panic Fear are correctly classified, 6% are misclassified as Shame and 6% are misclassified as Anger, Figure 9.2). Based on its perceptual characterization (in MLR classification), Panic Fear is weakly confused with Anger.

The confusion between Panic Fear and Shame is not bidirectional neither in RF classification nor in MLR classification; Shame is less frequently misclassified as Panic Fear than Panic Fear is misclassified as Shame.

The confusion between Panic Fear and Anger is also not bidirectional in both RF and MLR classifications. Indeed, Anger is weakly confused with Joy in both RF and MLR classifications.

Anger and Joy are not significantly confused in Emotion Perception task. Similarly Anger and Panic Fear are not significantly confused in Emotion Perception task. This result suggests that humans are able to distinguish Anger and Joy expressions, as well as Anger and Panic Fear expressions, but there is a need of more body cues to avoid their confusions in automatic classification.

Summary:

The confusion results presented above indicate the presence of different types of confusions; strong, weak, unidirectional and bidirectional. While some confusions are present in all recognition tasks, others are more related to the human perception, perceptual characterization or motion capture based classification. Strong confusions occurred only at the perceptual level ("Emotion Perception" and "Body Cues Rating" tasks). For instance, the unidirectional confusion of Anxiety with Neutral occurs only at Emotion Perception task. We explain the confusions that occur at the perception task as due to the lack of contextual factors. We also explain the confusions that occur at the classification task as due to the lack of further body features as discussed in Chapter 8.

9.4.2 Recognition per action (across all emotions) and per emotion (across all actions)

In the previous subsection (9.4.1), we focused on the comparison of human and automatic recognition of emotions across all the actions. We compared the classification rate and the confusion matrices. In the current subsection, we compare the classification rates of human and automatic recognition of emotions per action and per emotion.

<u>Per Action :</u>

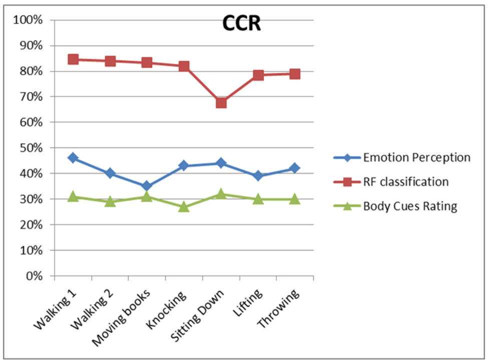
Firstly, we compare the classification rates for each action (See Figure 9.3, a)). The compared classification rates correspond to 1) the CCR of RF classification based on kinematic analysis for each action (red curve), 2) the CCR of MLR classification based on perceptual ratings of "Body Cues Rating" task (green curve) and 3) the average of emotion recognition in "Emotion Perception" task across all emotions (blue curve).

For each action, we find that the RF classification rates outperform the recognition rates in "Emotion Perception" and "Body Cues Rating" tasks (See Figure 9.3, a)). Besides, we observe that the classification rates are mostly similar across different actions. For instance, the classification rates achieved by RF are mostly around 80%, but they decrease to 68% for the classification of emotions in Sitting Down action.

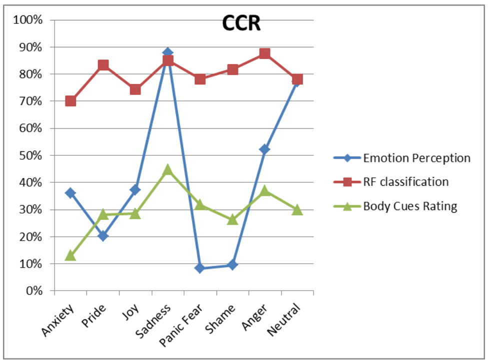
The above result means that the comparison of human and automatic classification rates for each action are congruent with their comparison across all the action (as discussed in the previous subsection: 9.4.1). No exception is found for all the actions. Thus, for all the actions, RF classification based on motion capture features outperforms the recognition of emotions in human perception and perceptual characterization(RF classification > Human Perception > Body Cues Rating as shown in Figure 9.3).

<u>Per Emotion :</u>

Secondly, we compare the classification rates of each emotion averaged across all the actions (See Figure 9.3, b)). The compared classification rates correspond to 1) the average of recall measures of emotions obtained from RF classification across all the actions (red curve), 2) the average of recall measures of emotions obtained from MLR classification of perceptual ratings ("Body Cues Rating" task), and 3) the average of emotion recognition in "Emotion Perception" task across all the actions (blue curve).



(a) Human recognition Vs RF classification per action



(b) Human recognition Vs RF classification per emotion

Figure 9.3: Human recognition vs RF classification of emotions

When averaged across all the actions, Sadness and Neutral are the best recognized in “Emotion Perception” task (See Figure 9.3, b)). Sadness is also the best classified in “Body Cues Rating”, but with lower score comparing to “Emotion Perception” task and RF classification. Unlike what have been observed for the recognition of all emotions for each action (RF classification > Human Perception > Body Cues Rating as shown in Figure), we observe few exceptions for some emotions. Indeed, the recognition of Sadness and Neutral in “Emotion Perception” task achieves their automatic classification. Besides, the classification of Pride, Panic Fear and Shame based on “Body Cues Rating” outperforms their recognition in ‘Emotion Perception” task. This result is mainly due to the strong confusions in the perception of these emotions (Pride is mostly perceived as Neutral, Panic Fear is mostly perceived as Anxiety and Shame is mostly perceived as Sadness, See Figure 9.3).

Anger is the best classified across all the actions in RF classification. It is also considered among the first most recognized emotions in “Emotion Perception” task: Sadness, followed by Neutral and Anger. However, Anger is still recognized with significantly lower score in “Emotion Perception” task than in RF classification. Indeed, the recognition of Anger in “Emotion Perception” task dependents on the performed action. We will illustrate with more details in the next subsection how the performed action affect the recognition of emotions.

9.4.3 Recognition of each emotion for each action

In order to get deeper insights into the comparison of emotion recognition in “Emotion Perception” task, “Body Cues Rating” and RF classification, we plot in Figure 9.5 the classification rates related to each emotion. The compared classification rates correspond to 1) the average of recall measures of emotions obtained from RF classification across all the actions (red curve), 2) the average of recall measures of emotions obtained from MLR classification of perceptual ratings ("Body Cues Rating" task), and 3) the average of emotion recognition in "Emotion Perception" task across all the actions (blue curve).

RF Vs "Body Cues Rating":
As it was observed for the classification rates of the model built across all the actions (See section 9.4.1) and for each action (See section 9.4.2), RF classification also outperforms MLR classification for each emotion (See Figure 9.5). Hence, motion capture based features outperforms perceptual body cues in emotion classification.

"Emotion Perception" Vs "Body Cues Rating":
The classification of Anxiety, Sadness, Anger and Neutral in "Emotion Perception" task totally outperforms their classification in "Body Cues Rating" task (See Figure 9.5).

Shame and Panic Fear are mostly perceived respectively as Sadness and Anxiety in "Emotion Perception" task. However, the classifications of Panic Fear and Shame are better based on their characterization in "Body Cues Rating" task (See Figure 9.5: Body Cues Rating > Emotion Perception). Consequently, the classification of Panic Fear and Shame based on Body Cues perceptual rating turns out to be better than their recognition in "Emotion Perception" task.

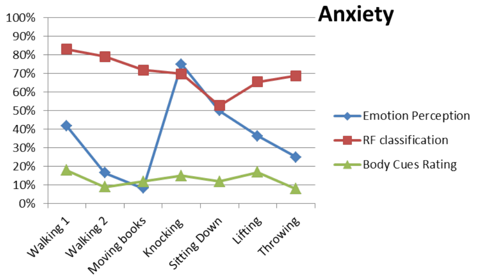
The same result is also found for Pride and Joy recognition except in few actions: Walking in Pride and walking, Moving Books and Throwing for Joy (See Figure 9.5: Body Cues Rating > Emotion Perception in Pride and Joy expressions except in these actions).

RF Vs "Emotion Perception":
For some actions, the classification of Anxiety, Sadness, Anger and Neutral achieve better scores in "Emotion Perception" task than in RF classification (See Figure 9.5).

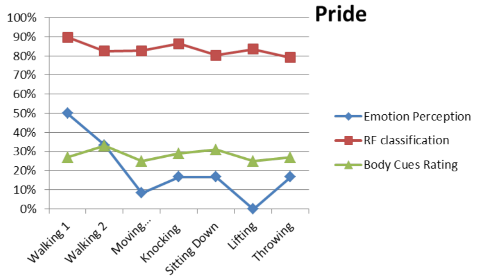
The observers correctly perceive Anxiety in Knocking and Sitting Down actions with similar scores as with RF model (75% and 70% in Knocking , 50% and 53% in Sitting Down). While Anxiety is the best perceived in Knocking action, it is the best classified in Walking action using motion capture data.

Sadness expression in Throwing and Sitting Down actions was correctly perceived in all the stimuli used in "Emotion Perception" task (Recognition rate = 100%). Using RF model, Sadness expression in these actions was classified with high but lower scores than human perception (76% and 80% respectively in Sitting Down and Throwing). While Sadness is the best perceived in Sitting Down and Throwing actions, it is the best classified in Walking action.

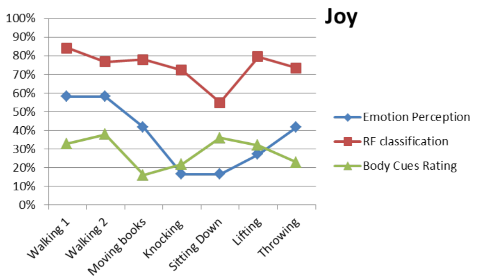
In Throwing action, the recognition of Anger in "Emotion Perception" task is also slightly better than its RF classification (100% vs 97%). It seems that the observers are more expert than RF model to recognize Sadness and Anger in Throwing action and Sadness in Sitting Down action. Besides, Anger is the best perceived and



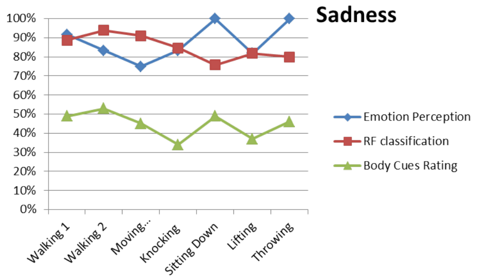
(a) Human recognition Vs RF classification of Anxiety per action



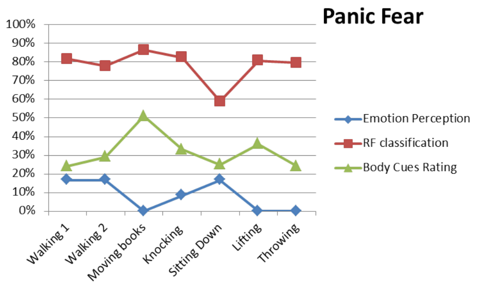
(b) Human recognition Vs RF classification of Pride per action



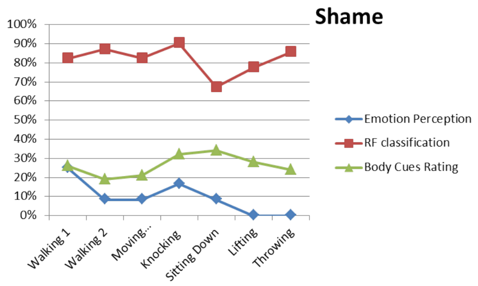
(c) Human recognition Vs RF classification of Joy per action



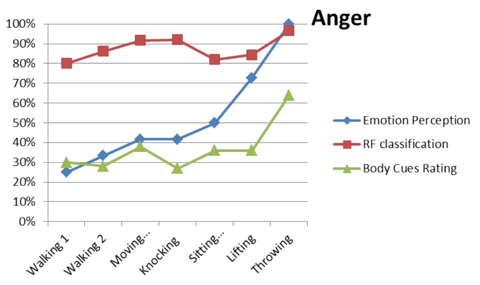
(d) Human recognition Vs RF classification of Sadness per action



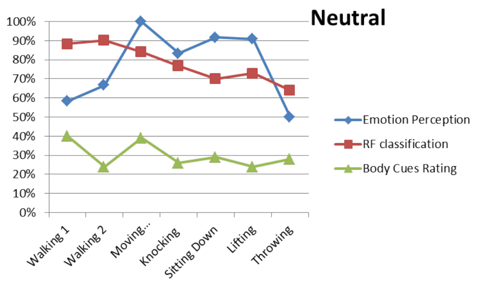
(e) Human recognition Vs RF classification of Panic Fear per action



(f) Human recognition Vs RF classification of Shame per action



(g) Human recognition Vs RF classification of Anger per action



(h) Human recognition Vs RF classification of Neutral per action

Figure 9.5: Human recognition vs RF classification of each emotion across all actions. Walking 1 and Walking 2 stands respectively for Simple Walking and Walking with an object in hand.

9

classified in Throwing action. It is also the best classified in Throwing action based on Body Cues Rating. As such, Anger expression seems strongly associated to Throwing action. This association is not restricted to the action tendencies and perceptual level (e.g. we tend to throw an object when expressing Anger). Indeed, based on its high classification rate in Throwing, we conclude that Anger expression in Throwing action is characterized highly differently from the other expressions. In chapter 8, we showed that the rating of movement Power was significantly higher in Anger expression. In chapter 11 we will illustrate the characterization of Anger expression based on motion capture features across all the actions, and in each action (Throwing in particular).

Finally, Neutral was also better recognized in "Emotion Perception" task than in automatic classification using RF model in most of the actions. However, we note that Neutral received a low Precision measure in "Emotion Perception" task (See Chapter 8). Indeed, the observers tend to perceive Neutral in a wide range of stimuli, including the ones that infer Neutral expression and other emotions (such as Pride and Anxiety).

9.4.4 Summary

In this section, we compared automatic recognition of emotions based on motion capture data and human recognition of emotions. During the perceptual task, the participants were asked to accomplish two tasks: "Emotion Perception" and "Body Cues Rating" tasks (See chapter 8).

We compared the recognition rates 1) across all the actions in subsection 9.4.1, 2) per action and per emotion in subsection 9.4.2, and finally 3) for each emotion in each action in subsection 9.4.3.

In each subsection, 3 measures of recognition rates were used to accomplish this comparison:

- Recognition rates in "Emotion Perception" task
- Recognition rates obtained from an MLR classification of emotion based on "Body Cues Rating" data
- Classification rates of RF classification of emotions based on motion capture features.

The results can be summarized as follows:

CCR: RF Vs "Emotion Perception" :

We found that RF classification based on motion capture data mostly outperforms the human recognition of emotions achieved in "Emotion Perception" task. Low recognition rate in "Emotion Perception" task were mostly due to the presence of strong confusions in the perceptual recognition; Pride was perceived as Neutral, Shame as Sadness and Panic Fear as Anxiety.

In few cases, emotion recognition in "Emotion Perception" outperforms RF classification based on motion capture data (e.g. Sadness recognition in Sitting Down action). Overall, in both human recognition and automatic classification, Sadness, Anger and Neutral receive the best scores.

We also found that the human recognition and the automatic classification of emotions depend somehow on the action being performed. Similar results were shown for both perceptual and automatic recognition of Anger, Pride and Joy. Indeed, Pride and Joy are the best perceived and classified in Walking action. Anger is the best perceived and classified in Throwing action. As such, positive emotions (Pride and Joy) are the best conveyed through walking in both human recognition and automatic classification. Anger is also the best conveyed in Throwing action in both human recognition and automatic classification.

However, different observations were found for Anxiety, Sadness and Neutral. For instance, Anxiety is the best perceived in Knocking action, but the best classified in Simple Walking using motion capture data. Sadness is the best perceived in Throwing and Sitting Down actions, but the best classified in walking actions using motion capture data.

CCR: RF Vs "Body Cues Rating" :

We found that RF classification based on motion capture data always outperforms MLR classification based on Body Cues Ratings. This result suggests that the set of motion capture features considered in our study allows the elimination of confusions that occur when classifying the expressed emotions using Body Cues Ratings (i.e. perceptual measures of 8 body cues rated on a 5 point-scale). On one hand, such perceptual ratings describe the motion of the whole body and do not consider the subtle changes in body posture and movement. For instance, we did not ask the participants to rate separately the speed of each body segment. The speed of movement was rather rated according to the whole body movement. Whereas motion capture features describe the posture and the movement of several body segments separately. On the other hand, Body Cues Rating task results in discrete measures ranging from 1 to 5 (1,2,3,4,5). Whereas motion capture features refer to rotational and positional values with a larger range of values. Such motion capture measures are more accurate for the classification task than 5-point scale perceptual measures.

CCR: "Emotion Perception" Vs "Body Cues Rating" :

We found that the recognition of emotion based on "Emotion Perception" is mostly better than the recognition of emotions based on "Body Cues Rating". the recognition of emotions based on "Body Cues Rating" outperforms that of "Emotion Perception" only for few emotions (Pride, Panic Fear and Shame) that are strongly confused with others in "Emotion Perception".

Confusions in emotion recognition :

The confusions occurred mainly between Sadness & Shame, Pride & Neutral, Pride & joy, Anxiety & Neutral, Anxiety & Panic Fear, Anger & Joy, .

The confusions of emotion recognition observed in "Emotion Perception" task concern Sadness & Shame, Pride & Neutral, Pride & joy, Anxiety & Neutral. Some of these confusions were also observed in automatic classification based on motion capture features or body cues rating (e.g. Sadness & Shame, Pride & joy). Such confusions can be explained as the similarity of emotion expression characterization. Other confusions were not observed in automatic classification (e.g. Anxiety & Neutral, Anxiety & Panic Fear). Such perceptual confusions may be due to the lack of contextual factors during the perceptual task. For instance, the participants were not provided with contextual factors regarding the identity of actors, the scenario used to elicit emotions. Such factors can have a considerable impact on the recognition of emotions [Kleinsmith and Bianchi-Berthouze, 2013].

We also note that some confusions occurred only in automatic classification of emotions (e.g. Anger & Joy). Thus, humans are more expert to distinguish these confused emotions and more body features are required to avoid their confusions.

While some of the confusions that occurred in emotion perception and automatic classification were unidirectional, others were bidirectional. For instance in "Emotion Perception" task, the observers perceived Neutral in Pride expression, but they did not perceive Pride in Neutral expression. An opposite result occurs in automatic classification using motion capture data (See Figure 9.2); in RF classification, Neutral was slightly confused with Pride but Pride was rarely confused with Neutral.

9.5 Comparison of the contribution of different description levels to the classification of expressed emotions

In section 9.3 and 9.4, we explored respectively the automatic classification of emotions using motion capture data and the comparison between automatic and human recognition of emotions. In this section, we give deeper insights into the automatic classification of emotions using motion capture data. Indeed, we explore the contribution of different types of features (e.g. Upper body features versus Lower body features).

We can note that all expressed emotions are best classified over few actions. These actions differ for all the emotions. For the sake of simplification, we compare the contribution of different description levels to the classification of each expressed emotion only in the "*associated*" action (we call an "*associated*" action for a particular expressed emotion the action for which this emotion was the best classified according to the results reported in Table 9.1). For instance, Throwing is the "*associated*" action of Anger.

In the following subsections, we compare the contribution of different types of body cues to the classification of each expressed emotion in "*associated*" action. The description levels are described in Chapter 6. The number of trees used in each RF model (for each description level and "*associated*" action) is chosen as explained in section 9.2.

9.5.1 Anatomical description levels

In this section, we compare the contribution of body cues from different anatomical description levels to the classification of each expressed emotions in the "*associated*" action. Each RF model is built using the data of emotion expression in a particular action and a particular anatomical description level; Global (16 features), Body parts (38 features), Semi Global (30 features), Local (46 features), Lower Body (29 features) and Upper Body (85 features) (Please note that some features can belong to different description levels). These anatomical description levels were explained in Chapter 6. The CCR_{OOB} of each expressed emotion classification are presented in Table 9.2.

The comparison between the contribution of Global features and Body Parts in the classification of each emotion is congruent with what we expected; using features of upper and lower body parts together (Body Parts features) gives better results than using features that describe the whole body (Global features). This result is observed in the classification of each expressed emotion in the "*associated*" action (See Table 9.2).

From Table 9.2, we can also conclude that considering only the coordination between body limbs (Semi Global features) does not lead to a good classification of expressive body movements as Local or Body Parts features do, but when considered all together they provide better results (See Table 9.1).

9

We also find that the classification of Panic Fear, Shame and Anger using Upper body features mostly gives better results than their classification using Lower body features. Knowing that the "*associated*" actions (Moving Books, Knocking and Throwing) mainly involve upper body movements, this result is congruent with what we expected. Interestingly, we find that Pride and Joy expressions in Simple Walking action are also better recognized using Upper Body feature than using Lower Body features. This result accentuates the importance of Upper Body features in positive emotion expression (Pride in particular) in Walking. Sadness expression in Walking with an object in the hand seems to be slightly better classified using Lower Body features (CCR_{OOB} is 88%) rather than Upper Body features (CCR_{OOB} is 87%). The latter finding highlights the need to focus on foot stride length and legs movement in the analysis/synthesis of Sadness expression in Walking motion. Finally, the classification of Anxiety and Neutral expressions leads to very similar results using Lower or Upper Body parts respectively in Walking and Walking with an object in the hand.

9.5.2 Directional description levels

In this section, we compare the contribution of movement direction to the classification of each expressed emotions in the "*associated*" action. Four directional levels are considered for this goal; Lateral, Sagittal, Vertical (Length) and Three dimensional direction (See Chapter 6 for more details about these description levels).

Table 9.2: Contribution of anatomical description levels: Recall Measure (%) of RF for each emotion expressed in "*associated*" action.

(%)	Global	Body Parts	Semi Global	Local	Lower Body	Upper Body
Ax (SW)	52%	**71**%	59%	**72**%	71%	72%
Pr (SW)	79%	**84**%	81%	**84**%	73%	**85**%
Jy (SW)	63%	**76**%	72%	**76**%	73%	**76**%
Sd (WH)	78%	**90**%	77%	**87**%	88%	87%
PF (MB)	59%	**79**%	74%	**81**%	73%	**85**%
Sh (KD)	70%	**86**%	74%	**88**%	71%	**88**%
Ag (Th)	77%	**92**%	91%	**94**%	72%	**96**%
Nt (WH)	78%	**87**%	79%	**85**%	85%	84%

The number of features considered in each directional description level ranges from 22 (Lateral description level) to 34 (Three dimensional description level).

Table 9.3 depicts the results of RF classification of each expressed emotion in the "*associated*" action using a different directional movement description level for each RF model.

We can observe that using Lateral direction features to classify Anxiety, Pride, Joy and Neutral mostly leads to the worst results compared to Sagittal, Vertical(Length) and ThreeD directions. This result suggests that the features that describe lateral movement are not enough to capture a high amount of expressiveness for Anxiety, Pride, Neutral and Joy expressions in walking action.

Anger expression classification in Throwing action and Sadness expression classification in Walking action receive similar CCR_{OOB} using Lateral and Sagittal direction features (the CCR_{OOB} are slightly better using Lateral direction features). Panic Fear expression in Moving Books action receives higher CCR_{OOB} using Lateral direction features (68%) than using Sagittal direction features (63%). This result suggests that Lateral and Sagittal directions infer similar amount of expressiveness of Sadness (in Walking), Panic Fear (in Moving Books), Shame (in Knocking) and Anger (in Throwing).

The use of Vertical(Length) direction features for Anxiety and Pride classification in Walking leads to the best classification rate. Vertical(Length) direction features describe the straightness/ collapse of body posture, but also to upward/ downward movement. This result highlights the important role of body limbs flexion and vertical arms movement in Anxiety and Pride expression in Walking actions. The use of Vertical(Length) direction features also leads to the best CCR_{OOB} of Panic Fear expression in Moving Books action (82%).

It has been widely shown in previous works that body limbs flexion plays an important role in bodily expression of Sadness (in particular collapsed posture) [Wallbott, 1998]. In our work, Sadness expression classification in Walking with an

9

Table 9.3: Contribution of directional description levels: Recall Measure (%) of RF for each emotion expressed in "*associated*" action.

(%)	Lateral	Sagittal	Vertical (Length)	ThreeD
Ax (SW)	47%	**63**%	**69**%	65%
Pr (SW)	72%	**78**%	**86**%	81%
Jy (SW)	61%	**69**%	**73**%	72%
Sd (WH)	78%	78%	82%	**85**%
PF (MB)	**68**%	63%	**82**%	76%
Sh (KD)	**72**%	70%	78%	**81**%
Ag (Th)	84%	85%	94%	**94**%
Nt (WH)	66%	**84**%	80%	**87** %

object in the hand receives slightly better CCR_{OOB} using ThreeD direction features (85%) compared to its classification using Vertical(Length) direction features (82%), but both CCR_{OOB} are high. ThreeD direction features refer to the extension and the coordination of body limbs. As such, the features that describe the body limbs extension and body limbs coordination contribute slightly better than body limbs flexion and vertical arms movement for Sadness expression classification in Walking with an object in the hand. The same result is found for Shame expression classification in Knocking action.

9

Three dimensional direction contributes significantly better than Vertical(Length) direction for Neutral expression classification in walking actions. However, Three dimensional direction and Vertical (Length) direction contribute similarly for Anger and Joy expressions classification respectively in Throwing and Walking actions.

9.5.3 Posture and Movement description levels

Posture/Movement description level categorizes the features according to their description of body posture or movement (See Chapter 6). In this section, we compare the contribution of features that describe 1) Postural Information (36 features), 2) Postural changes (26 features) and 3) the maximum values of speed and acceleration as well as the correlation between arms movement (52 features) in the classification of the expressed emotions in each "associated" action. Each feature belongs to one of the above description levels (36+52+26=114 features).

Overall, the classification of each expressed emotion in the "*associated*" action using features that refer to Postural information provides the best classification rates compared to the CCR_{OOB} obtained from RF models trained with Postural Changes or Movement Dynamics features (See Table 9.4). It is important to note that Movement Dynamics cues used in this work are a discretization form of dynamics cues. While 6 features represent the pearson correlation measure, the other 46 features

Table 9.4: Contribution of posture/movement description levels: Recall Measure (%) of RF for each emotion expressed in "*associated*" action.

(%)	Postural Information	Postural Changes	Max.S.A Corr
Ax (SW)	**79**%	61%	60%
Pr (SW)	**87**%	81%	81%
Jy (SW)	**77**%	71%	72%
Sd (WH)	**92**%	81%	**83**%
PF (MB)	**87**%	**70**%	65%
Sh (KD)	**89**%	**71**%	68%
Ag (Th)	**90**%	84%	**88**%
Nt (WH)	**89**%	**84**%	75%

represent the maximum values of speed and acceleration of movement (See Chapter 6). We refer to the latter as Max.S.A (Maximum values of Speed and Acceleration). They are computed by considering only the maximum value of velocity and acceleration. Such measures reduce a high amount of information regarding the dynamic aspect of movement to two discrete values.

Another interesting finding is that the CCR_{OOB} of Panic Fear and Neutral expressions classification in "*associated*" actions using only Postural features (87% and 89% respectively for Panic Fear and Neutral) is very similar to their classification using the whole set of Multi-Level features (86% and 90% respectively for Panic Fear and Neutral). As such, Panic Fear and Neutral expressions in "*associated*" actions are mainly described through body posture cues.

9

The contribution of Postural Changes cues and Movement Dynamics cues to the classification of Anxiety, Pride and joy in Simple Walking turn out to be very similar (See Table 9.4). The CCR_{OOB} of Sadness and Anger classification in "*associated*" actions based on Movement Dynamics cues (respectively 83% and 88%) are slightly better than the CCR_{OOB} of their classification in "*associated*" actions based on Postural Changes cues (CCR_{OOB} are respectively 81% and 84%). This result highlights the importance of Movement Dynamics properties in Sadness expression in Walking action and Anger expression in Throwing action. As the set of features used to describe Movement dynamics is restricted to the pearson correlation of arms movement and maximum values of speed and acceleration, we conclude that such statistical measures are enough to capture a high amount of Sadness and Anger expressions. Previous psychological studies point out that bodily expression of Sadness is often characterized with slow motion [Gross et al., 2010, Dahl and Friberg, 2007] while bodily expression of Anger involves fast and jerky movement [Gross et al., 2010, Dahl and Friberg, 2007]. Besides, previous studies on kinematic data analysis found that simple statistical measures such as the mean, the minimal and maximum values of speed and acceleration can infer the expression of emotions in body movement [Ka-

Table 9.5: Summary of results for each expressed emotion; "*associated*" action and ranking of description levels contributions. More important description levels>less important description levels; Slightly more important description levels>=slightly less important description levels; = means that description levels contribute similarly.

Emotion	Action	Description levels contributions ranking
Ax	SW	- UpperBody=Local>=LowerBody>SemiGlobal - Post.Inf>Post.Chg>=Max.S.A & Corr - VerticalLength>ThreeD>=Sagittal>Lateral
Pr	SW	- Local>UpperBody>SemiGlobal>LowerBody - Post.Inf>Post.Chg>=Max.S.A & Corr - VerticalLength>ThreeD>Sagittal>Lateral
Jy	SW	- UpperBody>=Local>LowerBody=SemiGlobal - Post.Inf>Max.S.A & Corr>=Post.Chg - VerticalLength=ThreeD>Sagittal>Lateral
Sd	WH	- LowerBody>=UpperBody>=Local>SemiGlobal - Post.Inf>Max.S.A & Corr>Post.Chg - ThreeD>VerticalLength>Sagittal>Lateral
PF	MB	- UpperBody>Local>SemiGlobal>=LowerBody - Post.Inf>Post.Chg>Max.S.A & Corr - VerticalLength>ThreeD>Sagittal>Lateral
Sh	KD	- Local>UpperBody>SemiGlobal>LowerBody - Post.Inf>Post.Chg>=Max.S.A & Corr - ThreeD>VerticalLength>Lateral>=Sagittal
Ag	Th	- UpperBody>Local>SemiGlobal>LowerBody - Post.Inf>Max.S.A & Corr>Post.Chg - ThreeD=VerticalLength>Lateral>=Sagittal
Nt	WH	- Local>UpperBody>=LowerBody>SemiGlobal - Post.Inf>Post.Chg>Max.S.A & Corr - ThreeD>Sagittal>VerticalLength>Lateral

pur et al., 2005] [Bernhardt and Robinson, 2007]. Finally, the classification of the expression of Neutral, Panic Fear and Shame in "*associated*" actions receive better CCR_{OOB} (only slightly better in Shame expression classification results) when using Postural Changes features compared to Movement Dynamics cues.

9.5.4 Summary

Based on our Multi-Level notation system described in Chapter 6, we studied the contributions of different description levels to the classification of emotions in "associated" actions. "Associated" actions are the actions in which the emotions are best classified. We compare the contributions of 1) Global, Body Part, Semi Global, Local, Lower Body and Upper Body levels from Anatomical dimension, 2) Lateral, Sagittal, Vertical (Length) and Three Dimensional levels from Directional dimension, and finally 3) Postural Information, Postural Changes and Movement Dynamics (maximum values of Speed and Acceleration and Correlation features) levels from Posture/ Movement dimension.

We found that features from detailed anatomical level provide better classification than features from Global anatomical level (i.e. features that describe the bounding box surrounding the whole body). Upper body features provide better classification rates than lower body features in arms movement based actions. However, the classification of Pride and Sadness in walking provide similar results using upper or lower body parts.

We also found that posture features provide better classification rates than postural changes and movement dynamics, but movement dynamics features are as important as postural features for Anger expression in Throwing action. The classification of emotions based on features from Vertical (Length) or 3D directions provide better classification rates than the classification based on features from Sagittal or Lateral directions.

Table 9.5 summarizes the results obtained for each expressed emotion; "*associated*" action and ranking of description levels contributions.

9.6 Analysis of the temporal profiles of motion cues

In the previous section, we compare the contribution of different description levels of our body movement notation system (described in details in Chapter 6). The set of Movement Dynamics features, in particular, is composed of 52 features: 6 features defined as the correlation of arms movement based on Pearson correlation measure and 46 features defined as the maximum values of speed and acceleration of movement (speed and acceleration are computed for different body segments as explained in Chapter 6).

In this section, we explore the contribution of further Movement Dynamics features that describe the temporal profile of motion cues. We follow the work of Castellano et al. [Castellano et al., 2008] to measure the temporal profile features.

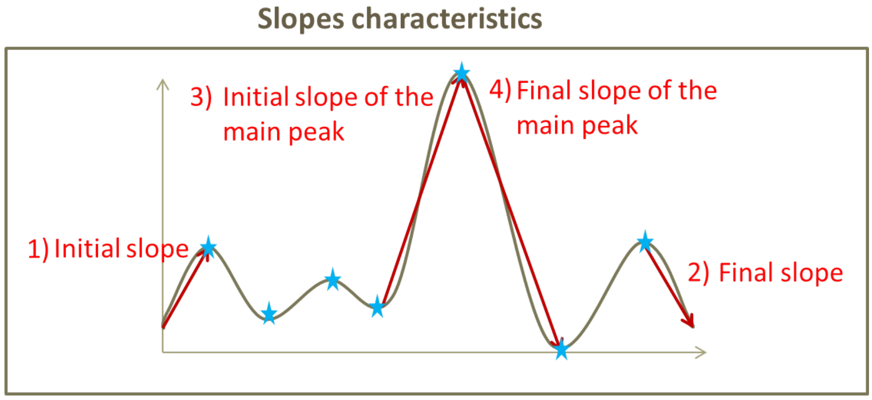


Figure 9.7: Slopes characteristics

Firstly, we describe the work of Castellano et al. [Castellano et al., 2008]. Secondly, we describe our work on the classification of expression emotion based on this set of temporal profile features and we compare the classification rates with those obtained with our 114 motion capture features (defined according to our body movement notation system, See Chapter 6).

9

9.6.1 Temporal Profile (TP) Features

Sixteen features are used to define the Temporal Profiles (TP) of specific motion cues in [Castellano et al., 2008]. Castellano et al. [Castellano et al., 2008] categorized the sixteen features into four categories; 1) Slopes characteristics (4 features), 2) Main peaks characteristics (4 features), 3) Overall characteristics of motion (4 features), and 4) Temporal regularity of motion structure (4 features).

$$Energy(f) = \frac{1}{2} \times m \times v(f)^2 \tag{9.1}$$

The sixteen features are defined as follow, but more details can be found in [Castellano et al., 2008]:

– **Slopes characteristics** (See Figure 9.7):

1. Initial slope: This feature is computed as the slope of the straight line joining the first relative maximum value and its initial value [Castellano et al., 2008]. According to Castellano et al. [Castellano et al., 2008], "this feature refers to a measure of the impulsiveness of the attack of a motion cue's temporal profile".

2. Final slope: This feature is computed as the slope of the straight line joining the last relative maximum value and its final value [Castellano et al., 2008]. According to Castellano et al. [Castellano et al., 2008],

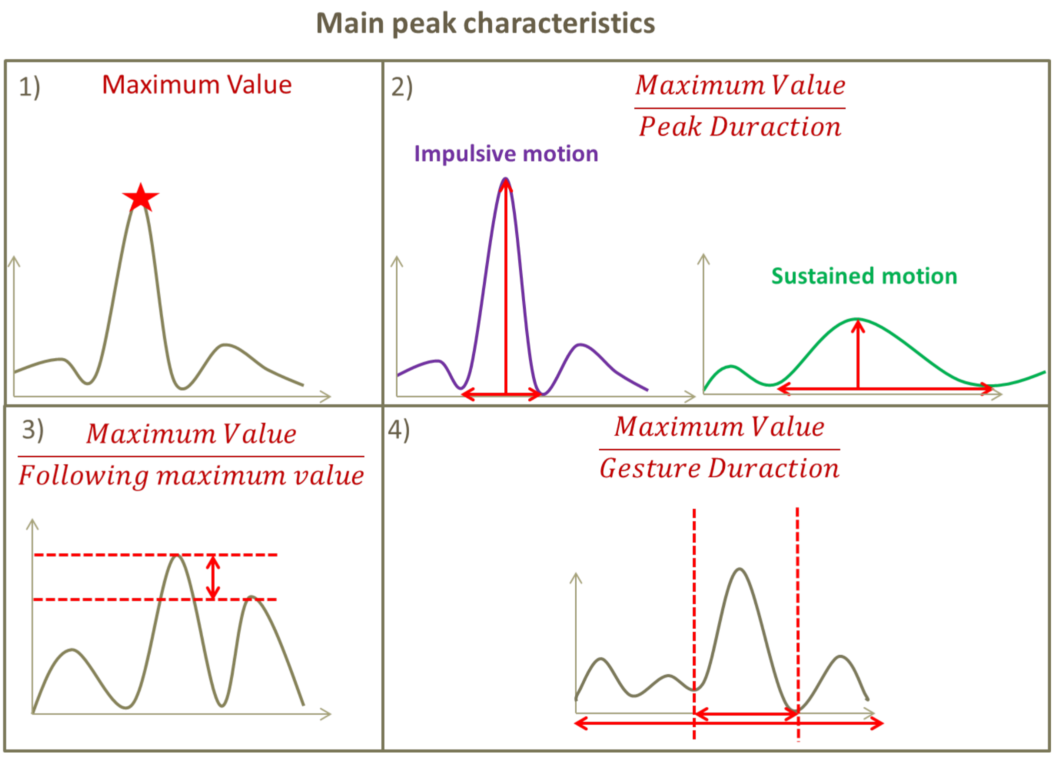


Figure 9.8: Main peak characteristics

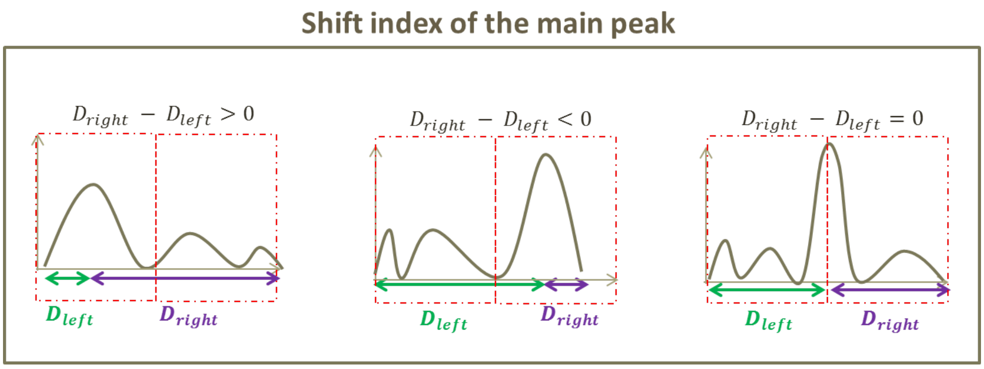


Figure 9.9: Shift index of the main peak

"this feature refers to a measure of the impulsiveness of the release of a motion cue's temporal profile".

3. Initial slope of the main peak This feature is computed as the slope of the straight line joining the absolute maximum value and the relative minimum value preceding it [Castellano et al., 2008]. According to Castellano et al. [Castellano et al., 2008], "this feature refers to a measure of the impulsiveness of the attack of a motion cue's main peak".
4. Final slope of the main peak: This feature is computed as the slope of the straight line joining the absolute maximum value and the minimum value following it [Castellano et al., 2008]. According to Castellano et al. [Castellano et al., 2008], "this feature refers to a measure of the impulsiveness of the release of a motion cue's main peak".

– **Main peak characteristics** (See Figure 9.8):

1. Maximum value:
   This feature is defined as the absolute maximum value.
2. $\frac{MaximumValue}{PeakDuration}$:
   This feature is defined as the ratio between the maximum value of the main peak and its duration (See Figure 9.8, 2)). According to Castellano et al. [Castellano et al., 2008], "this feature refers to the overall impulsiveness of a motion cue. A motion cue's temporal profile is impulsive when it is characterized by short peak duration with a high absolute maximum, while it is sustained when it is characterized by longer peak duration with a low absolute maximum". This is illustrated in Figure 9.8.
3. $\frac{MaximumValue}{Followingmaximumvalue}$:
   This feature is defined as the ratio between the absolute maximum value and the biggest relative maximum value. According to Castellano et al. [Castellano et al., 2008], "this feature explains how much the main peak is relevant with respect to the second one".
4. $\frac{MaximumValue}{GestureDuration}$:
   This feature is defined as the ratio between the duration of the main peak and the duration of the motion (See Figure 9.8, 4)).

– **Overall characteristics of motion**:

1. Mean value:
   This feature is defined as the mean value of a motion cue.
2. $\frac{MeanValue}{MaximumValue}$
   This feature is defined as the ratio between the mean and the maximum value of a motion cue [Castellano et al., 2008].
3. Number of peaks:
   This features is defined as the number of peaks that occur in a motion cue [Castellano et al., 2008].

9

4. Number of peaks preceding the main one

– **Temporal regularity of motion structure**:

1. Centroid of energy (CoE):
As in [Castellano et al., 2008], the centroid of energy is computed according to equation 9.2.

$$CoE = \frac{\sum_{t=1}^{N} t \times y_t}{\sum_{t=1}^{N} y_t} \tag{9.2}$$

2. Distance between absolute maximum and the centroid of energy:
This feature is defined as the temporal distance between the absolute maximum of a motion cue and its centroid of energy [Castellano et al., 2008].

3. Symmetry index:
This feature provides an estimation of the symmetry of the temporal profile of a given motion cue [Castellano et al., 2008] based on equation 9.3.

$$SymmetryIndex = \frac{|\sum_{i=\bar{x}}^{N} y_i - \sum_{i=1}^{\bar{x}} y_i|}{\sum_{i=1}^{N} y_i} \tag{9.3}$$

4. Shift index of the main peak (See Figure 9.9):
This features provides an estimation of main peak position regarding the center of the motion cue [Castellano et al., 2008]. The equation 9.4 and the Figure 9.9 illustrate this feature, where D_{right} denotes the duration on the right with respect to the position of the absolute maximum, D_{left} denotes the duration on the left with respect to the position of the absolute maximum and GD is the overall duration of the motion cue's temporal profile.

$$ShiftIndex = \frac{D_{right} - D_{left}}{GD} \tag{9.4}$$

The set of 16 Temporal Profile features just illustrated is used in the work of Castellano et al. [Castellano et al., 2008] to describe the temporal profile of two motion cues: the quantity of motion of upper body and the velocity of the head movements. As such, each motion cue constitutes a time-series. The 16 features are measured according to each time-series.

9.6.2 Classification of emotions using Temporal Profile features

In our work, we compute the set of 16 Temporal Profile features proposed in [Castellano et al., 2008] to describe the temporal profile of two groups of motion cues; 1) the 3D trajectory of the end-effectors related to hands, head and feet and

2) the energy of hands, head and feet motion (the energy of motion is calculated as illustrated in equation 9.1).

- First group of motion cues: The first group is composed of 5 motion cues which are the 3D trajectory of hands, head and feet end-effectors (leading to 80 features = 16 features * 5 trajectory based motion cues).
- Second group of motion cues: The second group is composed of 5 motion cues which are the energy of hands, head and feet motion (leading to 80 features = 16 features * 5 energy based motion cues).

Our goal is to compare the contribution of TP features in the classification task with respect to the initial set of multi-level features. For this purpose, we build a RF model using 6 combination of features;

1. using the set of 80 TP features based on the 3D trajectories of head, hands and feet (henceforth called TP.Trajectory),
2. using the set of 80 TP features based on the energy of head, hands and feet (henceforth called TP.Energy),
3. using the set of 160 TP features composed of 80 TP trajectory based features + 80 TP energy based features (henceforth called TP.Energy+TP.Trajectory),
4. using the set of 194 features composed of 114 features and 80 TP trajectory based features (henceforth called ML+TP.Trajectory),
5. using the set of 194 features composed of 114 features and 80 TP energy based features (henceforth called ML+TP.Energy),
6. using the set of 274 features composed of 114 features + 80 TP energy based features + 80 TP trajectory based features (henceforth called ML+ TP.Energy+ TP.Trajectory)

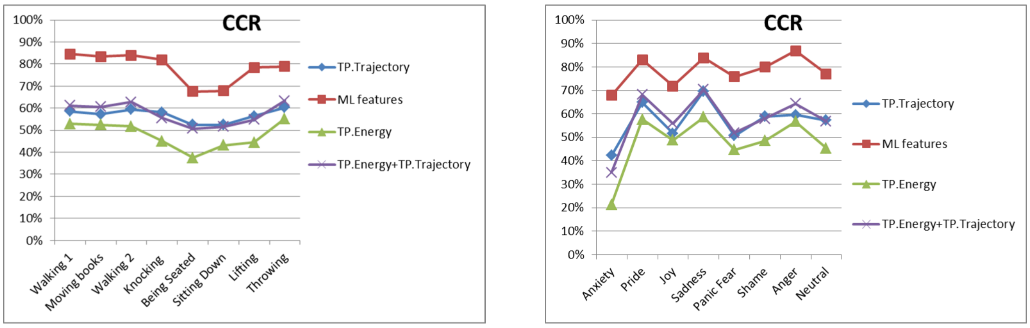
We build one Random Forest model per action and per combination of features.

Contribution of Temporal Profiles features per action and per emotion :

We found that, for each action (across all emotions) and for each emotion (across all actions), the classification rates are always better when using our set of 114 Multi-Level features than when using TP.Trajectory features, TP.Energy features or both (TP.Trajectory features+TP.Energy features) (Figure 9.10).

We also found that, for each action, the classification rates are better when using TP.Trajectory features than when using TP.Energy features (See Figure 9.10). When combined together, they mostly provide slightly better results. This result suggests that the 3D trajectories of end-effectors (head, hands and feet) is more likely to cover the expressive content of the motion than the measure of energy that is based on the velocity of head, hands and feet motion. However, we note that the classification rates of Anger and Joy using TP.Trajectory or TP.Energy are very similar. The combination of TP.Trajectory and TP.Energy also shows a significant improvement of the classification rates of Joy and Anger.

Contribution of Temporal Profiles + Multi-Level features per action :



(a) Contribution of Temporal Profiles features per action

(b) Contribution of Temporal Profiles features per emotion

Figure 9.10: Contribution of Temporal Profiles features

Figure 9.11 shows the classification rates of emotions classification using a combination of features: ML+ TP.Trajectory, ML+TP.Energy, ML+TP.Energy+TP.Position and finally ML alone. We found that the classification of emotions using Multi-Level features alone or combined with other Temporal Profiles features always lead to very similar rates. That means that the initial set of Multi-Level features contains a higher amount of expressive content than the features that describe the temporal profiles related to the end-effectors of the body.

Contribution of Temporal Profiles for each emotion in each action :

In order to give deeper insights into the contribution of Temporal Profiles features, we compare in Figure 9.12 the recognition rates of each emotion in individual actions using our set of 114 Multi-Level features and a subset of Temporal Profiles features: TP.Trajectory, TP.Energy. Similarly to what we found across all the actions and across all the emotions, the classification using our set of 114 Multi-Level features outperforms the classification using TP.Energy or TP.Trajectory features.

We observe few differences regarding the contribution of TP.Energy and TP.Trajectory features. Indeed, the classification of Joy, Panic Fear, and Anger using TP.Energy are somehow overlapped with the classification using TP.Trajectory features across different actions. For instance, the classification of Joy achieves better rates using TP.Energy features in Moving Books and Throwing actions. It also achieves highly similar rates to TP.Trajectory in Walking, Being Seated, Sitting Down and Lifting actions. The classification of Panic Fear shows better results using TP.Energy in Walking and Moving Books action than using TP.Trajectory features. Finally, the classification of Anger using TP.Energy achieves mostly similar or higher scores than the ones obtained from the classification using TP.Trajectory.

As such, the measure of motion energy appears to be as important as the trajectory of the motion in classifying Anger, Joy and Panic Fear expressions in several actions. These emotions are characterized with a high level of arousal. Previous psychological researches highlighted the relevance of energy in the characterization of Anger and Joy [Wallbott, 1998].

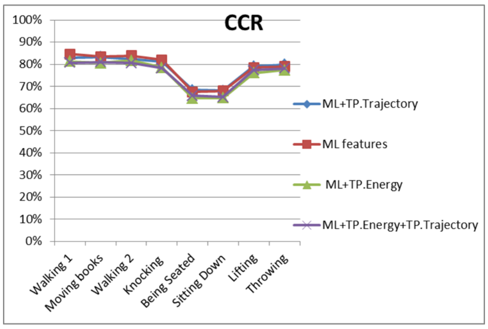


Figure 9.11: Contribution of Temporal Profiles+Multi-Level features per action

9.7 Conclusion

Considering both the variability of emotions and actions in the same database allows comparing the expression of emotions in different body actions. Using Emilya database, we have studied the classification of the expressed emotions in different actions based on Random Forest (RF) approach and using a wide range of low-level body cues. We provided the classification rates of each emotion expressed in each action. We found that Walking action is the most expressive as it receives the highest classification rate. However, the expressed emotions are not always the best classified in Walking action. For instance, Anger was the best classified in Throwing action.

We also found that Sit Down actions (Sitting Down and Being Seated) are the least expressive as it receives the lowest classification rate. We suggests that combining kinematic analysis (based on our set of 114 motion capture features) with kinetic data (e.g. provided from pressure sensors as performed in previous work) may enhance the classification of emotions in seated posture (Being Seated action).

We explored the problem of Inter-individual classification. We found that the classification of emotions decreases when classification is performed for an *unknown* actor. However, the classification of *unknown* actors expressions outperforms on average (across actors) the chance level score. Besides, Inter-individual classification shows a very similar score compared to the human perception of emotions. We conclude that RF classification model is robust enough to deal with the classification of emotions expressed by "unknown" actors. We also conclude that considering a prior knowledge of the individuality allows the increase of the recognition of expressed emotions.

We also compared RF classification using motion capture based body cues with MLR classification based on perceptual measures of "Body Cues Rating" task and human recognition of emotions in " Emotion Perception" task. We compare both the recognition rates and the confusions occurred in emotion recognition/classification. Strong confusions occurred mainly at the perceptual level (e.g. Shame mostly perceived as Sadness).

It appears that RF classification based on motion capture data provides better classification rates than MLR classification using perceptual body cues ratings. We conclude that using our set of 114 motion capture features captures in more details the expressiveness in body movement than our perceptual rating of 8 body cues defined in the perceptual study.

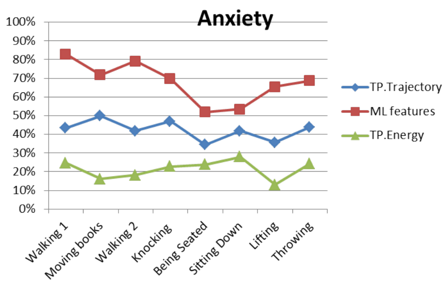
It also appears that RF classification mostly outperforms the percentage of human recognition of emotions. The relatively low recognition rate in emotion perception task is mainly due to the presence of strong confusions (e.g. Shame mainly perceived as Sadness, Pride as Neutral and Panic Fear as Anxiety). We also explain this result as the participants were not provided with contextual factors. Such contextual factors include for instance the individuality of the actors. However, the human recognition of emotions outperforms in few cases the motion capture based classification of emotions. Indeed, the participants recognized with 100% of accuracy Anger expression in Throwing, Sadness expression in Sitting Down and in Throwing and Neutral expression in Moving Books action. Besides, the participants recognized Sadness in Simple Walking, Knocking and Lifting with similar rates as the automatic classification. Across all the actions, Sadness is better perceived by the participants than classified using motion capture data by RF approach. Comparing the effect of action on human perception and automatic classification also shows that Joy and Pride are always the best recognized in Walking action, and that Anger is always the best recognized in Throwing action.

We also compared the contribution of different types of body cues (e.g. Lower body cues vs. Upper body cues) to the automatic classification of expressed emotions. We provide an analysis of the contribution of each description level of body cues to the classification of the expressed emotions. The next step is to identify, for each description level, the most relevant features for the correct classification of each expressed emotion for each action, and across the different actions. This task will be performed in the next chapter.

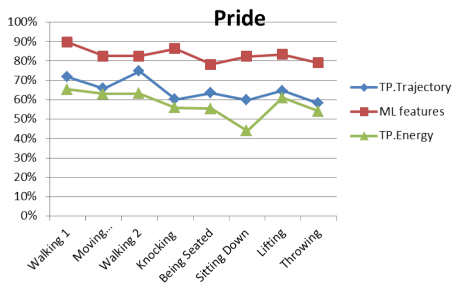
Finally, we explore the temporal profile of end-effectors motion. On one hand, we found that the initial set of Multi-Level features always provides the best result, only slightly better when coupled with Temporal Profile features. We conclude that Temporal Profiles features describing end effectors trajectories and energy do not fully capture the expressiveness of body movement as Multi-Level features do. On the other hand, we found that Temporal Profiles features that describe the energy of end effectors are as important as the Temporal Profiles features that describe the trajectory of end effectors for Anger, Joy and Panic Fear expressions in different actions.

9.8 Summary of chapter

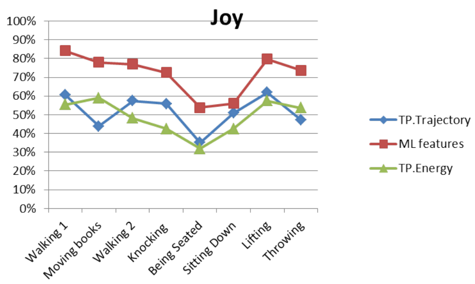
- The contribution of this chapter is multi-fold: 1) we use several low-level features for the classification of emotions expressed in different daily actions, 2) we provide a comparison of the classification of expressed emotions in different



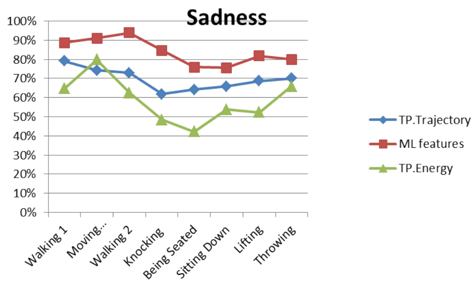
(a) Contribution of Temporal Profile features for Anxiety classification per action



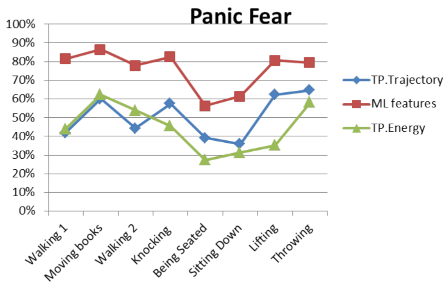
(b) Contribution of Temporal Profile features for Pride classification per action



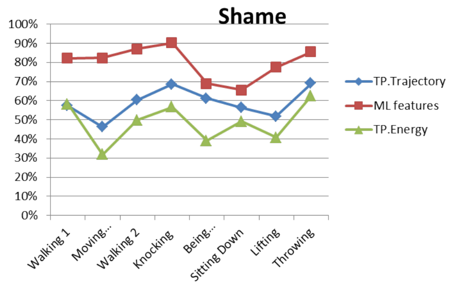
(c) Contribution of Temporal Profile features for Joy classification per action



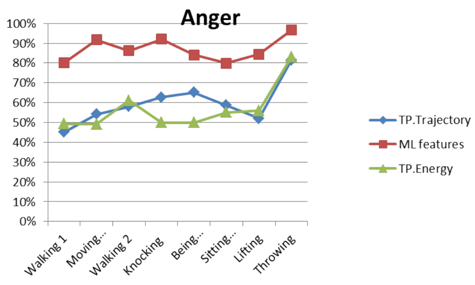
(d) Contribution of Temporal Profile features for Sadness classification per action



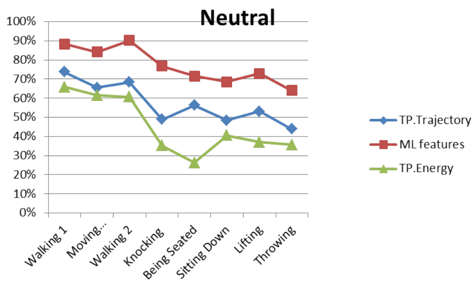
(e) Contribution of Temporal Profile features for Panic Fear classification per action



(f) Contribution of Temporal Profile features for Shame classification per action



(g) Contribution of Temporal Profile features for Anger classification per action



(h) Contribution of Temporal Profile features for Neutral classification per action

Figure 9.12: Contribution of Temporal Profile features for each emotion across all actions. Walking 1 and Walking 2 stands respectively for Simple Walking and Walking with an object in hand.

actions, 3) we compare automatic and human recognition of emotions in daily actions, 4) we illustrate the contribution of different body cues description levels on the classification of each expressed emotion, and 5) we explore the classification of emotions using a new set of features based on the temporal profiles of few body cues (mainly the motion of the head, the hands and the feet).

– For the comparison of Human Vs Automatic classification of emotions, we compare 1) RF classification of emotions using 114 motion capture based features, 2) MLR classification of emotions using 8 perceptual measures rates in “Body Cues Rating” task, 3) human recognition of emotions in “ Emotion Perception” task.
– The confusions in emotion recognition occur mainly between Sadness & Shame, Pride & Neutral, Pride & Joy, Anxiety & Neutral, Anxiety & Panic Fear, and Anger & Joy. Some of these confusions that occurred at the perceptual level are also present in the automatic classification of emotions (e.g. Sadness & Shame). Others occur only at the perceptual (e.g. Anxiety & Neutral) or automatic (Anger & Panic Fear) recognition.
– RF classification of emotions using Multi-Level features always provides better classification rates than MLR classification based on perceptual measures. This result can be explained as the perceptual measures we consider (8 body cues ratings) are not able to cover the expressive content of emotional body expression as motion capture based features do.
– RF classification of emotions using Multi-Level features mostly provides better classification rates than human recognition rates. This result is mainly due to the presence of strong confusions at the perceptual level: Pride is mostly perceived as Neutral, Shame as Sadness and Panic Fear as Anxiety. However, the participants recognized Sadness, Anger and Neutral above classification rate in few actions (Anger in Throwing, Sadness in Simple Walking, Sitting Down and Throwing, and Neutral in most of the actions). Besides, Sadness was better recognized in Emotion Perception task than in automatic classification across all the actions.
– The human perception of emotion as well as the automatic classification of emotions are affected by the performed action. In few cases, the effect of action on the recognition rate in human perception is congruent with its effect on the classification rate achieved by RF approach. For instance, the perception and the classification of Anger achieve the best score in Throwing action. Besides, Pride and Joy are the best perceived and classified in Walking actions.
– Based on our Multi-Level notation system, we studied the contributions of different description levels to the classification of emotions in “associated” actions. The results are summarized in Table 9.5. Overall, we found that posture features provide better classification rates than postural changes and movement dynamics, but movement dynamics features are as important as postural features for Anger expression in Throwing action. Upper body fea-

tures contributes better than Lower body features for Shame, Panic Fear and Anger and classifications in arms movement based actions. However, Upper and Lower body features contributes similarly to Pride and Sadness classification in Walking action. The classification of emotions based on features from Vertical (Length) or 3D directions provide better classification rates than the classification based on features from Sagittal or Lateral directions.

– The classification of emotions using the set of Multi-Level features achieve better results than their classification using features that describe the Temporal Profile of body end-effectors trajectory and energy. Comparing trajectory and energy based features, we found that Anger and Joy expressions are mostly better recognized using energy features than using trajectory features.

10 Expressive body cues selection using Random Forests

In this chapter, we aim to select the most relevant expressive features that characterize emotional body expression in daily actions. Multi-Level expressive features have been defined explicitly according to several description levels (see Chapter 6). In this chapter, our goal is to identify and discuss the most relevant features that contribute the best to the classification of emotions expressed in body movements. These features are measured using 3D motion capture data.

The selection of a reduced set of expressive body cues reported in previous works on emotional body expression characterization mostly relied on perceptual studies [Wallbott, 1998, Meijer, 1989, Dahl and Friberg, 2007]. Only few studies explored the selection of the most relevant body cues through automatic analyses. However, they are mostly based on statistical techniques such as Anova and Tukey-Kramer test [Camurri et al., 2003]. Recent studies explored the selection of the most important expressive body cues using machine learning techniques. Nevertheless, they were always focused on the identification of critical expressive features for the classification of emotions expressed in a single action. In this chapter, we use a machine learning based approach for the identification of critical expressive body cues across different daily actions.

The selection of the most relevant explicit expressive body cues gives better insights on how emotional body expressions are characterized and interpreted. We aim to apply feature selection techniques to explore the most relevant features considered during learning process. In chapter 4 we provide a brief summary of feature selection techniques commonly used by the machine learning community. We concluded that hybrid feature selection approaches combining Embedded feature ranking and Wrapper feature selection seem the most suitable for our purpose. Random Forest (RF) model is used for both embedded and wrapper strategies. RF FS approach is firstly applied to identify relevant features according to each action, then summarized to discuss the relevant features across various actions.

This chapter is organized as follows: Firstly we introduce the Data Selection step in section 10.1 which is useful to remove the outliers before starting feature selection process. Secondly, we present our approach of Feature Selection (FS) in section 10.2. Section 10.3 is devoted to the evaluation of our FS approach. Section 10.4 explains the selection of features across different actions. Section 10.5 shows the quantification of selected features across different actions. In section 10.6, we discuss the ranking of each selected feature according to its relevance measure for emotion expressions classification. Finally, we conclude and summarize the chapter in section 10.7 and 10.8.

10.1 Data selection

The presence of a high amount of noisy data in the dataset can influence the reliability of feature selection process. We hypothesize that “Outliers” observations (expressive movement considered as “outliers”) should not be considered in body cues selection process. We call data selection the process of removing the “Outliers” from the dataset of each daily action.

“Outliers” are defined as expressive movements that are not well recognized in automatic classification and in perceptual experiments by human participants (See Chapter 8). In order to maintain the variety of emotion expressions, an expressive movement is **NOT considered** as an “Outlier” if it is **Correctly** classified at least either in automatic classification or in “Emotion Perception” task.

We distinguish two types of “Outliers”: “inter-Outliers” (emotion expression of actor i is extremely different from emotion expression of other actors) and “intra-Outliers” (emotion expression of actor i in a given scenario is extremely different from his/her emotion expression in other scenarios). To determinate “inter-Outliers”, a cross validation scheme is adopted where the test dataset consists of one actor data and the training dataset consists of the other actors expressive movements. To determinate “intra-Outliers”, the OOB error is measured internally during the training of the model (as such, the test data are selected from the same actor expressive movements).

“Outliers” expressive movements can be determined individually or per subset across actions repetitions and emotion scenarios (e.g. a subset of 12 of expressive movements: 1 actor * 1 emotion * 1 action * 3 scenarios * 4 repetitions). In our perceptual study, participants have been asked to recognized the expressed emotion in only one sample (randomly selected) of each subset of expressive movements (See Chapter 8). Thus, one sample is randomly selected from the subset of 12 of expressive movements. In order to explore the consistency of intra-expression for each actor, we firstly explore the presence of “intra-Outliers” through the approach explained in the next subsection. A RF model is build for each subset of expressive movements and the OOB error is estimated. We found that very few “intra-Outliers” occur. These “intra-Outliers” reflect a low consistency in intra-expressions of the same actor. We discard these “intra-Outliers”. Thus, for the remaining of actors’ expressions, we suppose that the variability of scenarios of emotions and repetitions of actions do not affect the perception of emotion. As such, we suppose that we can remove “inter-Outliers” per subset of samples.

In the next subsection, we illustrate the process of Data Selection used to remove “inter-Outliers”.

10.1.1 Data selection process

Data selection process is obtained through the following steps:

– a) Remove "inter-Outliers": for one particular action and one particular actor, create a cross validation partition for data where the test dataset corresponds to his/her expressive movements and the training dataset corresponds to the expressive movements of all the other actors.

  1. Build a RF model with respect to the current cross validation partition for data
  2. Measure the classification error rate for expressive movements subset related to one particular emotion.
  3. The current subset of expressive movements is considered as "Outliers" if it is not correctly classified neither in automatic classification task (classification error > 0.5) nor in human perception task (the expressed emotion is not recognized by human participants in perceptual task). The threshold of classification error is fixed to 0.5 rather than 1- chance level to keep only the most relevant expressive movements.
  4. Repeat steps 2-3 for each emotion (8 emotions totally).

– b) Repeat (a) for each actor (12 actors totally) and for each action (8 actions totally).

Indeed, Data selection process is mainly intended to remove expressive movements that receive low recognition rates at the level of automatic classification and human perception.

10

10.1.2 Result

In the previous section, we illustrate our approach of Data Selection. Applying this approach allows the remove of "Outliers" from the database. That is, the expressive movements that receive low scores of recognition at the level of automatic classification and human perception are discarded.

Table 10.1 contains the quantification of expressive movements remaining after Data Selection process. We present the quantification of expressive movements per action and per emotion. The actions are 1) Simple Walking (SW), 2) Moving Books (MB), 3) Walking with an object in hands (WH), 4) Knocking at the door (KD), 5) Being Seated (BS), 6) Sitting Down (SD), 7) Lifting (Lf), and 8) Throwing (Th). The emotions are 1) Anxiety (Ax), 2) Pride (Pr), 3) Joy (Jy), 4) Sadness (Sd), 5) Panic Fear (PF), 6) Shame (Sh), 7) Anger (Ag) and 8) Neutral (Nt).

We observe from Table 10.1 that 61% of the database remains after Data Selection process. The discard of "Outliers" is related to low recognition rates in automatic classification and human recognition of emotions. We conclude that the quantification of the "Outliers" is congruent with the low recognition rates observed in some particular emotions and actions as highlighted in Chapter 9. For instance, a high amount of “Outliers” are discarded from Sitting Down and Being Seated actions (See Table 10.1). We showed in Chapter 9 that these actions receive the lowest classification rates. Besides, we observe from Table 10.1 that the highest percentage of data that remains after Data Selection process is shown in Simple Walking action (71%). In Chapter 9, we also showed that the emotions are the best classified in Walking actions. Finally, Table 10.1 also shows that the highest percentages of data that remains after Data Selection process are observed for Sadness, Anger and Neutral expressions (94%, 78% and 86%). In Chapter 9, we concluded that those emotions were the best recognized by humans and classified in automatic analysis.

Table G.1 in Appendix G precises the actors Id whose expressive movements are kept after the remove of “Outliers” from the data.

	Ax	Pr	Jy	Sd	PF	Sh	Ag	Nt	All Emo
SW	38%	100%	82%	91%	57%	64%	56%	88%	71%
MB	1%	46%	66%	94%	50%	38%	59%	100%	55%
WH	18%	62%	90%	90%	41%	53%	75%	89%	64%
KD	18%	62%	90%	90%	40%	51%	78%	90%	63%
BS	55%	56%	27%	100%	27%	26%	90%	92%	58%
SD	55%	38%	18%	100%	36%	55%	73%	92%	57%
Lf	43%	53%	72%	87%	37%	26%	91%	82%	60%
Th	23%	47%	39%	100%	35%	58%	100%	52%	56%
All Acti	31%	58%	61%	94%	41%	47%	78%	86%	**61%**

Table 10.1: Quantification of expressive movements after data selection process. All Emo and All Acti stands respectively to All Emotions and All Actions.

Table 10.2 illustrates the impact of data selection on the classification of Emilya expressive movements using the whole set of body cues. The classification rates are measured according to the OOB RF validation schema and they are averaged across 50 runs. The results show that removing “Outliers” data allows increasing the classification rates (See Table 10.2).

	SW	MB	WH	KD	BS	SD	Lf	Th
Original Data	84.63%	83.31%	83.99%	81.88%	67.55%	68.06%	78.51%	79.29%
“Selected” data	87.75%	90.63%	88.75%	87.76%	77.71%	76.89%	87.08%	89.63%

Table 10.2: Correct classification rates (%)of Random Forest before and after data selection process.

In the remaining of this chapter, we focus on the data obtained from Data Selection process to explore the most relevant features that contributes to the classification and the discrimination of bodily expression of emotion.

10.2 Feature selection approach

In Chapter 4, we reported that previous studies on emotional body expression mostly focused on perceptual rating of body cues to explore the most informative expressive body cues. In recent years, several works have been conducted to explore the most relevant expressive body cues through kinematic analysis. Several studies were focused on the statistical power of body cues using statistical tests. However, statistical techniques ignores body cues dependencies. Besides, they are useful to identify irrelevant body cues for which no significant main effect of expressed emotions is revealed, but they do not guarantee to find an optimal subset of body cues starting from a large set of features. Other studies used machine learning techniques to find the most expressive body cues.

In our work, we aim to identify a comprehensive set of body cues useful for the classification and the characterization of emotional body expression. To achieve this goal, we apply machine learning technique to select a subset of features starting from our set of 114 Multi-Level motion capture features defined in Chapter 6. Feature Selection is referred to as FS.

In chapter 4 we concluded that hybrid feature selection approaches combining Embedded feature ranking and Wrapper FS seem the most suitable for our purpose. Thus, our FS approach combines an embedded feature ranking and a wrapper FS approaches. The process of FS is applied for each action. The results are later combined to investigate relevant features across different actions.

As discussed in Chapter 4 and 9, Random Forest (RF) approach has been widely known as an efficient non-parametric ensemble method that provides both high performance and an embedded approach for feature selection. Thus, RF approach is adopted in our work to perform both an embedded and a wrapper FS strategies.

Our main purpose is to select the expressive body cues that are relevant for the characterization of emotional body expression in daily actions. In classification tasks, RF approach returns a measure of relevance for each feature considered during learning process. Thus, these features could be ranked according to their relevance measure. Several relevance measures have been used in previous works such as the Selection frequency, the Gini importance and the Permutation importance [Strobl et al., 2007, Altmann et al., 2010, Strobl et al., 2008, Hapfelmeier and Ulm, 2013] (See Appendix D). The Permutation importance has been in particular widely used to quantify the relevance of features during the training process of Random Forest approach. However, it has been shown that the permutation measure can be sensitive to the presence of a strong correlation between the features. Recently, Genuer et al. [Genuer et al., 2010] showed that the ranking of features according to their relevance measure based on the permutation measure can be robust to a strong correlation between the features. Following the approach presented in the work of Genuer et al. [Genuer et al., 2010], we use the permutation measure as the measure of relevance (See Appendix D). In order to select the most relevant features, we apply a wrapper algorithm of feature selection based on the ranking of features and a recursive backward elimination strategy. This approach is inspired from previous works on feature selection using RF in bioinformatics applications [Díaz-Uriarte and Alvarez de Andrés, 2006, Svetnik et al., 2004].

Figure 10.1 summarize the steps of our wrapper approach. The wrapper approach is described as follows:

1. Consider the whole set of features (114 in total) ,
2. Train RF model and save the ranking of features (according to their relevance measure) as well as the classification error across 50 runs.
3. Eliminate a batch of least relevant features. Unlike previous works where the batch of least relevant features is defined as a percentage of initial features number (such as 20% [Díaz-Uriarte and Alvarez de Andrés, 2006]), we propose to define this batch based on statistical tests; we remove the least relevant features for which the relevance measures are not statistically different based on Tukey-Kramer test (alpha=0.05).
4. Repeat steps 2-3 until no feature is left.
5. Starting from the lowest average error, we select the smallest set of features that does not lead to a significant increase of the error rate according to Tukey-Kramer test (alpha=0.05) (See Figure 10.1).

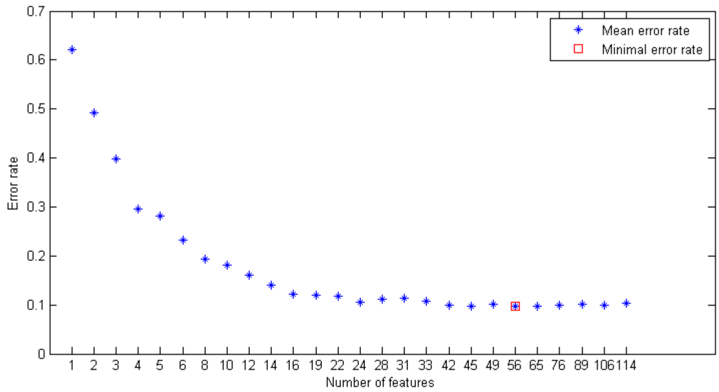
10

As previous works reported a difference between employing the OOB (out-of-bags) RF model error and Cross-validation error [Svetnik et al., 2004], we run the wrapper algorithm for each error measure and the set of features leading to the lowest error is chosen.

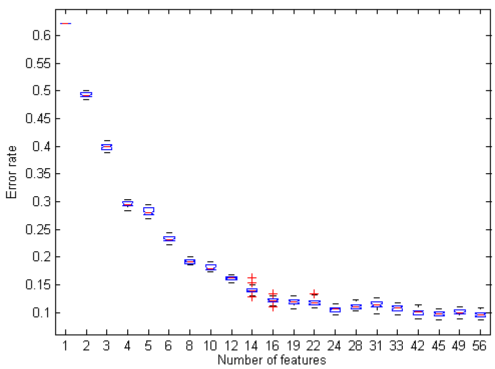
10.3 Evaluation of feature selection approach

In this section, we aim to evaluate our Feature Selection approach described in the previous section. Firstly, we provide the estimate of CCR_{OOB} before and after FS process when it is performed on the same training set of each action (See sectionD.4, Appendix D for more details about the CCR based on OOB error). Table 10.3 depicts the CCR_{OOB} in each action before (using the whole set of features) and after FS process (using selected features). As expected, the reduced set of features allows increasing the CCR_{OOB} as it has been chosen based on the same training dataset related to each action.

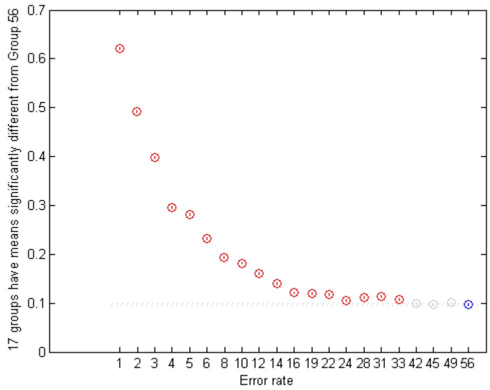
In the remaining of this section, we evaluate our FS approach using a training and testing datasets based on two criteria; 1) classification rates before and after FS process (subsection 10.3.1, and 2) the comparison of different feature sets selected in a training and a testing datasets (subsection 10.3.2).



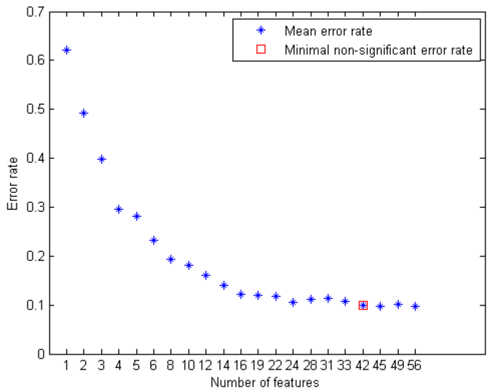
(a) The mean of error rate across 50 runs for each set of features and the minimum value.



(b) The boxplot of the classification error rates for 50 runs



(c) Result of Tukey-Kramer test for statistical difference measure.



(d) 42 Final selected features number according to the Tukey-Kramer test.

Figure 10.1: Wrapper Approach: Recursive Backward Elimination

	SW	MB	WH	KD	BS	SD	Lf	Th
Before FS	87,75%	90,63%	88,75%	87,76%	77,71%	76,89%	87,08%	89,63%
After FS	89,43%	91,53%	88,75%	89,03%	79,06%	78,29%	88,20%	90,46%

Table 10.3: Correct classification rates (%)of RF before and after feature selection (FS) process using "Selected" data.

10.3.1 Classification rates based evaluation

To evaluate FS approach based on classification rates, we apply the following process;

- We divide the database into a training and a testing datasets, where both the training and the testing datasets are based on the same set of 114 Multi-Level features.
  - The distribution of training and testing datasets is performed once using three fold cross-validation approach (see Table 10.4), and once using "Selected" data versus "Outliers" data (see Table 10.5).
- We measure the Correct Classification Rate (CCR) according to the testing dataset; we call CCR_{CV} the Correct Classification Rate based on a Cross-Validation scheme and we call CCR_{SO} the Correct Classification Rate measured while considering the "Selected" data as the training dataset and the 'Outliers" data as the testing dataset. Random Forest approach is used for the classification purpose, where; the whole set of features is used, 500 trees are used to build the forest (as the optimal number of trees is mostly around 500 for each action based on OOB error, see section 9.2, Chapter 9), the Correct Classification Rate is averaged across 20 runs.
- We apply FS process on the training dataset. The result is a subset of the most relevant features,
- The subset of the most relevant features is used to measure the Correct Classification Rate (CCR_{CV} or CCR_{SO}).
- We compare the Correct Classification Rate using the whole set of 114 Multi-Level features (before FS process) and the one using the most relevant features selected according to the training dataset (After FS process).

Table 10.4 depicts the mean of CCR_{CV} across 3 fold cross-validation approach, averaged across 20 runs. This table shows that FS approach does not always lead to significant increase of classification rates. The CCR_{CV} is slightly better using selected features for MB, KD, BS, and Lf actions. In other actions, the classification is slightly better using the whole set of features. This result can be explained as RF model is sufficiently robust to handle the classification task using a large set of features. In a previous work, Svetnik et al. [Svetnik et al., 2004] asserted that they "have never yet observed a case where the performance of RF actually improves as

(CCR_{CV})	SW	MB	WH	KD	BS	SD	Lf	Th
Before FS	80,80%	79,55%	80,19%	77,29%	64,90%	64,76%	74,69%	75,60%
After FS	80,57%	79,86%	79,76%	78,49%	65,70%	63,43%	75,04%	75,46%

Table 10.4: Correct classification rates (%)of RF before and after feature selection (FS) process based on three fold cross-validation approach.

(CCR_{SO})	SW	MB	WH	KD	BS	SD	Lf	Th
Before FS	21,49%	13,49%	9,77%	10,11%	4,26%	10,27%	11,61%	10,59%
After FS	23,62%	15,30%	11,18%	10,79%	5,40%	10,60%	12,86%	11,62%

Table 10.5: Correct classification rates (%)of RF before and after feature selection (FS) process using 'Outliers" data as test dataset and "Selected" data as training set.

variables are reduced".

When considering "Selected" data as training dataset and 'Outliers" data as testing dataset, we expected low classification rates since "Outliers" data are defined as data that are neither recognized in automatic classification nor in human recognition of emotions (See section 10.1). Table 10.5 shows that although the CCR_{SO} using "Outliers" data as the testing dataset is mostly lower than chance level (12.5%=1/8), it is always better after FS than before FS.

10.3.2 Features set based evaluation

Since our aim is not only to select an optimal set of features for classification purpose but also to interpret the most relevant features, we propose another strategy to evaluate our FS approach. This strategy is based on a categorical comparison of the subset of features selected using a training and a testing datasets. Similarly to section 10.3.1, the distribution of training and testing datasets is performed once using three fold cross-validation approach (see Table 10.6), and once using "Selected" data versus "Outliers" data (see Table 10.7). That is we evaluate the performance of our feature selection approach in two contexts: once when the training and testing datasets are somehow similar (cross-validation approach) and once when they are highly different ("Selected" data versus "Outliers" data).

The process of FS is performed separately in each training and testing datasets and the subsets of selected features are compared. The subsets of selected features are compared according to two criteria; the percentage of intersection and the Chi-squared test of independence. The latter indicates whether the subsets of selected features are independent. When p-value is smaller than the significance level (0.05, 0.01, or 0.001), we reject the null hypothesis that the subsets of features are independent.

	SW	MB	WH	KD	BS	SD	Lf	Th
FS intersection	56,94%	62,75%	59,12%	58,30%	44,75%	62,42%	56,22%	50,47%
Chi-Squared	***	***	***	***	***	***	***	***

Table 10.6: Comparison of selected subset of features using 3 fold cross validation; the percentage of intersection of selected features averaged according to the three-fold cross validation results and the result of Chi-squared test of independence. '***' stands for a significant difference with p<.001

	SW	MB	WH	KD	BS	SD	Lf	Th
FS intersection	37,88%	55,22%	43,66%	53,33%	49,15%	34,92%	48,57%	28,85%
Chi-Squared	**	***	***	***	***	**	***	*

Table 10.7: Comparison of selected subset of features using "Selected"s Vs "Outliers" data; the percentage of intersection of selected features and the result of Chi-squared test of independence. '***', '**' and '*' stand respectively for a significant difference with p<.001, p<.01, p<.05.

Table 10.6 indicates the percentage of intersection of the subsets of features averaged across 3 cross-validation approach. It also indicates the result of Chi-Squared test of independence. The same result of Chi-Squared test is found for each fold: p<.001. Hence, the Chi-squared test of independence indicates that the subset of features selected in training dataset is always significantly dependent (p<.001) of the subset of features selected in test dataset. The percentage of intersection of the two subsets of features ranges between 44.75% and 62.75% (See Table 10.6). Thus, using 3 fold cross-validation strategy, the results in Table 10.6 show that our FS approach is well stable to provide similar subsets of features when applied separately to training and testing datasets.

Table 10.7 contains the comparison of selected subsets of features using "Selected" Vs "Outliers" Data. Table 10.7 shows lower percentages of intersection using "Selected" and "Outliers" dataset comparing to 3 fold cross-validation strategy. However, the Chi-squared test of independence indicates that the subset of features selected in "Selected" data (training dataset) is always significantly dependent (p<.001, p<.01, p<.05) of the subset of features selected in "Outliers" dataset (test dataset). This result highlights the stability of RF based FS approach.

	All Features		Subset of Selected Features (SSF)	Actions				
				SW + WH	MB + KD	BS + SD	Lf + Th	All actions
Posture/Movement	CrossOcc	2	CrossOcc	0	0	0	0	0
	Corr	6	Corr	0	0	0	0	0
	SymOcc	2	SymOcc	1	0	0	0	1
	STD	26	STD	7	4	2	3	9
	Posture	32	Post	14	13	17	14	24
	Speed	23	Speed	9	10	10	9	18
	Acceleration	23	Acceleration	6	11	9	8	10
	Total:	114	Total:	37	38	38	34	62
Anatomical	Arms	54	Arms	15	21	13	16	28
	Lower Body	29	Lower Body	14	5	8	7	16
	Head	13	Head	3	4	8	3	10
	Torso	12	Torso	2	5	4	4	4
	TorsoHead	6	TorsoHead	3	3	5	4	4
	Total:	114	Total:	37	38	38	34	62
Directional	Lateral	22	Lateral	3	5	5	5	9
	Sagittal	20	Sagittal	9	4	6	5	10
	ThreeD	34	ThreeD	17	17	13	13	22
	VerticalLength	30	VerticalLength	6	11	10	10	16
	VerticalRotation	8	VerticalRotation	2	1	4	1	5
	Total:	114	Total:	37	38	38	34	62

Figure 10.2: Quantification of the set of 114 Multi-Level features and quantification of selected features according to Anatomical, Directional and Posture/Movement description level.

10.4 Feature selection across different actions

Feature Selection (FS) approach was described and validated in previous sections. We applied this FS approach to select the subset of relevant features for each action. We obtained 43, 55, 69, 52, 48, 59, 56, and 44 features (out of 114) respectively for Simple Walking (SW), Moving Books (MB), Walking with an object in hands (WH), Knocking at the door (KD), Being Seated (BS), Sitting Down (SD), Lifting (Lf), and Throwing (Th) actions. The presence of selected features regarding our initial set of 114 features is illustrated in Figures 10.4, 10.7, 10.6, 10.5 where each column represents a feature name and each line represents an action, a group of two actions, or the group of all the actions. We will discuss the content of these Figures in section 10.5.

Our aim is to identify the most relevant body cues for the characterization of bodily expression of emotion across different action. Thus, we do not aim to focus on action-related features; we are looking for the common relevant features selected across different actions. For the identification of the Subset of Selected Features (SSF) across different actions, we regroup and discuss the intersection of selected features firstly across "similar" actions and secondly across all the actions.

10.4.1 Intersection approach

The first approach that we used to identify the Subset of Selected Features (SSF) across different actions is the intersection of the SSFs obtained for each action. We call Int-SSF the subset of features obtained from the intersection of two (or more) subsets of selected features. The intersection approach is applied firstly to identify the SSF across "similar" actions then across all the actions.

Intersection of features across "similar" actions :

We define "similar" actions as actions that mainly involve the same body segments through similar bio-mechanical movement. We define four groups of actions; 1) Walking (SW + WH), 2) Repetitive Arms movement (KD + MB), 3) Non-repetitive Arm movement (Lf + Th) and finally 4) Sit Down movement (SD + BS). Intersection approach is applied in order to identify the intersection of relevant features across "similar" actions in order to filter out the action-dependent features (features that are tailored to the action). The Int-SSFs obtained by the intersections across "similar" actions are respectively composed of 37, 38, 38, and 34 features for (SW + WH), (KD + MB), (Lf + Th) and finally (SD + BS).

Selected features across all the actions :

Across all the actions, the intersection between the 8 subsets of features (for 8 actions) leads to an Int-SSF of 11 features. These 11 features are presented in the subset of selected feature for each action. They are highlighted in red in Figures 10.4, 10.5, 10.6, and 10.7. This Int-SSF consists in: **5** posture features (Backward torso posture, Forward torso posture, 3D openness of feet, Downward head flexion and Downward torso flexion; See Figure 10.4), **3** Speed features (the speed of absolute elbows extension, the speed of absolute hands extension, the speed of relative elbows extension; See Figure 10.5, **2** Acceleration features (the acceleration of relative elbows extension and the acceleration of relative right hand extension; See Figure 10.6) and **1** Postural Changes feature (postural change of backward torso movement; See Figure 10.7).

Based on the data of all the actions, the CCR_{OOB} error of a RF model built with this Int-SSF of 11 features is significantly lower (around 68%) than the CCR_{OOB} error of a RF model built with the whole set of 114 features (around 81%). Thus, it seems that this reduced subset of feature does not achieve the same classification rate based on the whole set of features. That is, the intersection between the subsets of features from all the actions provide a reduced subset of features at the cost of the classification accuracy. In the following subsection, we focus on an optimal subset of features across all the actions that provides a good compromise between the classification accuracy and a reduced subset of features.

10.4.2 Feature Selection approach across all the actions

In order to obtain an optimal subset of features across all the actions, we apply our feature selection approach to the data containing expressive movements of all the actions. That is grouping the data of different actions and applying Feature Selection. We obtain a subset of 62 features. We call FS-SSF the subset of selected features obtained from Feature Selection approach. Based on the data of all the actions, the CCR_{OOB} error of a RF model built with this FS-SSF of 62 features is 82%, which is slightly better than the CCR_{OOB} error of a RF model built with the whole set of 114 features (81%).

In the following sections, we discuss the Int-SSFs identified across "similar" actions (4 Int-SSFs for 4 groups of actions) and the FS-SSF obtained across all the actions. The discussion refers firstly to the features quantification according to each description level (e.g. how many selected features correspond to upper or lower body). This quantification will be discussed in section 10.5). Secondly, the obtained Int-SSFs and FS-SSF will be described in term of features ranking according to their relevance measure returned by RF model (in section 10.6). RF model provides the ranking of features according to the classification task, but it also provides the ranking of features for each class (emotion in our case) considered during the learning process. Based on this output, we are able to provide a visual presentation of the role of each relevant feature according to each expressed emotion across different actions.

10.5 Quantification and presence of selected features

In this section, we compare the quantification of the subsets of selected features according to the different description levels of our body movement notation system (this notation system is described in Chapter 6). The goal from this comparison is to provide first insights into the distribution of relevant features.

Figure 10.2 depicts the quantification of the whole set of 114 Multi-Level features as well as the quantification of Int-SSF for each group of similar actions and the quantification of FS-SSF across all the actions. The quantification of features is presented according to the description levels considered in our body movement notation system (See Chapter 6). STD, SymOcc, Corr and CrossOcc refer respectively to Standard Deviation (postural changes), Occurrence of Symmetry of arms posture, Correlation of arms motion, and finally the Occurrence of Crossing limbs (arms and legs). This Figure provides a summary of the presence of selected features (detailed in Figures 10.4, 10.7, 10.6, 10.5) according to each description level.

We note that the features that represent the Pearson correlation of arms motion (elbows motion and hands motion) along the three axes as well as the index of crossing limbs (hands and feet) do not occur as relevant features in any group of "similar" actions (See Figure 10.2). This may also suggest that the proposed measure of crossing limbs and arms motion correlation are not well appropriate for the study of emotional body expression in the actions considered in our database: SW, WH, MB, KD, SD, BS, Lf and Th.

For each group of "similar" actions, we observe that postural cues are more present than Speed, Acceleration and Postural Changes features. However, when grouping the features that describe movement dynamics (speed and acceleration), we find that movement dynamics features are as present as postural features (See Figure 10.2).

Knowing that arms features are more present than lower body features in the initial set of 114 Multi-Level features (54 vs 29, See Figure 10.2), we find that arms and lower body features of the SSF are equally present in walking actions (SW+WH). This result can be explained as walking pattern involves both arms and legs movement (See Figure 10.2). However, arms features are more present than lower body features in Sit Down actions (SD+BS) although Sitting Down and Being Seated involve both upper and lower body segments.

As repetitive (MB+KD) and non-repetitive Arms movement (Lf+Th) mainly involve arms movement, we expected that arms features are more present than lower body features in the Int-SSF. The result is congruent with this expectation; 21 and 16 arms features against 5 and 7 lower features are selected respectively in MB+KD and Lf+Th actions (See Figure 10.2).

Across all the actions (See Figure 10.2, last column), we observe that movement dynamics features are as present as postural features (24 postural and 28 speed and acceleration features). Although speed and acceleration features are equally distributed in the initial set of 114 Multi-Level features (23 vs 23, See Figure 10.2), speed features are more present than acceleration features across all the actions (18 vs 10, See Figure 10.2). We also observe that arms features are more present than lower body features in the FS-SSF across all the actions(28 vs 16, See Figure 10.2). This is mainly due to the presence of arms movement based actions: Moving Books, Knocking, Lifting and Throwing.

Figures 10.4, 10.5, 10.6 and 10.7 illustrate in details the features considered in the SSF: according to each action (FS-SSF), across similar actions (Int-SSF), and across all the actions (FS-SSF). While common features occur in the FS-SSF across different actions (e.g. backward arms posture, See Figure 10.4), some features appear to be more tailored to a specific action such as lateral torso posture which was only selected for Throwing action (See Figure 10.4).

10.6 Ranking of selected features

In section 10.5, we explored the quantification of the subset of selected features (SSF) according to different description levels of our body movement notation system. In this section, we explore the predictive power of each selected feature. Indeed, we aim to give more insights into the relevance of each feature according to the classification of all the emotions and also according to the classification of each emotion.

Thanks to the relevance measure returned by RF model for each feature, we are able to rank the features and to explore the predictive power of each one according to the classification task and according to the classification of each emotion. For this purpose, we train 4 RF models based on the Int-SSFs obtained across "similar" actions. These 4 RF models correspond to the groups of "similar" actions (walking: SW+WH, repetitive arms movement: MB+KD, Sit Down: BS+SD and non-repetitive arm movement: Lf+Th). We also train one RF model based on the FS-SSF obtained across all the actions. The rankings of features provided in this section are based on the relevance measure returned by these RF models and averaged across 50 runs. Each RF model returns the ranking of features according to the classification task as well as their ranking according to each class considered during learning process (a class represents an emotion in our work).

In order to ease this discussion, we make use of a color-based graphical representation of features ranking (See Figures 10.3, 10.10, 10.8, 10.9, 10.11). These Figures are presented in table form to give a clear visualization of the ranking of features, where the columns represent the features names.

We give an explanation on how to each such tables. Features names are illustrated in the columns. We follow a composition approach. For instance in Figure 10.3, the lateral postural closeness feature (the first postural feature) is defined as: Posture->Lateral->Inside->Arms. In each table, each cell depicts the ranking of a given feature. The darker the color is, the lower ranking value is, and the more relevant the feature is. For instance, a ranking value equal to 1 indicates the most relevant feature. The ranking of features are provided according to the permutation measure returned by RF model for each emotion (See Appendix D). When no ranking value is provided, we can deduce that this feature was not selected among the SSF. For instance Figure 10.3 shows that the lateral closeness of arms (Lateral->Inside->Arms) is not selected as a relevant feature for walking actions (SW+WH). Thus, we do not provide any ranking value of this feature in walking actions.

We also train RF models for each action. The rankings of features according to each action are provided in Appendix H (Figures H.21, H.24, H.22, H.23, H.25). In this chapter, we only focus on the features selected across "similar" actions and across all the actions.

In the following subsections, we discuss firstly the ranking of features according to the classification task, then their ranking according to each expressed emotion.

10.6.1 Ranking of features according to the classification task

Figure 10.3 illustrates the ranking of features for each group of similar actions and across all the actions. The ranking results are split into 4 sub-figures for posture, acceleration, speed and standard deviation features (See Figure 10.3). For each of these sub-figures, the lines represent the actions labels and the columns represent the features names.

We can observe from Figure 10.3 that few common features appear to be relevant

(a) Ranking of postural features

Posture/ Movement:	Posture																										Sym Occ
Direction:	Lateral					Sagittal					ThreeD								Vertical Length							Vertical Rotatio	ThreeD
Subdirection:	Inside		Lateral	Outside		Backward		Forward											Down ward	Downward. Flexion			Upward		Upward. Flexion		
Modality:	Arms	Lower Body	Head	Arms	Lower Body	Lower Body	Torso Head	Arms	Lower Body	Torso Head	Arms						Lower Body	Lower Body	Lower Body	Head	Lower Body	Torso	Arms		Head	Head	Arms
SubModality:											Elbows. Hips	Hands. Hips	Lelbow. Relbow	Lhand. Head	Lhand. RHand	Rhand. Head	Feet. Hips	Lfoot. Rfoot			Knees						Lelbow. Relbow
SW + WH		23		36		21	15	37	9	1	33		34		32			12		4		26			14		16
MB + KD					21		18	36		1	35			29		22	8	30		6	20	4		37			
BS + SD	7	1			10		5			8	26	12	25				34	19	23	2	32	16	20		27	13	
Lf + Th		26		33			23			5			29		6		28	12	34	19	20	4	10			32	
All Actions	57	14	60	39	17	41	6	62	24	1	56	61	29	53	46		27	35	42	3	23	11	40		18	19	44

(b) Ranking of acceleration features

Posture/ Movement:	Acceleration															
Direction:	Lateral		Sagittal	ThreeD								Vertical Length				Vertical Rotation
Subdirection:												Flexion			Vertical	
Modality:	Head	Torso	Torso Head	Arms						Lower Body		Arms	Head	Torso	Arms	Head
SubModality:				Elbows. Hips	Hands. Hips	Lelbow. Relbow	Lhand. Head	Lhand. RHand	Rhand. Head	Feet. Hips	Lfoot. Rfoot	Elbows				
SW + WH						5	35		31	19	17					28
MB + KD	12	25		5	2	3	27	11	34			16	31			10
BS + SD	28		18	11	17	3			15				14	33		6
Lf + Th	17	24		3	7	9			1			21			8	
All Actions	15			5	10	2		21	16			31	45		49	7

(c) Ranking of speed features

Posture/ Movement:	Speed																		
Direction:	Lateral				Sagittal			ThreeD							Vertical Length				Vertical Rotation
Subdirection:															Flexion			Vertical	
Modality:	Arms	Head	Lower Body	Torso	Arms	Lower Body	Torso Head	Arms					Lower Body		Arms	Lower Body	Torso	Arms	Head
SubModality:								Elbows. Hips	Hands. Hips	Lelbow. Relbow	Lhand. RHand	Rhand. Head	Feet. Hips	Lfoot. Rfoot	Elbows	Knees			
SW + WH			27		30	3		8	20	11			6	2		24			
MB + KD	19			7				23	17	9	14	38			15		13	28	
BS + SD				37		36	30	9	22	29			31				38	35	4
Lf + Th				16		31	13	18	15	22		11					30	2	
All Actions	28	52		32	50	30	34	4	12	8	36	22	38	33	20	54	48	25	9

(d) Ranking of postural changes features

Posture/ Movement:	Postural Changes (standard deviation)											
Direction:	Sagittal		ThreeD				Vertical Length				Vertical Rotation	
Subdirection:	Backward						Flexion		Vertical		LRRotation	
Modality:	Lower Body	Torso Head	Arms		Lower Body		Lower Body	Torso	Arms	Lower Body	Head	Torso
SubModality:			Elbows. Hips	Rhand. Head	Feet. Hips	Lfoot. Rfoot	Knees					
SW + WH	13	22	25			18			29	10		7
MB + KD		26		33				32	24			
BS + SD		24									21	
Lf + Th		25			27				14			
All Actions	51	13	55		58		59		43	47	26	37

Figure 10.3: Ranking of the subset of relevant feature based on RF relevance measure returned according to the classification task

across “similar” actions and across all the actions. In particular, the forward posture of torso and head appears to be highly important for emotions expression in various actions (its ranking value is mostly around 1, meaning that it is the most important feature). The downward flexion of the head is also mostly considered among the first relevant features.

In walking actions (SW+WH), the first relevant features involve the forward leaning of torso and head, the downward flexion of the head, but also the speed of lower body limbs. The acceleration of arms swing is also considered as highly relevant for the classification of emotions in walking actions (SW+WH). Arms swing is described through the relative elbows extension (Acceleration->ThreeD -> Arms -> Lelbow.Relbow). Another relevant features related to SW+WH is the postural changes of torso orientation. The vertical orientation of the torso is measured according to the shoulders swing around the vertical axis. Thus, the change in shoulders swing during walking is mainly resulted from arms swing.

Looking in more details to the ranking of features related to the classification of emotion in Sit Down group of actions (BS+SD), we observe that the first relevant feature refers to the lateral closeness of lower body limbs. The speed/acceleration of head and arms movement and to the posture of head and torso are also relevant for the classification of emotions in Siting Down and Being Seated actions.

As such, lower body limbs features turn out to be highly relevant for the classification of emotions in Walking (SW+WH), Sitting Down and Being Seated actions (BS+SD).

In addition to the forward leaning of torso and head, the features describing the acceleration of arms movement are ranked as the most relevant features for the classification of emotions in arms movement actions (MB+KD and Lf+Th). As Lifting and Throwing actions (Lf+Th) are mainly performed through the right hand, the acceleration of the relative right hand extension appears to be the first most relevant feature for (Lf+Th) group of actions (Acceleration-> ThreeD -> Arms -> Rhand.Head).

10.6.2 Ranking of features according to the classification of each emotion

In the previous subsection, we study the ranking of relevant features according to the actions. We now provide more insights into the ranking of relevant features according to each expressed emotion. To be able to present all the results, we split them into 4 figures; Figure 10.8 represents the ranking of posture features, Figure 10.9 reflects the ranking of postural changes (standard deviation) features, Figure 10.10 represents the ranking of speed features, finally Figure 10.11 contains the ranking of acceleration features.

For each of these figures, we present the ranking of features according to each emotion. Similarly to Figure 10.3, these figures are presented in tables form where the lines represent emotion and action labels and the columns represent the features name.

Posture features :

Lateral closeness of body posture (hands and feet) mostly receives the highest relevance measure for the classification of each emotion in (BS+SD) group of actions (See Figure 10.8, first two columns). We showed previously that the lateral closeness of feet is particularly considered as an important expressive feature for the classification of emotions in BS+SD group of actions as it receives a good ranking for classification task (ranking value =1, See Figure 10.3). Lateral closeness of feet is also considered as relevant for each emotion (ranking value ranging from 1 to 11, See second column in Figure 10.8). Lateral openness of feet also contributes with high relevance measure to the classification of emotions in Sit Down (BS+SD) group of actions, specifically for Anxiety, Panic Fear and Shame expressions.

However, the relevance of lateral closeness and openness of feet is not restricted to Sit Sown actions (BS+SD). For instance, the lateral closeness of feet also receives good ranking for Pride and Sadness in Walking group of actions (SW+WH) (See Figure 10.8). Although repetitive arm movement group of actions (MB+KD) are mainly based on arms motion, the lateral openness of feet also appear to be highly relevant for the expression of Anxiety, Pride and Shame in (MB+KD) group of actions (See Figure 10.8).

Across all the actions, the lateral openness of feet is highly relevant for the expression of Anxiety and Shame (See Figure 10.8). However, the fact that the extension/openness of limbs appear as relevant for the expression of a given emotion does not inform us on how this emotional body expression is characterized. That is, on what value does this feature have. In chapter 11, we will discuss the characterization of emotional body expression through the Int-SSF identified in this chapter across all the actions (as explained in section 10.4.1). For instance, we will show that Shame expression is characterized with higher closeness of feet than other emotion expressions across all the actions.

The Three dimensional openness of the feet (Posture-> ThreeD -> Lower Body -> Lfoot.Rfoot) receives different ranking values according to the expressed emotion and the performed action (See Figure 10.8). While the 3D openness of feet is the most relevant for Anger expression in (SW+WH) group of actions, it is the most relevant for Panic Fear expression in (BS+SD) group of actions. Interestingly, the relevance of the 3D openness of feet is not restricted to the actions that involve lower body parts such as walking and sitting down, but it also appears to be relevant for the characterization of expressive body postures in arms movement actions (such as Lifting and Throwing). For instance, the 3D openness of feet is ranked among the first relevant features for Sadness and Shame expression in non-repetitive arm movement (Lf+Th).

Knee flexion (Vertical Length-> Downward. Flexion-> Lower Body ->Knees) also appears to be an important expressive feature in arms movement based actions. It is considered among the first relevant features for the classification of Anxiety and Panic Fear in repetitive arms movement (Knocking+Moving Books) and for

10

the classification of Anxiety, Sadness, and Neutral in non-repetitive arm movement (Lifting+Throwing). Across all the actions, Knee Flexion feature is highly important for Neutral expression.

Forward torso posture is mostly considered as a highly relevant feature for the classification of each emotion. While mostly considered among the first relevant feature for Anxiety, Pride and Joy expressions, backward torso posture is particularly the most relevant feature for Joy and Pride expression in Sit Down action (See Figure 10.8).

Downward head and torso flexion appear to be highly relevant for the expression of Anxiety, Pride, Sadness, Panic Fear and Neutral across different actions. Head downward flexion is ranked as the best relevant feature for Sadness expression in Sit Down actions (BS+SD), but also across all the actions. Except in walking action, torso flexion posture contributes better than head flexion for the classification of Pride expression.

Left/Right head rotation (Posture->Vertical Rotation-> Head) feature plays an important role in particular for Pride expression in Sitting Down and repetitive arms movement actions (MB+KD). It is also ranked among the first relevant features for Joy, Sadness and Neutral expressions in Sitting Down action (BS+SD).

The amount of upward arms posture turns out to be the most relevant feature for Anxiety, Pride, Sadness and Shame in non-repetitive arms movement (Lifting+Throwing). Finally, the occurrence of elbows posture symmetry appears to be important in particular for Anxiety, Pride and Neutral expressions in walking actions (SW+WH).

Postural changes features :

Figure 10.9 represent the ranking of postural changes features according to each emotion and each group of actions.

The feature that describes arms movement in vertical direction is highly relevant for each emotion expression in non-repetitive arm movement (Lifting+Throwing). These actions are mainly based on arm movement in vertical direction. However, this feature is also considered as highly relevant to classify Neutral expression in walking action (in walking, this feature refers to arms swinging). It also appears to be highly relevant for the classification of Anxiety and Panic Fear in repetitive arms movement (Moving Books+ Knocking).

The feature of backward movement of upper body parts is considered as highly relevant mainly for Anxiety, Pride and Joy expression across various actions. The flexion of the torso appears to be particularly highly relevant for Pride and Neutral expressions in repetitive arms movement (Knocking and Moving Books).

Vertical head rotation receives the best ranking for Pride expression in Sit Down action, followed by Sadness, Joy and Anxiety expressions. During walking, the feature of torso movement in vertical rotation is described through shoulders swing around the vertical axis. Ir receives the best ranking in Pride and Neutral expressions in Walking action, followed by Anxiety and Sadness expressions.

Speed features : The ranking of speed features according to each expressed emotion across various actions is illustrated in Figure 10.10.

Lateral arms movement speed is more relevant for the classification of Anxiety, Joy, Sadness and Shame expressions in repetitive arms movement actions (MB+KD) than for the other emotions expressions. However, the speed of lateral torso movement receives better ranking value than the speed of lateral arm movement in (MB+KD) actions (See Figure 10.3). In particular, the speed of lateral torso movement is ranked as the first relevant feature for Joy expression in repetitive arms movement. We observed that the expression of Joy in repetitive arms movement implies lateral torso swing. Indeed, the movement of arms is spread into the torso during Joy expression in repetitive arms movement.

The speed of lower body movement plays an important role in the classification of emotions expressed in walking actions (SW+WH); the speed of sagittal legs movement and the speed of foot stride are ranked as highly relevant for the expression of most of the emotions in walking actions (SW+WH).

The features that describe the speed of arms movement are ranked differently depending on the emotion, the action and the feature. For instance, the speed of the relative hands extension (Lhand.Rhand) appears to be relevant only for Sadness and Anger expression in repetitive arms movement actions (MB+KD). Overall, the features that describe the speed of absolute hands and elbows extensions (Elbows.Hips and Hands.Hips) as well as the relative elbows extension (Lelbow.Relbow) are the most relevant across various actions. In particular, the speed of the absolute hands extension (Hands.Hips) appears to be the most relevant feature for Sadness expression in repetitive and non-repetitive arms movement actions. As such, the way (specifically the speed) the hands are moving in relation to the body center seems to be one of the main characteristics of Sadness expression in arms movement actions (MB+KD and Lf+Th).

Similarly to postural changes features (See Figure 10.9), the speed of vertical arms movement appears to be mostly relevant in non-repetitive arm movement actions (Lf+Th) (See Figure 10.10). Similarly, the speed of head orientation appears to be mostly relevant in Sit Down actions (SD+BS) (See Figure 10.10).

Acceleration features :

The ranking of acceleration features according to each expressed emotion across various actions is illustrated in Figure 10.11.

Similarly to the speed features (See Figure 10.10), we observe that the acceleration of foot stride play an important role in the classification of emotions in walking action, particularly for Anxiety, Pride and Panic Fear expressions. Except Joy, Panic Fear and Neutral expressions, the acceleration of vertical arms movement is also considered among the first relevant features for classification of emotions in non-repetitive arm movement (Lf+Th). The acceleration of head rotation has been also found to be relevant for the classification of each emotion in Sit Down action. The acceleration of head rotation has also been found to be relevant for the classification of emotions across all the actions, in particular for Panic Fear expression.

We note that most of the acceleration features, in particular those of arms movement, receive the best ranking value in Anger expression across various actions. This result highlights the importance of movement dynamics features for the characterization of Anger expression in various actions.

Summary :

We discussed the relevance of posture, speed, acceleration and postural changes features. We explored in details the relevance of each feature according to the expressed emotions and the performed actions. Several findings have been discussed.

Across all the actions, we found that the lateral openness of feet is highly relevant for Anxiety and Shame expressions. The relevance of features that describe lower body limbs is not limited to the actions that involve lower body movement. For instance, the 3D openness of feet is ranked as highly relevant for Sadness and Shame expressions in Lifting and Throwing actions.

The forward/ backward torso posture and head flexion are considered as highly relevant across all the emotions. While downward head flexion is considered as more relevant than torso flexion for Sadness expression, the straightness of the torso is considered as more important than head flexion for Pride expression. The speed of lateral torso swing is particularly relevant for Joy expression in repetitive arms movement action.

The relevance of arms movement speed receives different scores across actions and emotions. For instance, the speed of hands movement according to the body center appears to be the most relevant feature for Sadness expression in arms movement actions. We also note that the acceleration of arms movement always receive the best relevance scores in Anger expression across various actions.

Overall, we highlight two findings based on the ranking of features according to their relevance measures: Action related features and Emotion related features.

- **Action related features:** For a given expressed emotion or across all the emotions, some features receive different relevance measure according to the action. For instance, for Anger expression, the lateral closeness of the legs is particularly relevant in Sitting Down and Being Seated actions. Across all the emotions, the speed of sagittal legs movement is particularly important for Walking actions.
- **Emotion related features:** For a given action or across all the actions, the relevant measure of features can be particularly high in some emotions. For instance, the lateral openness of feet is relevant for Anxiety and Shame expressions across all the actions.

10.7 Conclusion

Random Forest (RF) approach is used to perform the feature selection (FS) task in order to identify the most relevant expressive features for the characterization of emotional body expression. The evaluation of this Feature selection approach highlights its stability to changes in the dataset. Feature selection process results in a slight increase of the RF classification rates. This result is explained as Random

Forest approach is robust enough to handle high dimensionality [Svetnik et al., 2004].

The use of RF approach allows us achieving two goals: 1) identify the most relevant features for the characterization of emotional body expression through a classification task and 2) rank the set of relevant features according to their relevance measure. In this chapter, we discuss the quantification and the ranking of the selected features according to the actions. The ranking of features is provided through a color-based graphical representation where the darkest color correspond to the most relevant feature.

We found that the most relevant features do not necessary reflect only the main body segments involved in the action (such as right hand movement in Throwing action). For instance, lateral openness of lower body limbs appears to be highly relevant for the classification of Anxiety, Pride and Shame in Moving Books and Knocking actions. This finding highlights the need to go beyond action-related features to classify emotions and to explore whole body movement based features.

We compare the quantification of features in our initial set of 114 features and the quantification of selected features according to each description level. In each subset of relevant features (across similar actions or all the actions), we found that the features that describe the speed and the acceleration of movement are always more present that the features describing body posture. This result is congruent with their initial distribution in the set of 114 Multi-Level features as we define more speed and acceleration features (23+23) that posture features (32). Although equally distributed in the set of 114 Multi-Level features (23 vs 23), the features that refer to the speed of movement are more present than the ones that refer to the acceleration of movement in the subset of relevant features across all the actions.

Except for walking actions, we also found that arms features are more present than lower body features in the subset of selected features. However, lower body features appear among the first relevant features for the classification of emotions in walking, Sitting Down and Being Seated actions. We also observed that speed and acceleration features are more present than postural features. But across all the actions, the posture of head and torso mostly appear to be more relevant than speed and acceleration features. These results highlight the need to explore the predictive power of each feature in addition to the quantification of features presence according to each description level.

We explore the presence and the ranking of features according to the actions. While some features receive high relevance measure for the expression of each emotion across different actions (e.g. Forward torso and head posture), other features are more specific to some particular actions (e.g. arms posture symmetry appears to be relevant in Walking actions). As such, our approach allows disentangling which features are specifically characteristic of emotions (the ones that are selected and considered as relevant across all the actions) and which are considered as action-related features. Indeed, some features are only selected in specific actions such as the acceleration of lower body movement selected for Walking actions. Others are selected in different actions but considered as relevant in specific actions (e.g. legs extension in repetitive arms movement actions: Moving Books and Knocking).

As we also explore the ranking of selected features according to each emotion, our approach also allows disentangling which features are specifically characteristic of all emotions (e.g. Forward torso and head posture), and which features are considered as emotion-related features (such as the speed of lateral torso swing for Joy expression in repetitive arms movement).

While the discussion of features ranking is useful to interpret the role of each feature according to each emotion, it is not enough to understand how each emotion is characterized based on solely ranking of features. Indeed, if a given feature is considered as relevant for a given emotion (e.g. ranking value is equal to 1), it is not possible to understand how and why this feature was deemed relevant. For instance, the fact that head downward flexion receives high relevance measure for a given emotion does not tell us how it contributes for the characterization of this emotion. Besides, we saw in previous sections that a given feature can receive high relevance measure for more than one emotion, but we still don't know how this feature contributes differently for the characterization of each emotion. In the next chapter, we will discuss in details the patterns of emotional body expression characterization based on the subset of selected features.

10.8 Summary of chapter

- The process of expressive features selection adopted in our work can be summarized into two steps; A) Data selection: remove "Outliers" from data, and B) Random Forest based Feature Selection: reduce the set of features for interpretation purpose.
- We combine RF based Embedded feature ranking and RF based Wrapper feature selection approaches to select the subset of relevant features useful for the discrimination between expressed emotion in each action.
- Feature selection (FS) approach is evaluated according to two criteria; 1) the classification rates before and after FS process and 2) the dependency of two subsets of selected features obtained respectively from a training and a testing datasets. Random Forest (RF) based Feature Selection (FS) approach leads to slightly better results in term of the classification rates. However, we show that RF based FS approach is stable enough to produce significantly similar subsets of features when applied on a training and a testing datasets.
- We combine the results of selected features across "Similar" actions through an intersection of features.
- We explore the quantification of selected features according to each description level and their ranking according to their relevance measure returned by RF model.
- Finally we explore the ranking of selected features for each emotion expressed across various actions. Some features appear to be relevant for the classification of all emotions in all actions (e.g. forward/backward torso posture),

others turn out to be specific to the expression of some emotions in some actions (e.g. elbows posture symmetry). We also identified relevant expressive features that do not necessary describe the main movement involved in the action. For instance, the posture of lower body limbs appears to be highly relevant for the expression of Shame and Sadness in Lifting and Throwing actions.

Posture/ Movement:	Posture																																SymOcc	
Direction:	Lateral						Sagittal						ThreeD								Vertical Length										Vertical Rotation		ThreeD	
Subdirection:	Inside		Lateral		Outside		Backward			Forward			ThreeD								Downward		Downward. Flexion				Upward	Upward	Upward. Flexion	Upward. Flexion			ThreeD	
Modality:	Arms	Lower Body	Head	Torso	Arms	Lower Body	Arms	Lower Body	Torso Head	Arms	Lower Body	Torso Head	Arms						Lower Body	Lower Body	Arms	Lower Body	Arms	Head	Lower Body	Torso	Arms	Lower Body	Arms	Head	Head	Torso	Arms	
SubModality:													Elbows. Hips	Hands. Hips	Lelbow. Relbow	Lhand. Head	Lhand. RHand	Rhand. Head	Feet. Hips	Lfoot. Rfoot			Elbows		Knees		Arms	Lower Body	Elbows				Lelbow. Relbow	Lhand. RHand
SW		x			x			x	x	x	x	x	x		x	x	x			x				x		x				x	x		x	
MB						x			x	x	x	x	x	x		x		x	x	x				x	x	x	x	x						
WH		x	x		x		x	x	x	x	x	x	x	x	x		x	x	x	x	x	x		x	x	x	x			x			x	x
KD	x	x			x	x			x	x		x	x		x	x	x	x	x	x				x	x	x		x		x			x	
BS	x	x	x			x			x	x	x	x	x	x	x	x		x	x	x		x		x	x	x	x			x	x			
SD	x	x			x	x	x		x			x	x	x	x				x	x	x	x		x	x	x	x			x	x			
Lf	x	x	x		x	x	x		x			x			x	x	x		x	x		x		x	x	x	x	x			x			
Th		x		x	x				x			x		x	x		x		x	x	x	x		x	x	x	x				x			
SW + WH		x			x			x	x	x	x	x	x		x		x			x				x		x				x			x	
MB + KD						x			x	x		x	x			x		x	x	x				x	x	x		x						
BS + SD	x	x				x			x			x	x	x	x				x	x		x		x	x	x	x			x	x			
Lf + Th		x			x				x			x			x		x		x	x		x		x	x	x	x				x			
All Actions	x	x	x		x	x		x	x	x	x	x	x	x	x	x	x		x	x		x		x	x	x	x			x	x		x	

Figure 10.4: Presence of Posture selected features

10

Posture/ Movement:	Speed																						
Direction:	Lateral				Sagittal			ThreeD									Vertical Length						Vertical Rotation	
Subdirection:			Lateral	Lateral				ThreeD									Flexion				Vertical			
Modality:	Arms	Head	Lower Body	Torso	Arms	Lower Body	Torso Head	Arms						Lower Body		Arms	Head	Lower Body	Torso	Arms	Lower Body	Head	Torso
SubModality:								Elbows. Hips	Hands. Hips	Lelbow. Relbow	Lhand. Head	Lhand. RHand	Rhand. Head	Feet. Hips	Lfoot. Rfoot	Elbows		Knees					
SW			x		x	x		x	x	x	x		x	x	x			x					
MB	x	x		x			x	x	x	x		x	x			x	x		x	x		x	x
WH	x		x	x	x	x	x	x	x	x		x		x	x	x		x		x	x		
KD	x			x	x			x	x	x		x	x			x			x	x			
BS			x	x		x	x	x	x	x				x					x	x		x	
SD				x	x	x	x	x	x	x			x	x		x	x	x	x	x	x	x	
Lf	x			x		x	x	x	x	x	x	x	x	x			x		x	x	x	x	x
Th				x	x	x	x	x	x	x			x		x	x		x	x	x			
SW + WH			x		x	x		x	x	x				x	x			x					
MB + KD	x			x				x	x	x		x	x			x			x	x			
BS + SD				x		x	x	x	x	x				x					x	x		x	
Lf + Th				x		x	x	x	x	x			x						x	x			
All Actions	x	x		x	x	x	x	x	x	x		x	x	x	x	x		x	x	x		x	

Figure 10.5: Presence of Speed selected features

Posture/ Movement:	Acceleration																						
Direction:	Lateral				Sagittal			ThreeD								Vertical Length						Vertical Rotation	
Subdirection:																Flexion				Vertical			
Modality:	Arms	Head	Lower Body	Torso	Arms	Lower Body	Torso Head	Arms						Lower Body		Arms	Head	Lower Body	Torso	Arms	Lower Body	Head	Torso
SubModality:								Elbows. Hips	Hands. Hips	Lelbow. Relbow	Lhand. Head	Lhand. RHand	Rhand. Head	Feet. Hips	Lfoot. Rfoot	Elbows		Knees					
SW					x					x	x		x	x	x							x	
MB	x	x		x				x	x	x	x	x	x			x	x			x		x	
WH		x		x		x	x	x	x	x	x	x	x	x	x		x	x				x	
KD		x		x				x	x	x	x	x	x			x	x		x			x	
BS	x	x					x	x	x	x			x				x		x			x	
SD		x		x			x	x	x	x	x		x	x		x	x	x	x	x		x	
Lf		x		x			x	x	x	x		x	x			x				x		x	x
Th		x		x				x	x	x			x			x			x	x			
SW + WH										x	x		x	x	x							x	
MB + KD		x		x				x	x	x	x	x	x			x	x					x	
BS + SD		x					x	x	x	x			x				x		x			x	
Lf + Th		x		x				x	x	x			x			x				x			
All Actions		x						x	x	x		x	x			x	x			x		x	

Figure 10.6: Presence of Acceleration selected features

Posture/ Movement:	Postural Changes (standard deviation)																									
Direction:	Lateral				Sagittal						ThreeD								Vertical Length						Vertical Rotation	
Subdirection:	Lateral				Backward			Forward			ThreeD								Flexion				Vertical		LRRotation	
Modality:	Arms	Head	Lower Body	Torso	Arms	Lower Body	Torso Head	Arms	Lower Body	Torso Head	Arms						Lower Body	Lower Body	Arms	Head	Lower Body	Torso	Arms	Lower Body	Head	Torso
SubModality:											Elbows. Hips	Hands. Hips	Lelbow. Relbow	Lhand. Head	Lhand. RHand	Rhand. Head	Feet. Hips	Lfoot. Rfoot	Elbows		Knees					
SW						x	x				x					x		x					x	x		x
MB				x			x			x						x				x		x	x	x	x	
WH						x	x		x		x	x	x				x	x			x		x	x	x	x
KD							x	x					x	x	x	x			x			x	x			
BS							x		x	x											x				x	
SD							x				x	x	x	x								x	x		x	
Lf	x						x							x			x						x		x	x
Th							x									x	x				x		x			
SW + WH						x	x				x							x					x	x		x
MB + KD							x									x						x	x			
BS + SD							x																		x	
Lf + Th							x										x						x			
All Actions						x	x				x						x				x		x	x	x	x

Figure 10.7: Presence of Postural changes selected features

Posture/Movement:		Posture																										SymOcc
Direction:		Lateral					Sagittal					ThreeD								Vertical Length							Vertical Rotation	ThreeD
Subdirection:		Inside		Lateral	Outside		Backward		Forward											Down ward	Downward. Flexion			Upward		Upward. Flexion		
Modality:		Arms	Lower Body	Head	Arms	Lower Body	Lower Body	Torso Head	Arms	Lower Body	Torso Head	Arms						Lower Body		Lower Body	Head	Lower Body	Torso	Arms	Lower Body	Head	Head	Arms
SubModality:												Elbows. Hips	Hands. Hips	Lelbow. Relbow	Lhand. Head	Lhand. RHand	Rhand. Head	Feet. Hips	Lfoot. Rfoot			Knees						Lelbow. Relbow
Anxiety	SW + WH		32		30		13	16	25	12	1	35		36		15			10		11		19			17		9
	MB + KD					7		18	27		1	16			22		17	2	28		4	5	3		30			
	BS + SD	7	1			2		5			14	19	8	17				24	6	3	4	16	18	22		10	12	
	Lf + Th		13		34			2			12			22		27		16	8	28	10	9	6	1			19	
	All Acti	44	5	38	30	6	29	2	39	19	1	47	36	21	37	22		10	23	11	3	13	8	34		12	26	25
Pride	SW + WH		14		29		27	7	16	19	1	30		26		24			32		2		4			9		5
	MB + KD					10		3	7		2	34			33		22	13	6		9	27	1		32			
	BS + SD	23	7			10		1			2	15	35	16				27	18	11	12	29	9	8		22	5	
	Lf + Th		13		22			8			1			21		7		19	14	30	25	31	3	2			6	
	All Acti	59	7	56	46	10	43	3	24	37	1	54	62	49	30	32		52	21	35	6	29	2	16		19	5	20
Joy	SW + WH		37		30		17	7	29	26	1	21		25		35			11		19		34			8		32
	MB + KD					14		33	36		13	30			23		31	38	25		12	7	4		9			
	BS + SD	24	15			29		1			5	7	6	18				30	27	17	16	32	9	19		10	3	
	Lf + Th		32		21			6			23			20		29		34	28	26	33	18	5	15			24	
	All Acti	60	50	62	23	40	17	3	44	21	1	46	56	35	59	53		51	33	36	19	25	8	37		18	48	49
Sadness	SW + WH		7		19		12	16	37	8	5	25		21		24			15		4		29			27		18
	MB + KD					17		23	29		6	19			25		22	9	24		3	16	15		27			
	BS + SD	2	11			22		13			6	35	5	19				34	17	21	1	31	4	30		12	9	
	Lf + Th		20		14			32			28			16		7		15	4	23	8	5	27	29			26	
	All Acti	33	10	57	15	23	43	14	59	48	4	51	60	13	44	21		35	31	36	1	34	29	54		11	9	22
Panic Fear	SW + WH		25		36		34	24	26	5	8	31		32		37			18		9		16			20		22
	MB + KD					30		14	36		13	29			24		28	3	35		10	7	1		25			
	BS + SD	8	3			2		6			12	30	10	17				24	1	32	7	38	13	22		4	33	
	Lf + Th		32		34			7			3			29		18		30	16	27	2	15	1	4			31	
	All Acti	53	28	58	60	21	30	7	36	18	2	50	57	42	35	48		23	51	26	10	20	5	32		33	37	55
Shame	SW + WH		18		22		8	27	24	19	6	21		20		29			25		2		11			13		15
	MB + KD					5		6	22		1	25			23		15	2	19		17	20	7		35			
	BS + SD	3	1			2		5			7	10	6	13				32	22	19	21	12	23	18		28	36	
	Lf + Th		2		20			24			10			18		11		12	1	31	16	21	5	4			30	
	All Acti	36	4	16	21	3	30	7	32	38	2	39	22	13	49	59		27	18	47	8	46	12	56		14	33	54
Anger	SW + WH		35		34		25	23	36	6	1	32		26		29			2		19		28			21		30
	MB + KD					32		31	35		14	29			34		36	13	30		9	37	17		25			
	BS + SD	35	9			15		24			25	11	23	17				22	18	31	27	20	28	19		33	32	
	Lf + Th		33		29			20			10			19		9		17	23	32	30	25	21	13			22	
	All Acti	62	40	61	56	24	38	29	60	17	7	26	57	32	59	48		19	23	34	16	21	25	44		43	54	58
Neutral	SW + WH		29		30		21	25	27	24	1	20		32		15			22		9		31			13		10
	MB + KD					30		15	32		1	34			11		2	3	35		6	19	13		38			
	BS + SD	2	1			20		5			6	36	16	28				38	30	15	3	29	23	12		21	9	
	Lf + Th		13		33			14			1			34		10		23	12	22	9	2	16	21			18	
	All Acti	33	6	29	54	37	42	4	56	30	1	58	43	51	28	35		26	41	50	2	5	20	10		9	18	27

Figure 10.8: Ranking of postural features for each emotion

10

Posture/Movement:		Postural Changes (standard deviation)											
Direction:		Sagittal		ThreeD				Vertical Length				Vertical Rotation	
Subdirection:		Backward						Flexion		Vertical			
Modality:		Lower Body	Torso Head	Arms		Lower Body		Lower Body	Torso	Arms	Lower Body	Head	Torso
SubModality:				Elbows. Hips	Rhand. Head	Feet. Hips	Lfoot. Rfoot	Knees					
Anxiety	SW + WH	14	21	34			18			37	27		6
	MB + KD		19		33				35	11			
	BS + SD		11									15	
	Lf + Th		5			17				4			
	All Acti	52	4	58		40		51		35	54	28	31
Pride	SW + WH	13	12	23			33			37	28		3
	MB + KD		4		31				5	23			
	BS + SD		6									4	
	Lf + Th		10			24				5			
	All Acti	42	4	39		51		53		23	50	11	13
Joy	SW + WH	18	3	10			24			27	4		22
	MB + KD		37		27				17	15			
	BS + SD		4									14	
	Lf + Th		2			22				14			
	All Acti	26	2	32		43		45		29	13	31	28
Sadness	SW + WH	13	22	32			9			34	11		6
	MB + KD		28		30				37	26			
	BS + SD		27									7	
	Lf + Th		33			12				31			
	All Acti	56	24	52		41		47		61	46	12	20
Panic Fear	SW + WH	35	29	27			28			33	14		15
	MB + KD		27		23				38	11			
	BS + SD		9									21	
	Lf + Th		12			25				5			
	All Acti	46	12	47		34		39		25	45	27	49
Shame	SW + WH	4	26	23			31			17	16		10
	MB + KD		11		32				37	26			
	BS + SD		29									37	
	Lf + Th		28			26				7			
	All Acti	43	9	60		58		61		53	62	50	41
Anger	SW + WH	20	31	27			8			37	9		22
	MB + KD		33		28				38	27			
	BS + SD		36									34	
	Lf + Th		28			34				11			
	All Acti	45	55	39		49		41		36	31	47	51
Neutral	SW + WH	12	23	8			34			4	14		3
	MB + KD		27		17				8	18			
	BS + SD		17									19	
	Lf + Th		19			3				26			
	All Acti	21	7	38		55		47		17	48	46	24

Figure 10.9: Ranking of postural changes features for each emotion

Posture/Movement:		Speed																		
Direction:		Lateral				Sagittal			ThreeD							Vertical Length				Vertical Rotation
Subdirection:																Flexion			Vertical	
Modality:		Arms	Head	Lower Body	Torso	Arms	Lower Body	Torso Head	Arms					Lower Body		Arms	Lower Body	Torso	Arms	Head
SubModality:									Elbows. Hips	Hands. Hips	Lelbow. Relbow	Lhand. RHand	Rhand. Head	Feet. Hips	Lfoot. Rfoot	Elbows	Knees			
Anxiety	SW + WH			33		23	5		20	31	24			4	2		3			
	MB + KD	8			12				26	24	23	36	38			37		10	25	
	BS + SD				36		35	31	21	27	33			32				38	34	9
	Lf + Th				21		26	33	25	32	20		30					31	3	
	All Acti	24	56		53	50	46	45	9	14	16	59	62	27	41	61	32	57	48	17
Pride	SW + WH			21		17	8		20	25	11			10	6		22			
	MB + KD	21			12				17	20	11	18	37			35		8	26	
	BS + SD				21		28	24	13	20	32			19				36	38	3
	Lf + Th				26		34	12	20	27	23		16					28	4	
	All Acti	25	60		33	48	36	40	12	18	15	34	45	44	38	47	58	55	31	8
Joy	SW + WH			20		36	13		6	14	5			28	9		31			
	MB + KD	8			1				28	32	11	21	29			24		2	34	
	BS + SD				31		12	11	22	20	26			13				28	38	2
	Lf + Th				7		17	1	16	19	12		31					3	9	
	All Acti	12	30		22	52	39	24	7	9	5	15	47	57	34	54	41	16	38	11
Sadness	SW + WH			26		36	3		14	31	30			1	2		17			
	MB + KD	11			8				7	1	2	4	35			31		10	21	
	BS + SD				37		20	16	14	28	26			15				33	38	3
	Lf + Th				3		19	9	6	1	13		21					24	2	
	All Acti	30	49		26	40	27	19	5	3	6	18	37	17	25	55	39	28	45	2
Panic Fear	SW + WH			11		30	2		10	21	17			3	1		13			
	MB + KD	33			12				19	18	17	22	32			37		20	31	
	BS + SD				15		18	29	16	19	25			11				23	28	14
	Lf + Th				17		14	8	9	10	19		24					28	22	
	All Acti	56	29		24	59	13	38	6	14	15	43	40	16	17	62	22	44	54	8
Shame	SW + WH			28		9	3		12	14	1			34	7		33			
	MB + KD	9			14				24	34	3	28	36			12		13	27	
	BS + SD				34		33	9	11	8	24			26				38	20	31
	Lf + Th				33		17	29	13	15	8		14					32	3	
	All Acti	15	37		52	24	35	25	6	11	1	44	48	42	40	26	57	45	28	23
Anger	SW + WH			14		12	10		5	7	11			15	4		13			
	MB + KD	26			8				19	24	12	7	23			4		21	15	
	BS + SD				30		38	29	3	14	10			37				26	21	12
	Lf + Th				15		24	27	14	31	18		6					26	4	
	All Acti	33	35		18	28	50	37	6	22	12	30	10	53	46	5	52	42	13	20
Neutral	SW + WH			37		36	2		11	17	6			26	5		33			
	MB + KD	29			12				33	22	31	37	36			28		14	26	
	BS + SD				31		37	32	22	13	35			18				34	26	4
	Lf + Th				17		30	15	25	32	24		27					28	7	
	All Acti	40	53		49	61	14	32	11	13	15	39	52	36	22	62	44	57	34	12

Figure 10.10: Ranking of speed features for each emotion

Posture/Movement:		Acceleration															
Direction:		Lateral		Sagittal	ThreeD								Vertical Length				Vertical Rotation
Subdirection:													Flexion			Vertical	
Modality:		Head	Torso	Torso Head	Arms						Lower Body		Arms	Head	Torso	Arms	Head
SubModality:					Elbows. Hips	Hands. Hips	Lelbow. Relbow	Lhand. Head	Lhand. RHand	Rhand. Head	Feet. Hips	Lfoot. Rfoot	Elbows				
Anxiety	SW + WH						26	7		22	29	8					28
	MB + KD	14	21		13	15	6	32	29	20			31	34			9
	BS + SD	29		23	25	26	28			13				30	37		20
	Lf + Th	7	11		15	29	24			23			14			18	
	All Acti	18			20	42	15		43	33			49	55		60	7
Pride	SW + WH						18	31		35	34	15					36
	MB + KD	25	28		19	14	15	38	24	30			29	36			16
	BS + SD	31		30	25	26	14			34				33	37		17
	Lf + Th	33	29		11	15	18			9			32			17	
	All Acti	27			14	22	9		41	26			61	28		57	17
Joy	SW + WH						2	12		23	33	16					15
	MB + KD	6	16		20	19	5	10	26	18			35	22			3
	BS + SD	34		8	21	25	35			36				33	37		23
	Lf + Th	4	11		8	25	13			10			30			27	
	All Acti	10			20	42	4		55	27			61	14		58	6
Sadness	SW + WH						23	35		33	10	20					28
	MB + KD	14	18		13	20	12	34	32	38			33	36			5
	BS + SD	36		23	25	32	10			18				24	29		8
	Lf + Th	10	25		17	30	22			18			34			11	
	All Acti	32			16	42	7		53	38			62	58		50	8
Panic Fear	SW + WH						7	19		23	6	4					12
	MB + KD	4	21		9	5	2	16	8	26			34	15			6
	BS + SD	26		35	20	37	5			36				31	34		27
	Lf + Th	23	6		11	21	26			13			20			33	
	All Acti	11			3	9	4		41	19			52	31		61	1
Shame	SW + WH						5	32		37	35	30					36
	MB + KD	18	33		8	10	4	29	21	31			16	38			30
	BS + SD	30		25	17	14	4			27				15	35		16
	Lf + Th	34	25		19	23	27			9			22			6	
	All Acti	20			19	17	5		29	51			31	55		34	10
Anger	SW + WH						3	33		16	24	18					17
	MB + KD	10	11		2	1	3	18	6	22			5	20			16
	BS + SD	8		16	4	6	1			5				7	13		2
	Lf + Th	16	12		3	2	7			1			5			8	
	All Acti	14			2	3	1		11	4			8	27		15	9
Neutral	SW + WH						7	35		18	16	28					19
	MB + KD	9	25		10	4	5	23	7	21			16	20			24
	BS + SD	24		7	14	25	8			27				11	33		10
	Lf + Th	6	8		5	20	11			4			31			29	
	All Acti	16			19	23	3		25	31			60	45		59	8

Figure 10.11: Ranking of acceleration features for each emotion

11 Motion capture characterization of emotional body expression

Using Random Forest (RF) approach, we were able to select a subset of relevant features for the classification of emotions. Based on the obtained subset of relevant features, we aim to provide deeper insights on how each expressed emotion is characterized based on motion capture data. Ideally, the interpretation of expressed emotions characterization would be achieved through the interpretation of RF approach used previously for the classification of expressed emotions. However, the interpretation of RF model is a very complex task due to its "black-box" nature. Although we discussed the predictive power of the most relevant features for each emotion across different actions in chapter 10, we need to go further to explore how each relevant feature contributes to the characterization of emotional body expression.

In this chapter, we use the subset of selected features obtained in chapter 10 to explore the motion capture characterization of emotional body expressions using two approaches. The first approach consists in exploring the patterns of emotion expression through the mean of each feature value. This process allows us comparing the characterization of emotional body expression performed during the perceptual study (See Chapter 8) with their motion capture characterization. It also allows us comparing the expression of emotions in each action. We call this process motion capture characterization of emotional body expression. It is achieved through a graphical representation of emotion expression patterns. The second approach consists in driving a set of rules that characterize motion captured expressive movement in Emilya database. This is achieved by building a decision tree model based on the subset of selected features provided in Chapter 10.

This chapter is organized as follows: Firstly, we present the patterns of motion capture characterization of expressed emotion in secion 11.1. In this section, we discuss the characterization of expressed emotions across all the actions, we compare them with perceptual characterization and we discuss the characterization of expressed emotions in each action, finally we provide a summary. Secondly, we present the decision rules derived from Decision Tree models to characterize expressed emotions in section 11.2. In this section, we describe the process of tree pruning, then we interpret the rules derived from pruned trees. Finally, we conclude and summarize this chapter in section 11.3 and 11.4.

11.1 Patterns of motion capture characterization of Emilya database

Firstly, we compare the characterization of expressed emotions across all the actions. This comparison is useful to explore the differences of characterization patterns of different expressed emotions. Secondly, we compare their characterizations in different actions. The latter comparison is useful to explore individually the characterization of each expressed emotion and the impact of action on its characterization pattern. In the following of this chapter, features names are represented as follows: Posture (Post)/ Standard Deviation (STD)/Movement Dynamics (Speed/Acceleration/Correlation) _ Direction _ Body segment. For instance the downward flexion of the head is described as Post_ Downward.Flexion_ Head. The ' outliers" described and removed for feature selection process in Chapter 10 are also not used for the characterization of emotional body expression.

11.1.1 Patterns of emotion characterization cross all the actions

Figure 11.1 represents the patterns of motion capture characterization of all the expressed emotions across all the actions. In this figure, the patterns of characterization are based on the subset of selected features (FS-SSF) deduced from Feature Selection approach as explained in section 10.4.2 in Chapter 10. Each pattern is drawn in different color for each emotion. In our body movement notation system (See Chapter 6), we distinguish posture features, postural changes features and movement dynamics features. To clarify the representation, we split the patterns into three figures according to the type of features; Posture features, Postural changes features (standard deviation) and finally movement dynamics features (speed and acceleration) (See Figure 11.1). The patterns corresponds to the characterization of Anxiety (Ax), Pride (Pr), Joy (Jy), Sadness (Sd), Panic Fear (PF), Shame (Sh), Anger (Ag) and Neutral (Nt). Each pattern is defined through the value of features averaged across all the actions. The error bars indicate 95% confidence interval. All the patterns are normalized between 0 and 1 according to the maximal and minimal values of each feature averaged across all the actors.

Posture features :
Based on posture features, we observe high differences in the characterization patterns at the level of some features. For instance, the upward and downward flexion of the head receive significantly the highest values for Sadness and Shame patterns (See Figure 11.1). In fact, downward flexion is measured through positive rotation. Thus, the highest value of downward flexion reflects a high collapse of head posture. The upward flexion of the head is defined through a negative rotation measure (the more upward flexion occur, the more the measure is negative). Thus, the highest value of upward flexion (towards 0) reflects the smallest upward head movement. The expression of Pride (followed by the expression of Joy) receives the lowest amount of head downward flexion. Upward head flexion also occur significantly more in Pride

and Joy expressions than in Sadness and Shame expressions ($p<.001$).

In chapter 10 we showed that the downward head flexion is considered as the most relevant feature for Sadness expression across all the actions. Figure 11.1 shows that the highest mean rating of downward head flexion is attributed to Sadness expression ($p<.05$). As such, downward head flexion is considered as one of the most relevant features for Sadness expression across all the actions.

We also observe that the highest amount of posture openness is often attributed to the expression of Joy or Anger. For instance, the lateral openness of lower body (Post_ Outside_ LowerBody) and the 3D relative openness of elbows (Post_ ThreeD_ Lelbow.Relbow) receive the highest value for Anger expression, while the lateral openness of arms (Post_ Outside_ Arms) and the 3D openness of feet (Post_ ThreeD_ Lfoot.Rfoot) receive the highest values for Joy expression followed by Anger expression.

The lowest average value of posture openness/extension features are mostly observed for Sadness and Shame expressions. the lateral closeness of lower body (Post_ Outside_ LowerBody) has shown lower averaged value for Shame than for Sadness. In chapter 10, we indicated that the lateral closeness of lower body limbs is highly relevant for Shame classification. In chapter 9, we showed that Shame expression is mostly perceived as Sadness expression in human recognition of emotions. We also showed that Shame expression is slightly confused with Sadness expression in automatic classification (See chapter 9). Figure 11.1 also shows a high similarity between Shame and Sadness patterns except along two features: head downward flexion which receives the highest mean rating for Sadness, and the lateral openness of lower body limbs which receives the lowest mean rating for Shame expression. As such, it seems that the lateral openness of lower body limbs is one of the most relevant features that characterizes Shame expression and distinguishes it from Sadness in automatic classification of emotions.

Figure 11.1 shows that the highest mean rating of left/right vertical head rotation (Post_ LRRtation _ Head) is attributed to Pride expression ($p<.05$). In chapter 10, we found that the left/right vertical head rotation receives the highest relevance measure for Pride expression across all the actions. Thus, the left/right head rotation appears to be a frequent and an important characteristic of Pride expression.

Postural changes features :

The patterns of emotions characterization based on postural changes features are presented in Figure 11.1. Except the mean rating values of few features, we can observe two groups of patterns; the first group of patterns involve the characterization of Joy, Anger and Pride. The second group refers to the characterization of Sadness, Shame and Neutral (See Figure 11.1). Indeed, the patterns of Joy, Pride and Anger characterizations receive the highest values of postural changes features (See Figure 11.1). The patterns of Sadness, Neutral and Shame characterizations are overlapped, but they are well differentiated from the first group of patterns as they mostly receive the lowest averaged values of features. The patterns of Anxiety and Panic Fear characterization are mostly situated between these two groups.

Except Anxiety and Panic Fear characterization, the simplified graphical representation of emotions characterization patterns based on postural changes features provides somehow a discrimination between "high arousal" (Pride, Joy, Anger) and "low arousal" (Shame, Sadness, Neutral) expressions.

Speed/Acceleration features :
The last subfigure in Figure 11.1 represents the patterns of emotions characterization through the maximum values of speed/acceleration features. While the pattern of Joy characterization receives the highest averaged values of lower body movement speed, the pattern of Anger characterization receives the highest averaged values of upper body movement dynamics (speed and acceleration of arms, head and torso movements). As more upper body movement dynamics features are present in the subset of selected features, the pattern of Anger characterization seems highly differentiated.

Based on speed and acceleration features, the pattern of Panic Fear characterization turns out to be highly overlapped with the pattern of Joy characterization. This result was not observed for posture and postural changes based patterns. The pattern of Pride expression is highly similar to Joy expression pattern, but it is characterized with lower mean ratings.

Similarly to what we found in postural changes features based patterns, we can clearly dissociate two groups of patterns based on speed and acceleration features; the first group of patterns includes the characterization of "high arousal" emotions (Anger, Pride, Joy and Panic Fear) while the second group includes the characterization of "low arousal" emotions (Sadness, Shame and Neutral). Pride and Anxiety characterizations patterns are mainly situated between these two groups of patterns.

In chapter 10, we highlighted the importance of speed and acceleration features for the classification of Anger expression across different actions. As shown in Figure 11.1, Anger expression is mostly characterized with the highest mean ratings of speed and acceleration features ($p<.001$ for most of the features), in particular for the speed and the acceleration of arms movement. Thus, high speed and acceleration of movement are considered as highly relevant features for the characterization and the discrimination of Anger expression over the other emotions considered in the Emilya database.

11.1.2 Human vs motion capture patterns for the characterization of emotional body expression across all the actions

In order to compare motion capture and perceptual emotion characterization patterns, we focus on a reduced subset of selected features. This reduced SSF was introduced in Chapter 10 (section 10.4.1) and it consists of the intersection of SSF

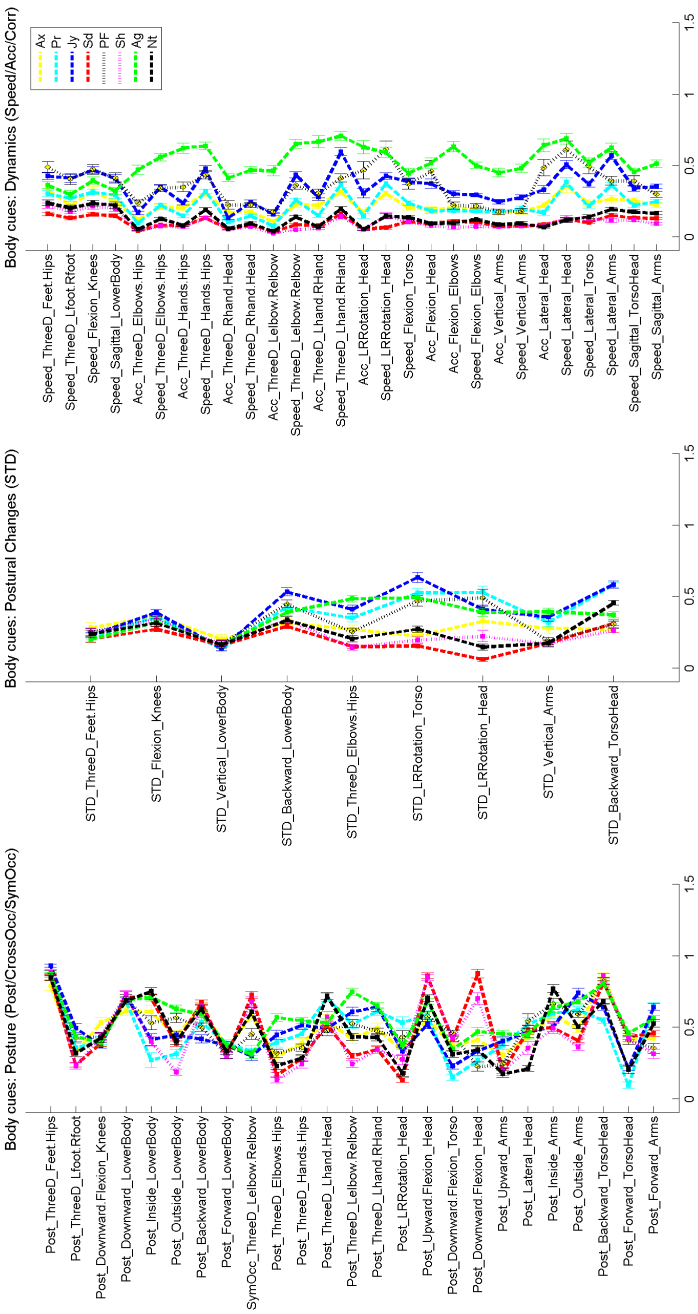


Figure 11.1: Motion capture characterization of expressed emotions across all the actions based on the FS-SSF obtained by our FS approach

across all the actions. It is called Int-SSF. This intersection leads to 11 features (See Figure 11.2b).

Using this Int-SSF of 11 features, the classification of emotions across all the actions has shown a significantly reduced CCR_{OOB} (around 68%) comparing to the CCR_{OOB} obtained using FS-SSF (62 features) (82%) (See Chapter 10). However, we choose to focus on the reduced subset of features to ease the comparison of patterns with the perceptual characterization of emotions achieved in Chapter 8. Figure 11.2 represents the patterns obtained from the motion capture characterization of emotions across all the actions (using the Int-SSF of 11 features) versus the patterns obtained from the perceptual characterization of emotions (using 8 perceptual body cues). The error bars indicate 95% confidence interval. The mean rating of motion capture characterization is normalized within the interval [0,1]. The mean rating of perceptual characterization ranges from 1 to 5.

In term of movement dynamics properties, no motion capture feature corresponds to the fluidity and the regularity of movement in our initial set of features, neither in the Int-SSF of 11 selected features (See Figure 11.2b). Thus, we are not able to compare the perceptual and the motion capture rating of fluidity and regularity features.

We now illustrate the comparison between perceptual and motion capture characterizations based on the patterns showed in Figure 11.2.

Movement dynamics cues :

While the perceptual rating of speed and acceleration features are rated according to the whole body movement, the speed and acceleration features of the motion capture SSF are specific to arms movement. However, we can observe a similar configuration of patterns; perceptual and motion capture mean ratings of speed and acceleration receive the highest values for Anger, Joy and Panic Fear expressions, followed by Pride/Anxiety, Neutral and finally by Sadness and Shame expressions. Besides, in both perceptual and motion capture characterizations patterns, the mean rating of movement Power and arms movements acceleration receive significantly the highest value for Anger expression. The motion capture mean rating of arms movement speed also receives significantly the highest value for Anger expression. However, no significant difference is found between the perceptual rating of speed in Anger, Joy and Panic Fear across all the actions (See Figure 11.2a). We note that more significant differences in perceptual mean rating of movement speed are shown for Anger expression in specific actions, Throwing in particular (See Figure H.19 in Chapter H).

Postural changes cues :

The perceptual rating of the quantity of movement concerns arms movement. However, the only motion capture feature that refers to postural changes in the reduced SSF is the standard deviation of backward torso and head movement (See Figure 11.2b).

The perceptual rating of arms movement quantity receives the highest value in

Joy and Anger expressions, followed by Panic Fear, Pride, Anxiety and Neutral. The motion capture rating of torso postural change receives the highest value in Joy and Pride expressions, followed by Neutral, Anger and Panic Fear.

Although the perceptual rating of movement quantity and the motion capture rating of postural change are not based on the same body segments, they do somehow show similarity at the level of patterns configuration. Indeed, in both perceptual and motion capture characterizations, the lowest ratings of postural changes features are attributed to Sadness and Shame expressions (and Anxiety in motion capture rating) while the highest value is attributed to Joy expressions.

Openness body cues :

The perceptual rating of the openness (the extension) of body limbs concerns the whole body posture. The openness of body posture in the reduced SSF is reduced to the distance between the feet (See Figure 11.2b).

Although not based on the same body segments, the perceptual rating and the motion capture rating of openness show a certain similarity at the level of patterns configuration. Indeed, in both perceptual and motion capture rating of openness, Joy expression receives the highest mean rating while Shame and Sadness receive the lowest mean rating. In perceptual characterization patterns, the mean rating of openness in Joy expression is followed by Pride and Anger expressions. In motion capture characterization patterns, the mean rating of openness in Joy expression is followed by Anger and Panic Fear.

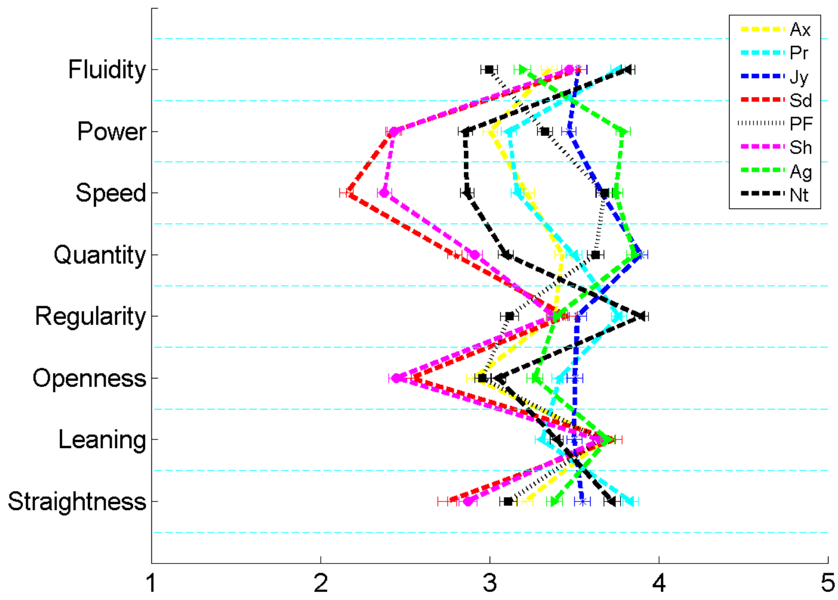
In both perceptual and motion capture ratings, we can cluster two groups of emotions based on the statistical difference of openness mean ratings ($p<.001$); the first group includes Joy, Anger and Panic Fear expressions and refers to the highest openness mean ratings, the second group includes Shame and Sadness expressions and refers to the lowest openness mean ratings.

Trunk leaning and flexion cues :

The perceptual rating of forward/backward leaning mainly concerns the trunk of the body (torso and head). Thus, it can be directly compared to the motion capture characterization of forward/backward leaning of body trunk (torso and head). A small perceptual rating of leaning reflects a backward leaning, while a high perceptual rating reflects a forward leaning. In motion capture characterization, two features are considered to describe forward and backward leaning. As motion capture backward leaning is defined through negative rotation values, a high mean rating (toward 0) reflects a straight posture, while a low (negative) mean rating reflects a high backward leaning. As motion capture forward leaning is defined through positive rotation values, a high (positive) mean rating reflects a high forward leaning, while a low (toward 0) mean rating reflects a straight posture.

In both perceptual and motion capture characterization of trunk leaning, we observe that the cluster of emotions composed of Joy, Pride and Neutral receives significantly ($p<.001$) different mean ratings from other emotions. This cluster is

(a) Perceptual characterization of expressed emotions across all the actions



(b) Motion capture characterization of expressed emotions across all the actions based on the intersection of selected features

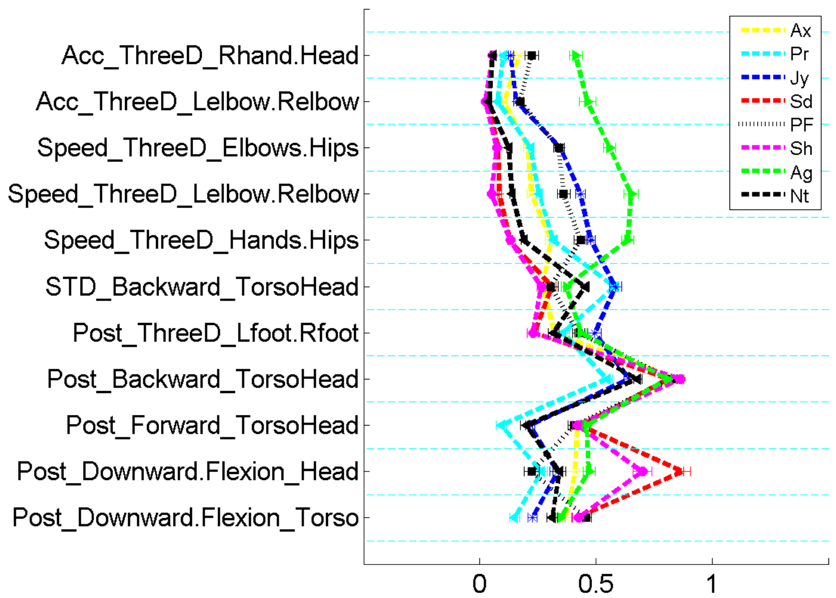


Figure 11.2: Perceptual and motion capture characterization of expressed emotions across all the actions.

characterized by the highest backward leaning and the smallest forward leaning. As Joy and Pride are the only "positive" emotions considered in our database, this result suggests that the forward/backward leaning of the trunk play an important role in discriminating "positive" and "negative" emotional body expressions in the Emilya database.

In motion capture characterization, the mean rating of torso downward flexion is significantly lower for Pride and Joy expressions than for the other emotions. In perceptual characterization, the highest mean rating of straightness is attributed for Pride expression, followed by Neutral and Joy expressions. However, both Pride and Panic Fear expressions receive the lowest mean rating of head downward flexion.

Summary : Although not directly compared, we report high similarity between perceptual and motion capture characterization patterns across all the actions. For instance, the power of body movement receives the highest perceptual mean rating for Anger expression. Similarly, the motion capture features describing movement acceleration receive the highest value for Anger expression. The lowest mean ratings of the quantity of movement and the postural changes of torso movement are attributed to Sadness and Shame, while the highest mean ratings are attributed to Joy expression. A similar result is found for the mean rating of body openness features; the highest values are attributed to Joy expression, the lowest values are observed for Sadness and Shame expressions. The highest mean rating of posture straightness is attributed to Pride, Joy and Neutral expressions in both perceptual and motion capture characterization patterns.

In the following subsection, we compare perceptual and motion capture characterization patterns across different actions.

11.1.3 Emotion characterization for each action

In the previous sections, we discuss and compare the pattern of expressed emotions characterization across all the actions. In this section, we focus on each expressed emotion characterization and explore this characterization for each action. To compare motion capture and perceptual ratings we use the reduced Int-SSF of 11 selected features, which represent the common features selected across all the actions (See section 10.4.1, Chapter 10).

Figure 11.3, 11.4,11.5, 11.6,11.7, 11.8, 11.9, 11.10 illustrate the mean ratings of perceptual and motion capture characterization of each expressed emotion for each action. The error bars indicate 95% confidence interval. The mean rating of motion capture characterization is normalized within the interval [0,1]. The mean rating of perceptual characterization ranges from 1 to 5. The actions are 1) Simple Walking (SW), 2) Moving Books (MB), 3) Walking with an object in hands (WH), 4) Knocking at the door (KD), 5) Being Seated (BS), 6) Sitting Down (SD), 7) Lifting (Lf), and 8) Throwing (Th).

We proceed to one-way ANOVA to explore the statistical difference in rating a given feature in a given expressed emotion across various actions. We found that actions have mostly a significant effect on motion capture features rating for each emotion (mostly p<.001). That is, there is a significant difference in mean ratings of each motion capture feature for each emotion in different actions. An exception is found for the rating of the speed of relative elbows extension (Speed_ ThreeD_ Lelbow.Relbow) in Panic Fear expression (See Figure 11.4b). We also found a significant difference in perceptual ratings of each of each body cue for each emotion in different actions. Few exceptions are reported. For instance, no significant difference is found for the perceptual rating of Speed and Power of Shame expression across different actions (See Figure 11.8a). Besides, no significant effect of action is found for the perceptual rating of Fluidity in Pride, Joy and Neutral expression (See Figures 11.5a, 11.6a 11.10a).

We now focus on the comparison between motion capture and perceptual ratings across different actions. We discuss the results for each emotion: Anxiety, Panic Fear, Pride, Joy, Sadness, Shame, Anger and Neutral.

Anxiety :

Figure 11.3 represents the perceptual and the motion capture characterization of Anxiety in each action. Across all the actions, the perceptual characterization of Anxiety appears to be neither strong, smooth, and fast nor light, jerky and slow. The quantity of arms movement is higher in arms movement actions (Moving Books, Lifting and Throwing) than in the other actions. In motion capture ratings, Anxiety expression mostly receives the lowest mean ratings of the speed and acceleration of arms movement, except in Knocking (KD) and Throwing actions (Th).

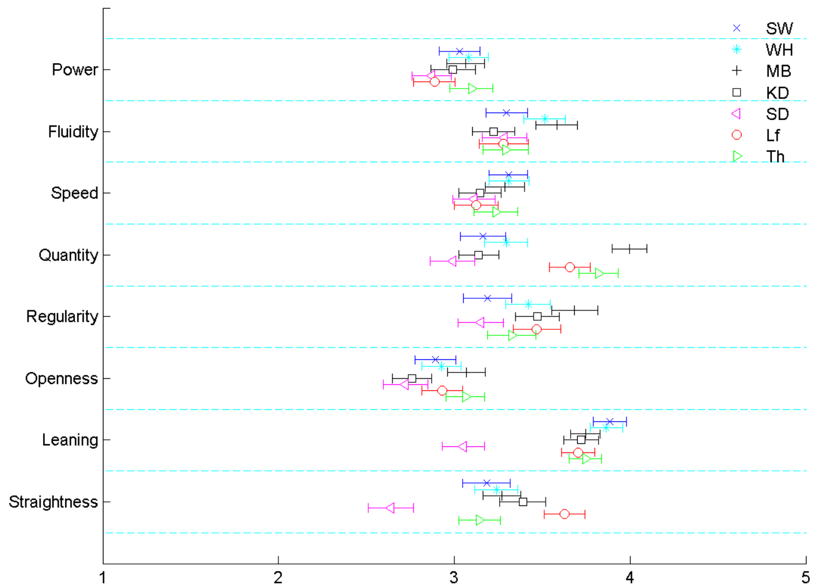
11

In perceptual mean rating, Anxiety expression is characterized with slightly contracted posture (the mean rating of Openness is slightly lower than the medium value). Based on the motion capture mean rating, Anxiety expression is also characterized with contracted openness of feet, except in Walking actions (SW and WH) where the distance between the feet is higher due to the walking pattern. Anxiety expression is characterized by forward posture in both perceptual and motion capture ratings. The torso is rated as collapsed in particular in Sitting Down action.

Panic Fear :

The perceptual and the motion capture characterization of Panic Fear for each action is illustrated in Figure 11.4. In perceptual rating, Panic Fear expression is characterized as neither smooth nor jerky movement. The acceleration and the speed of movement in Panic Fear expression seem depending on the performed action. For instance, in perceptual ratings, Panic Fear expression is characterized with stronger movement in Knocking action than in walking action. Similarly in motion capture characterization, the acceleration of relative right hand movement (Rhand.Head) is significantly higher in Knocking action than in the other actions (p<.001). Panic Fear expression is also perceptually rated as having high speed, in particular in the actions Moving Books (MB), Knocking (KD) and Lifting (Lf). However, the motion capture rating of speed features did not receive significantly higher rating in

(a) Perceptual characterization of Anxiety for each action



(b) Motion capture characterization of Anxiety for each action based on the intersection of selected features across all the actions

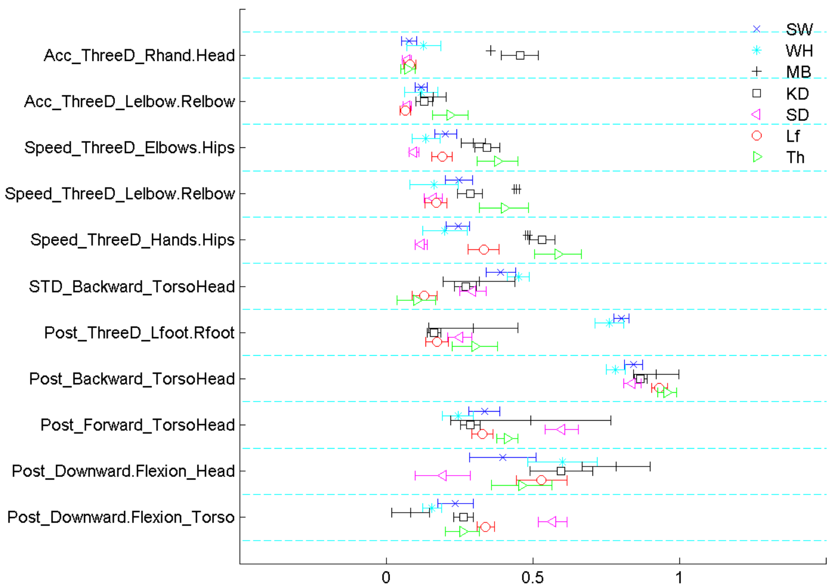


Figure 11.3: Perceptual and motion capture characterization of Anxiety for each action.

Lifting action. Taking a deeper look at the motion capture characterization of Panic Fear in Lifting action (See Figure H.12, Chapter H), we found that fast movement is mainly observed for head movement. This feature is not considered among the reduced subset of selected features (Int-SSF) as only arms movement speed features are present among this subset of 11 features (See Figure 11.4).

Panic Fear expression mostly receives high perceptual rating of arms movement quantity, in particular in arms movement actions: Moving Books (MB), Lifting (Lf), Throwing (Th) and Knocking (KD).

In perceptual ratings, the openness feature in Panic Fear expression is rated as neither expanded nor contracted across all the actions. In motion capture ratings, the openness of lower body limbs receives significantly higher value in walking actions.

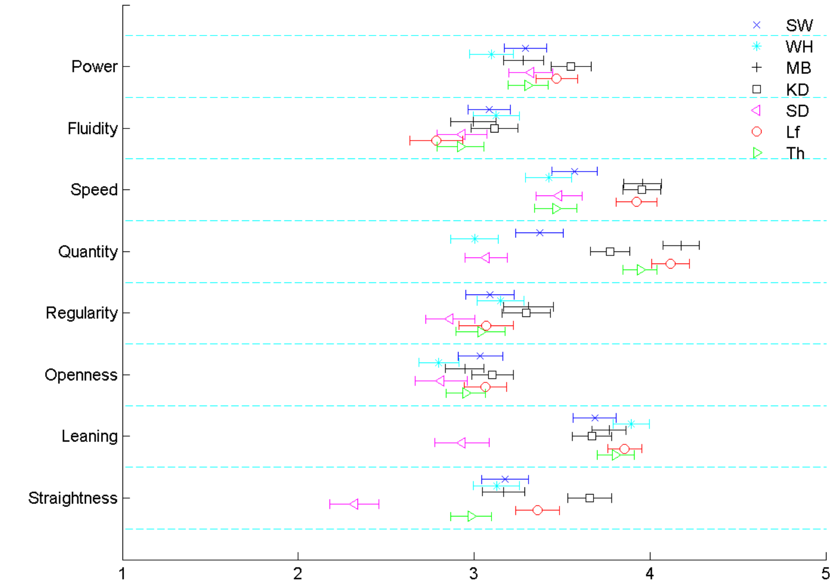
More differences between the actions are observed regarding the perceptual ratings of Leaning and Straightness. Indeed, the leaning of the torso is perceptually rated as neither forward nor backward posture in Sitting Down action, but with high amount of forward leaning in the other actions. In motion capture ratings and in perceptual rating, the torso is considered as significantly more collapsed in Sitting Down action.

Pride :

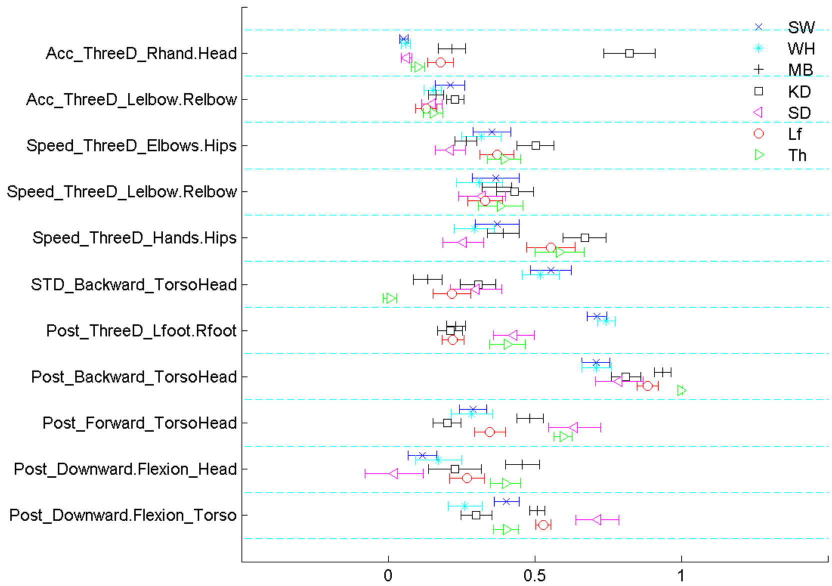
Figure 11.5 represents the perceptual and the motion capture characterization of Pride expression. In perceptual rating, body movement of Pride expression is rated as neither jerky nor light, slightly fast and highly smooth across all the actions (See Figure 11.5a).

In motion capture characterization, hands movements of Pride expression are mostly rated as significantly faster in Knocking action. In both perceptual and motion capture ratings, the openness of body posture receives the highest mean rating in walking actions (SW and WH). Body movements of Pride expression are mostly perceived as highly regular and with high quantity of arms movement except in Sitting Down which receives significantly lower mean rating of arms movement quantity ($p<.001$).

In motion capture characterization, Pride expression in Sitting Down (SD) action receives the highest mean rating of postural changes of backward torso movement. In both perceptual and motion capture ratings, Pride expression is characterized with the highest backward leaning of the torso in Sitting Down action. However, a forward torso posture with a large error bar is also observed in motion capture characterization. This result is explained as we measure the mean of torso posture across Sitting Down action and Being Seated action in the motion capture characterization. Taking a deeper look at the motion capture characterization in respectively Sitting Down and Being Seated actions (See Figure H.9 and Figure H.11), we observe that forward torso posture mainly occurs in Sitting Down action, while backward torso posture mainly occurs in Being Seated action. Forward behavior of the torso during



(a) Perceptual characterization of Fear for each action



(b) Motion capture characterization of Fear for each action based on the intersection of selected features across all the actions

Figure 11.4: Perceptual and motion capture characterization of Fear for each action.

Sitting Down action also affects the perceptual mean rating of straightness (Figure 11.5a) and the motion capture mean rating of torso flexion (Figure 11.5b). Pride expression is characterized with high straightness across the other actions in both perceptual and motion capture characterization.

In Lifting, Throwing, and Moving Books actions, Pride expression is characterized by head downward flexion. Viewing the video of the actors, we observed that the occurrence of head downward flexion - during Pride expression in these actions - is mainly due to the staring at the object hold in hands. During Pride expression, the gaze of the actors tend to focus on the books and on the piece of paper respectively in Moving Books, Lifting and Throwing actions.

Joy :

Figure 11.6 depicts the perceptual and motion capture characterizations of Joy expression in each action. The perceptual ratings of Joy expression shows high mean ratings of Power (strong), Fluidity (smooth), Speed (fast), Regularity and arms movement Quantity. The latter receives significantly high mean rating particularly in Moving Books, Lifting and Throwing actions. In motion capture rating, Joy is characterized with low acceleration of arms movement except in Knocking action with receives significantly higher mean rating ($p<.001$). However, the speed of arms movement receives higher mean rating than the acceleration of arms movement in motion capture ratings, which is congruent with the perceptual ratings.

The openness of the whole body in perceptual ratings along with the openness of feet in motion capture ratings receive the highest mean ratings in walking actions (SW and WH).

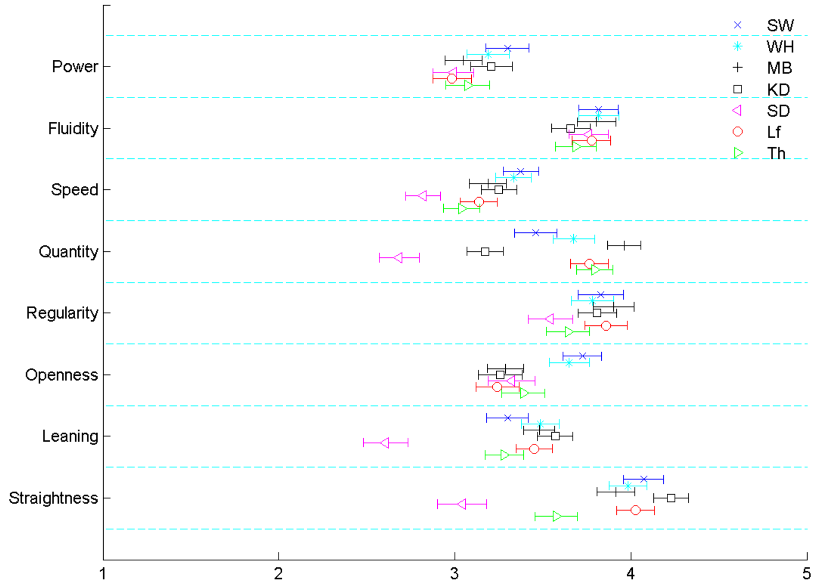
Similarly to what has been found in Pride expression, it seems that Joy expression is characterized with high backward leaning in Being Seated action and high forward leaning in Sitting Down actions. As in Pride expression characterization, we also found that the downward flexion of the head during Joy expression is affected by “the staring at object” behavior in Moving Books, Lifting and Throwing actions.

Sadness :

Figure 11.7 contains the perceptual and the motion capture characterization of Sadness expression. The perceptual ratings of movement dynamics of Sadness expression is mostly homogeneous across all the actions. It is described through light (low Power), smooth (High Fluidity), regular (high Regularity) and slow (low Speed) movement. Similarly, motion capture ratings of Sadness expression show low mean ratings of arms movement speed and acceleration, but the acceleration of the hands movement seems significantly higher in Knocking action.

In perceptual ratings of Sadness expression, the quantity of arms movement is more frequent in actions involving arms movement (Knocking, Moving Books, Lifting and Throwing) than in Sitting Down (SD) and Walking actions (SW and

(a) Perceptual characterization of Pride for each action



(b) Motion capture characterization of Pride for each action based on the intersection of selected features across all the actions

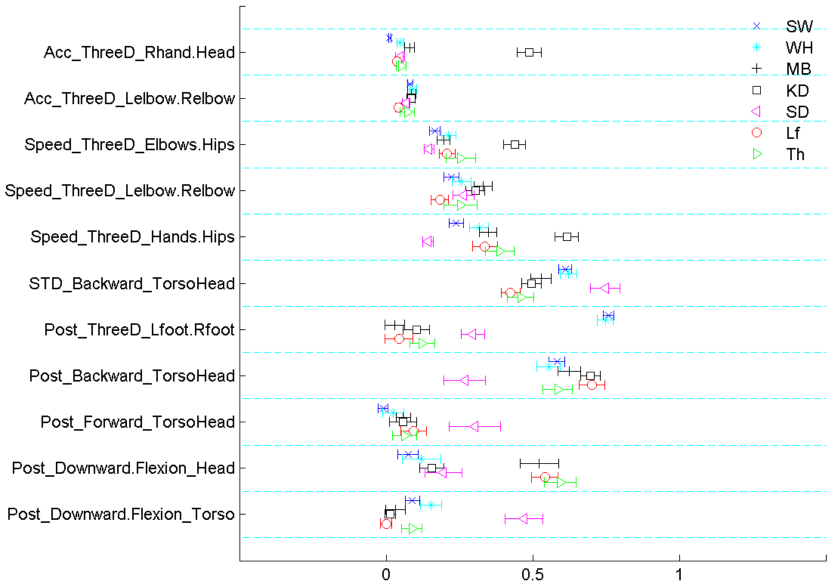
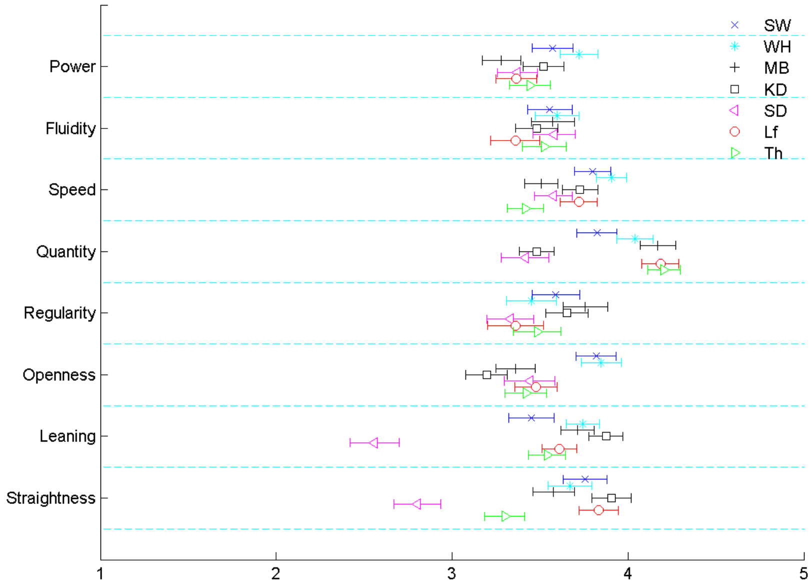


Figure 11.5: Perceptual and motion capture characterization of Pride for each action.

(a) Perceptual characterization of Joy for each action



(b) Motion capture characterization of Joy for each action based on the intersection of selected features across all the actions

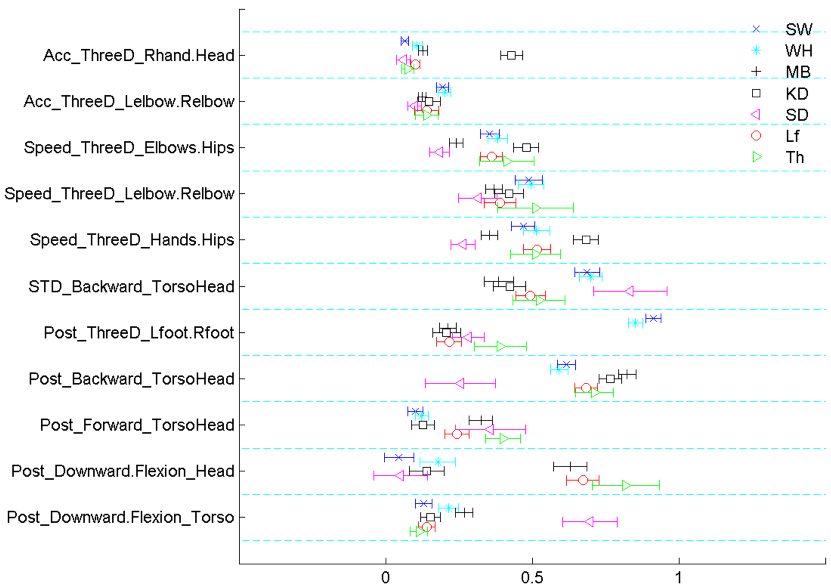


Figure 11.6: Perceptual and motion capture characterization of Joy for each action.

WH). In motion capture ratings, Sadness expression is characterized by low mean rating of the postural changes of backward torso movement across all the actions.

Sadness expression is mostly characterized with low mean rating of Openness in perceptual ratings for all the actions. This result is also observed in motion capture mean rating of feet openness except in Walking actions which receives significantly higher mean rating due to the walking patterns.

While Sadness expression is mostly characterized with forward torso leaning in the perceptual ratings, this result is particularly observed in Sitting Down action with motion capture ratings. The downward flexion of the torso receives significantly higher mean rating in Sitting Down action in both perceptual and motion capture ratings. The downward flexion of the head also receives the highest mean rating across all the actions. Unlike Pride and Joy expressions, no backward torso leaning is observed for Sadness expression.

Shame :

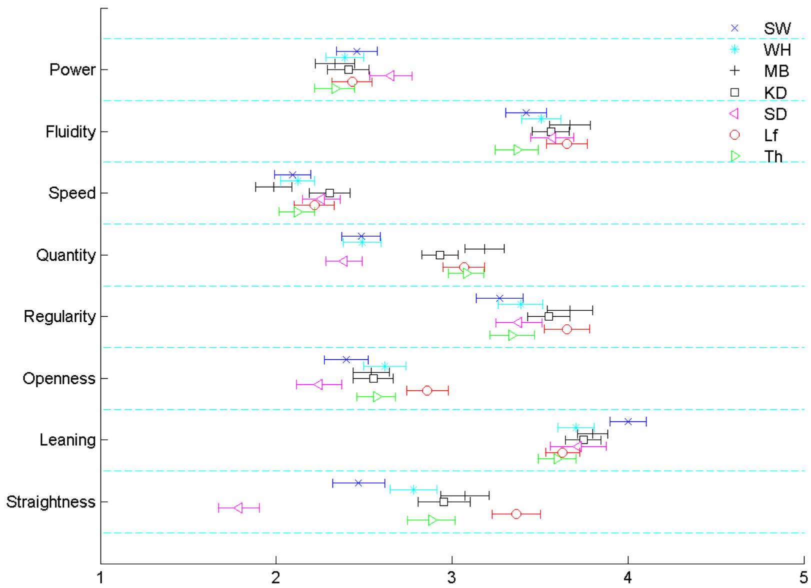
Figure 11.8 depicts the perceptual and the motion capture characterization of Shame expression. We showed in previous sections that, across all the actions, the pattern of Shame characterization appears to be highly similar to the pattern of Sadness expression. Comparing Figure 11.8 and Figure 11.7, we also observe that Shame and Sadness characterization show highly similar mean ratings for each action. The differences mainly occur at the level of motion capture ratings of head and torso flexions. Shame expression is characterized by lower downward flexion in Sitting Down and Throwing actions compared to Sadness expression. Besides, Shame expression is characterized by lower torso flexion in Sitting Down action but with higher torso flexion in the other actions.

Anger :

Figure 11.9 shows the motion capture and the perceptual characterizations of Anger expression in all the actions. The perceptual ratings of Anger expression show high mean ratings of Speed and Power. A significantly higher mean rating of Power is observed in Throwing action ($p<.001$). In motion capture ratings, the acceleration of right hand movement is significantly higher in Knocking action, but the acceleration of the relative elbows extension is significantly higher in Throwing action (Lelbow.Relbow). While the speed of the absolute extension of elbows (Elbows.Hips) is significantly higher in Throwing action, the speed of the relative elbows extension (Lelbow.Relbow) is higher in both Throwing and Walking actions. The latter result is due to the arms swing behavior occurred in Anger expression during Walking.

High arms movement quantity is observed in perceptual ratings of Anger expression in Walking and in arms movement actions. A medium backward postural changes of torso movement is observed in motion capture characterization of Anger expression in Sitting Down and Walking actions.

(a) Perceptual characterization of Sadness for each action



(b) Motion capture characterization of Sadness for each action based on the intersection of selected features across all the actions

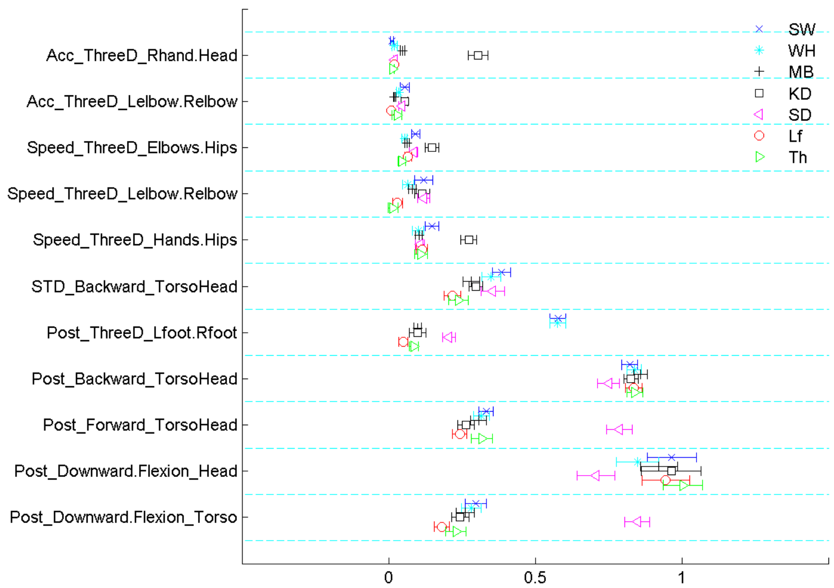
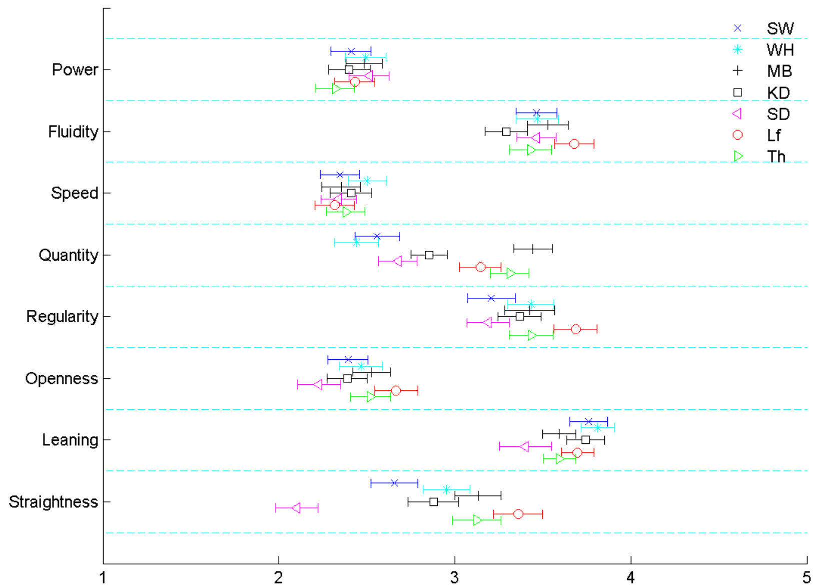


Figure 11.7: Perceptual and motion capture characterization of Sadness for each action.

(a) Perceptual characterization of Shame for each action



(b) Motion capture characterization of Shame for each action based on the intersection of selected features across all the actions

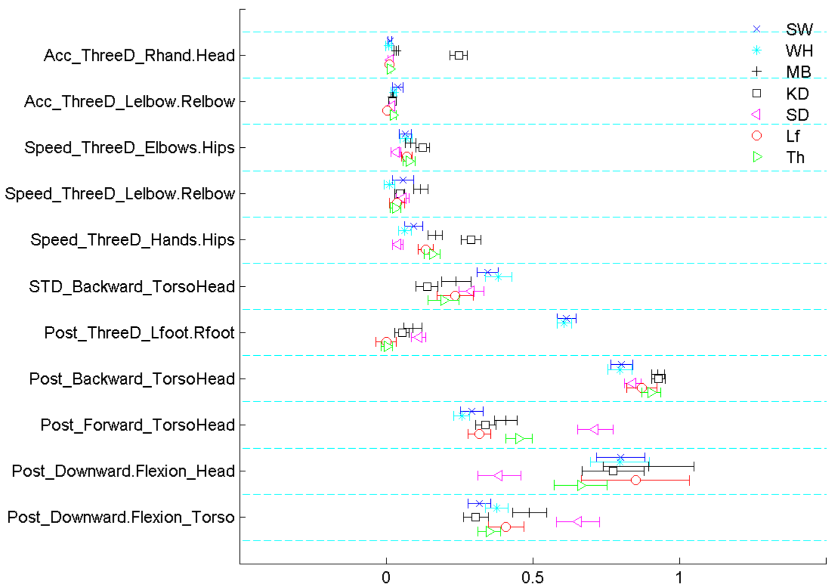


Figure 11.8: Perceptual and motion capture characterization of Shame for each action.

11

The leaning and the straightness of the torso appear to be affected by Sitting Down action. Indeed, a significantly higher amount of collapsed and more forward torso posture are observed for Anger expression in Sitting Down action than in other actions.

Neutral :

Figure 11.10 shows the motion capture and the perceptual characterizations of Neutral expression in all the actions. In perceptual ratings, Neutral expression is rated as neither strong nor light, neither fast nor slow, but highly smooth across all the actions. In motion capture characterization, Neutral expression is mostly rated as slow. Depending on arms movement feature, Neutral expression is sometimes rated as significantly faster in Knocking action.

In perceptual ratings, the rating of the quantity of movement is highly dependent on the action: significantly low mean ratings are attributed to Sitting Down action and significantly high mean ratings are attributed to Moving Books, Lifting and Throwing actions. In motion capture characterizations, the postural changes of torso movement is also dependent on the performed action: significantly higher mean ratings are attributed to Sitting Down and Walking actions.

In perceptual ratings, the openness of body movement is rated as slightly higher in Walking actions. In motion capture characterization, the openness of feet is rated as significantly higher in Walking action.

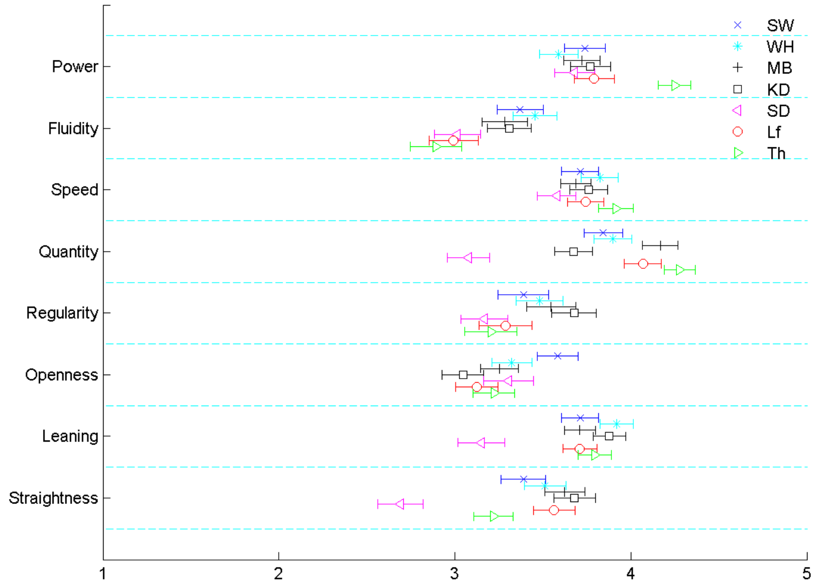
Similarly to the other expressed emotions, the leaning and the flexion of the torso appear to be highly different in Sitting Down action. Neutral expression is also characterized by higher backward torso posture in Sitting Down action than in the other actions in motion capture characterization. However, comparing to Neutral characterization, Pride and Joy are characterized with higher backward torso posture.

11.1.4 Discussion

Based on the Int-SSF of the 11 selected features obtained in the previous chapter, we explored both the motion capture characterization of expressed emotions through their patterns across all the actions and their motion capture characterizations for each action. Besides, we compared motion capture and perceptual characterization patterns.

Emotion expression characterization across all the actions :
Across all the actions, we were able to differentiate the characterization patterns of different emotions in both perceptual and motion capture characterizations. Joy and Anger expressions are mostly characterized with the highest mean ratings of body posture openness. The downward flexion of the torso and of the head received significantly the highest mean rating for Sadness expression. Left/Right head rotation behavior occurs during Pride expression across all the actions.

(a) Perceptual characterization of Anger for each action



(b) Motion capture characterization of Anger for each action based on the intersection of selected features across all the actions

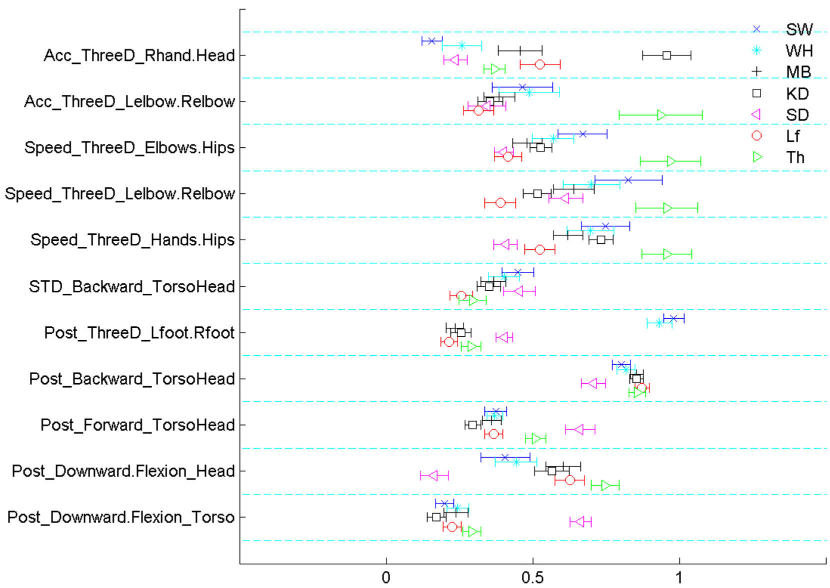
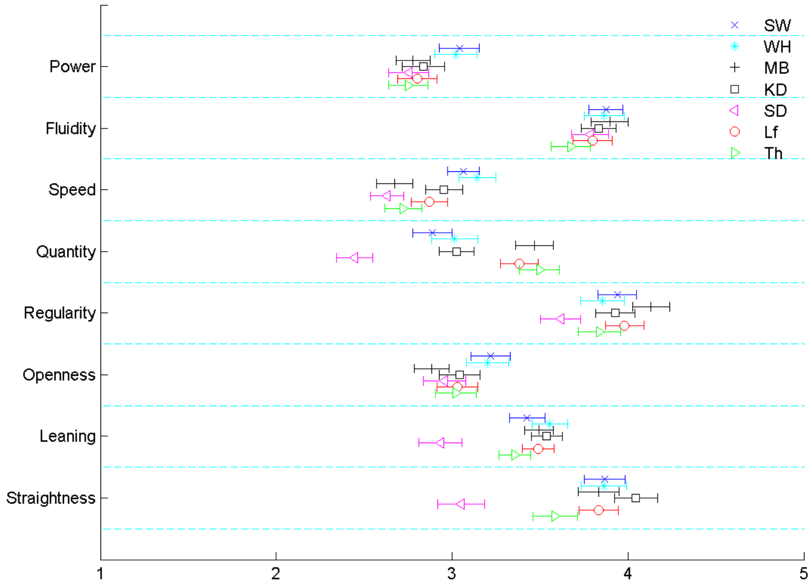


Figure 11.9: Perceptual and motion capture characterization of Anger for each action.

(a) Perceptual characterization of Neutral for each action



(b) Motion capture characterization of Neutral for each action based on the intersection of selected features across all the actions

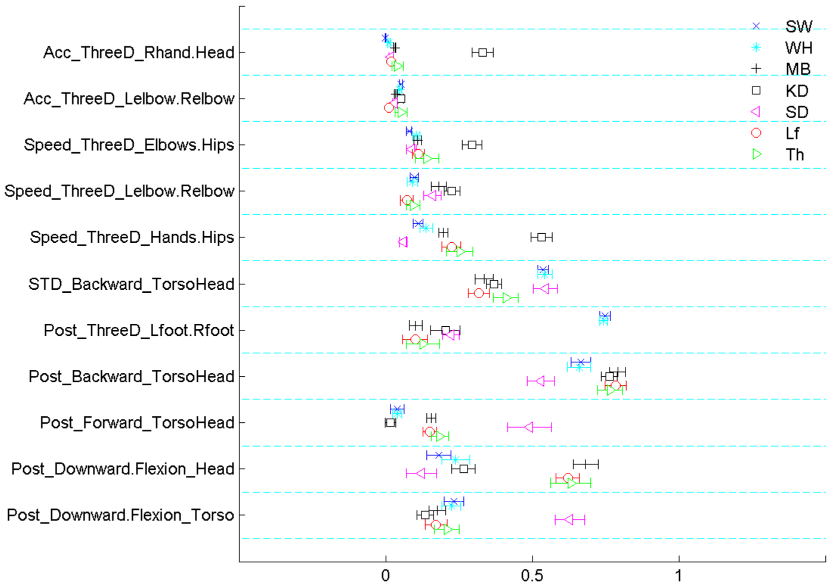


Figure 11.10: Perceptual and motion capture characterization of Neutral for each action.

However, we observed that the patterns of Sadness and Shame characterization as well as the patterns of Anxiety and Panic Fear characterization are highly similar. In Chapter 9, we showed that Shame and Sadness were confused in human and automatic recognition of expressed emotions. However, Anxiety and Panic Fear were only confused in perceptual study. As such, the simplified graphical representation of emotion characterization patterns does not always reflect the discrimination between emotions as achieved by Random Forest classification.

We also showed that the patterns of Anger and Joy expressions characterizations are highly differentiated from the other patterns based on postural changes and movement dynamics properties. The pattern of Anger expression is particularly clearly distinguishable from the other patterns based on the speed and the acceleration of upper body movement (arms, head and torso).

We also discussed the mapping between the patterns of characterization of emotions and their arousal and valence levels. Except few features, the patterns of characterization based on postural changes, speed and acceleration features mostly allows discriminating between emotions of “high” (Anger and Joy) and “low” arousal (Sadness and Shame) in the Emilya database. The forward/backward leaning and the straightness of the torso allows somehow discriminating between “positive” emotions (Pride, Joy, but Neutral is mostly grouped with them) and “negative” emotions in the Emilya database.

In Appendix H we give the detailed patterns of expressed emotions characterization across “similar” actions in Figures H.1, H.2, H.3, and H.4. These figures are based on the same graphical representation adopted in Figure 11.1. However, each Figure is based on the expressive sequences related to specific actions. For instance the patterns of characterization represented in Figure H.1 are based on the data related to walking actions (SW+WH). Consequently, the value of a given feature averaged across a group of “similar” actions (e.g. SW+WH) does not correspond to the feature value averaged across all the actions. It is more specific to a group of actions. A feature that belongs to the FS-SSF obtained across all the actions may not appear in the Int-SSF related to a group of “similar” actions and vice versa.

Emotion expression characterization for each action :
We also studied the effect of the performed actions on the characterization of each expressed emotion. On one hand, we found that some features are affected by the actions in the same way for each expressed emotion. For instance, in perceptual ratings, we found that the quantity of arms movement always receives the highest mean ratings in actions based on arms movement (e.g. Moving Books, Throwing..). In motion capture characterization, the speed and/or the acceleration of arms movement is mostly significantly higher in Knocking action. This result is may be due to the discontinuity of motion that occur at the Knocking phase. The openness of feet is also highly affected by Walking actions; it receives the highest motion capture mean rating for the expression of each emotion in Walking action. In both perceptual and motion capture ratings, the forward/backward leaning and the downward/upward flexion of the torso seems significantly affected by Sitting Down action.

On the other hand, we found that the expression of a given emotion can be differently affected by a particular action. For instance, the expression of Anger is characterized with a particular pattern in Throwing action. In Chapter 9, we showed that Anger expression is better recognized by humans in Throwing action. Anger expression is also better classified in motion capture based classification in Throwing action (See Figure 9.5 in Chapter 9).

11.2 Decision tree based modeling of emotional body expression

In the previous section, we discuss the different patterns of emotional body expression based on the mean ratings of features. In this section, we provide deeper insights into the discrimination of emotional body expression based on a machine learning model.

We adopt a Decision Tree model which provides a good compromise between classification interpretation and performance. Decision Tree is one of the most well known techniques for supervised learning [Narsky and Porter, 2014a]. An algorithm of Decision Tree consists in a greedy partition of the observations in the feature space into disjoint blocks. We use the CART (classification and regression tree) algorithm which is a well known algorithm of Decision Tree proposed by Breiman et al. [Breiman et al., 1984].

An advantage of Decision Tree models is the possibility to easily interpret the model. A single Decision Tree can be easily visualized. In our work, a Decision Tree model allows us to easily interpret the relationship between expressive body features and the expressed emotions. The objective to use a Decision Tree model is to explore the decision rules that could characterize bodily expression of emotions in the Emilya database. These decision rules are useful to explore what common and what different characteristics occur between the different bodily expressions of emotions in the Emilya database.

One of the issues of Decision tree models is the instability to changes in the dataset, and also to the features selected to perform the splits during the growing of the tree. In fact, Decision Tree models suffer from overfitting the training dataset. Besides, they are highly sensitive to slight changes in the training dataset. In order to prevent the Decision Tree model from selecting features that are tailored to the learning dataset, we build the Decision Tree models using the subset of features selected using Random Forest model (that is the features selected in Chapter 10). Int-SSF is used across "similar" actions and FS-SSF is used across all the actions.

Even with the restriction of using a subset of the most relevant features, a Decision Tree model may overfit the training dataset. Avoiding the overfit of the training dataset while achieving maximum accuracy is a well known problem when using Decision Tree models for classification purpose; a large tree risks overfitting the training data while a small tree may be insufficient to capture the complexity of data structure.

Previous studies have attempted to find criteria to stop growing the tree before it overfit the training dataset [Narsky and Porter, 2014a]. These studies impose a stopping criterion to prevent the grow of deep and inaccurate tree. When reached, a stopping criterion prevents further split of nodes. For instance, a stopping criterion could be "*the maximal number of leaf nodes has been reached*" [Narsky and Porter, 2014a].

Other studies propose to grow a deep fully-split tree, and then remove the leaf nodes with poor classification accuracy [Narsky and Porter, 2014a]. The process of reducing a deep fully-split tree is known as **pruning**. In the context of building optimal trees, pruning receives more interest than the approach based on stopping criteria, due to its simplicity and its success. As said in [Narsky and Porter, 2014a] (pages 312-313), "*A popular belief is that trees produced by pruning tend to be more accurate than trees produced by applying a conservative stopping rule*". In our work, we adopt pruning strategy to avoid overfitting issue and to obtain an acceptable balance between the accuracy and the interpretability of the tree.

In the following subsections, we firstly introduce few definitions related to pruning process. Then we describe two pruning strategies and we explain our pruning process. In section 11.2.5 we present the results of pruning in terms of classification rates and the size of the tree. Finally, we discuss the decision rules obtained from the pruned trees in section 11.2.6. These decision rules are useful to discuss the characterization of emotional body expressions.

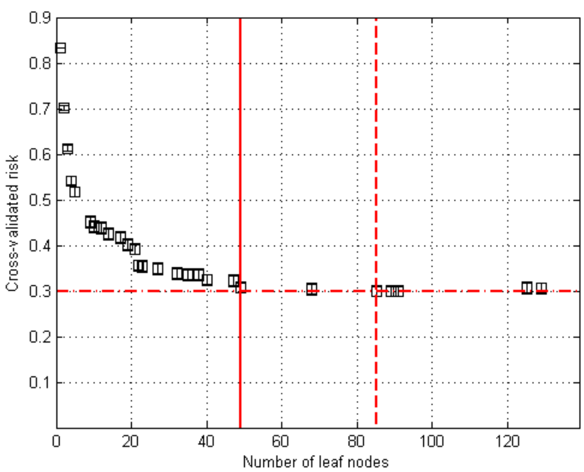


Figure 11.11: 10 fold cross-validation error with one standard error confidence bounds of each $T^*(\alpha_i)$ versus leaf nodes for each α_i. Dashed lines refer to the minimal cross-validation error (optimal pruning level) while solid line refers to the minimal cross-validation error within one standard error (final optimal pruning level).

11

11.2.1 Pruning: Definitions

- **Leaves, leaf nodes, or terminal nodes**: The last nodes in the tree are known as leaves or also as terminal nodes. They contain the class name.
- **Pruning**: Pruning consists in reducing tree depth by merging leaves on the same tree branch (See Figure 11.13).
- **Optimal pruning sequence**: The optimal pruning sequence is defined in [Narsky and Porter, 2014a] as a set of trees $T^*(\alpha_0)$ ... $T^*(\alpha_M)$ ordered by a parameter α_i, where α_0 refers to the initial pruning level, $T^*(\alpha_0)$ refers to the initial tree (minimal pruning), α_M refers to the last pruning level, $T^*(\alpha_M)$ refers to the root alone (maximum pruning). If $\alpha_1 \leq \alpha_2$, optimal $T^*(\alpha_1)$ contains $T^*(\alpha_2)$ as a subtree [Narsky and Porter, 2014a]. In other words, if $\alpha_1 \leq \alpha_2$, $T^*(\alpha_1)$ is deeper than $T^*(\alpha_2)$.
- **"Level based pruning"**: We call "Level based pruning" the process of tree pruning where the tree is pruned to the best pruning level. This process is explained in details in [Narsky and Porter, 2014a]. It is composed of two steps. First of all, an *optimal pruning sequence* is measured. Secondly, the optimal pruning level is selected and the tree is pruned to that level.
- **"Weakly relevant leaves"**: We define "Weakly relevant leaves" as the leaf nodes that leads to a small proportion of observations or a low probability of its main class. We consider that the probability of the main class is low if it is < 0.5, meaning that there are few chances that the rule associated to this leaf node leads to its main class. We consider that the proportion of observations of its main class is relatively low if it is $<$ than $\frac{1}{(NbLeafMainClass)}$, where *NbLeafMainClass* stands for the number of leaf nodes leading to the same main class as the current leaf node.
- **"Node based pruning"**: We define "Node based pruning" as the process of pruning "Weakly relevant leaves". We define "Node based pruning" as a pruning strategy based on the class membership of each leaf node.

11

11.2.2 Pruning: Level based pruning

We describe the pruning process as proposed in [Narsky and Porter, 2014a]. Figure 11.13 summarizes the two steps of "Level based pruning": 1) Finding the optimal pruning sequence and 2) choosing the optimal pruning level.

Find the optimal pruning sequence :

The construction of the optimal pruning sequence is a recursive process where the branches giving less improvement in error cost are removed at each pruning level. Figure 11.13 presents an example of an optimal pruning sequence. Starting from $T^*(\alpha_0)$, the subtree T_{Node} defined with a root node $Node$ that minimizes α_{Node} is removed [Narsky and Porter, 2014a]. α_{Node} is defined according to the equation

11.1, where

- $r(Node)$ is the node risk: the node error multiplied by the node probability,
- $r(T_{Node})$ is the branch risk: the sum over risk values for all leaves descending from this node,
- $|L(T_{Node})|$ is the number of descendant leaves.

$$\alpha_{Node} = \frac{r(Node) - r(T_{Node})}{|L(T_{Node})| - 1} \tag{11.1}$$

In the example presented in Figure 11.13, α_{Node} corresponding to $Node_6$, $Node_4$, $Node_3$, $Node_1$ are respectively equal to 0.0131, 0.0214, 0.1008 and 0.1590. These measures are obtained by substituting the corresponding values of the variables of each $Node$ in equation 11.1. As α_{Node} of $Node6$ receives the lowest value, $Node_6$ is the pruned node in the first pruning level, followed by $Node_4$, $Node_3$ and finally $Node_1$ respectively in the second, third and fourth pruning level.

Find the optimal pruning level :

Once the optimal pruning sequence is constructed, we need to know to what optimal pruning level the tree should be pruned. One possible approach is to chose the pruning level that leads to the minimal classification error measured on a testing dataset [Narsky and Porter, 2014a]. Unfortunately, we do not dispose of another independent testing dataset. Thus, we use the alternative approach based on cross-validation. However, as reported and explained in [Narsky and Porter, 2014a], the optimal pruning sequence can be slightly different for each fold. For instance, we note T the tree grown on the whole dataset. We suppose that we grow n T_k trees for n-fold cross-validation and we build the optimal pruning sequence for each (n-1)/n of the data (where k=1..n). We cannot ensure that we obtain, for each T_k, the same optimal pruning sequence as the one obtained from T.

It has been argued in [Narsky and Porter, 2014a] that we can assume that each T_k is "reasonably" close to the tree built on the entire dataset if the number of folds is sufficiently large (e.g. 10 folds). That is, it is assumed that we can compute the cross-validated error of T by growing n T_k trees and apply each T_k to the $\frac{1}{n}$ not used for the training of this K^{th} tree.

As such, we can compute the cross-validated error at each m^{th} pruning level (α_m) in order to estimate the error of each $T^*(\alpha_m)$. Instead of pruning each T_k to α_m, Breiman et al. [Breiman et al., 1984] argued that each T_k should be pruned to the geometric averaged level defined as $\alpha'_m = \sqrt{\alpha_m \alpha_{m+1}}$.

Finally, we can chose the optimal tree that leads to the minimal cross-validation error within one standard error as proposed in [Narsky and Porter, 2014a]. Figure 11.11 presents an example of the plot of 10 folds cross-validation error measured for each $T^*(\alpha_m)$ versus the corresponding number of leaf nodes starting from the initial pruning level α_0 to the last pruning level α_M. In this example, the tree corresponding to α_0 contains 129 leaf nodes. The tree corresponding to α_M is always restricted to

1 leaf node: the root of the tree. As the minimum of the cross-validation error is usually flat, choosing the minimal cross-validation error within one standard error allows us to obtain a smaller tree without significantly increasing the error.

11.2.3 Pruning: Node based pruning

Level based pruning approach described above allows the optimization of the tree depth while maintaining the best classification accuracy. However, Level based pruning approach does not prevent the presence of “weakly relevant leaves”. In our work, we aim to interpret the final pruned tree to explore the decision rules that may be associated with bodily expression of emotions. Indeed, we desire to remove as much as possible “weakly relevant leaves” which we consider less representative for interpretation purpose. For instance, a “weakly relevant leave” can lead to 25% of Shame, 35% of Sadness, 40% of Neutral and 0% for the other emotions. In this case, Neutral is considered as the main class of this leaf node as it receives the maximal probability. However, we want to focus on leaf nodes for which the main class is dominant. Besides, a “weakly relevant leaf” can lead to a small proportion of observations related to the main class, which may refer to a particular actor.

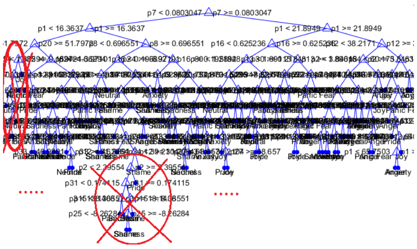
While Node based pruning reduces the depth of the tree by pruning “weakly relevant leaves”, it may result in a significant decrease of its accuracy. We propose to prune “weakly relevant leaves” as long as the cross-validation error of the obtained tree is not significantly higher than the one measured on the precedent tree. After an experimental step, we observed that the cross-validation error of the tree is often significantly higher after the second application of Node based pruning. Thus, only one Node based pruning is applied.

We note that the goal from applying Node based pruning is to reduce and not to completely eliminate “weakly relevant leaves”. Thus, the tree resulted from a Node based pruning contains less “weakly relevant leaves”.

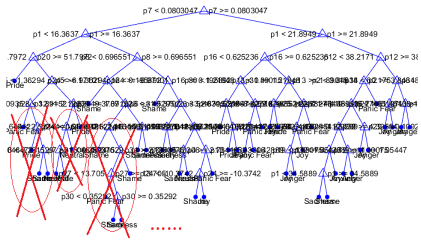
11.2.4 Our pruning process

Our pruning approach is based on the two strategies of pruning presented above: Level based pruning and Node based pruning. This pruning process can be summarized in the following three steps. An example of applying this pruning process is described in Figure 11.12.

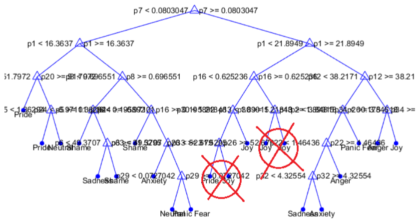
1. **First Level based Pruning**: The tree is pruned to the best level according to the optimal pruning sequence and the cross-validation error.
2. **Node based Pruning**: The tree is pruned to the best rules according to the probability and the percentage of observation of the main class of each leaf node.
3. **Second Level based Pruning**: As we changed the configuration of the tree after applying Node based Pruning, we propose to apply again a Level based



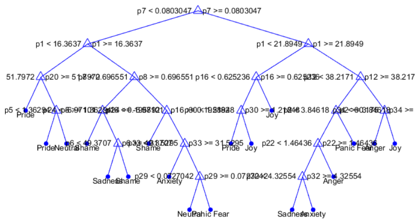
(a) Original Tree: 129 terminal nodes



(b) Tree resulted from the First level Pruning: 49 terminal nodes



(c) Tree resulted from the Node bases Pruning: 21 terminal nodes



(d) Tree resulted from the Second level Pruning: 19 terminal nodes

Figure 11.12: Pruning process: successions of trees built on Walking actions (SW+WH)

pruning to make sure that we obtain an optimal tree at the end of the pruning process. Thus, the tree obtained from Node based Pruning is pruned again to the best level according to the optimal pruning sequence and the cross-validation error.

We will show in the following subsection (11.2.5) that this pruning process results in a good compromise between the the size and classification accuracy of the tree.

11.2.5 Results

As we aim to characterize emotional body expressions across different actions, we present the characterization of emotional body expression in similar actions and across all the actions. Thus, the process of pruning is applied to each dataset composed of the expressive movements of similar actions and to the dataset composed of the expressive movements of all the actions. The groups of similar actions (as described in Chapter 10) are: 1) Walking (SW + WH), 2) Repetitive Arms movement (KD + MB), 3) Non-repetitive Arm movement (Lf + Th) and finally 4) Sit Down movement (SD + BS).

Tables 11.1 presents the 10 folds cross-validation errors measured after each step of our pruning process. Each error is averaged across 50 runs to smooth the results. Tables 11.1 shows that we achieved a desired balance between the accuracy and the interpretability of each Decision Tree model. In fact, the classification accuracy that we obtain after the pruning process is highly similar to the classification accuracy of the initial fully grown tree. Besides, the number of leaf nodes obtained after the pruning process is highly lower than the leaf nodes of the initial fully grown tree.

11.2.6 Decision rules

In this section, we discuss the set of rules resulted from the pruning process applied on each Decision Tree model. Five Decision Tree models are built: 4 models for each group of similar actions and 1 model across all the actions. Figures 11.15, 11.16, 11.17, 11.18 show the set of decision rules of expressed emotions for each group of similar actions. Figures 11.19, 11.20, 11.21 show the decision rules of each expressed emotion across all the actions. The latter result is split into three Figures only for the sake of clarity as the three figures does not fit into one page. The set of rules described in these three Figures (11.19, 11.20, 11.21) are derived from the same Tree.

In each of these Figures (11.15, 11.16, 11.17, 11.18, 11.19, 11.20, 11.21), we provide the set of decision rules as resulted from a tree; we start from the split at the root and we finish with the terminal nodes, where each terminal node is associated to an expressed emotion. We also provide the percentage of observations (i.e. samples) resulted from the rule that leads to each terminal node and we provide the probability to obtain the expressed emotion associated to a terminal node (e.g. 70% of Anger).

The number of decision rules in each Tree corresponds to the number of terminal nodes of the final pruned tree as showed in Table 11.1. For instance, we observe that Figure 11.17 contains 9 rules which correspond to 9 terminal nodes as showed in Table 11.1.

11.2.6.1 For each group of similar action

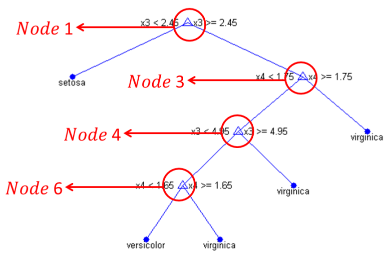
We discuss the set of rules generated for each group of "similar" actions.

Walking actions; SW+WH :

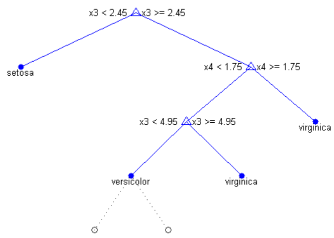
Figure 11.15 shows the decision rules that describe the characterization of each emotion in Walking actions (SW+WH). These rules correspond to the final Decision Tree model obtained after applying the pruning process described in section 11.2.4.

The first feature considered at the root of the Tree is the acceleration of the relative elbows extension (Acc_ ThreeD _Lelbow.Relbow) followed by the forward torso posture. These features are ranked by the Random Forest model as the first most relevant for the classification of emotions in Walking actions (See Chapter 10, Figure 10.3).

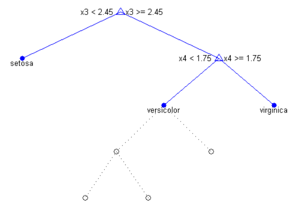
We can observe from Figure 11.15 that a small amount of forward torso movement is associated to Joy, Pride and Neutral. In section 11.1.2, we showed that forward



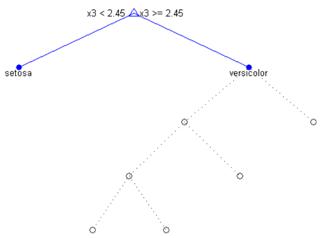
(a) $T^*(\alpha_0)$



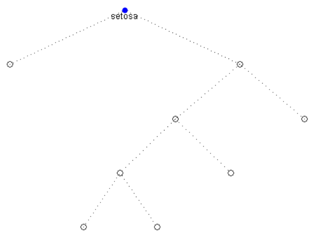
(b) $T^*(\alpha_1)$



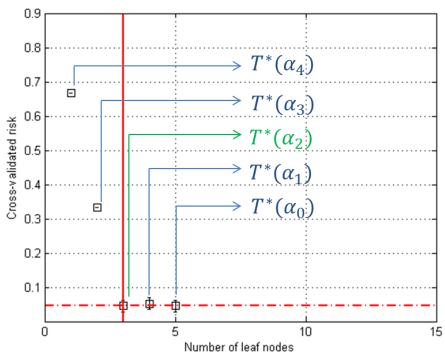
(c) $T^*(\alpha_2)$



(d) $T^*(\alpha_3)$



(e) $T^*(\alpha_4)$



(f) Cross-validation error of each pruning level: $T^*(\alpha_2)$ is the optimal pruning level according to the cross-validation error of each pruning level

Figure 11.13: Example of the construction of an optimal pruning sequence and the selection of the optimal pruning level

Table 11.1: For each group of actions; Cross-validation errors for initial tree, pruned tree according to the first level based pruning, node based pruning and second level based pruning

Group of Actions	CCR and Nb Leaves	Initial tree	First Level pruning	Nodes pruning	Second Level pruning
SW+WH	CCR_{CV}	68,04%	67,93%	68,83%	68,85%
	Nb Nodes	129	49	21	19
MB+KD	CCR_{CV}	61,67%	62,09%	60,77%	59,36%
	Nb Nodes	142	75	29	26
SD+BS	CCR_{CV}	56,30%	55,46%	55,50%	55,72%
	Nb Nodes	143	36	11	9
Lf+Th	CCR_{CV}	66,59%	66,63%	67,20%	66,81%
	Nb Nodes	116	63	27	23
AllActi	CCR_{CV}	57,59%	58,98%	57,71%	56,80%
	Nb Nodes	543	206	62	51

torso posture receives the lowest mean ratings for Pride, Joy and Pride expressions.

Shame and Sadness are both characterized with low acceleration of arms movement and high forward torso posture. Three decision rules are associated with Shame expression. The first part of Shame expression samples is differentiated from Sadness due to faster legs movement than Sadness expression (20.76% of Shame observations and 78.57% of probability to obtain Shame expression). The second part of Shame expression samples (23.27% of observations, 78.72% of probability to obtain Shame expression) is differentiated from Sadness due to higher lateral closeness of lateral lower body movement. In Chapter 10, we showed that lateral closeness of lower body movement is considered among the most relevant features for Shame expression across all the actions. Finally, the third part of Shame expression samples (only 10.06% of observations, 66,67% of probability to obtain Shame expression) is differentiated from Sadness expression through larger steps (higher forward lower body movement).

Overall, Sadness is mainly characterized with low acceleration of elbows movement, high forward torso posture, low speed of legs movement and finally lower closeness of legs and smaller steps (3D openness of feet) comparing to Shame expression.

The main decision rule associated with Anger expression (62.28% of observations, 87.40% of probability of obtain Anger) is defined through high acceleration of elbows movements, high forward torso posture, large steps (Post_ ThreeD _ Lfoot.Rfoot $>=$ 38.21) and lower lateral openness of arms posture than Joy expression. A smaller amount of Anger expression observations (15.57%) is characterized with smaller steps (Post_ ThreeD _ Lfoot.Rfoot $<$ 38.21), slower legs movement (Speed_ ThreeD _ Lfoot.Rfoot $<$ 3.84) compared to Panic Fear expression, and higher speed of hands movement (Speed _ ThreeD _ Hands.Hips>=1.46) comparing to Anxiety and Sadness.

Repetitive Arms Movement actions; MB+KD :

Figure 11.16 shows the decision rules that describe the characterization of each emotion in Repetitive arms movement actions (MB+KD). These rules correspond to the final Decision Tree model obtained after applying the pruning process described in section 11.2.4.

The acceleration of hands movement (Acc_ ThreeD_ Hands.Hips) is used at the root of the tree associated with MB+KD group of actions. This feature was considered as the second most relevant feature for emotions classification in MB+KD group of actions (See Chapter 10, Table 10.3).

Anger expression is mainly characterized (71.20% of Anger observations and 83.97% probability of Anger) with high acceleration of hands movement, straight lower body posture (Post_ ThreeD_ Feet.Hips>=95.76%), high downward head flexion and high acceleration of elbows movement. Anger expression is also obtained by replacing the last two characteristics with lower downward head flexion, high extension of right hand and high acceleration of head movement. However, a lower percentage of observations (11,96%) and a lower probability of Anger (59.46%) are obtained with this rule.

A part of Pride expression observations (39.73%) is characterized with high probability of Pride (79.54%) by low acceleration of hands movement, high speed of elbows movement, low speed of lateral torso movement, high postural changes of torso flexion and high speed of vertical arms movement. However, another smaller part of Pride expression observations (19.18%) is characterized with high speed of lateral torso movement and low torso flexion (66.67% of probability to obtain Pride).

A part of Sadness expression observations (35.19%) is described with a high probability of Sadness (88.17%) by low acceleration of hands movement, low speed of elbows movement, straight lower body limbs, high downward head flexion and low speed of lateral arms movement.

A part of the observations of Panic Fear expression (21.23%) is described with high probability (81.58%) by a high acceleration of hands movements and a bending lower body posture. The characterizations of the other expressed emotions are somehow scattered within the branches of the Tree.

Sit Down actions; SD+BS :

Figure 11.17 shows the decision rules that describe the characterization of each emotion in Sit Down actions (SD+BS). Nine rules are described and they correspond to the final Decision Tree model obtained after applying the pruning process described in section 11.2.4.

The feature used at the root of the Tree is the acceleration of elbows movement (Acc_ ThreeD_ Lelbow.Relbow), which was considered among the first relevant fea-

Level1	Level2	Level3	Level4	Level5	Level6	Emotion	Observ	Prob
Acc_ ThreeD_ Lelbow.Relbow<0.08	Post_ Forward_ TorsoHead<16.3637	SymOcc_ ThreeD_ Lelbow.Relbow<51.79				Pride	29,22%	84,21%
		SymOcc_ ThreeD_ Lelbow.Relbow>=51.79	STD_ LRRotation_ Torso<1.36			Pride	6,39%	26,42%
			STD_ LRRotation_ Torso>=1.36			Neutral	75,30%	69,44%
	Post_ Forward_ TorsoHead>=16.3637	Speed_ ThreeD_ Feet.Hips<0.69	Post_ Inside_ LowerBody<-6.97			Shame	23,27%	78,72%
			Post_ Inside_ LowerBody>=-6.97	Post_ Forward_ LowerBody<49.37		Sadness	69,13%	79,90%
				Post_ Forward_ LowerBody>=49.37		Shame	10,06%	66,67%
		Speed_ ThreeD_ Feet.Hips>=0.69	Speed_ ThreeD_ Lelbow.Relbow<0.19			Shame	20,76%	78,57%
			Speed_ ThreeD_ Lelbow.Relbow>=0.19	Post_ ThreeD_ Lhand.RHand<31.52		Anxiety	24,00%	50,00%
				Post_ ThreeD_ Lhand.RHand>=31.52	Acc_ LRRotation_ Head<0.072	Neutral	10,24%	42,50%
					Acc_ LRRotation_ Head>=0.072	Panic Fear	5,97%	30,77%
Acc_ ThreeD_ Lelbow.Relbow>=0.08	Post_ Forward_ TorsoHead<21.8949	Speed_ ThreeD_ Lelbow.Relbow<0.62	Speed_ Lateral_ LowerBody<1.21			Pride	26,94%	65,56%
			Speed_ Lateral_ LowerBody>=1.21			Joy	12,12%	43,08%
		Speed_ ThreeD_ Lelbow.Relbow>=0.62				Joy	53,68%	75,15%
	Post_ Forward_ TorsoHead>=21.8949	Post_ ThreeD_ Lfoot.Rfoot<38.21	Speed_ ThreeD_ Lfoot.Rfoot<3.84	Speed_ ThreeD_ Hands.Hips<1.46	Speed_ Flexion_ Knees<4.32	Sadness	7,83%	32,73%
					Speed_ Flexion_ Knees>=4.32	Anxiety	30,67%	56,10%
				Speed_ ThreeD_ Hands.Hips>=1.46		Anger	15,57%	74,29%
			Speed_ ThreeD_ Lfoot.Rfoot>=3.84			Panic Fear	45,52%	83,56%
		Post_ ThreeD_ Lfoot.Rfoot>=38.21	Post_ Outside_ Arms<60.1			Anger	62,28%	87,40%
			Post_ Outside_ Arms>=60.1			Joy	3,90%	60,00%

Figure 11.15: Decision tree rules for the characterization of expressed emotions in walking actions (SW and WH)

Level1	Level2	Level3	Level4	Level5	Level6	Level7	Emotion	Observ	Prob
Acc_ ThreeD_ Hands.Hips<0.30419	Speed_ ThreeD_ Lelbow.Relbow<0.37191	Post_ ThreeD_ Feet.Hips<97.0591	Post_ Backward_ TorsoHead<-0.52435	Post_ Forward_ TorsoHead<18.3482	Post_ Downward.Flexion_ Torso<2.3123		Neutral	7,34%	44,83%
					Post_ Downward.Flexion_ Torso>=2.3123		Anxiety	15,84%	59,26%
				Post_ Forward_ TorsoHead>=18.3482	Speed_ Flexion_ Elbows<0.038365		Shame	34,75%	49,00%
					Speed_ Flexion_ Elbows>=0.038365		Sadness	11,16%	86,67%
			Post_ Backward_ TorsoHead>=-0.52435				Shame	38,30%	79,41%
		Post_ ThreeD_ Feet.Hips>=97.0591	Post_ Downward.Flexion_ Head<21.6649	STD_ Flexion_ Torso<0.14607			Neutral	11,30%	86,96%
				STD_ Flexion_ Torso>=0.14607	Post_ Forward_ TorsoHead<14.8846		Pride	6,16%	45,00%
					Post_ Forward_ TorsoHead>=14.8846	Post_ Outside_ LowerBody<14.9262	Sadness	7,73%	60,00%
						Post_ Outside_ LowerBody>=14.9262	Panic Fear	5,48%	66,67%
			Post_ Downward.Flexion_ Head>=21.6649	Speed_ Lateral_ Arms<0.8953			Sadness	35,19%	88,17%
				Speed_ Lateral_ Arms>=0.8953			Neutral	11,30%	52,63%
	Speed_ ThreeD_ Lelbow.Relbow>=0.37191	Speed_ Lateral_ Torso<0.10546	STD_ Flexion_ Torso<0.31239				Neutral	36,16%	53,78%
			STD_ Flexion_ Torso>=0.31239	Speed_ Vertical_ Arms<0.86081			Neutral	2,82%	50,00%
				Speed_ Vertical_ Arms>=0.86081			Pride	39,73%	79,45%
		Speed_ Lateral_ Torso>=0.10546	Post_ Downward.Flexion_ Torso<5.2527	Post_ Downward.Flexion_ Torso<0.96541			Pride	19,18%	66,67%
				Post_ Downward.Flexion_ Torso>=0.96541	Acc_ LRRotation_ Head<0.054238		Neutral	9,60%	29,31%
					Acc_ LRRotation_ Head>=0.054238		Joy	54,14%	61,15%
			Post_ Downward.Flexion_ Torso>=5.2527				Panic Fear	23,97%	81,40%
Acc_ ThreeD_ Hands.Hips>=0.30419	Post_ ThreeD_ Feet.Hips<95.7688						Panic Fear	21,23%	81,58%
	Post_ ThreeD_ Feet.Hips>=95.7688	Post_ Downward.Flexion_ Head<18.6354	Post_ ThreeD_ Rhand.Head<46.171	Acc_ Lateral_ Head<0.10057			Joy	8,28%	65,00%
				Acc_ Lateral_ Head>=0.10057			Panic Fear	15,07%	70,97%
			Post_ ThreeD_ Rhand.Head>=46.171	Acc_ LRRotation_ Head<0.083626			Pride	8,90%	59,09%
				Acc_ LRRotation_ Head>=0.083626			Anger	11,96%	59,46%
		Post_ Downward.Flexion_ Head>=18.6354	Acc_ ThreeD_ Elbows.Hips<0.13445	Speed_ Lateral_ Torso<0.15627			Anxiety	6,93%	58,33%
				Speed_ Lateral_ Torso>=0.15627			Joy	5,73%	60,00%
			Acc_ ThreeD_ Elbows.Hips>=0.13445				Anger	71,20%	83,97%

Figure 11.16: Decision tree rules for the characterization of expressed emotions in Knocking and Moving books actions

tures for the classification of emotions in Sit Down actions (SD+BS) using Random Forest approach (See Chapter 10, Figure 10.3).

On one hand, slightly more than half of Sadness observations (52.57%) are described through low acceleration of elbows movement (Acc_ ThreeD_ Lelbow.Relbow < 0.10) and high head downward flexion movement (Post_ Downward.Flexion_ Head >=22.99). However, this characterization is not restricted to Sadness expression. Indeed, the probability to obtain Sadness with such a rule is 65.30%. Taking a deeper look at the tree, we find out that Anxiety, Pride and Shame can be also obtained based on this characterization with respectively 8%, 6% and 10% of probability.

On the other hand, slightly less than the half of Anger observations (42.73%) are described through high acceleration (Acc_ ThreeD_ Lelbow.Relbow >= 0.10) and high speed (Speed_ ThreeD_ Elbows.Hips >= 0.76) of elbows movement. This characterization allows obtaining Anger with a probability of 83.19%.

The characterization rule associated with Pride expression allows obtaining 51.56% of Pride observations. It is described through low acceleration of elbows movement, low downward head flexion, high speed of head movement and backward torso posture. However, this characterization rule is not restricted to Pride expression as the probability to obtain Pride is 54.10%. Taking a deeper look at the class membership associated with the terminal node related to Pride, we found that Joy and Neutral are also involved with respectively 19% and 8% of probabilities.

11

Non-Repetitive arms movement actions: Th+Lf :

Figure 11.18 shows the decision rules that describe the characterization of each emotion in Lifting and Throwing actions (Lf+Th). 23 rules are described and they correspond to the final Decision Tree model obtained after applying the pruning process described in section 11.2.4.

The feature used at the root of the Tree is the acceleration of right hand movement movement (Acc_ ThreeD_ Rhand.Head), which was considered as the first relevant feature for the classification of emotions in Lf+Th group of actions using Random Forest approach (See Chapter 10, Figure 10.3).

Most of Anger observations (78.86%) are characterized with high probability of Anger (89.82%) by high acceleration of right hand movement, high speed of vertical arms movement and high openness of hands posture.

A part of Panic Fear observations (14.29%) is characterized by high acceleration of right hand movement, low speed of vertical arms movement and high downward flexion of the torso. This characterization rule seems to be restricted to Panic Fear expression (100% of probability to obtain Panic Fear). Another part of Panic Fear observations (14.29%) is also characterized with downward flexion of the torso, but also with high speed of sagittal torso movement, high speed of vertical arms movement and low acceleration of right hand movement. Based on this characterization rule, we obtain 60.87% of Panic Fear, but also 30% of Anger. A similar characterization rule with low downward torso flexion leads to 54.73% of Joy observations, but

Level1	Level2	Level3	Level4	Level5	Emotion	Observ	Prob
Acc_ThreeD_Lelbow.Relbow<0.10	Post_Downward.Flexion_Head<22.99	Speed_LRRotation_Head<0.31	Post_Inside_LowerBody<-2.55		Shame	51,79%	31,02%
			Post_Inside_LowerBody>=-2.55	Post_Inside_Arms<-14.65	Sadness	13,60%	50,00%
				Post_Inside_Arms>=-14.65	Neutral	56,25%	59,28%
		Speed_LRRotation_Head>=0.31	Post_Backward_TorsoHead<-9.17		Pride	51,56%	54,10%
			Post_Backward_TorsoHead>=-9.17		Anxiety	28,08%	32,80%
	Post_Downward.Flexion_Head>=22.99				Sadness	52,57%	65,30%
Acc_ThreeD_Lelbow.Relbow>=0.10	Speed_ThreeD_Elbows.Hips<0.76	Post_Inside_LowerBody<-4.95			Anxiety	11,64%	33,33%
		Post_Inside_LowerBody>=-4.95			Anger	35,46%	55,71%
	Speed_ThreeD_Elbows.Hips>=0.76				Anger	42,73%	83,19%

Figure 11.17: Decision tree rules for the characterization of expressed emotions in Sitting Down and Being Seated actions

the probabilities are distributed between Joy, Pride, Anger and Anxiety (respectively 58.70%, 12%, 10% and 10%).

A part of Anxiety observations (16.85%) shares common characterization rules with a part of Panic Fear observations (28.57%). Both are characterized by low acceleration of right hand movement, low speed of vertical arms movement, high bending of lower body limbs, relatively high feet openness and low backward torso posture. Anxiety is distinguished from Panic Fear with lower values of forward torso posture.

A part of Pride observations also shares common characteristics with a part of Neutral observations. The main differences occur at the level of upward arms posture (more expanded in Pride expression) and the postural changes of lower body movement (less changes are observed for Neutral expression).

A part of Sadness expression (38.70%) is described with high probability of Sadness (90.82%) by low acceleration of right hand movement, low speed of vertical arms movement, low knee flexion and high downward head flexion. However, other parts of Sadness expression share common characterization rules with Neutral and Shame. The main differences occur at the level of lateral closeness of lower body limbs and downward flexion of the torso.

11.2.6.2 Across all the actions

The decision rules that describe the characterization of each emotion across all the actions are shown in Figures 11.19, 11.20 and 11.21. The decision rules are split into three figures only for the sake of representation as 51 rules are obtained. These 51 rules correspond to the same Decision Tree model obtained after applying the pruning process described in section 11.2.4.

We can observe from Figures 11.19, 11.20 and 11.21 that the first two features used to build the tree correspond to the acceleration of elbows movement (Post_ ThreeD_ Lelbow.Relbow) and forward torso posture (Post_ Forward_ TorsoHead). These features are considered, according to Random Forest model, as the most two relevant features for the classification of emotions across all the actions (See Chapter 10, Table 10.3).

Despite the pruning process applied on the tree obtained across all the actions, we observe a highly deep tree (composed of 51 terminal nodes). Besides, we can also observe the presence of several rules that lead to a small subset of observations. This is due to the fact that Decision Tree models try to fit at best the training dataset. Besides, the training dataset used to build this tree consists in the expressive behaviors of all the actions considered in the Emilya database. Thus, a Decision Tree model is more likely to overfit such data to consider as much as possible the variations present in the data. The classification of bodily expression of emotions across all the actions achieve highly better results with Random Forest model (80%) than with a single decision tree (56.80%).

Figure 11.19 shows the decision rules that are characterized by low acceleration

Level1	Level2	Level3	Level4	Level5	Level6	Level7	Emotion	Observ	Prob
Acc_ ThreeD_Rhand.Head<0.49	Speed_ Vertical_Arms<0.96	Post_ Downward.Flexion_Knees<14.92	Post_ Downward.Flexion_Head<28.25	Post_ Backward_TorsoHead<-2.35	Post_ ThreeD_Lelbow.Relbow<49.97	Post_ Downward.Flexion_Head<22.94	Neutral	8,33%	47,62%
						Post_ Downward.Flexion_Head>=22.94	Sadness	10,44%	88,89%
					Post_ ThreeD_Lelbow.Relbow>=49.97	Post_ Downward.Flexion_Torso<1.94	Sadness	2,61%	50,00%
						Post_ Downward.Flexion_Torso>=1.94	Neutral	28,33%	72,34%
				Post_ Backward_TorsoHead>=-2.35	Post_ Inside_LowerBody<-3.64		Shame	24,07%	60,47%
					Post_ Inside_LowerBody>=-3.64		Sadness	20,87%	72,73%
			Post_ Downward.Flexion_Head>=28.25				Sadness	38,70%	90,82%
		Post_ Downward.Flexion_Knees>=14.92	Post_ ThreeD_Lfoot.Rfoot<12.38				Shame	56,48%	57,55%
			Post_ ThreeD_Lfoot.Rfoot>=12.38	Post_ Backward_TorsoHead<-1.25			Joy	12,16%	33,96%
				Post_ Backward_TorsoHead>=-1.25	Post_ Forward_TorsoHead<34.22		Anxiety	16,85%	83,33%
					Post_ Forward_TorsoHead>=34.22		Panic Fear	28,57%	70,00%
	Speed_ Vertical_Arms>=0.96	Speed_ Sagittal_TorsoHead<0.26	Post_ Forward_TorsoHead<26.13	STD_ ThreeD_Feet.Hips<0.038			Neutral	10,83%	76,47%
				STD_ ThreeD_Feet.Hips>=0.038	Post_ Upward_Arms<36.24	Post_ ThreeD_Feet.Hips<96.67	Pride	15,44%	70,00%
						Post_ ThreeD_Feet.Hips>=96.67	Neutral	10,83%	50,00%
					Post_ Upward_Arms>=36.24		Pride	56,62%	85,56%
			Post_ Forward_TorsoHead>=26.13	Post_ Downward.Flexion_Head<21.99			Anxiety	15,73%	77,78%
				Post_ Downward.Flexion_Head>=21.99			Shame	3,70%	28,57%
		Speed_ Sagittal_TorsoHead>=0.26	Post_ Downward.Flexion_Torso<4.95				Joy	54,73%	58,70%
			Post_ Downward.Flexion_Torso>=4.95				Panic Fear	14,29%	60,87%
Acc_ ThreeD_Rhand.Head>=0.49	Speed_ Vertical_Arms<1.77	Post_ Downward.Flexion_Torso<5.85					Anger	6,91%	38,64%
		Post_ Downward.Flexion_Torso>=5.85					Panic Fear	14,29%	100,00%
	Speed_ Vertical_Arms>=1.77	Post_ ThreeD_Lhand.RHand<42.47					Panic Fear	9,18%	64,29%
		Post_ ThreeD_Lhand.RHand>=42.47					Anger	78,86%	89,82%

Figure 11.18: Decision tree rules for the characterization of expressed emotions in Lifting and Throwing actions

11

of elbows movement and low forward torso and head posture. We observe from this Figure that most of the decision rules - based on these two characteristics - leads to the classification of samples as Neutral or Pride. Thus, Pride and Neutral expressions across all the actions share common characteristics. In chapter 8, we showed that Pride and Neutral were confused in "Emotion Perception" task.

Figure 11.20 shows the decision rules that are described through low acceleration of elbows movement and high forward torso and head posture. We observe that most of the decision rules - based on these two characteristics - lead to the classification of samples as Shame, Sadness or Neutral. Few rules are associated to Joy or Panic Fear, but the last split(s) mostly involve high ThreeD openness of elbows posture, high lateral openness of lower body posture and high speed/ acceleration of head movement. Indeed, Sadness and Shame expressions share common characteristics at the level of the acceleration and the speed of elbows motion, forward torso posture and downward flexion of the head. In chapter 8, we also showed that Sadness and Shame were confused in "Emotion Perception" task. Besides, we showed in section 11.1.2 that the characterization patterns of Sadness and Shame are highly similar.

Figure 11.21 shows the decision rules that are described through high acceleration of elbows movement and high or low forward torso and head posture. We observe that most of the decision rules presented in this figure refer to the classification of samples as Anger, Joy and Panic Fear. Several decision rules are associated with Anger expression. Most of them represent a small subset of Anger observations. However, about one third of Anger observations (33.91%) are characterized with high probability of Anger (83.94%) by high acceleration of elbows movement, high forward torso posture and high speed of elbows flexion.

11.2.6.3 Discussion

We note that the classification of expressed emotions achieve better results using an ensemble of trees: the CCR_{OOB} obtained from Random Forest model is equal to 88%, 87%, 77% and 87% respectively for SW+WH, MB+KD, SD+BS and Lf+Th group of actions against 68%, 62%, 56% and 67% obtained from a 10 folds cross-validation applied on a single Decision Tree model. However, we used a single Decision Tree model as a "white box" model that allows us providing few insights into what can go behind the classification of bodily expression of emotions.

Decision Tree models are known to be unstable to the changes in the training dataset. Thus, we did not focus on the exact value at which the split is performed. Besides, due to their instability, the variable chosen for the split of any branch can change if the tree is grown on a slightly different dataset. To overcome this issue, we used a subset of the most relevant features returned by Random Forest model as described in Chapter 10. Besides, we observed that the selection of features performed by Decision Tree model is congruent with the ranking of features provided by Random Forest model (the ranking of features were presented and discussed in Chapter 10). Indeed, the feature selected at the root of Decision Tree models always correspond to one of the most relevant features assigned by Random Forest model. Besides, the features used for the last splits to be associated with a particular emotion are congruent with the features that are the most relevant for this emotion (e.g. lateral closeness of lower body limbs which is relevant for Shame expression).

Decision Tree models also mostly suffer from the risk of overfitting the training dataset. We applied a post-pruning process in order to discuss the decision rules generated from a pruned tree. Despite the post-pruning process, we obtained a highly forked tree for the classification of emotions across all the actions. This can be explained as the tree tries to fit at best the characterization of each expressed emotion which may somehow depend upon the action (as discussed in section 11.1.3). Based on the pruned tree built across all the actions, we observed that some of the emotions that were confused in "Emotion Perception" task (See Chapter 8) share common characterization rules such as Shame and Sadness, and Pride and Neutral.

The straightness, the least forward leaning, and the frequent backward leaning of torso movement are often associated with the expression of Pride, Joy but also with Neutral. The bending and the forward leaning of torso movement are often associated with the expression of Anxiety, Sadness, Shame, Anger and Panic Fear. In section 11.1.2, in both perceptual and motion capture characterization, we also showed that the highest mean ratings of the collapse and the leaning of the torso are attributed to "negative" emotions in the Emilya database, while the highest mean ratings of the straightness and the backward leaning of the torso are attributed to "positive" and neutral emotions. This result suggests that the straightness and the leaning of torso posture can be somehow associated to the valence dimension to discriminate between bodily expression of "negative" and "positive" emotions when performing daily actions. As the set of "positive" emotions that we considered in Emilya database is restricted to Pride and Joy, the study of emotional body expression of more "positive" emotions is needed to generalize this hypothesis.

Anger expression is mostly characterized and differentiated from other emotions based on high values of speed and acceleration features. In chapter 10 (Figure 10.11), we showed that the acceleration of arms movement is always considered as highly relevant for Anger expression across different actions.

11.3 Conclusion

In this chapter, we proposed two approaches to explore the interpretation of emotions characterization based on the selected subset of features obtained in the previous chapter. The first approach refers to a simplified graphical representation of emotions characterization patterns based on the mean ratings of features. This approach allows us exploring the difference in the characterization of expressed emotions using both perceptual and motion capture mean-ratings. It also allows us exploring the effect of action on the characterization of each expressed emotion. The second approach refers to a set of rules derived from a Decision Tree model. This approach allows us exploring the common features that expressed emotions can share as well as the features that allow discriminating them.

The patterns of expressed emotion characterization based on the mean ratings of features reveals interesting finding. Across all the actions, the patterns of Shame and Sadness are highly similar. Besides, we found tha Shame and Sadness share common characterization rules based on the decision rules inferred from each Decision Tree model according to each group of "similar" actions and across all the actions. These results may explain the confusion that occurred in "Emotion Perception task" and in automatic classification of Sadness and Shame (See Chapter 9). However, Shame expression is slightly distinguishable from Sadness by few features such as the lateral closeness of lower body limbs (higher value is attributed to Shame expression) and downward head flexion (a lower value is attributed to Shame expression).

While Anger expression is significantly dissociated from the other expressed emotions based on the mean rating of posture Openness, posture Straightness, movement Fluidity and Power in perceptual characterizations, it is mainly dissociated from the other expressed emotions based on the speed and acceleration of arms movement in motion capture characterizations (based on mean ratings and decision rules). Although restricted to the maximal value of speed and acceleration, this finding highlights the importance of movement dynamics in Anger expression. Further analyses of movement dynamics properties are needed to better explore the dynamics of bodily expression of Anger.

In both perceptual and motion capture characterizations, we observe the presence of certain clusters of emotions. For instance, the straightness and the forward/backward leaning always differentiate Pride, Joy and Neutral from the other expressed emotions in the Emilya database. As Pride and Joy are the only "positive" emotions considered in the Emilya database, we suggest that the straightness and the leaning of the torso could be used to dissociate bodily expression of "positive" and "negative" emotions, but further bodily expressions of "positive" emotions (such as amusement) must be explored to confirm or deny this hypothesis.

In addition to what we found regarding the expression of emotions across all the actions, we also explored the characterization of emotions expressed in each action. We found that some actions affect all the expressed emotions in the same way. For instance, the mean rating of feet openness is always higher in walking actions for all the emotions expressed in the Emilya database. We also found that the expression of some emotion seems more associated to some specific actions. For instance, Anger expression receives a particular patterns of characterization in Throwing action.

11.4 Summary of chapter

- We provide the patterns of motion capture characterization of expressed emotions across different actions based on the subset of the most relevant features. Expressed emotions can be clearly distinguished based on the mean ratings of features.
- We also compare the perceptual and the motion capture characterizations of

expressed emotions across all the actions. A high amount of similarity between perceptual and motion capture characterizations is observed.

- We discuss the characterization of expressed emotions in each action. While some actions affect the patterns of certain features in the same manner across different emotions, some actions receive a particular pattern of characterization only in some specific emotions (e.g. the characterization pattern of Anger expression in Throwing action is particularly different from the other actions).
- We propose a three-steps based pruning process to obtain an optimal Decision Tree model. The final goal is to discuss the decision rules inferred from such an optimal tree. We discuss what features are specifically characteristics of one emotion (e.g. lateral openness of feet receives the lowest value for Shame expression) and which are common to a group of emotions (e.g. the straightness of body posture differentiate Pride, Neutral and Joy from the other expressed emotions).

Level1	Level2	Level3	Level4	Level5	Level6	Level7	Level8	Level9	Level10	Emotion	Observ	Prob
Acc_ ThreeD_ Lelbow.Relbow < 0.10224	Post_ Forward_ TorsoHead < 17.1541	Post_ LRRotation_ Head < 8.6893	Acc_ Lateral_ Head < 0.024402							Neutral	18,62%	63,30%
			Acc_ Lateral_ Head>=0.024402	Post_ Outside_ LowerBody < 8.28						Pride	8,59%	83,08%
				Post_ Outside_ LowerBody>=8.28	SymOcc_ ThreeD_ Lelbow.Relbow < 51.8032	Post_ Forward_ TorsoHead < 4.2689				Pride	3,82%	100,00%
						Post_ Forward_ TorsoHead>=4.2689	Speed_ Flexion_ Elbows < 0.045			Pride	5,25%	55,93%
							Speed_ Flexion_ Elbows>=0.045			Neutral	2,82%	41,86%
					SymOcc_ ThreeD_ Lelbow.Relbow>=51. 8032	Post_ Inside_ Arms < -0.28631	Speed_ Lateral_ Arms < 0.154			Shame	2,31%	66,67%
							Speed_ Lateral_ Arms>=0.154	Post_ Backward_ TorsoHead < - 8.7014	STD_ LRRotation_ Torso < 1.5945	Sadness	1,87%	42,86%
									STD_ LRRotation_ Torso>=1.5945	Pride	5,09%	49,23%
								Post_ Backward_ TorsoHead>=- 8.7014		Neutral	8,45%	64,29%
						Post_ Inside_ Arms>=-0.28631				Neutral	13,93%	80,18%
		Post_ LRRotation_ Head>=8.6893	Post_ Forward_ TorsoHead < 12.65	Post_ Downward. Flexion_ Torso < 8.9355						Pride	28,78%	79,74%
				Post_ Downward. Flexion_ Torso>=8.9355						Joy	2,18%	38,24%
			Post_ Forward_ TorsoHead>=12.65	Post_ Downward.Flexion_ Torso < 2.148						Pride	4,61%	63,04%
				Post_ Downward. Flexion_ Torso>=2.148						Neutral	4,69%	37,98%

Figure 11.19: Part 1: Decision tree rules for the characterization of expressed emotions across all the actions

11

Level1	Level2	Level3	Level4	Level5	Level6	Level7	Level8	Level9	Emotion	Observ	Prob
Acc_ ThreeD_ Lelbow.Relbow < 0.10224	Post_ Forward_ TorsoHead >=17.1541	Post_ Downward.Flexion_ Head < 24.5532	Speed_ ThreeD_ Lelbow.Relbow < 0.27	Post_ Inside_ LowerBody < -2.4646					Shame	37,50%	49,62%
				Post_ Inside_ LowerBody>=-2.4646	Post_ ThreeD_ Lhand.Head <				Sadness	5,60%	50,47%
					Post_ ThreeD_ Lhand.Head>=80.39	Speed_ ThreeD_ Rhand.Head < 0.84854			Neutral	7,04%	52,94%
						Speed_ ThreeD_ Rhand.Head>=0.84854			Shame	1,54%	33,33%
			Speed_ThreeD_ Lelbow.Relbow>=0.27	Acc_ LRRotation_ Head < 0.056091	Post_ Inside_ Arms < -12.9185				Sadness	4,87%	32,64%
					Post_ Inside_ Arms>=-12.9185	Post_ Backward_ TorsoHead < -1.3996	Acc_ Flexion_ Elbows < 0.0045132		Neutral	6,73%	58,90%
							Acc_ Flexion_ Elbows>=0.0045132	Post_ Outside_ Arms < 44.8273	Neutral	5,01%	43,84%
								Post_ Outside_ Arms>=44.8273	Pride	3,82%	58,54%
						Post_ Backward_ TorsoHead>=-1.3996			Anxiety	6,33%	30,23%
				Acc_ LRRotation_ Head>=0.056091	Post_ Downward.Flexion_ Torso < 5.5843				Joy	13,76%	32,54%
					Post_ Downward.Flexion_ Torso>=5.5843				Panic Fear	6,06%	30,11%
		Post_ Downward.Flexion_ Head>=24.5532	Speed_ LRRotation_ Head < 0.39042	Post_ Inside_ LowerBody < -6.2996					Shame	8,85%	52,27%
				Post_ Inside_ LowerBody>=-6.2996	Speed_ ThreeD_ Hands.Hips < 0.60633	Speed_ ThreeD_ Feet.Hips < 0.69416	Post_ LRRotation_ Head < 7.5256		Sadness	38,55%	86,31%
							Post_ LRRotation_ Head>=7.5256	STD_ ThreeD_ Feet.Hips < 0.5302	Sadness	3,42%	73,33%
								STD_ ThreeD_ Feet.Hips>=0.530	Shame	3,27%	70,83%
						Speed_ ThreeD_ Feet.Hips>=0.69416			Shame	2,31%	57,14%
					Speed_ ThreeD_ Hands.Hips>=0.6063 3	Post_ Backward_ TorsoHead < -0.9096			Sadness	6,53%	55,26%
						Post_ Backward_ TorsoHead>=-0.9096			Shame	3,46%	50,00%
			Speed_ LRRotation_ Head>=0.39042	Post_ Outside_ LowerBody < 12.9838					Shame	9,23%	57,14%
				Post_ Outside_ LowerBody>=12.9838	Post_ ThreeD_ Elbows.Hips < 47.3064				Neutral	2,35%	39,47%
					Post_ ThreeD_ Elbows.Hips>=47.30				Joy	2,35%	34,15%

Figure 11.20: Part 2: Decision tree rules for the characterization of expressed emotions across all the actions

11

Level1	Level2	Level3	Level4	Level5	Level6	Level7	Level8	Level9	Emotion	Observ	Prob
Acc_ ThreeD_ Lelbow.Relbow>=0.10224	Post_ Forward_ TorsoHead < 21.8379	STD_ Vertical_ LowerBody < 1.5903	Acc_ ThreeD_ Hands.Hips < 0.47349	STD_ Backward_ TorsoHead < 1.4718	Acc_ ThreeD_ Lhand.RHand < 0.46789				Sadness	2,07%	41,67%
					Acc_ ThreeD_ Lhand.RHand>=0.46789				Anger	1,96%	29,63%
				STD_ Backward_ TorsoHead>=1.4718	Post_ Downward.Flexion_ Torso < 5.8506	Post_ Outside_ LowerBody < 11.5083			Pride	4,13%	63,42%
						Post_ Outside_ LowerBody>=11.5083			Joy	4,70%	56,00%
					Post_ Downward.Flexion_ Torso>=5.8506				Panic Fear	1,52%	43,75%
			Acc_ ThreeD_ Hands.Hips>=0.47349						Anger	7,59%	71,26%
		STD_ Vertical_ LowerBody>=1.5903							Joy	26,85%	72,07%
	Post_ Forward_ TorsoHead >=21.8379	Speed_ Flexion_ Elbows < 0.086444	Post_ Forward_ LowerBody < 49.1578	Speed_ Sagittal_ LowerBody < 3.6609	Post_ Downward.Flexion_ Knees < 7.7146				Anger	6,73%	83,33%
					Post_ Downward.Flexion_ Knees>=7.7146	Post_ Backward_ TorsoHead < -4.578			Anger	5,39%	36,98%
						Post_ Backward_ TorsoHead>=-4.578	Post_ Downward.Flexion_ Torso < 4.395	Post_ ThreeD_ Feet.Hips < 96.7961	Joy	3,52%	28,77%
								Post_ ThreeD_ Feet.Hips>=96.7961	Anger	4,77%	68,42%
							Post_ Downward.Flexion_ Torso>=4.395	Speed_ LRRotation_ Head < 0.30441	Anxiety	3,16%	44,83%
								Speed_ LRRotation_ Head>=0.30441	Panic Fear	20,56%	69,85%
				Speed_ Sagittal_ LowerBody>=3.6609					Panic Fear	11,26%	100,00%
			Post_ Forward_ LowerBody>=49.1578						Anger	20,56%	59,79%
		Speed_ Flexion_ Elbows>=0.086444							Anger	33,91%	83,94%

Figure 11.21: Part 3: Decision tree rules for the characterization of expressed emotions across all the actions

Part V : Conclusion

12 Contributions and future work

This thesis falls within the framework of Affective Computing (AC). AC combines several disciplines including psychology and computer science to study the recognition, the interpretation and the synthesis of affect. In this thesis, we focus on the recognition and the interpretation of emotions expressed through body movement. For a long time, research in AC has focused on vocal and facial expressions of emotions. Despite the growing interest in bodily expression of emotions, AC research has been widely focused on the communication of prototypical expression of emotions. Implicit expression of emotions during daily movement has been so far limited to specific actions such as Knocking.

Four main contributions are achieved in this thesis. Firstly, we proposed a body movement notation system that allows the description of expressive body movement across various body actions. Second, we collected a new database of emotional body expression in daily actions. This database constitutes a large repository of bodily expression of emotions including the expression of 8 emotions in 7 actions, combining video and motion capture recordings and resulting in more than 8000 sequences of expressive behaviors. Thirdly, we explored the classification of emotions based on this notation system. We also compared the automatic classification of emotions with human perception of emotions expressed in different actions. Finally, we extracted the most relevant features that capture the expressive content of the motion. We discussed their relevance according to the actions and emotions, and we explored the characterization of emotional body expression based on the most relevant body cues.

In the following of this chapter, we summarize the work conducted in this thesis and we highlight our contributions. We end by reporting the limitations of our work and discussing several future directions of it.

12.1 Summary of thesis

In this section, we provide a summary of this thesis.

12.1.1 Body movement notation system

Our aim is to characterize expressive body movement in different actions. That is, we aim to describe implicit bodily expression of emotions (e.g. emotion expression during walking, knocking...). Different body movement notation systems have been proposed in the literature to describe body movement as discussed in Chapter 2. Due to the complexity of body structure and the variety of applications that make use of body movement notation system (e.g. dance, social interaction, emotion expression), there is a lack of consensus on a common notation system for the description of body movement. Besides, we concluded in Chapter 2 that there is a need to offer an acceptable compromise between 1) a fine-grained description of bodily expression as proposed in BAP notation system [Dael et al., 2012] and 2) a description of movement quality as offered in Laban notation system [Laban, 1988] to deeply study how expressed emotion modulates body movement.

In our work, we proposed a Multi-Level body movement notation system (MLBNS) that is inspired from Laban [Laban, 1988] and BAP [Dael et al., 2012] coding systems. Our Multi-Level notation system is intended to characterize emotional body expressions in daily actions, while offering a trade-off between a detailed description of the whole body movement and a comprehensive illustration of the way emotion expression modulates body movement. We distinguish 3 main description levels in our Multi-Level body movement notation: Anatomical, Directional and Posture/ Movement. **Anatomical** level describes the body segments considered for the characterization of expressive posture or movement. We distinguish three anatomical description levels: Global (e.g. bounding box surrounding the whole body/ specific body parts), Semi-Global (e.g. coordination between two body segments) and Local (e.g. rotation of a particular body segment). **Directional** level describes the directions of movement. We differentiate 5 directions: Sagittal (forward/ backward), Lateral (left/ right sides), Vertical Length (upward/ downward), Vertical Rotation (to the left/ to the right) and finally Three dimensional direction. **Posture/ Movement** level differentiates posture features, postural changes and movement dynamics features.

12

Based on our Multi-Level body movement notation system, we defined a large set of 114 motion capture features. A first study of expressive walking analysis showed the ability of this notation system to discriminate between different styles of walk. Besides, as discussed in section 12.1.3, we showed that this set of 114 motion capture features allows the classification of emotions expressed in different daily actions.

The set of 114 motion capture features is described in details in Appendix E. The description of each body feature is performed according to the description levels considered in our Multi-Level notation system. Besides, we provide a visual description of each feature according to a 3D virtual puppet (See Appendix E).

A subset of our Multi-Level notation system was also used in Chapter 8 to define a set of perceptual body cues that can be rated by humans during a perceptual study.

The contribution of our Multi-Level notation system is four-fold:

– Firstly, it provides a good compromise between the accuracy of expressive movement representation (through the description of body segments of the whole body) and the ability to describe movement quality for the characterization of implicit expression of emotion (that is how emotion expression affects the way we perform a movement).
– Secondly, it can be used to describe expressive movement in both perceptual studies (e.g. perceptual body cues rating, See Chapter 8) and kinematic analyses (e.g. using motion capture data, See Chapter 9).
– Thirdly, it can be used to describe emotional body expression in various daily actions such as walking, knocking, and sitting down (See Chapter 9).
– Fourthly, it provides a comprehensive description of expressive body movement allowing the interpretation of emotional body expression characterization (See Chapter 10 and Chapter 11).

12.1.2 Emilya database collection and validation

Our work relies on the analysis of emotional body expressions. In Chapter 3, we discussed the methodologies and the issues of collecting bodily expression of emotions and we provided an overview of existing databases. We concluded that content and the methodology of recording in existing databases are mostly constrained by specific research directions. Besides, existing databases are reduced to a small subset of emotion categories and/ or a limited range of motions (often reduced to one particular movement task). As we aim to study the expressions of emotions in different daily actions, we created a new database of expressive movements. The collection of this database is described in Chapter 7.

Our database, called Emilya for EMotional body expression In daiLY Actions, constitutes a new rich repository of expressive body movements of more than 8000 video and motion capture files. Eleven actors participated in collecting 8 emotions expressions in 7 daily actions.

The emotions are: Anxiety, Pride, Joy, Sadness, Panic Fear, Shame, Anger and Neutral. These emotions are selected to cover the arousal and the valence levels. We focus on 8 emotions composed of a subset of basic emotions and of pride, shame and anxiety. A scenario-based approach was used to elicit these emotions, where three scenarios are proposed for each emotion (except for Neutral for which only two scenarios were proposed).

The actions performed by the actors are: Simple Walking, Walking with an object in the hands, Knocking, Moving Books on a table, Sit Down (split into Sitting Down and Being Seated in kinematic analysis), lifting and Throwing. These actions cover body movements that involve the whole body movement (walking and sitting down), repetitive arms movements (moving books and knocking) and non-repetitive arm movement (lifting and throwing). The actors were asked to repeat each action 4 times in order to capture a large set of data.

Besides, in order to avoid exaggerated behaviors, the actors worked with a professional acting director. Seven training sessions were given to the actors where the goal was to make the actors aware about how to use their body to express emotions in daily actions. However, no specific instructions were provided to the actors. Each actor was free to express the proposed emotions as s/he wished to.

During the recording sessions, we recorded both audio-visual data from 4 viewpoints (using 4 Canon cameras) and motion capture data using Xsens motion capture system [Roetenberg et al., 2009]. Besides, video and motion capture recordings were synchronized. Long motion sequences were recorded for each emotion expression in order to foster the continuity of expression at the cost of extensive post-processing steps. Several post-processing steps were performed to structure the database. Post-processing steps consist in the segmentation of each emotion expression sequence into individual segments according to each action (e.g. separate walk from sit down action) and each repetition (e.g. separate individual repetitions of the same action). These post-processing steps produced sorted motion sequences that are appropriate for perceptual studies and kinematic analyses.

After the collection and the post-processing of Emilya database, a perceptual study was conducted to validate the reliability of emotions expressions. The procedure and the results of this perceptual study are described in Chapter 8. More than one thousand participants took part in this study to conduct two tasks: “Emotion Perception” and “Body Cues Rating” tasks . The first task is useful to ensure the matching between the expressed and the perceived emotion. The second task allows the characterization of emotional body expressions. Two critical features were addressed for this perceptual study: the labeling approach and the creation of stimuli. A multi-labeling approach was adopted allowing a detailed description of emotions perception. Stimuli consists in movies of a virtual puppet reproducing motion capture data. We offered different solutions for the automatic creation of these stimuli as described in Appendix C. The solution adopted to create the stimuli for our perceptual study allows the perception of the whole sequence of motion from a side viewpoint set automatically according to the gaze direction of the virtual puppet. This solution also allows perceiving the walking motion of the virtual puppet at the same distance.

The contribution of Emilya database is multi-fold:

- Emilya database can be used to build computational models based on machine learning techniques as it offers a large variety of expressions (i.e. 3 scenarios * 4 repetitions for each expression acted by 1 actor in a single action).
- Emilya database encompasses a larger variety of emotion categories (8 emotions) and actions (7 actions) compared to previous existing databases as they offered either a large set of emotions or actions. The variety of emotions and actions provided in the same database is useful to conduct comparative experiments based on the performances of the same actors.
- Emilya is a multi-media database that contains synchronized video and motion capture recordings.
- Emilya database is labeled through a perceptual study where the participants

were asked 1) to recognize the emotions expressed by the actors and reproduced on a virtual puppet (i.e. "Emotion Perception" task, See Chapter 8) and 2) to characterize the expressive movement through perceptual rating of 8 body cues (i.e. "Body Cues Rating" task, See Chapter 8).

12.1.3 Multi-Level classification of emotional body expressions

Based on a database of emotional body expression in various daily actions and a Multi-Level notation system, we were able to study the classification of emotions in various movement tasks. Random Forest (RF) approach is applied to classify expressed emotions using our set of 114 motion capture features. The use of Random Forest approach in our work is double-fold. Firstly, RF is known as a reliable and robust ensemble classification model [Breiman, 2001]. Secondly, RF returns a measure of relevance for each feature, which has been used in our work to select the most relevant expressive features and to study their usefulness according to each expressed emotion (See Section 12.1.4).

Using RF approach, we conducted a series of Multi-Level classification and comparison: 1) Classification of expressed emotions across different actions, 2) Comparison of automatic classification and human recognition of emotions, 3) Comparison of the contribution of different description levels to the classification of emotions and 4) Analysis of the temporal profiles of body cues.

Classification of expressed emotions across different actions:

We compared the classification rates of emotions classification across various actions. We found that the classification achieves the best scores in walking action, followed by arms movement actions (knocking, moving books, lifting and throwing). The classification of emotions showed the worst (but above chance level) scores in Sitting Down and Being Seated actions. Previous studies on seated posture analysis used to be based on Kinetic data provided from pressure sensors. However, such data do not achieve high recognition scores when dealing with multi-class problem. Thus we conclude that combining Kinematic and Kinetic data of emotion expression in seated posture would provide better classification accuracy.

We also compare the classification rates of each expressed emotion across various actions. We found that most of the emotion expressions are the best classified in Walking action (Simple Walking or Walking with an object in hand). However, we also found that Panic Fear, Shame and Anger are respectively the best classified in Moving Books, Knocking and Throwing. Throwing action also received the best classification rate for Anger expression. Thus, we conclude that Anger expression is the best conveyed in Throwing action and that the motor behavior of Throwing action foster at best the expression of Anger in the Emilya database.

We also studied the effect of individuality on the classification of expressed emotions. To achieve this purpose, we repeat, across all the actors, the classification of one actor' expressions while training the model with the other actors' expressions. We found that inter-individual classification always achieve above chance scores. We concluded that the classification of expressed emotions in the Emilya database can be generalized to expressions of unknown actors above chance level.

Comparison of automatic classification and human recognition of emotions:

We compared three classification approaches: 1) RF classification of expressed emotions (based on motion capture data), 2) human recognition of emotions in "Emotion Perception" task, and 3) MLR (Multinomial Logistic Regression) classification of emotions perceptually characterized in "Body Cues Rating" task.

We found that RF classification based on 114 motion capture features outperforms MLR classification based on the perceptual rating of 8 body cues. Perceptual ratings are intended to describe the motion of the whole body (they do not consider the subtle changes in body posture and movement) whereas motion capture features describe the posture and the movement of several body segments. Besides, "Body Cues Rating" task results in discrete measures ranging from 1 to 5 (1,2,3,4,5), whereas motion capture features refer to rotational and positional values with a larger range of values.

We also found that, across all the expressed emotions, RF classification based on motion capture data mostly outperforms the human recognition of emotions achieved in "Emotion Perception" task. A similar result was found across all the actions except for Sadness and Neutral as their recognition in perception task achieve similar scores as obtained by RF classification. Overall, in both human recognition and automatic classification, Sadness, Anger and Neutral receive the best scores. Across all the actions, low recognition rates in "Emotion Perception" task were mostly due to the presence of strong confusions in the perceptual recognition; Pride was perceived as Neutral, Shame as Sadness and Panic Fear as Anxiety.

We observed that the recognition of emotion based on "Emotion Perception" is better than the recognition of emotions based on "Body Cues Rating" except in few emotions (Pride, Panic Fear and Shame) that are strongly confused with others in "Emotion Perception" task.

Comparison of the contribution of different description levels to the classification of emotions:

Based on our Multi-Level notation system, we studied the contributions of different description levels to the classification of emotions in "associated" actions (that is the actions in which the emotions are best classified). We compared the contribution of different description levels from Anatomical, Directional and Posture/ Movement level. We found that upper body features provide better classification rates in arms movement based actions. However, we also observed that the classification of Pride and Sadness in walking provide similar results using upper or lower body parts. We also found that posture features provide better classification rates than postural changes and movement dynamics, but movement dynamics features are as important as postural features for Anger expression in Throwing action. The classification of emotions based on features from Vertical (Length) or 3D directions provide better classification rates than the classification based on features from Sagittal or Lateral directions.

Analysis of the Temporal Profiles of body cues:

Finally, we analyzed the Temporal Profiles (TP) of end effectors trajectory and energy. Temporal Profile features are measured as proposed in the work of Castellano et al. [Castellano et al., 2008]. We compared the classification rates provided with our set of 114 Multi-Level features and the classification rates provided with the new set of TP features. We also studied the combination of our set of 114 Multi-Level features with the new set of TP features. We found that TP features do not lead to a significant increase of the classification rates obtained with our set of 114 Multi-Level features. We also found that, except for Anger and Joy, the TP based on the trajectory of end-effectors motion contributes better than the TP based on the energy of end-effectors motion.

Summary:

The results of Multi-Level classification can be summarized as follows:

– Across all the expressed emotions, Walking is the most expressive action in the Emilya database. It received the best classification rate for the 8 emotions taken together.
– Across all the expressed emotions, Sitting Down and Being Seated are the least expressive actions in the Emilya database. They received the least classification rate of the 8 emotions classification. Combining Kinematic and Kinetic data would be useful to obtain better classification accuracy.
– The emotions Anxiety, Pride, Joy, Sadness and Neutral are the best classified in Walking action; but Panic Fear is the best classified in Moving Books action, Shame in Knocking and Anger in Throwing action.
– The classification of expressed emotions in the Emilya database can be generalized to expressions of unknown actors above chance level.
– The classification of emotions based on 114 motion capture features outperforms the classification of emotions based on 8 perceptual body cues rating.
– Except for Sadness and Neutral recognition, motion capture classification of emotions outperforms human recognition of emotions in “Emotion Perception” task across all the actions.
– Except Panic Fear, Shame and Pride, emotion recognition in “Emotion Perception” task outperforms emotion classification based on perceptual body cues ratings (“Body Cues Rating” task) across all the actions.
– Anger is the best recognized in Throwing action according to Kinematic analysis, emotion perception task, and emotion classification through perceptual body cues rating.
– Pride and Joy are the best recognized in Walking actions according to Kinematic analysis, emotion perception task, and emotion classification through perceptual body cues rating.
– Upper and Lower body features contribute similarly to the classification of Sadness and Pride in Walking.
– Except for Anger, posture features contribute highly better than maximum

12

values of speed and acceleration to the classification of emotions in "associated" actions.

- The classification of emotions based on features from Vertical (Length) or 3D directions provide better classification rates than the classification based on features from Sagittal or Lateral directions.
- The classification of emotions using the set of Multi-Level features achieve better results than their classification using features that describe the Temporal Profile of body end-effectors trajectory and energy. Comparing trajectory and energy based features, we found that Anger and Joy expressions are mostly better recognized using energy features than using trajectory features.

12.1.4 Identification of the most relevant expressive body cues

Based on Random Forest (RF) approach, we combined an Embedded feature ranking method and a Wrapper feature selection method to select the most relevant expressive body cues. This approach is described in Chapter 10. Feature selection approach was evaluated according to two criteria; 1) the classification rates before and after FS process and 2) the dependency of two subsets of selected features obtained respectively from a training and a testing datasets. Our feature selection approach leads to slightly better results in term of the classification rates. However, we show that RF based FS approach is stable enough to produce significantly similar subsets of features when applied on a training and a testing datasets.

As we aim to study the relevance of body features across different actions, we combine the results of selected features across "Similar" actions through an intersection of features. We also conduct our feature selection approach across all the actions to discuss the relevance of each selected feature to expressed emotions. The permutation measure returned by RF approach was used to assign a relevance measure to the selected features. We discussed features ranking according to their relevance measure.

We found that some features appear to be relevant for the classification of all emotions in all actions (e.g. forward/backward torso posture), others turn out to be specific to the expression of some emotions in some actions (e.g. elbows posture symmetry). We also identified relevant expressive features that do not necessary describe the main movement involved in the action. For instance, the posture of lower body limbs appears to be highly relevant for the expression of Shame and Sadness in Lifting and Throwing actions. This result highlights the need to go beyond action-dependent notation systems.

Based on the subset of features obtained from our feature selection approach, we explored the characterization of emotional body expressions through two approaches: 1) mean ratings based approach that allows the comparison between motion capture and perceptual characterization, and 2) a Decision Tree model that allows the discussion of the decision rules that characterize expressed emotions. We proposed a three step-based pruning process to obtain an optimal Decision Tree model. The final goal is to discuss the decision rules inferred from such an optimal tree.

The results of expressed emotion characterization can be summarized as follows:

- Relevant features are not restricted to the main segments involved in the movement task. For instance, the posture of lower body limbs appears to be highly relevant for the expression of Shame and Sadness in Lifting and Throwing actions.
- The patterns of perceptual and the motion capture characterizations of expressed emotions across all the actions are highly similar
- Based on the subset of the most relevant features, the patterns of expressed emotions can be clearly distinguished based on the mean ratings of motion capture and perceptual features. However, Sadness and Shame patterns are highly overlapped. They are mainly distinguishable through two features: head downward flexion (higher flexion observed for Sadness) and lateral closeness of feet (higher closeness observed for Shame).
- Two groups of patterns were clearly dissociated based on speed and acceleration features: the first group of patterns includes the characterization of "high arousal" emotions (Anger, Pride, Joy and Panic Fear) while the second group includes the characterization of "low arousal" emotions (Sadness, Shame and Neutral). Similarly, emotions characterization patterns based on postural changes features provides somehow a discrimination between "high arousal" (Pride, Joy, Anger) and "low arousal" (Shame, Sadness, Neutral) expressions.
- The forward/backward leaning and the straightness of the torso allows discriminating between "positive" emotions (Pride, Joy, but Neutral is mostly grouped with them) and "negative" emotions in the Emilya database.
- We discussed the characterization of expressed emotions in each action. While some actions affect the patterns of certain features in the same manner across different emotions (e.g. the speed of arm movement is mostly the highest in Knocking action), some actions receive a particular pattern of characterization only in some specific emotions (e.g. the characterization pattern of Anger expression in Throwing action is particularly different from the other actions).
- We discussed what features are specifically characteristics of one emotion (e.g. lateral openness of feet receives the lowest value for Shame expression) and which common to a group of emotions (e.g. the straightness of body posture differentiate Pride, Neutral and Joy from the other expressed emotions).

12

12.2 Limitations of our work

In the previous section, we summarized the work conducted in this thesis. While several contributions were achieved, a number of limitations can be revealed as discussed in this section.

12.2.1 Acted data

The collection of spontaneous expressions in daily actions is a challenging task, particularly due to the difficulty to record naturalistic emotions in specific movement tasks. Our work relies on acted expressions of emotions. The main advantage of acted data is the ability to clearly define the emotions expressed by the actors and to record multiple emotions expressions in controlled environment, which fosters the high quality of data structuring. The reliability of acted data has been explored through a perceptual study. However, the performance of an emotion recognition model based on acted data in real life situations remains unknown. Indeed, the analysis of acted data limits the understanding of emotion expression characterization as no clear insights are provided to enable the comparison between acted and naturalistic expressions. The connection between acted and naturalistic bodily expressions remains an open research problem.

Besides, another challenge in transferring the recognition of emotions from acted to real life scenarios concerns the measurement of body cues. Indeed, our set of motion capture features describes body movement in a fine-grained fashion thanks to the accuracy offered by a 3D motion capture system. In real world situations, there will not be possible to acquire data using motion capture, even when markerless motion capture systems are used. Besides, sensors-based motion capture systems still provide more accurate motion reconstruction than markerless motion capture techniques.

12.2.2 Feature Selection

In Chapter 10, we presented our Feature Selection approach that aims to identify a subset of the most relevant expressive features. This Feature selection approach is based on the ranking of features according to their relevance measure returned by Random Forest (RF) model. The permutation measure is used to quantify the relevance of features. However, a widely common issue that has been reported is the impact of features correlation on the permutation measure. Recent studies have shown that the presence of highly correlated features can seriously affect the measure of variable importance in RF [Strobl et al., 2008], [Tolosi and Lengauer, 2011] [Genuer et al., 2010]. Indeed, the subset of features, selected based on the permutation measure, includes the most relevant features, but it may also include other "weakly relevant" features that are correlated with the most relevant ones. As such, the subset of features selected by our approach may include correlated and redundant features.

12.2.3 Body movement notation system

Our body movement notation system allows the description of body movement inside the personal space that surrounds the body center. However, it does not include the description of body displacement in the environment. The displacement of the body in space is a critical body cue for the analysis of expressive body movement, in particular for the movement tasks that involve the displacement of the body such as walking and dancing. Besides, movement dynamics features are restricted to the maximum value of speed and acceleration and 6 Pearson correlation measures. Thus, our notation system is somehow coarse regarding the dynamics features of the movement as well as the features related to the spacial description level.

12.3 Perspectives

We present some perspectives we aim to undertake to overcome some of the limits we just described, as well as to continue new research directions.

12.3.1 Short term perspectives

12.3.1.1 Synthesis of bodily expression of emotions for 3D virtual characters

In order to give better insights of the characterization of bodily expression of emotions in the Emilya database, we studied the decision rules obtained from Decision Tree models based on the most relevant features (See Chapter 11). In addition to the interpretation purpose, another possible use of these rules is to explicitly describe how expressed emotions are characterized in the Emilya database for generating the animation of virtual characters. The set of rules derived from Decision Tree models could be used to model the expression of emotions in virtual agents. While postural features can be directly mapped into the generation model for the virtual agent's behaviors, the mapping between the maximum values of the speed, the acceleration of body movement and the dynamics properties of the generated movement is a challenging task. In fact, the maximum values of the speed and the acceleration are not enough to reproduce the movement dynamics of emotion expression. These statistical measures should be coupled with further movement dynamics features to allow reproducing the characteristics of motion capture dynamics. For instance, the instant in which occurs the maximal speed/ acceleration, the regularity of the speed/ acceleration during the motion sequence and the fluidity of the movement represents crucial features for the synthesis of virtual character's motion.

12.3.1.2 Recognition of emotions from multimodal expressions

Bodily expressions of emotions in Emilya database are collected through both video and motion capture techniques. So far, we only made use of motion capture data. One possible future direction of our work is to explore the recognition of the expressed emotions from facial and bodily expressions. It has been shown in previous psychological researches that combining facial and bodily expressions increases the

12

emotion recognition rates [De Gelder, 2006]. It is crucial to study whether the perception of facial expressions allows reducing the confusions of Shame, Pride, Anxiety and Panic Fear expressions as they received low recognition scores from the perception of bodily expressions (Anxiety was confused with Neutral, Panic Fear with Anxiety, Shame with Sadness and Pride with Neutral, See Chapter 8).

12.3.1.3 Feature selection

Our future selection approach combines an embedded and a wrapper strategies based on Random Forest model. Random Forest approach was used as an embedded feature ranking technique to rank the features according to their relevance measures. However, recent works reported the impact of features correlation on the permutation measure that we used for features ranking. Motivated by the sensitivity of permutation importance to highly correlated features, Strobl et al. [Strobl et al., 2008] proposed a new conditional variable importance for Random Forests. Hence, one future direction of this work is to repeat our future selection approach using a conditional variable importance measure as suggested in [Strobl et al., 2008]. In future work, it is also interesting to compare our feature selection techniques with others, such as greedy forward selection or greedy backward elimination [Narsky and Porter, 2014b].

12.3.2 Long term perspectives

12

12.3.2.1 Naturalistic expressions

Another future direction of our work concerns the collection of naturalistic expressions. One possible approach to record spontaneous bodily expressions could be to use computer games scenarios. The latter should include different movement tasks (such as dancing, playing tennis...) in order to enable the generalization of emotional body expression recognition and characterization across different movement tasks. The recognition and the characterization of naturalistic expressions can be then compared to the recognition and the characterization of acted data recorded in Emilya database.

12.3.2.2 Extension of body movement notation system

Our body movement notation system can be extended to include further components. For instance, Functional and Spacial description levels can be added in parallel with Anatomical, Directional and Posture/Movement description levels. The Spacial level allows the description of the dynamics of body movements when moving in the space that surrounds the environment (See Space Laban component described in Chapter 2). For instance, it allows the description of the speed of walk. Functional level allows including functional description of gestures as explained in Chapter 2.

As such, grouping functional description level with the other description levels allows the description of both explicit and implicit expression of emotions. Indeed, communicative gestures conveying emotional states (such as making a fist to express Anger or rubbing the head to express Shame) can occur during implicit expression of emotions (e.g. during walking).

The description of movement dynamics provided in our body movement notation system is very limited. It is mainly reduced to discrete features (i.e. maximum value of speed and acceleration). However, further movement dynamics body cues must be considered in addition to the speed and acceleration of movement such as the fluidity of movement. Discrete features of movement dynamics can be extended to include an extensive description of movement trajectory. In Chapter 9, we provided a comparative study between the classification of emotions using our body movement notation system and the classification of emotions using temporal profiles features proposed in [Castellano et al., 2008]. We found that combining features describing the temporal profiles of end effectors motion with features of our Multi-Level body movement notation system does not lead to a significant improvement of the classification rate. However, the features describing temporal profiles considered in this study were measured according to the trajectories and the energy of 5 end-effectors which are: head, left and right hand and left and right foot. Thus, temporal profiles features can be computed on other body cues. One possible feature work concerns the computing of Temporal Profiles features according to the most relevant body cues that were selected in Chapter 10. For instance, downward head flexion was considered as a relevant feature for the classification of emotions across various actions. Thus, we can compute the Temporal Profiles features proposed in [Castellano et al., 2008] to describe the dynamic of downward head rotation. Besides, other descriptors of expressive body movements can be used to characterize emotional body expression. For instance, the fluidity of body movement has been widely reported in previous perceptual studies to describe how emotions are expressed in body movement [Dahl and Friberg, 2007] [Montepare et al., 1999]. However, only few studies computed the Fluidity measure based on motion capture data, and when computed, they are mostly focused on the fluidity of the motion in a 2D plane [Castellano et al., 2012] [Castellano et al., 2007] [Glowinski et al., 2011]. Considering more descriptors of the movement dynamic (e.g. fluidity, regularity of the motion, the temporal profiles features of specific body cues) will allow a deeper analysis of the most relevant body cues that characterize emotional body expression.

Part VI : Appendices

A Head, Shoulders and Hips Behaviors during Turning

A.1 Introduction

The turning task, and more generally the change of direction in locomotion, was studied in many fields of research including clinical domain, neuroscience, computer animation, locomotion behavior annotation, analysis and synthesis. Turning can be considered as a part of the basic library of motor synergies characterized with an adaptive and stable behavior [Olivier and Cretual, 2007], [Hicheur and Berthoz, 2005]. Turning behavior involves a complex interplay between the different body parts. Previous researches revealed interesting founding that concern the interaction between the trunk, the head, the eyes [Imai et al., 2001], and also the coordination between lower body parts [Hicheur and Berthoz, 2005] during turning task. However, the coordination between shoulders and hips during turning movement has not been widely studied. The interaction between head and trunk movement during turn was mostly studied in turns within a restricted range around 90˚. Only few studies investigated the head and trunk behavior during turning task from different angles [Sreenivasa et al., 2008]. Since the head and trunk movement were studied on neutral turning behavior, it is still unclear whether they depends on the style of walk during turning task.

In the following section, we will discuss the different properties of turning task found in previous works. Section 3 is devoted to the description of the databases that we used in this study and the segmentation of walking sequences. In section 4, we will investigate the relationship between shoulders and hips during turning with different angles along with different styles of walk. Finally we will study the head and trunk behavior during turning with different angles and walk styles.

A.2 Turning behavior properties studied in previous works

Turning behaviors have been studied from a biomechanical point of view. Several approaches studied the trajectory of one or several limbs (mainly pelvis and head) in the space. They focused on the curvature-velocity relationship. Based on the walking trajectories, these studies showed that a power law controls the relation between radius of curvature and velocity of the followed path in cyclic trajectories

[Hicheur et al., 2005], [Vieilledent et al., 2001]. This rule is probably the most known assumption used in turning behavior studies from curved trajectories, but it has recently been extended to a single turning task [Olivier and Cretual, 2007]. However, the displacement of the body in the space is not enough to understand the behavior of turning. Thus it is important to study the coordination between the different body parts during turning.

The coordination of lower body limbs has also received a high level of interest in the studies aiming to understand the turning task. This is because the locomotion behavior is mainly and not only based on the production of a motor pattern via lower limbs coordination. It has been shown that the coordination at the level of lower limbs shifts from a symmetric to an asymmetric mode when the person change the direction of walk [Hicheur and Berthoz, 2005]. The asymmetric mode of coordination at the level of the lower limb was explained with the stabilizing mechanisms for postural control [Hicheur and Berthoz, 2005].

Other studies focused on the properties related to upper body parts during turning task. It has been shown that upper body parts anticipate the movement toward the direction of turn before lower body parts and that the head systematically shifts toward turns direction before the trunk does [Sreenivasa et al., 2008]. This indicates the role of the head as an inertial guidance platform to which are referred the movements of the other body segments [Hicheur and Berthoz, 2005]. This anticipation turned out to be occurred at a constant distance rather than at a constant duration before a turn [Sreenivasa et al., 2008]. The anticipation of head direction change to the direction of turn has been used as a mean to detect turning onset and offset [Li et al., 2012]. It was also found that there is a strong relationship between head, eyes and body orientation in the locomotion as well as turning behaviors [Hicheur and Berthoz, 2005], [Imai et al., 2001],[Sreenivasa et al., 2008].

Apart from the anticipatory orientation of the head, the stabilization of head orientation during turning is considered as a second main factor affecting head movement during turning behavior [Sreenivasa et al., 2008], [Hicheur and Berthoz, 2005], [Imai et al., 2001]. The anticipation and stabilization mechanisms are both combined in turning behavior [Hicheur and Berthoz, 2005]. It has been shown that the head can stabilize itself by looking toward the new direction of walk during turning task. Only few studies investigate the effect of different turning angles on the head behavior during turning [Sreenivasa et al., 2008]. Besides, the effect of the walk style on the head behavior is still unclear. In our work, we study the effect of the angle of turn and the style of walk on the head behavior during turning. Previous works described the head behavior during turning through the maximum orientation of the head around the yaw axis. Two references were used to determine this measure; the heading and the trunk [Sreenivasa et al., 2008]. We used these measure to study the head orientation during turning with respect to the trunk and to the trajectory

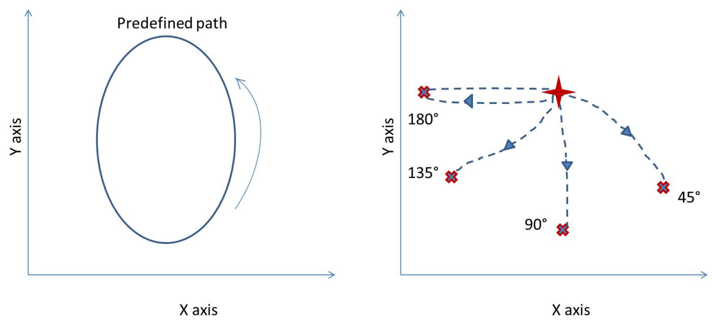


Figure A.1: Two types of turns: a) Constrained turn with predefined path , b) Unconstrained turns

of walk.

Although several properties of turn were studied previously, studying the relationship between lower and upper body parts to get some insights into how lower and upper body parts are coordinated received little consideration in the studies related to the analysis of walking and turning behavior. In this paper we also studied the relationship between shoulders and hips movement during walk and turning behavior with different angles and different styles of locomotion.

A.3 Databases description

In this work, two different databases recorded with two different inertial motion capture systems were used. The first database (eNTERFACE'08 3D) is devoted to the study of neutral turning behavior with different angles while the second database (Emilya Database, See Chapter 7) is used to study the effect of the style of walk on walking and turning behavior with 180 ˚ turn angle.

In general, in databases of turning behaviors two types of turns are differentiated: Constrained [Hicheur et al., 2005] turns and Unconstrained [Tilmanne et al., 2009] turns (See Fig. A.1). In the first type of turn, the person is asked to follow a specific trajectory predefined in advance (See Fig. A.1 a)). The advantage of this recording is that there are more accurate information about the angle of turn and the trajectory of the body limbs in space. However, it can result in a non-natural behavior of walking and turning tasks since the person is always looking at the predefined trajectory. In unconstrained turns recording, the person is asked to turn around some obstacles that define the turn angle(See Fig. A.1 b)). Both databases that we used in our study are based on the recording of unconstrained turns.

A.3.1 eNTERFACE'08 3D database

The eNTERFACE'08 3D database is described in details in [Tilmanne et al., 2009]. The movement of the body was recorded through the inertial motion capture system Animazoo [Ani,]. This database contains, among others, walking sequences containing turning task with different turn angles for 41 subjects. A single turn was performed for each walking sequence. For turning task, the subjects were asked to walk straight starting from a point drawn on the floor until they reached a line (which was also drawn on the floor) from which they must change the direction of walk to reach another point (See Fig. A.1 b)). The angles of the direction change were 45 ˚ , 90 ˚ , 135 ˚ and 180 ˚ .

A.3.2 Emilya database

The Emilya database is also used for this study. It is described in details in Chapter 7. In this study, only the walking sequences of seven subjects (4 female and 3 male) that are the more expressive among all the subjects were selected. Four styles corresponding to the expression of four emotions: Pride, Anger, Anxiety and Sadness were selected. Previous researches showed that the expression of those styles of walk has a significant influence on the head movement [Dael et al., 2011]. The actors were not aware of the study of turning behavior. They were rather focused on the style of walk.

A.3.3 Walking sequences segmentation

The segmentation of walking sequences into straight walks sequences and turning movement is a primordial step in our work. It requires to find out the start (onset) and end (offset) of a turn. The study of the relationship between shoulders and hips during turn requires the reduction of the turn interval time in order to focus on turning behavior and to remove straight steps or transitions steps between straight and turning behavior. Thus, we define the onset and offset of turn for this study as the first two foot events that surround the turning movement such as the Swing Heel Off (SW HO) or the heel contact with the floor [Mickelborough et al., 2000]. The segmentation of walking sequences was first based on the automatic detection of turn instant, and second on the automatic detection of onset and offset of turn. Both the detection of turn instant and turn boundaries (onset and offset) were based on turning angles. Turning angles were measured from hips and shoulders vector orientation around the yaw axis (See Fig. B.1).

- **Turn instant detection**;

  Since all the turns were performed with 180 ˚ in the emotional walking database, the turn instant was detected as the frame in which the turning angle reaches

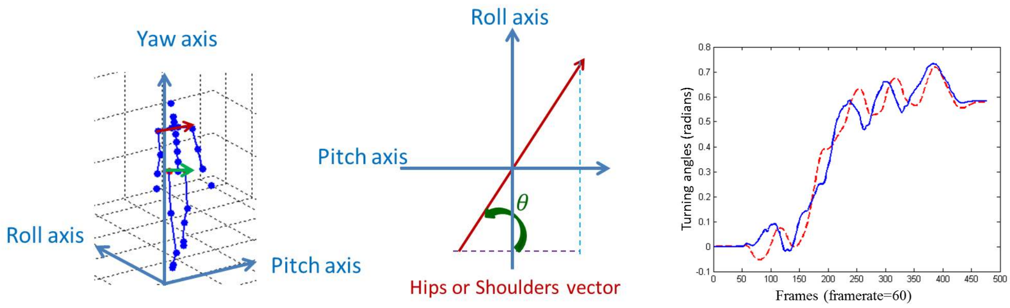


Figure A.2: The measure of Turning angles from hips and shoulders vector orientation around the yaw axis : a) Hips and shoulders vectors based on global positions of joints, b) Turning angle of Hips or Shoulders vector measured at a specific frame, c) Turning angles of hips and shoulders vector for all the sequence

90 ˚ . Using the prior knowledge that there is only one turning task in each walking sequence of eNTERFACE'08 3D database, the turn instant was defined as the frame corresponding to the average of turning angle values related to shoulders or hips vector orientation.

- **Turn onset and offset determination**;

  The determination of turn onset and offset is based on the turn instant:
  - First we detect the occurrence of all the foot events in the walking sequence,
  - Second we detect the turning boundaries frames defined as the first local maximum/minimum of turning angles that occur before and after the turn instant,
  - Finally we define the turn onset as the foot event that occurs before the first turning boundary and the turn offset as the foot event that occurs after the second turning boundary. Only few false detection of turn offset and onset instant were detected and corrected by hand (the percentage of false detection was: 24% for 45 ˚ , 4.87% for 90 ˚ , 2.43% for 135 ˚ and 0% for 180 ˚).

A.4 The relationship between shoulders and hips during walk and turn

The relationship between upper and lower parts and their coordination received little consideration in the studies related to walking and turning behavior. In this section, we investigate the relationship between shoulders and hips during walking and turning with different angles and different emotions. We focus on the movement of shoulders and hips in the horizontal plane, that is the orientation of the vectors around the yaw axis. We use the global positions defined in the spatial coordinate system (that reflects the real world) to measure the orientation of the shoulders and hips vectors defined respectively with left and right shoulders positions and left and right hips positions (See Fig. B.1 a)). At each frame of the motion sequence, the turning angle of the shoulders vector is deduced from the cosine and the sinus of this vector with respect to Pitch and Roll axes (See Fig. B.1 b)). The same procedure is adopted to measure hips turning angles based on the positions of right and left hips.

In the following, we will look at the coordination between shoulders and hips movement during turning task (between the onset and the offset of the turn) studying different angles and different styles of turning behavior. Based on the assumption that the behavior of the pelvis is more linear in walking along curved path than in walking along straight path [Anne-Helene Olivier and Cretual, 2009], we tried to see whether the shoulders movement follow the same pattern as the pelvis during a turning behavior and if maintains the same properties for different angles and different styles of walk.

A.4.1 Shoulders and Hips relationship during walk and turn with different angles

We used the eNTERFACE'08 3D database to study the behavior of shoulders with regards to the behavior of hips in turning task with different angles. We tried to see whether there is a linear relation between those two different body parts for different turn angles. Using the results of motion sequences segmentation, we fit a linear model to the data projected in shoulders and hips turning angle space between the onset and offset of turn. We found that the shoulders and hips movement are related through a positive linear relationship during turning behavior (see Fig. A.3). The coefficient of determination (R^2) was measured for each linear model fitted to the relationship between shoulders and hips between turn onset and turn offset. After applying One-way Anova, we found that the angle that characterized the turning movement had a significant effect on the value of R^2 ($p < 0.001$). The higher is the turning angle, the higher is the average value of R^2 for all the samples (See Table A.1).

A

To measure the global properties of the linear relationship between shoulders and pelvis movement, we fitted the linear model to all the samples of the eNTERFACE'08 3D database for each of the four sets of turning behavior (related to 45°, 90°, 135° and 180° angles). Figure A.3 illustrates the linear regression models for the samples related to each set of turning behavior. The corresponding R^2 for each linear model are shown in Table A.1. They indicate a strong linear relationship between shoulders and hips movement during turning task with different angles. The low value of R^2 related to 45° turns is mainly due to the variation of turning behavior with 45° among the subjects ; since the turning task was unconstrained,

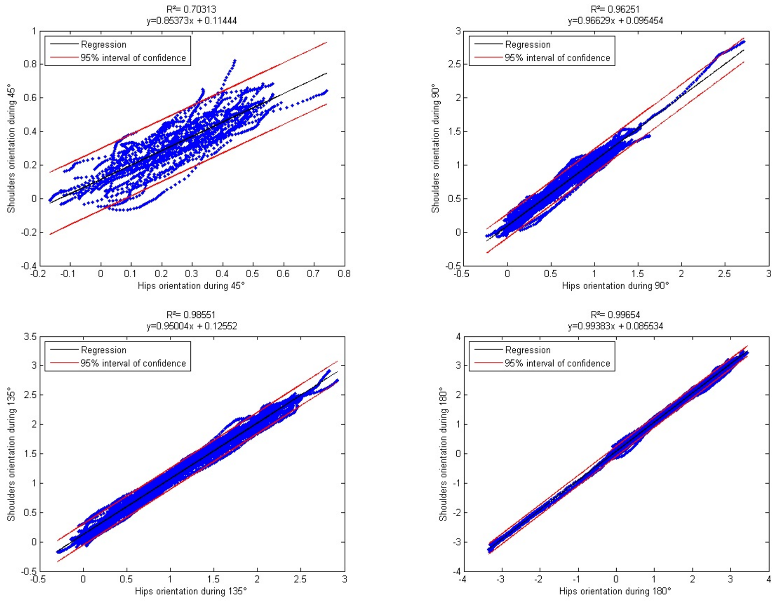


Figure A.3: Positive linear relationship between hips and shoulders movement for a) 45 ° turn angle, b) 90 ° turn angle, c)135 ° turn angle, d)180 ° turn angle

subjects tend to turn with angles different from 45 ° (mostly lower than 45 °). Overall, the higher was the turn angle the stronger was the relationship between hips and shoulders with slight differences in the coefficient of the model between 90 °, 135 ° and 180 ° (See Fig. A.3 where the 95% interval of confidence diminishes as turn angle increases).

A.4.2 Shoulders and Hips relationship during walk and turn back with different emotions

Although we previously showed the strength of the linear relationship between shoulders and pelvis movement during turning with different angles and especially

Parameters	**45 °**	**90 °**	**135 °**	**180 °**
Average of R^2	0.9057	0.9783	0.9927	0.9973
Linear model coefficients	(0.8537, 0.1144)	(0.9963, 0.0955)	(0.95, 0.1255)	(0.9938, 0.0855)
R^2 of the linear model fitted to all the samples	0.7031	0.9625	0.9855	0.9965

Table A.1: The results of fitting a linear model to the shoulders-hips relationship in eNTERFACE'08 3D database

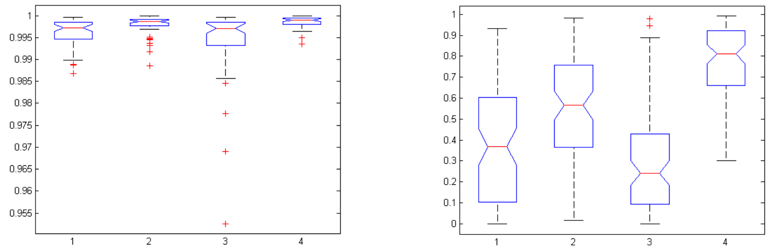


Figure A.4: The distribution and the medians of R^2 in a)emotional turning behavior and b) straight emotional walks, for Anger, Anxiety, Pride and Sadness

for 180 $^\circ$, it is important to see whether we obtain the same results when studying different styles of walking and turning behavior. The Emilya database was used for this purpose. After applying One-way Anova, we found that there was a significant effect of the emotions and all the subjects on the coefficient of determination R^2 and on the slope of the linear model ($P<0.001$). However, the standard deviation of R^2 for all the samples was 0.0047 (R^2 ranged from 0.9526 to 0.9999) which means that the variance that explained the linear model was always small. Figure A.4 (a) presents the distribution and the median of R^2 in emotional turning movement for the four sets of turning task.

The hips as well as the shoulders vector movement follow a nonlinear pattern (mostly sinusoidal) during straight walk due to the movement of the legs and the arms. This nonlinear behavior is due to the opposite movement between each two successive steps. In order to study the correlation between upper and lower body parts during straight walk with different emotions (Anger, Anxiety, Pride and Sadness), we applied One-way Anova on the coefficient R^2 for the four emotions studied in our database. We found that emotions have a significant main effect on the correlation between shoulders and pelvis movement during straight walks. Figure A.4 (b) shows that the differences between the medians and the distribution of R^2 between emotions are highly significant. The higher median value of R^2 was assigned to the sad emotional walks while the lower median value of R^2 was related to the pride emotional walks. This can be explained as actors tend to walk slowly with less variability of upper body parts and small foot strides. This behavior results in similar patterns between shoulders and hips. Unlike sadness expression, the actors express pride with large foot strides and a large amount of arms swings which leads to an opposite pattern of shoulders with respect to the hips. Anger and Anxiety expression included a significant change at the level of upper body parts as well as lower body parts.

A.5 Head behavior during turning

To be able to compare our results with the results found in previous works, we employ the same parameters used to study the stabilization of head during turn [Sreenivasa et al., 2008]. Thus, the analysis of head behavior was based on three features. The first and second feature represent the relative maximum relative head yaw during turning with respect to two references: heading and trunk yaw. Heading was measured based on the trajectory of body center of mass (CoM) in the space as described in [Imai et al., 2001]. In the following sections we use the term head-trunk yaw and head-heading yaw to refer respectively to the maximum orientation of the head with respect to the trunk and the heading. The third feature corresponds to the occurrence of the maximum relative head yaw (in seconds) with respect to the onset of turn. As the SW HO is considered as the first event that occurs during a step, it was used as a common reference that represents the onset of turn in this study. We conduct two studies; the first one is focused on the head behavior during turning with different angles while the second one is devoted to the analysis of head behavior during turning with different styles of walk.

A.5.1 Head behavior during turning with different angles

In this study, we focus on the orientation of the head around the yaw axis between the onset and the offset of the turn. We measured the maximum head orientation that occur between turn boundaries with respect to the trunk and the heading as explained in section 5. After applying One-way Anova, we found that the turn angle had a significant effect on the relative head orientation ($P < 0.001$) both for heading-head and trunk-head). Similarly to the results presented in [Sreenivasa et al., 2008], we found the same pattern of results for heading and trunk references. The relative head orientation using the trunk or heading as a reference shows a continuous increase for the turn angles up to 135 ° and a leveling off after that. The estimation of the magnitude of the relative head orientation was different between the two references. The average of the head orientation measured respectively using the heading and the trunk as references are shown in table A.2. The trunk-head turned out to be smaller than the heading-head for all the turn angles (See table A.2). This is explained by the orientation of the trunk that occurs with the orientation of the head during turning task.

We also measured the instant when the head orientation reaches the maximum value relative to the SW HO event involving in the first turning steps. The main effect of of turn angle was significant ($P<0.001$) leading to a linear increase from 45 ° to 180 ° (See Table A.2). This indicates that the instant in which occurs the maximal head orientation toward the turn direction with respect to the onset of turn is also affected by the turn angle.

A

Parameters	45 °	90 °	135 °	180 °
Heading-head	10.1 °	16.1 °	23.4 °	21.6 °
Trunk-head	5.9 °	9.9 °	15.1 °	14.7 °
The average of the occurrence of maximum head yaw (in seconds)	0.4375	0.7047	0.9613	1.1812

Table A.2: Head behavior stabilization measures in eNTERFACE'08 3D database

A.5.2 Head behavior during turning with different emotions

The trunk-head and heading-head features were measured between the onset and the offset of 180 ° turns detected in the emotional walks. The expression of emotions through walking action had a significant effect on the head-trunk orientation as well as on the heading-head orientation ($P<0.001$). Similarly to the study presented in [Sreenivasa et al., 2008], we found that the head orientation relative to the trunk was smaller than with the heading as a reference. This result shows that the trunk was also rotated to the turn direction regardless of the style of walking and turning actions. Since the heading reference strongly depends on the trajectory of the CoM, we focus only on the maximum head yaw relative to the trunk to compare the head behavior between the different styles of walk and turn.

No significant difference was found between the trunk-head in turning related to the expression of Anger and Pride ($P=0.33$). The maximum head yaw relative to the trunk ranged from 0.22 ° to 35.22 ° . The effect of the expression of emotion on the trunk-head yaw increase while taking into account Anxiety expression ($P<0.001$) but the expression of Sadness showed the most significant difference among the other emotion expressions. The median (4.94 °) as well as the range (from 0.1 ° to 14.8 °) of trunk-head yaw in turning while expressing sadness turned out to be smaller than with the other emotion expressions. These results showed that the maximum head orientation relative to the trunk during turning task is not only depending on the angle of turn but also on the style of walk.

Studying the effect of emotions on the occurrence of maximum head yaw (in seconds) with One-way Anova showed a significant differences ($P<0.001$) among the four groups representing the four emotional behavior. Interestingly, the median related to the occurrence of maximum head yaw showed higher values for Pride ans Sadness (1.5s and 2.1s) and lower values for Anger and Anxiety (1.2s and 1.1s). That means that the maximum head yaw occur earlier with Anger and Anxiety expression than with Pride and Sadness expression. This result is congruent with the activation dimension of the studied emotional states; head orientation is slower for less active emotional states and faster for active emotional states.

A

A.6 Conclusion and future work

In this paper, we investigate the relationship between upper and lower body parts through the relationship between shoulders and hips movement. We found that those two modalities follow a strong linear relationship during turning task for different turn angles (45 ˚, 90 ˚, 135 ˚and 180 ˚) and different styles of walk and turn (while expressing Anger, Pride, Anxiety and Sadness). We found that the linear relationship is stronger for higher turn angles. In the future work, we aim to study whether this relationship is maintained for turn angles lower than 45 ˚. The relationship between shoulders and hips movement must also be studied in cyclic curved trajectories to see whether it follows the same linear behavior found in turning around a corner. Like hips movement, shoulders movement is characterized with a non linear (sinusoidal) behavior during straight walk. However, shoulders and hips are in opposite of phase, mainly due to arm swings. But we showed that this opposite behavior strongly depends on the style of walk for Anger, Anxiety, Pride and Sadness expression through walking. That is the relationship of phase opposition between shoulders and hips is no more maintained during the sad expression as the body movement is less energetic and arms balance less.

In our work, we also investigate head behavior during turning with different angles and different styles of walk. We found that the maximum head yaw relative to the trunk and heading shows a continuous increase for the turn angles up to 135 ˚and a leveling off after that. This result was found previously in [Sreenivasa et al., 2008] with similar values of maximum head yaw for the different angles. The occurrence of the maximum head yaw was also affected by the turn angle leading to a linear increase from 35 ˚to 180 ˚. However, this behavior was not stable while changing the style of walk. The head orientation with respect to the trunk showed a significant decrease in turning task while expressing Sadness than while expressing more active emotional states like Anxiety and Anger. We showed that the occurrence of the maximum head yaw during turning task is also affected by the expression of emotion. The maximum head yaw occurs earlier for active emotional states (Anger and Anxiety) and later for less active emotional states (Pride and Sadness).

B Walking motion analysis based on shoulders and hips turning angles

B.1 Introduction

Walking behavior has been widely studied from the perspective of human motion analysis and virtual human motion synthesis. Most of these researches rely on the study of natural human walking data mostly consisting in three-dimensional walking data. Different motion capture techniques are available nowadays for the recording of three-dimensional human motion [Ani,] [MVN BIOMECH system. Xsens website,]. In order to use the data collected efficiently, it is necessary to post-process the resulted motion data. The post-processing of the data involves the correction of rotational and positional errors, the annotation and the segmentation of motion sequences.

In previous databases aiming to record walking motion, the resulted motion sequences contain often both straight walking and curved walking (or turning) [Ma et al., 2006] [Tilmanne and Dutoit, 2010b]. Recording both straight walking and turning behavior in the same sequence is often used to have a good compromise between the amount of data and the time it takes to record the database. Previous researches have shown the difference of locomotor patterns that exists between walking along a straight path and walking along a curved path. This difference concerns the behavior of lower body limbs [Hicheur and Berthoz, 2005], the behavior of upper body limbs [Sreenivasa et al., 2008] as well as the relationship between upper and lower body parts (See Appendix A). Thus, the analysis of walking sequence may provide different results if it includes a turning behavior. Hence the segmentation of walking motion sequences into straight walk sequences and turning movement is required for a large set of goals using the database (Motion concatenation techniques, stylized motion analysis..).

In this chapter, we propose a new method for walking sequences segmentation aiming to separate straight walks from turn movement. Our method is based on the relationship between shoulders and hips movement during walking and turning behaviors. We showed in Appendix A that shoulders and hips follow a strong linear relationship during turning with different turn angles while the movement of both shoulders and hips follow an opposite or conform sinusoidal behavior during straight walk depending on the style of walk. Thus, after detecting the presence of a turn, we define the turning behavior as the time interval in which the relationship between shoulders and hips movement is the most linear.

B.2 Related work

The analysis of turning behavior in motion capture data involves the detection of turn presence, turn instant, turning steps, or all of them. The determination of the start (onset) and the end (offset) of a turn is often more difficult than the detection of turning behavior presence or turning task instant. The determination of the onset and offset of turns can be achieved with different approaches: using specific equipment, based on manual annotation or based on motion data analysis.

To detect turn in locomotion behavior, several previous works were based on some specific external equipment or sensors such as uni-axial gyroscopes or angular velocity transducers mainly attached to the trunk or to the heel [Tong and Granat, 1999] [Salarian et al., 2009] [Visser et al., 2007]. Such devices has been mainly used to study the behavior of people who suffer from some specific disease or need a reeducation walk in order to collect accurate measures while avoiding to equip the person with a whole body motion capture system. Visser et al. [Visser et al., 2007] study the turn behavior in Parkinson's disease using two angular velocity transducers attached to the lower back of the patient. However, using specific devices to analyze turning behavior provides only a small range of data and does not allow to collect other basic information related to the motion of the whole body. In the particular task of walking, we need to have more subtle information especially to detect the foot events. A foot support pattern can be used in a machine learning algorithm to distinguish straight and turning steps [Arikan et al., 2003]. This method requires to manually annotate the different types of turning steps in the database. This annotation stage is strongly dependent on the quality of user annotation as well as the training data. Thus the machine learning algorithm can differentiate only the turning steps that were presented on the training database.

From the perspective of data analysis, automatic turn detection approaches can be mainly divided into two categories: global approaches and step based approaches. Global approaches consider the animation sequence as a unique entry. Thus they aim to detect the frames that define approximately turning boundaries. Several approaches are based on the application on some threshold on turning angles (along the yaw axis) to detect the start and the end of turning task. However turning angle value is not equal to zero during straight steps. This is due to the constant continuous oscillations of the pelvis during straight walking [Olivier et al., 2009]. Furthermore, the threshold applied on turning angles values is strongly dependent on the velocity of walking since the oscillation aspect depends on walking speed. The application of a low-pass filter can resolve this problem but makes harder the detection of slight turn. Thus the application of a threshold to detect turns cannot always provide satisfying results. Besides, applying a threshold on turning angle cannot distinguish accurately straight steps from turning steps. Previous works showed that the head anticipates the rotation along the yaw axis into the turn direction before the beginning of turning task [Hicheur and Berthoz, 2005]. This anticipation of head direction change has been used as a mean to detect turning onset and offset [Li et al., 2012]. Walking trajectory is also widely used in automatic turn detection techniques. Walking path based approaches are based on the 3D trajectory of some markers in space like the 3D positions of the center of mass (CoM) or the head. A geometric method is described in the work of Olivier et al [Olivier and Cretual, 2007] to delimit the turning steps in curved walking. The authors used the two axes that represent the straight lines surrounding a 90 $^\circ$ turn to compute the quadrant that delimits the turn. Their method is based on the prior knowledge that only one turn is present in the sequence of motion.

Step based approaches require first the determination of foot strike to deduce turning steps. Those approaches are mainly based on the comparison of some properties between successive steps. Olivier et al [Olivier et al., 2009] used the power law between curvature and velocity of CoM trajectory to detect turning steps in a locomotion sequence. Based on the constant oscillations of the CoM along the trajectory, they assumed that two successive straight steps can be modeled with two asymmetric arcs of circle. All the steps of the walking sequence are first defined by the pair (mean of velocity, mean of curvature of CoM model). Turning steps are represented by the points that are outside the 95% interval of confidence of the linear model that define the relationship between velocity and curvature during straight walks in the log-log space. It was showed that this approach enables the detection of turning steps related to a small angle turn behavior. However the modeling of CoM trajectory is a crucial step for this approach. Measuring curvature and velocity values from original data will provide unsatisfying results. Thus the accuracy of the results is strongly dependent on the model that represents the CoM trajectory.

While the performance of step based approaches is strongly related to the accuracy of steps detection approach, global approaches are easier since they require less preliminary steps. Turning steps can be deduced from global approaches by looking for the foot events that are close to the global detected frames. In our work, we adopted a semi-global approach where foot events were used in a recursive algorithm based on the study of shoulders-hips relationship during walking. Turn instant is first detected based on the distribution of shoulders and hips turning angles. Then several foot events are detected and used to define an incremental window initialized through the foot events that surround the turn instant. The defined window is incremented based on the linearity of shoulders-hips relationship to deduce finally turning steps when the behavior of shoulders-hips is no more linear.

B

B.3 Experimental motion databases

In this study, we used two databases (eNTERFACE08 3D and Mockey databases) recorded with an inertial motion capture system to analyze the 3D body movement of walking behavior . The eNTERFACE08 3D database [Tilmanne et al., 2009] was used to test the accuracy of turn movement analysis with different turn angles.

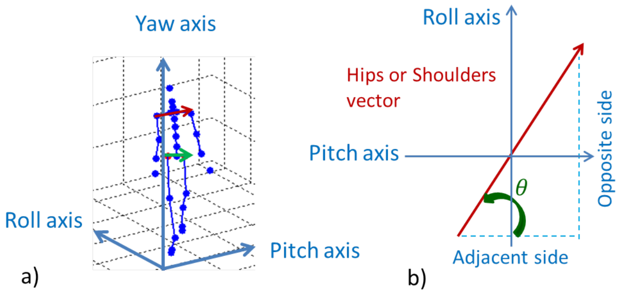


Figure B.1: The measure of Turning angles from hips and shoulders vector orientation around the yaw axis : a) Hips and shoulders vectors based on global positions of joints, b) Turning angle of Hips or Shoulders vector measured at a specific frame

This database is described in detail in [Tilmanne et al., 2009]. The 3D movement of the body was recorded with the inertial motion capture system Animazoo [Ani,] with 41 subjects. Among other tasks, the subjects were asked to walk straight until they reached an obstacle from which they had to change the direction of walk to reach another point. In our work, we used the walking sequences containing 0 ˚, 45 ˚, 90 ˚, 135 ˚ and 180 ˚ direction change for all the subjects. The Mocket database [Tilmanne and Dutoit, 2010b] was used to study several turns recorded in the same walking sequences. The 3D body movements of one subject were recorded when he walks and turns with different styles of walk. The number of turns performed in each walking sequence recorded with this database ranges between 4 and 6. Totally, 61 180 ˚ turns were performed by the subject in this database.

B

B.4 The relationship between shoulders and hips during the locomotion behavior

Turning task involves a complex interplay between the different body parts. Previous researches focused on turning behavior revealed interesting founding that concern the behavior of upper and lower body parts during turn. We have shown recently that shoulders and hips follow a strong linear relationship during turning while they follow a sinusoidal behavior during straight walk (See Appendix A). In our work, the measure of turning angles from shoulders and hips is based on the 3D positions of right and left hips and shoulders in the spatial coordinate system. Thus, turning angles are computed from hips and shoulders vector orientation around the yaw axis as shown in Fig. B.1. This is performed using the adjacent and the opposite side of the angle formed by the vector (of shoulders or hips) and pitch axis. We did not fix a predefined interval in which the turning angle must belong to in order to avoid the discontinuity of the turning angle curve.

In our method, we only make use of turning angles measured from shoulders and hips to detect the presence of turns. Turning steps are deduced from the linear

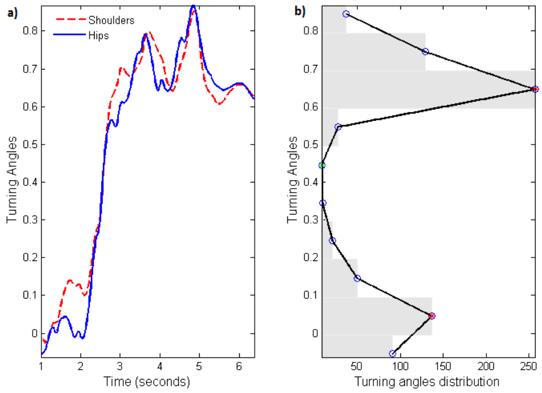


Figure B.2: Turning angles modeling; a) Shoulders and hips turning angles, b) Linear interpolation for data distribution modeling

relationship existing between shoulders and hips during turning (See Appendix A). Our approach enables the accurate detection of all turning steps involved in the turning task.

B.5 Turn presence detection based on shoulders and hips turning angle

Based on turning angles, the turning task can be easily detected if we have a prior knowledge about the angle of turn and if only a single turn exists in the walking sequence. However this method fails to detect several turns with different turn angles. In our method, the detection of turn was based on the distribution of the turning angles during the whole walking sequence. Due to the different patterns of turning angles during walking and turning, the distribution of the turning angles data reaches the higher values in straight walks and the lower values in turns. Since the distribution of turning angles is quite sensitive to the pauses in walking sequences, we removed the pauses that occur mostly at the beginning and at the end of walking sequences. This step was based on the analysis of turning angles velocity.

The first part of Fig. B.2 depicts the turning angles measured separately from shoulders and hips vector while the second part illustrates the distribution of turning angles by the mean of the histogram of the two turning angles curves. We modeled the turning angles distribution with a linear interpolation that fits the peaks of the histogram. The model of the histogram was used to detect the turn in the walking sequence. First we detected all the local minima and maxima of the distribution model. All the local minima can be considered as possible indicators of turn detection. However, small variation of turning angles during straight walk can cause the presence of local minimums as well. Thus we checked for each local minimum

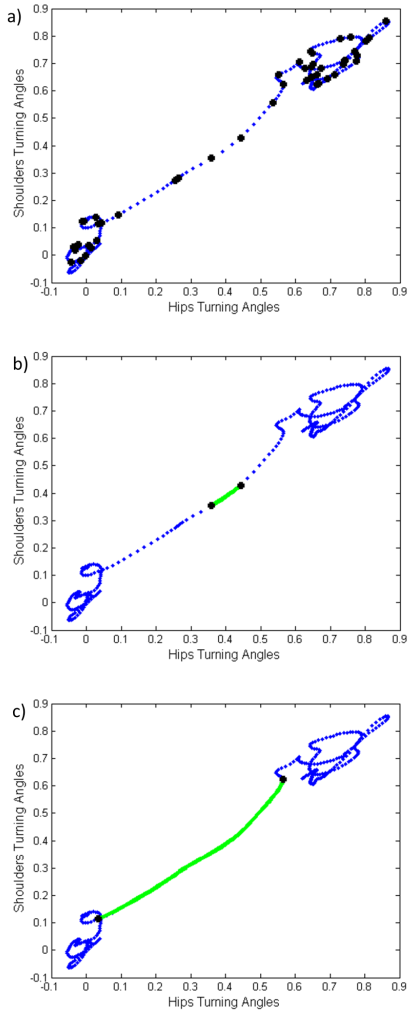


Figure B.3: Turning steps detection based on the linear relationship between shoulders and hips; a) All gait events presented in the walking sequence, b) First window defining turning steps, c) Last window defining turning steps

Turn Angles	**45 °**	**90 °**	**135 °**	**180 °**
Correct detection (%)	80	95.12	90.24	95.1

Table B.1: Percentage of turn detection based on shoulders and hips turning angles in the eNTERFACE08 3D database

whether the corresponding hole was deep enough by comparing the ratio composed of the current local minimum and the previous (and next) local maximum with a predefined threshold. The turns instants detected in the walking sequences are then defined by the frames in which the corresponding distribution model of turning angles is low enough.

B.6 Turning steps detection based on shoulders and hips relationship

We adopted a semi-global approach to detect the turning steps based on the turn instant along with the gait events detected in the whole walking sequence. Most of the gait events are detected as explained in [Mickelborough et al., 2000]. We also measure the local minima and maxima of the distance between heels as additional gait events. Starting from the turn instant, we defined an incremental window based on the gait events occurrences that surround the turn instant. The increase of the window size is conditioned by the linear relationship between shoulders and hips. Furthermore, the size of the incremental window stopped increasing when the relationship between shoulders and hips is no more linear (See Fig. B.3).

The linearity of the shoulders-hips relationship was defined using two conditions. In the first condition, we compared the slope of the linear model that fits the hips-shoulders relationship in the current window with the slope of the linear model that fits the hips-shoulders relationship in the next window. The relationship between shoulders and hips was considered as no more linear (with respect to the first condition) if the difference between the slopes was higher than a predefined threshold. The second condition concerns the intersection of the hips-shoulders relationship with the horizontal and vertical line for each frame. The relationship was then considered as non linear (with respect to the second condition) if the intersection of the curve with the horizontal or vertical lines resulted in two or more values.

B.7 Results and discussion

The evaluation of our method is based on two experimental databases. As we know the actions that were performed in the databases that we used, we were able to evaluate the results of turn presence detection and turning steps determination.

B.7.1 Turn presence detection

Due to the unconstrained radius condition of turn in the eNTERFACE08 3D database, the subjects tend to turn differently even for the same turn angle. As a result, the turning task was sometimes performed within two stages separated with a slight straight step. Thus, to evaluate the results of turn presence detection based on turning angles distribution model, we consider that a turn presence is correctly identified if the algorithm detects a turn instant for each stage of turning task. Table B.1 refers to the results of turn detection in the eNTERFACE08 3D database. The percentages of correct turn detection are 80%, 95.12%, 90.24% and 95.1% respectively for 45 ˚, 90 ˚, 135 ˚, 180 ˚ turn angles. Most of the false detection are either due to the detection of the performed turn with other non-existent turns or to the inaccurate detection of turn instant. Only one turning task (with 45 ˚) was not detected.

We also tested the turn detection method based on shoulders and hips turning angles distribution on walking sequences where the direction change corresponds to 0 ˚ in the eNTERFACE08 3D database. The percentage of turn detection in the studied 0 ˚ walking animation is 19.51%. Nevertheless, taking a deep look at these walking behaviors, we found that the subjects tend to change the direction of walk unconsciously to reach the end of the path, which explains most of turn detection results in 0 ˚ walking animation.

The Mockey database was used to test the reliability of turn instant detection based on turning angles distribution when several turns with the same turn angles occur (180 ˚ in this database) in the same walking sequence. The percentage of correct detection is 93.44% which indicates that only few false detection occurred. All the false detections of turn were due to the misclassification of straight walk as a turn, but all the performed turns were correctly detected.

Although the results of turn detection in a sequence of walking containing several turns were evaluated only with 180 ˚ turn angle, shoulders and hips turning angles during walking and turning showed the same pattern for turn angles higher then 45 ˚. Thus, turns can be detected as well using turning angles distribution in walking sequence with several turns having the same turn angle.

B.7.2 Turning steps detection

In order to evaluate the results of turning steps detection, we discussed only the results related to a correct turn detection. The evaluation of the automatic detection of onset and offset is a difficult task. Indeed, turning onset and offset do not reflect sudden events but they are rather considered as slow transitions between straight walk and turn [Salarian et al., 2009]. Turning steps detection must be compared to a reliable reference (which can be a manual annotation or another efficient automatic method). In addition, walking steps must be defined through a particular foot event in order to compare the onset and offset of turn with the results of another reference.

Turn Angles	**45°**	**90°**	**135°**	**180°**
Onset correct detection (%)	91.42	100	97.5	100
Offset correct detection (%)	88.57	92.68	90.34	97.56

Table B.2: Percentage of turns onset and offset detection in the eNTERFACE08 3D database

In our work we defined the onset of turns as the Swing Hell Off (SW HO) that occurs before the foot event that was attributed to the first turning step. Similarly, the offset of turns was defined as the SW HO that occurs after the the foot event attributed to the last turning step. Based on the empirical evidence that shoulders and hips movement follow a sinusoidal pattern and linear pattern respectively during straight walks and turning movement, we compared the results of turning steps detection with transition steps between straight walk and turn as a reference. Indeed, the transition between straight walk and turn was defined as the nearest SW HO to the last peak of turning angles curves related to shoulders and hips. Similarly the transition between turn and straight walk was defined as the nearest SW HO to the first peak of turning angles curves. The automatic detection of transition steps was also based on the detection of turn instant. The false results of the transition steps detection were corrected manually in order to obtain a reliable reference to evaluate the results of our method. The evaluation of turning steps was performed separately for the onset and offset of turn. Table B.2 illustrates the results of turning steps detection based on shoulders-hips linearity after their comparison with the transition steps between walk and turn movement. Results show that more that 85% of turning steps for each set of turn angle were correctly identified. Turning steps related to correct turn instant in Mockey database were all correctly identified based on the same approach.

Intuitively, we can think that the same results of turning steps detection could be obtained based only on the linearity of shoulders movement or hips movement. We compared the results of turning steps detection based only on shoulders movement, based only on hips movement and turning steps detection based on shoulders-hips relationship from walking sequences presented in the eNTERFACE08 3D database. We found that the turning boundaries detected based only on shoulders movement were mostly different from turning boundaries based only on hips movement. But most interestingly, we found that the turn onset and offset based on hips turning angle were randomly detected before or after the turn onset and offset based on shoulders turning angle. That means that shoulders and hips turning angles are both important to delimit the start and the end of turning task.

Studying the linearity of shoulders-hips relationship allowed us also to detect the split of turning task into two stages of turns separated by a straight step. Turn split occurred mainly with 45° (5%) and 90° (19.51%) turns. This behavior is mainly due to the unconstrained radius of turns recorded in the eNTERFACE08 3D database.

B.8 Conclusion

In our work, we proposed an efficient method for the segmentation of walking sequences into straight walks and turning task. Our method is based on the pattern of turning angles measured from shoulders and hips vectors. This approach was evaluated through two databases to study the detection of turns onset and offset with different turn angles and with several turns in the same walking sequence. We showed that turning steps involved in turning task are accurately identified based on the linearity of shoulders-hips relationship. Using this approach, we are able to annotate walking sequences recorded in the databases that contains motion sequences composed of walking and turning behaviors.

Turn detection based on turning angles distribution was tested and evaluated for turn detection from walking sequences containing a single turn or several turns with the same angle. In future work, we aim to test the efficiency of this method with walking sequences containing different turns with different turn angles presented in the same walking sequence, and with turn angles lower than 45 ˚ .

Turning steps determination based on the linearity of shoulders-hips relationship showed satisfying results for turn angles between 45 ˚ and 180 ˚ and for turning with 180 ˚ with different styles of walk. We found that turning steps detection based only on shoulders turning angles or hips turning angles cannot provide satisfying results since shoulders are not always perfectly synchronized with hips movement. In the future work, we aim also to study whether turning with turn angles lower than 45 ˚ can be easily detected based on the linearity of shoulders and hips movement.

B

C Dynamic stimuli visualization for experimental studies of body language

C.1 Introduction

The studies of the human perception of body language and motion patterns received a wide range of interest since a long time for different fields of research like the recognition of affect in body movement [Kleinsmith et al., 2011] and the identification of body cues that contribute to the attribution of emotions and affects [Meijer, 1989, Dahl and Friberg, 2007].

The content of the stimuli that the observers are asked to judge depends on the research question that needed to be answered from the results of the perception study. One can use the raw videos (videotaped) that depict the real visual content of body movement of the "actors" [Meijer, 1989]. Digital modifications can also be done on the original movies or pictures to abstract some bodily information [Dahl and Friberg, 2007, Atkinson et al., 2004]. For many other purposes, it is required or preferable to use computer avatar as the content of the stimulus.

To visualize stimuli of an avatar wandering around an environment (walking, turning, etc) we can choose to have a static or a dynamic camera that follows the avatar in its displacement. However, when the camera is static the distance between the avatar and the camera varies and this may affect the perception of the avatar body movement. Similarly if the orientation of the camera and the avatar body varies, it may affect how body movement is perceived. To overcome such biases in perception studies, we propose tools to parameterize the camera movement and orientation. For example, we can control the trajectory of the camera and its orientation so that it maintains an equal distance and orientation with the avatar.

C.2 Related work

The use of computer avatars in perception studies of body movement has widely emerged recently [Coulson, 2004, Hicheur et al., 2013, Kleinsmith et al., 2011, Roether et al., 2009]. Depending on the goal of the study, the movements reproduced with the avatar may be the result of motion capture data [Kleinsmith et al., 2011] or the results of a model that provides the synthesize of new body movements [Hicheur et al., 2013].

Previous discussions related to the perception studies of body movement were mostly around the body model of the computer avatar. Point-light display of body movements was primarily used for the studies related to the perception of biological motion [Johansson, 1973, Dekeyser et al., 2002]. Other body models were used for the studies that rely on the perception of both body posture and the dynamics of movement. Those models are mostly based on body skeleton model through specific geometric shape primitives [Griffin et al., 2013b, Kleinsmith et al., 2011, Kleinsmith et al., 2006a, Roether et al., 2009] or a virtual animated character [Hicheur et al., 2013].

As body posture involves a three-dimentional presence, human perception of body postures and body movements reproduced on a three-dimentional avatar may depend on the viewing angle [Coulson, 2004, Daems and Verfaillie, 1999], especially for viewpoint that result in occlusion of some body parts by others. In the studies based on the perception of body movement, the viewpoint is defined according to the goal of the study. Kleinsmith et al. [Kleinsmith et al., 2011] reproduced expressive postures on computed avatar and simulate a frontal view for the perception of emotion from body posture. Hicheur et al [Hicheur et al., 2013] chose a side viewpoint to create the videos depicting walking behaviors reproduced on an animated character. Roether et al. [Roether et al., 2009] used movies of a animated virtual avatars turned 20 degrees from the frontal view. However, it could be interesting to study the effect of different viewpoint on the perception of body behavior [Coulson, 2004].

C.3 The description of the proposed approaches

C

Two different types of virtual camera must be distinguished: free camera and target camera. While the orientation of free camera requires the definition of the 3D rotation, target camera is, by default, facing its target. Most often, the target refers to the center of interest of the object to be followed (the avatar). We assign the target of the camera to the pelvis in order to perceive the whole body posture, but the choice of the joint associated with the target could change from one study to another.

C.3.1 The position and the orientation of the camera

The definition of the viewpoint of the avatar refers to the determination of the position of a virtual camera that looks toward the avatar.

The viewpoint determined by the virtual camera has to be defined based on the orientation of the object (here the avatar body). We define the orientation of the whole body based on the orientation of the pelvis.

C.3.1.1 The position of the camera

The desired viewpoint of the avatar may differ from one study to another. Our goal is to provide a solution that can be controlled through a set of parameters. The determination of the 3D camera position is based on three parameters: the distance between the camera and the target, the height of the camera, and the angle that defines the viewpoint of the avatar. By default, the height of the camera and the distance between the camera and the target could be proportional to the height of the pelvis. As a result, the attribution of the desired viewpoint relies on the determination of X and Y components of the camera position.

The determination of the camera position turn out to be a geometric problem that involves both the vector orthogonal to the direction of the whole body and the vector between the target and the camera position. When considering the pelvis posture as the indication of the body orientation, the geometric problem involves the vector defined with the Left Hip Position and the Right Hip Position and the vector defined with the Pelvis Position and the Camera Position. Knowing the positions of Right Hip and Pelvis, the distance between the pelvis and the camera, and the angle that defines the angle of viewpoint, we are able to determine the position of the camera.

C.3.1.2 The orientation of the camera

The target of the camera is used to define the orientation of the camera towards an object. Assigning the target to the pelvis position makes the camera point to the center of the body structure. However, defining the target as the pelvis itself could affect the perception of pelvis motion. In fact, a target camera will not only be oriented toward its target, but it will also follow (without changing the position) all the motions performed by its target, including the more subtle motions. For instance, if the avatar is jumping up and down, the camera motion will follow the same motion (up and down). As a result, in the related video, we will perceive the floor as a moving object and the pelvis as a static object, which is the opposite of the result that we are expecting. For this reason, we define the target as an approximation of the pelvis position. For body movements that involve small body displacement (where the avatar can be still visible to the camera), the target position can be set to the first position of the pelvis, and still static for the whole animation. However, for body movements that involve considerable body displacement in the space, the target has to move according to the pelvis motion. In the next section, we introduce some solutions for the motion of the camera as well the target following the avatar motion.

C.3.2 The control of virtual camera motion

Up to our knowledge, previous perception based studies that rely on the perception of body movements tend to use movies where the viewpoint as well as the position of the camera is static while the avatar is moving in the 3D space. While this approach could be a good solution when the whole body movement is relatively small, it has the limitation of loosing the details of body motion during the perception if the animation involves turning behavior or walking along long distance. In this section, we introduce some solutions for the control of the camera trajectory and the target motion when the animation involves a considerable displacement of the whole body in space.

One principal issue that could affect our perception of body movements is the desynchronisation of the camera motion with the avatar displacement, which creates an effect of zoom in and out. Another issue that can create the same effect is the change of the distance between the camera and the target. So the first motivation for the solutions that we propose to control the camera path is the non-uniform motion of the camera following as much as possible the same change of velocity and acceleration in the avatar displacement. And the first motivation for the solution proposed to control the path of the target is to keep as much as possible the same distance between the camera and the avatar.

C.3.2.1 The path of the target

As we explained previously, the target position has to follow the targeted joint. We project the pelvis positions along straight lines defined through the positions of the pelvis in two successive time steps. Figure C.1 (4) depicts the path of the target.

C

C.3.2.2 The path of the camera

For perception based studies, there is a lack of discussions on the control of the path of virtual camera. Thus, we based our work on the assumption that the virtual camera motion can influence the perception of body movements. Our aim is to create camera with less potential influence on the perception of body movements.

A first intuitive solution is to update the camera position in each frame during the whole animation. The result of this solution can be visualized in Figure C.1 (1). This method results in a perfect synchronisation between the motion of the avatar and the camera, while handling the same viewpoint during the whole animation (based on the angle between Left Hip - Right Hip vector and Pelvis-camera vector). The algorithm that controls the path of the camera is as follows; for each frame, we get the positions of Right Hip and Pelvis and update the position of the camera according to their current positions as described in section C.3.1.1. However, one should bear in mind that walking motion give rise to non linear pattern of body segments, including the pelvis [Olivier et al., 2009]. Hence, this camera motion may

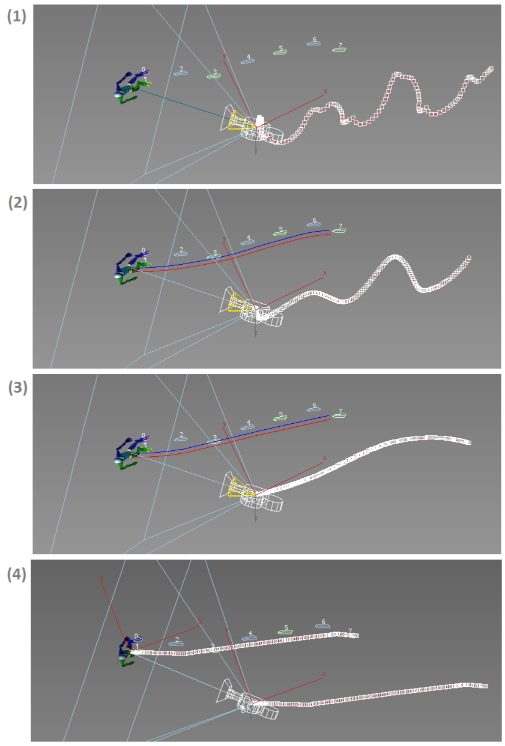


Figure C.1: Camera trajectories when 1) Updating the camera position each frame based on pelvis and right hip positions, 2) Updating the camera position each frame based on the estimation of Pelvis and Right Hip motion through spline curve where the control points are the time steps, 3) Updating the camera position each frame based on the estimation of Pelvis and Right Hip motion through spline curve where the control points are the timing of all the right steps, 4) Defining the camera motion as the translation of its target trajectory.

affect the perception of body movements since the camera is shaking from the left to the right due to the non linear motion of the pelvis.

In the following, we propose some different solutions for the control of virtual camera path according to the avatar movement.

– Synchronized non uniform non linear style: Update the camera position in each frame based on the approximation of pelvis motion
  One solution to control the motion of the camera is to approximate first the trajectory of pelvis and the right hip (or in more general way the joints that determine the position of the camera) and then to update the position of the camera (as explained in section C.3.1.1) at each frame. In this way, the viewpoint is updated at each frame according to the approximation of pelvis posture for straight walk and turning behavior (See Figure C.2).
  The approximation of pelvis and right hip positions is set on a spline curve (the

red and blue curves in Figure C.1 (2) and (3)). However, this approximation is strongly based on the control points. Defining the control points along a fixed time window (for example each 30 frames) or along the time step results in a sinusoïdal form of the camera motion (See Figure C.1 (2)). This is due to the opposite posture of the pelvis in two successive steps (left step and right step). This problem can be resolved by defining the control points on the steps of one side (all the right steps or all the left steps), which result in a more linear camera motion (see Figure C.1 (3)). This solution provide a good trade-off between the smoothness of the camera motion and the conservation of the same viewpoint during the animation (as a result the synchronization between the avatar and the camera motion).

– Synchronized non uniform semi-linear style: Make the camera follow the avatar motion without keeping the same viewpoint.
  Another solution for the control of the camera motion is to maintain a perfect synchronization between the camera motion and the avatar displacement without updating the viewpoint (See Figure C.1 (4) and Figure C.2 (1)). Comparing to the results in Figure C.2 (2), the camera position in Figure C.2 (1) does not provide the same viewpoint during the whole animation, but this might be interesting for the studies based on the perception of turning behavior. The camera motion is obtained by the translation of its target trajectory. The latter is based on the projection of pelvis positions along straight lines. Each straight line is defined through the positions of the pelvis in two successive steps timing. In this way, the camera motion is defined as a succession of small straight lines according to the successive steps.
– Walking steps based style: Update the camera motion differently for straight walking steps and turning steps
  Finally another solution that aims to maintain the same viewpoint on the avatar and a good synchronization between the avatar and the camera motion is to combine the synchronized non uniform semi-linear style for straight walking steps and the update of the camera position in each frame for turning steps. This approach requires the annotation of walking steps into straight walking steps and turning steps. However, this solution needs smoothing the camera path during the transition between straight and turning steps.

C.4 Conclusion

In this chapter, we propose some solutions to control the position and the trajectory of the virtual camera used to visualize stimuli for experimental studies. Our solutions allow to automatically convert a database of body movement animation files into a database of movies for the use in a perception study. Furthermore, we propose automatic control of the virtual camera position and motion in perceptive studies.

For future work, we aim to compare the visualization of stimuli using moving

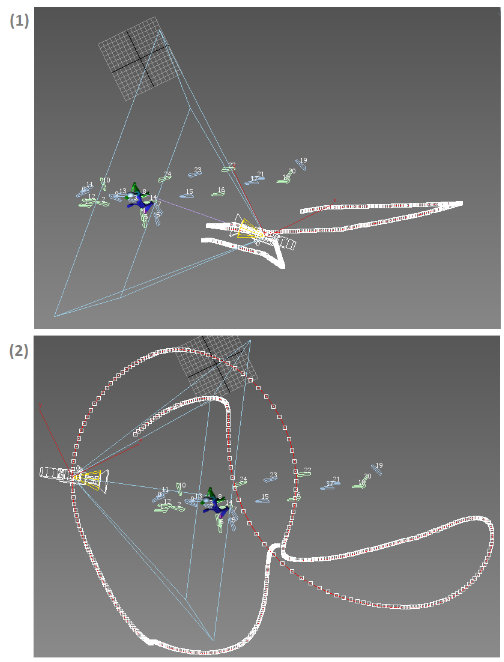


Figure C.2: Camera trajectories in straight walking and (180 ˚) turning behaviors; 1) without keeping the same viewpoint (Synchronized non uniform semi-linear style), 2) while keeping the same viewpoint (Synchronized non uniform non linear style). The screen shot corresponds to the same frame in the animation.

virtual camera with those created using static virtual camera through a perception based study. We also aim to compare the stimuli displayed with the different solutions that we proposed through a perception based studies for different body movements (walking, turning, sitting down...).

D

Random Forest approach

Random Forest (RF) approach is an increasingly popular ensemble method that was introduced a couple of years ago by Breiman [Breiman, 2001]. We should note that variants of Random Forest can be proposed. The Random Forest algorithm proposed by Breiman [Breiman, 2001] is a particular form of Random Forest (See [Genuer, 2010] for the description of the other possible variants of Random Forest). In our work, we are interested in the classical Random Forest (RF henceforth) proposed by Breiman [Breiman, 2001] as it is the most used in the literature.

RF is an ensemble of unpruned trees for non-parametric classification and regression [Cutler et al., 2012]. It has been shown that significant improvements in classification accuracy can be achieved by growing an ensemble of classifiers (e.g. trees) that consists of multiple version of a classifier rather than a single classifier [Breiman, 1996] [Freund and Schapire, 1996]. RF combines several decision tree predictors and includes two aspects of random selection in the training process: Bootstrap aggregation (bagging) and features random selection. The idea behind those random selections is to build small, efficient and uncorrelated ensemble of trees. While each tree by itself is unstable to changes in the learning set, combining the prediction of several trees turns out to be reliable. In classification, the most popular class across all the trees is attributed to a new sample (See Figure D.1). In regression, the average over the trees is attributed to a new sample as the outcome is numerical. In our work, we focus only on classification problem.

The trees of the forest are all unpruned. That means that the trees are fully grown and no algorithm is applied to reduce the size of the trees which may contains selection that provide little power for classification. Building unpruned trees is expensive in term of computational cost [Genuer et al., 2010]. For the prediction performance, the average of trees votes eliminate the negative effect of individual over-fitting [Genuer et al., 2010].

D.1 Bagging

Bagging (Bootstrap and Aggregating) approach consists of building a number of independent predictors, each one based on a subset of training samples build through a random sampling from the original training set [Breiman, 1996]. There are two possible strategies to build a subset of training samples (to perform a random sampling). The first strategy consists of sampling n samples with replacement from the original learning set, while n is the number of samples in the original training set. That means that, if we consider the samples as balls, to construct the subset of training samples we pick up balls n times from a bag, and we return the balls each time. Using this strategy, one sample may appear one or several times. The second strategy consists of sampling without replacement k samples from the original learning set, where k is strictly inferior to n. Again if we consider the samples as balls, that means that we do not return the ball to the bag once it was picked up. One of those two strategies can be applied to build all the subsets of training samples.

Unlike Boosting, each subset of learning data is chosen independently from the previous one. As such, the predictors are built in the same way, but with different learning subsets of data (of the same size); hence they may differ in prediction. In RF approach, each predictor is an unpruned classification tree fitted to a single subset of the learning set. Unlike boosting, the successive trees are not correlated and the final prediction combines the prediction of all the trees using a non-weighted vote (all the trees are considered similar). A single classification tree is known as an unstable procedure [Breiman, 1996]. That means that a small change in the training set can result in large changes in the prediction of the decision tree [Strobl et al., 2009]. Growing an ensemble of trees using Bagging allow adjusting the instability of individual trees [Breiman, 1996] [Strobl et al., 2009].

D.2 Random selection of features

Ho [Ho, 1998] was the first to introduce the Random Subspace method in 1998. Breiman [Breiman, 2001] proposed to consider both bagging and random subspace method to build a tree-based ensemble method. Using Bagging, a single tree in RF is trained using a single learning subset. Similarly, the variable selection process performed at each node for each classification tree (i.e. the process of node splitting) is based on a small random subset of features. Given a small subset of learning samples and a small random subset of predictor variables used split each node for each tree, the problem of “small n large p” (few observations and several features) is avoided.

D.3 CART: Classification And Regression Trees

CART is a decision tree algorithm proposed by Breiman et al. in 1986 [Breiman et al., 1984]. The Random Forest approach is usually basd on CART methodology (the forest is constituted with an ensemble of CART predictors).

D.4 OOB: Out-of-Bag Predictions

When using Random Forest approach, it is not necessary to separate the dataset into a training and testing dataset through n-fold cross-validation partition. It is possible to estimate an unbiased estimate of the error of the model during the learning process. This is explained as follow. Each tree in the forest is grown on a bootstrap sample of the data. For a given $Tree_t$ in the forest, only two-third are used in the construction of the $Tree_t$. One-third of the observations are left out of the bootstrap sample; they are called "out-of-bag" (OOB) data. OOB data are useful to obtain the prediction performance of $Tree_t$. Each observation considered in OOB data of $Tree_t$ is used to get a classification using $Tree_t$.

In this way, each observation $observation_o$ in the dataset will be approximately left out (considered as OOB data) of one-third of the trees in the forest [Svetnik et al., 2004]. That is, $observation_o$ is considered as a test set of one-third of the trees. The prediction of $observation_o$ is defined as the most of the votes that is assigned to it based on the trees that consider $observation_o$ as a test set.

Thus, the OOB error estimate of the full ensemble is based on all the OOB predictions; it corresponds to the proportion of times that an OOB prediction is not equal to the true class.

As the number of trees increases, the OOB error estimate has been proven to be an unbiased measure [Verikas et al., 2011]. The OOB error estimate is similar to a cross-validation performance estimate. The computation of OOB error estimate is performed at a much lower computational cost than cross-validation performance estimate [Svetnik et al., 2004].

D.5 Parameters of RF

RF is considered as a very user friendly model as it has only three parameters. Furthermore, parameter optimization step is preferred but it can be skipped as the default values are mostly enough to achieve good performance. Svetnik et al. [Svetnik et al., 2004] showed that RF can produce good prediction performance when used "off the shelf". Díaz-Uriarte et al. [Díaz-Uriarte and Alvarez de Andrés, 2006] also found that RF is robust to changes in parameters values.

The parameters are 1) ntree: the number of trees in the forest, 2) mtry: the number of variables used for the split of each node in a single tree and 3) nodesize: the minimum node size, which is the minimum size of nodes below which no split will be attempted [Svetnik et al., 2004].

D.5.1 Minimum node size

The default value for the minimum node size is 1 in classification. It has been shown that the prediction error does not change whether the nodesize change from 1 to 5 in [Díaz-Uriarte and Alvarez de Andrés, 2006].

D

D.5.2 Number of features (mtry)

The default value of mtry is $p^{\frac{1}{2}}$ for classification [Lunetta et al., 2004] and $\frac{1}{3}$ for regression [Svetnik et al., 2004] [Breiman, 2001]. It has been shown that the model can achieve good prediction accuracy while considering this default value for the number of variables to consider for each split [Svetnik et al., 2004] [Díaz-Uriarte and Alvarez de Andrés, 2006].

D.5.3 Number of trees

The number of trees in the forest (ntree) has to be sufficiently large that the prediction of the ensemble method has stabilized [Svetnik et al., 2004]. Breiman encourages the user to assign a high number of trees [Breiman, 2002]. In previous works, the number of trees can range for instance from 1000 [Svetnik et al., 2004] to 40000 [Díaz-Uriarte and Alvarez de Andrés, 2006]. Svetnik et al. [Svetnik et al., 2004] found that a higher number of trees is needed (e.g. 10 000 trees) if the user is interested in the proportion of votes for a given class, but a smaller number of trees may be sufficient (e.g. 1000) when the user is interested in performance accuracy and variable importance measures [Svetnik et al., 2004]. The authors in [Díaz-Uriarte and Alvarez de Andrés, 2006] found that the time of execution of the training increases almost linearly with the number of trees and that 2000 trees in the forest seems to be sufficient to have stable results that receive negligible changes with a higher trees number. However, the computational cost increases linearly with the number of trees D.2. Choosing arbitrary a high number of trees to guarantee a good performance is a good approach only if a huge computational environment is available [Oshiro et al., 2012].

So far, except few recent works [Oshiro et al., 2012], the literature associated with the RF approach has given few assumptions on the optimal number of trees to achieve a good tradeoff between the computational time and the model prediction accuracy.

Oshiro et al. [Oshiro et al., 2012] were based on the statistical tests (Freidman and Benjamini-Hochberg tests) to analyze whether the decrease of the OOB error rate is significant when increasing the trees number in exponential rates using base two. In their work, they compared the OOB error rates corresponding to several dataset. Inspired from the work of Oshiro et al. [Oshiro et al., 2012], we assume that the search for the optimal number of trees can be based on a statistical test that compare the significance of OOB error rate decrease according to the number of trees. To do so, for each trees number (the trees number can increase in exponential rates or simply linearly), we run the RF model several times in order to study the significant difference of the OOB error rate sets (See Figure D.4). The optimal number of trees can be considered as the one for which the corresponding set of OOB error rates is no more significantly different from the set of OOB error rates corresponding to a high trees number. In other words, the optimal number of trees is chosen in such a way that the OOB error rate does not decrease significantly anymore with the increase of the trees number.

D.6 Applications for the use of RF

RF has become a popular machine learning method for the numerous advantages provided by the model. The basic features that distinguish RF from the traditional statistical models is that it can be used in several non-linear and complex prediction problems along with the capacity to measure the relevance score for each feature [Strobl et al., 2008].

RF approach has been recently widely used in bioinformatics and medical fields as a powerful predictor for its ability to handle high-dimensional data and to cope with “small n large p” problems [Pang et al., 2012], and also as a means to select relevant variables [Pang et al., 2012]. For instance, RF has been used for the analysis of gene expression data generated using DNA microarrays[Pang et al., 2012] [Díaz-Uriarte and Alvarez de Andrés, 2006], and generally in genetic epidemiology and microbiology [Strobl et al., 2007]. RF approach was used to predict disease risks from highly imbalanced data [Khalilia et al., 2011]. RF was also applied to the problem of Quantitative Structure-Activity Relationship (QSAR) modeling for pharmaceutical molecules [Svetnik et al., 2004]. Due to its performance and advantages, Díaz-Uriarte et al. [Díaz-Uriarte and Alvarez de Andrés, 2006] claimed that RF approach and feature selection (gene selection in particular) using RF should probably become part of the "standard tool-box" of methods for class prediction and gene selection with microarray data.

Due to its ability to handle high-dimensional data and its high generalization power, Random Forest approach was also successfully applied by the computer vision and image processing community, in particular in real-time video sequences analysis, for instance to localize and track eye pupil [Markuš et al., 2014] or to recognize facial expression [Abd El Meguid and Levine, 2014].

Up to our knowledge, Random Forest approach has been less used in the context of body movement analysis [Griffin et al., 2013a]. Griffin et al. [Griffin et al., 2013a] compare the performance of several supervised learning models for the automatic prediction of the category of laughing using a predefined set of body movement cues. The predictors were Random Forest, Multi Layer Perceptron with Softmax, k-Nearest Neighbour, Linear and Kernel Ridge Regression, Linear and Kernel Support Vector Regression. Random Forest showed the best performing scores. The authors [Griffin et al., 2013a] highlighted the need to a deeper analysis of the Random Forest model to get insight into the most relevant body cues that could constitute an optimal subset of feature.

D.7 Variable importance (VI) measure in Random Forest

The quantification of model-based relevance measure is derived from the learning process and can be defined as *the general concept of the impact of a predictor variable in predicting the response* [Strobl et al., 2008]. Different measures were proposed in previous works to quantify the relevance of features.

D.7.1 Selection frequency

In a single classification tree, the features selected to build the tree are considered as relevant features for the classification tree as it includes to an embedded feature selection approach. One could define a naive measure of feature relevance in RF approach as the number of times each feature is selected by all the individual trees [Strobl et al., 2007]. However, this measure is not considered as efficient [Strobl et al., 2007].

D.7.2 Gini importance

The Gini importance measure is based on the relevance of features according to the decrease of Gini Index (impurity) that follows the corresponding node split [Altmann et al., 2010]. As such, the importance of the feature and the decrease of Gini Index are linked through a positive linear relationship (the larger is the decrease of impurity after a split, the more relevant is the corresponding feature).

D.7.3 Permutation importance

D

The permutation importance measure is based on the following principle: randomly permuting a predictor variable that is relevant to the prediction accuracy with another one will lead to a decrease in prediction accuracy. Contrary, randomly permuting a predictor variable that is irrelevant to the prediction with another one should have subtle effect on the performance [Svetnik et al., 2004]. The permutation importance of a particular feature refers to the difference in prediction accuracy performed with and without this feature [Strobl et al., 2008, Strobl et al., 2007]. The prediction accuracy that defines the permutation importance is based on the "out-of-bag"(OOB) samples. Permutation importance was widely used as a measure of feature relevance in previous works [Genuer et al., 2010] [Hapfelmeier and Ulm, 2013] [Díaz-Uriarte and Alvarez de Andrés, 2006].

D.7.4 The use of VI in previous works

Strobl et al. [Strobl et al., 2007] conducted a series of simulation studies to investigate the reliability of selection frequency, Gini importance and permutation importance measures of relevance. In particular, they considered situations where potential features vary in their scale of measurement or their number of categories [Strobl et al., 2007]. In fact, they considered five variables; the first variable is continuous while the other variables are categorical. The number of categories is different for each categorical variable (the number of categories ranges between 2 and 20). They found that the three variable importance measures (selection frequency, Gini importance and permutation importance) of the original Random Forest approach are affected by the number of categories in categorical variables and by the scale of measurement of the variables. However, it has been shown that selection frequency and permutation importance measures are more robust to changes in categories number and the scale of measurement when conditional inference trees are used [Hothorn et al., 2006]. In our work, we aim to use RF VI to quantify the relevance of continuous features.

A widely common issue that should be considered with caution in variable importance measure is the impact of features correlation on variable importance measure. Recent works have been focused on the impact of correlated features on the measure of variable importance quantified using RF (in particular permutation importance) [Strobl et al., 2008], [Tolosi and Lengauer, 2011] [Genuer et al., 2010]. Most of them showed that the presence of highly correlated features can seriously affect the measure of VI in RF.

Tolosi et al. [Tolosi and Lengauer, 2011] investigated the impact of features correlation on Random Forest variable importance (RF-VI). They simulate artificial data containing unbalanced groups of correlated features as well as a group of irrelevant features. They found that RF-VI of irrelevant features always receives the lowest values (tend to zero) regardless the amount of correlated features, which is a good result. However, they found that the RF-VI values of less relevant features become higher that RF-VI values of the most relevant features when the cardinality of the group containing the most relevant correlated features is much higher than the cardinality of the group containing the less relevant and correlated features (e.g. 180 relevant correlated features versus 20 less relevant correlated features). Thus, a naïve quantification of VI using RF in the presence of highly imbalanced groups of correlated features can affect the ranking of features, then can lead to a confounding in the interpretation of the model [Tolosi and Lengauer, 2011].

Motivated by the sensitivity of permutation importance to highly correlated features, Strobl et al. [Strobl et al., 2008] proposed a new conditional variable importance for Random Forest. However, recent studies indicate that the original permutation importance is still extremely popular and desired for its unconditional properties [Hapfelmeier and Ulm, 2013].

In a recent study, Genuer et al. [Genuer et al., 2010] simulate the presence of correlated variables to study the impact of variables correlation on the permutation importance measure. In the first simulation, they added a number of correlated variables to the variable that has the highest permutation importance measure. They showed that the permutation importance of this variable decrease with the increase of correlated variables. However, the highest permutation importance was always assigned to one of those correlated variables (and not another uncorrelated variable). Besides, when adding new groups of variables that are respectively correlated to two original (uncorrelated) variables, they found that the relative position between the two original variables is preserved in the two groups of variables. Besides, as a result to their proposed feature selection technique (dedicated to enhance the prediction performance) based on the ranking of the variables according to their importance, they found that it is possible to eliminate correlated variables and to keep only the most important and uncorrelated ones.

Motivated by the work of Genuer et al. [Genuer et al., 2010], we used the original permutation measure as a measure of the relevance of features during the training of Random Forest model. However, further analyses based on the conditional variable importance measure can be conducted to explore the effect of features correlation on the selection of the most relevant features. Indeed, using the conditional variable importance measure may lead to a smallest subset of features selected according to their relevance measure.

Figure D.1: A general architecture of a Random Forest [Verikas et al., 2011].

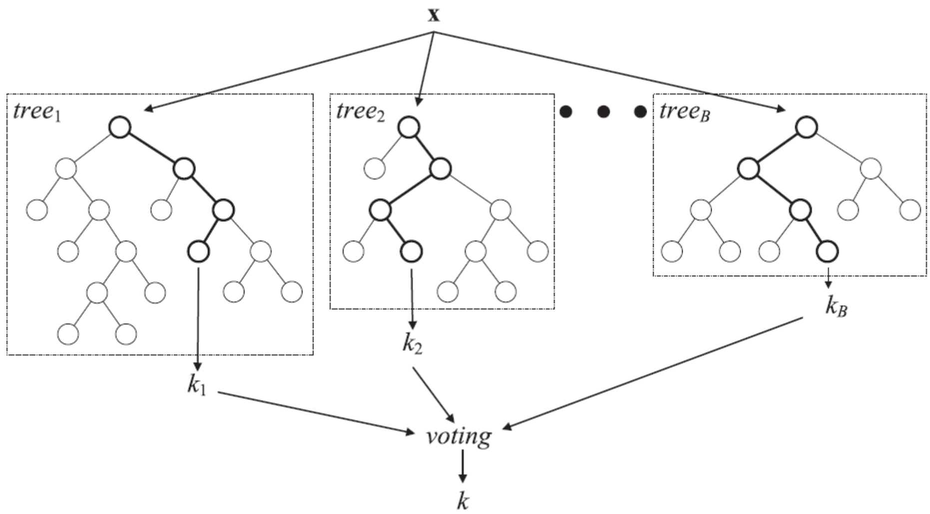
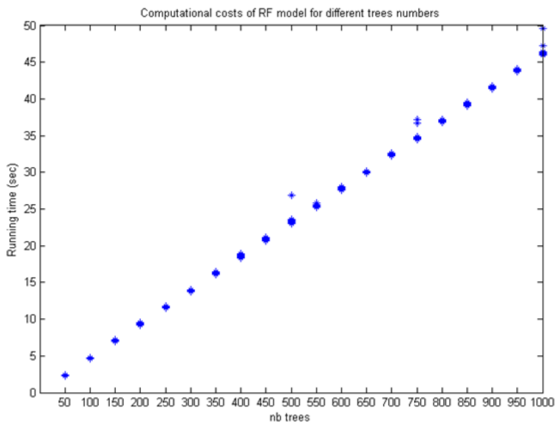


Figure D.2: The linear increase of RF computational cost. The RF model was run 20 times for each trees number. The RF model was run on a dataset constituted with 1022 rows, 8 classes and 110 features.

(a) Mean of running times in seconds



(b) Boxplot of running times in seconds.

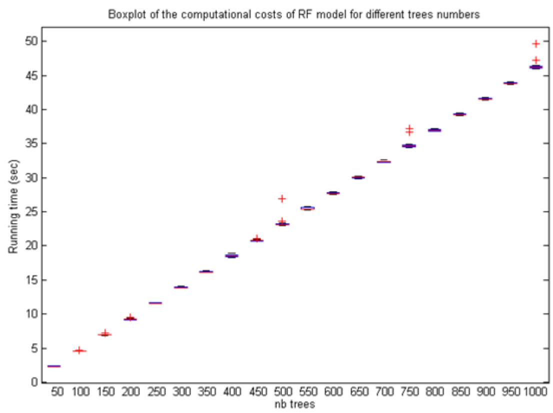
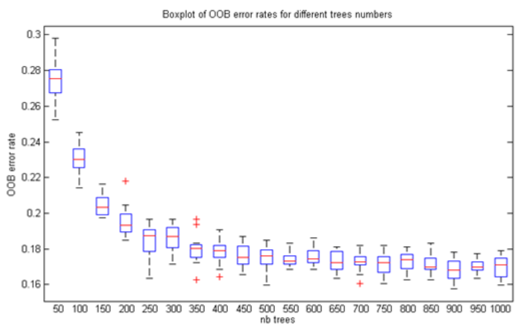
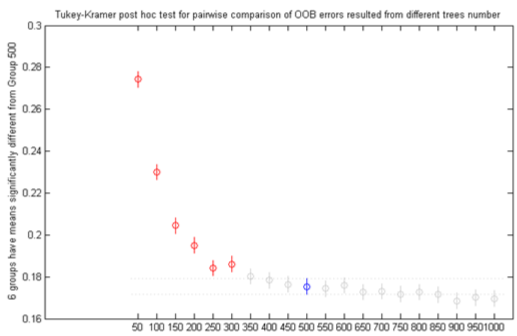


Figure D.4: The decrease of OOB error rates for each RF model build with a different trees number and run 20 times. The RF model was run on a dataset constituted with 1022 rows, 8 classes and 110 features.

(a) The boxplot of the OOB error rates for each RF model run 20 times for each trees number



(b) The result of a multiple comparison test ('tukey-kramer') to compare the significance of difference between OOB rates of different trees number; building the model with more than 500 trees is pointless as there is no significant difference in OOB rates starting from the model build with 500 trees until the model build with 1000 trees (while increasing the trees number by a step of 50 trees)



E Detailed description of motion capture features derived from MLBNS

In the following tables, we provide a fine-grained description of motion capture features derived from the MLBNS (Multi-Level Body Notation System) introduced in Chapter 6. For the sake of clarity, the description of features is split into 4 sections. Each section refers to features from a different anatomical description level; Global, Body Parts, Semi-Global and Local.

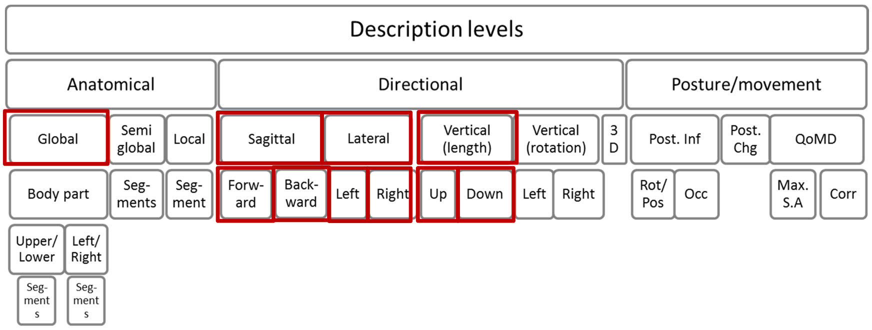
In each section, we firstly provide a fine-grained description of each body cue while considering 7 main aspects as presented just next. Then we summarize the list of features defined according to that anatomical description level (See Appendix I for more details about the acronym used for the features names). The 7 main aspects considered to describe body features are:

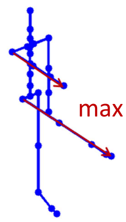
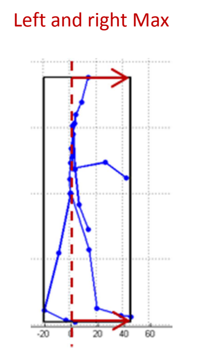
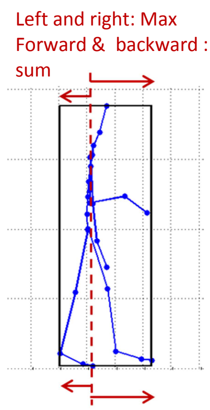
1. the name of the body cue as well as the number of derived features,
2. the body segments concerned by the body cue,
3. the corresponding directional description level,
4. the corresponding type of feature (e.g. postural feature, speed or acceleration feature),
5. a short description of each feature. Position stands for a feature based on the position of the body segment(s) according to the hips. Rotation stands for a feature based on the rotation of the body segment(s).
6. a graphical representation of the postural reference (if existing)
7. a graphical representation of the body cue.

Section E.1 describes in detail the 16 motion capture features associated to the whole body movement. This set of 16 features is defined along the Global Anatomical level in the MLBNS. Section E.2 illustrates the 38 motion capture features that describe the bounding boxes surrounding different body parts (mainly the arms, the Torso+Head and lower body). These 38 features are defined along the Body Part Anatomical level in the MLBNS. Section E.3 illustrates the 30 motion capture features associated to the Semi-Global Anatomical description level defined in the MLBNS. Last but not least, section E.4 illustrates the 46 motion capture features associated to the Local Anatomical description level defined in the MLBNS. Finally we clarify few important points regarding the way we measure these body features.

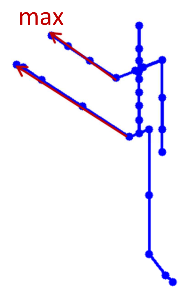
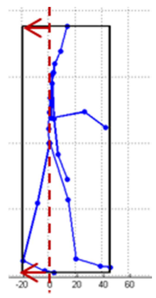
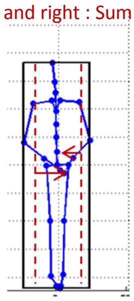
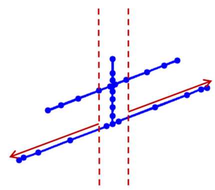
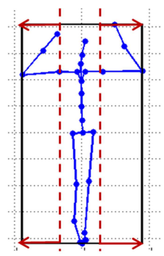
E.1 GLOBAL BODY CUES

Anatomical description level: Global

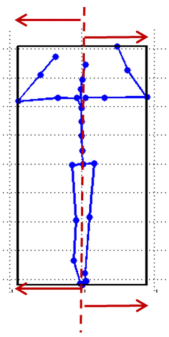
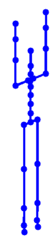
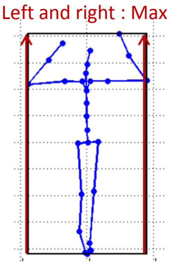
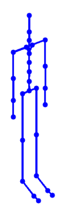
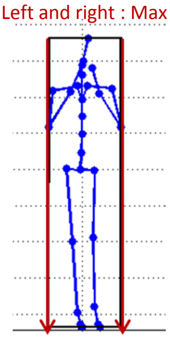


Body Cue	Directional level	Short description	Post/ Occ/ STD/ Speed/ Acc/ Corr	Graphical representation of the Reference	Graphical representation of Feature
Forward Whole Body (#2)	Sagittal (Forward)	- Position - Relative forward whole body movement	Post		
		- Position - The standard deviation of forward whole body movement	STD		
Sagittal Whole Body Max.S.A (#2)	Sagittal	- Position - The speed of sagittal whole body movement defined as the sum of forward and backward movement	Speed		
		- Position - The acceleration of sagittal whole body movement defined as the sum of forward and backward movement	Acc		

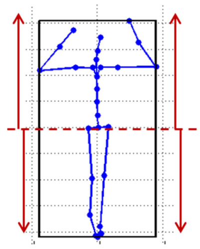
E

Backward Whole Body (#2)	Sagittal Backward	- Position - Relative backward whole body movement	Post		Left and right : Max 
		- Position - Standard deviation of backward whole body movement	STD		
Inside Whole Body (#1)	Lateral (Inside)	-Position - Lateral closeness body movement defined as the sum of right and left lateral closeness	Post	-	Left and right : Sum 
Outside Whole Body (#1)	Lateral (Outisde)	-Position - Relative lateral openness of body movement defined as the sum of right and left lateral openness	Post		Left and right : Sum 

E

Lateral Whole Body STD (#1)	Lateral	- Position - The standard deviation of lateral whole body movement defined as the sum of right and left lateral openness	STD		
Lateral Whole Body Max.S.A (#2)	Lateral	- Position - The speed of lateral whole body movement defined as the sum of right and left lateral openness	Speed		
		- Position - The acceleration of lateral whole body movement defined as the sum of right and left lateral openness	Acc		
Upward Whole Body (#1)	Vertical Length: Upward	- Position - Relative vertical body movement	Post		
Downward Whole Body (#1)	Vertical Length: Downward	- Position - Relative MeanMin of vertical body movement	Post		

APPENDIX E. DETAILED DESCRIPTION OF MOTION CAPTURE FEATURES DERIVED FROM MLBNS

Vertical Whole Body STD (#1)	Vertical Length	- Position - Standard deviation of vertical body movement defined as the sum of upper body and lower body vertical extensions	STD		Upper and Lower : Sum Left and Right : Max 
Vertical Whole Body Max.S.A (#2)	Vertical Length	- Position - Speed of vertical body movement defined as the sum of upper body and lower body vertical extensions	Speed		
		- Position - Acceleration of vertical body movement defined as the sum of upper body and lower body vertical extensions	Acc		

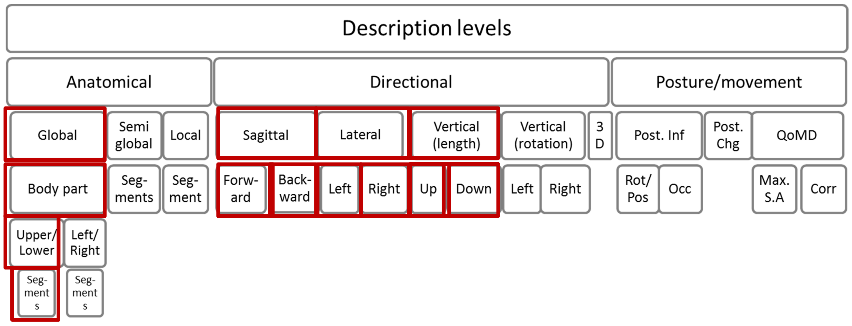
E

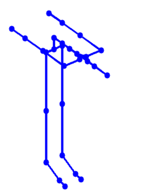
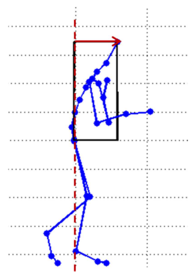
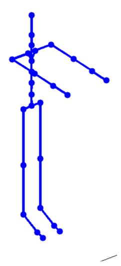
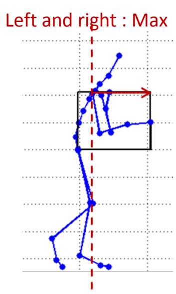
Table E.1: A summary of the 16 features describing Global description level in Anatomical dimension: Sagittal, Vertical (length) and Lateral directions of the bounding box that surround the whole body

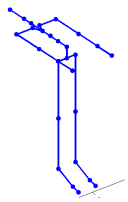
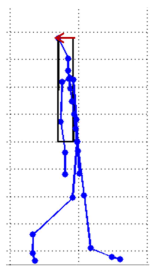
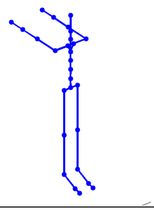
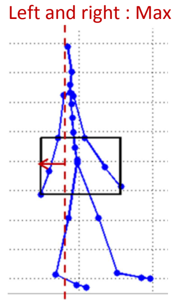
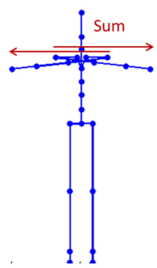
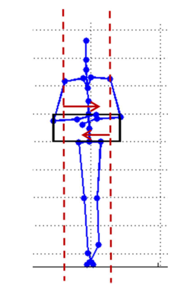
Body cues	Features
Global (#16)	Post-Outside-WholeBody
	Post-Inside-WholeBody
	Post-Forward-WholeBody
	Post-Backward-WholeBody
	Post-Upward-WholeBody
	Post-Downward-WholeBody
	STD-Outside-WholeBody
	STD-Forward-WholeBody
	STD-Backward-WholeBody
	STD-Upward-WholeBody
	Speed-Lateral-WholeBody
	Acc-Lateral-WholeBody
	Speed-Vertical-WholeBody
	Acc-Vertical-WholeBody
	Speed-Sagittal-WholeBody
	Acc-Sagittal-WholeBody

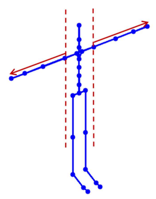
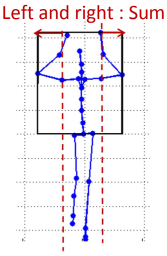
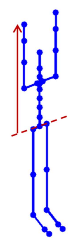
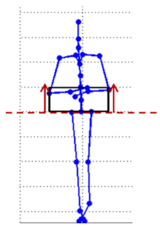
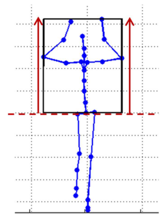
E.2 Body Parts body cues

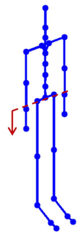
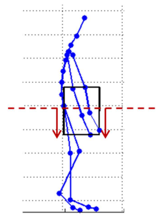
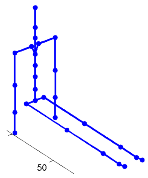
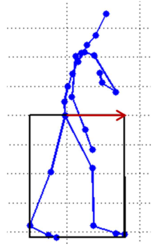
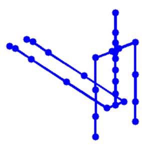
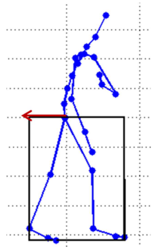
Anatomical description level: Body Parts

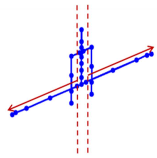
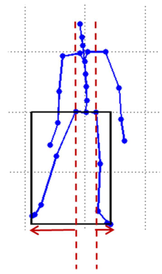
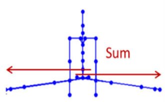
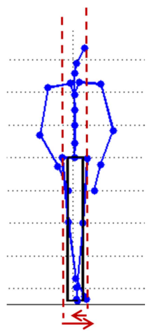
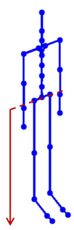
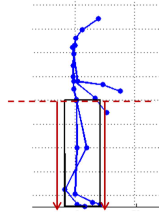


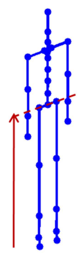
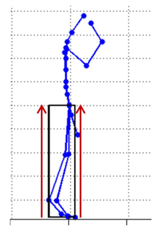
Body Cue	Anatomical Level	Directional Level	Post/ Occ/ STD/ Speed / Acc/ Corr	Short description	Graphical representation of the Reference	Graphical representation of Feature
Froward Torso (#2)	TorsoHead: Shoulders, Head end effector	Sagittal (Forward)	Post	- Position - Relative forward torso movement		
			STD	- Position - The standard deviation of forward torso movement		
Forward Torso Max.S.A (#2)	TorsoHead: Shoulders, Head end effector	Sagittal	Speed	- Position - The speed of sagittal torso movement		
			Acc	- Position - The acceleration of sagittal torso movement		
Froward Arms (#2)	Arms: Elbows, Wrists, Wrist end effectors	Sagittal (Forward)	Post	- Position - Relative forward arms movement		

			STD	- Position - The standard deviation of forward arms movement		
Froward Arms Max.S.A (#2)	Arms: Elbows, Wrists, Wrist end effectors	Sagittal	Speed	- Position - The speed of sagittal arms movement		
			Acc	- Position - The acceleration of sagittal arms movement		
Backward Torso (#2)	TorsoHead: Shoulders, Head end effector	Sagittal (Backward)	Post	- Position - Relative backward torso movement		
			STD	- Position - The standard deviation of backward torso movement		
Backward Arms (#2)	Arms: Elbows, Wrists, Wrist end effectors	Sagittal (Backward)	Post	- Position - Relative backward arms movement		Left and right : Max 
			STD	- Position - The standard deviation of backward arms movement		
Inside Arms (#1)	Arms: Shoulders, Elbows, Wrists, Wrist end effectors	Lateral (Inside)	Post	- Position - Relative lateral closeness of arms movement with respect to shoulders position		Left and right : Sum 

Lateral Arms STD (#1)	Arms: Shoulders, Elbows, Wrists, Wrist end effectors	Lateral	STD	- Position - The standard deviation of lateral arms movement		
Arms Lateral Max.S.A (#2)	Arms: Shoulders, Elbows, Wrists, Wrist end effectors	Lateral	Speed	- Position - The speed of lateral arms movement		
			Acc	- Position - The acceleration of lateral arms movement		
Outside Arms (#1)	Arms: Shoulders, Elbows, Wrists, Wrist end effectors	Lateral (Outside)	Post	- Position - Relative lateral openness of arms movement		Left and right : Sum 
Upward Arms (#1)	Arms: Elbows, Wrists, Wrist end effectors	Vertical: length (Upward)	Post	- Position - Relative upward arms movement		Left and right : Max  
Vertical Arms STD (#1)	Arms: Elbows, Wrists, Wrist end effectors	Vertical: length	STD	- Position - The standard deviation of vertical arms movement		
Vertical Arms Max.S.A (#2)	Arms: Elbows, Wrists, Wrist end effectors	Vertical: length	Speed	- Position - The speed of vertical arms movement		
			Acc	- Position - The acceleration of vertical arms movement		

Downward Arms (#1)	Arms: Elbows, Wrists, Wrist end effectors	Vertical: length (Down-ward)	Post	- Position - Relative downward arms movement		Left and right : Max 
Forward Lower Body (#2)	Lower Body: Hips, Knees, Ankles, Ankle end effectors	Sagittal (Forward)	Post	- Position - Relative forward legs movement		Left and right : Max 
			STD	- Position - The standard deviation of forward legs movement		
Sagittal Lower Body Max.S.A (#2)	Lower Body: Hips, Knees, Ankles, Ankle end effectors	Sagittal	Speed	- Position - The speed of sagittal legs movement		
			Acc	- Position - The acceleration of sagittal legs movement		
Backward Lower Body (#2)	Lower Body: Hips, Knees, Ankles, Ankle end effectors	Sagittal (Backward)	Post	- Position - Relative backward legs movement		Left and right : Max 
			STD	- Position - The standard deviation of backward legs movement		

Outside Lower Body (#1)	Lower Body: Hips, Knees, Ankles, Ankle end effectors	Lateral (Outside)	Post	- Position - Relative lateral openness of legs movement		Left and right : Sum 
Lateral Lower Body STD (#1)	Lower Body: Hips, Knees, Ankles, Ankle end effectors	Lateral	STD	- Position - The standard deviation of lateral legs movement		
Lateral Lower Body Max.S.A (#2)	Lower Body: Hips, Knees, Ankles, Ankle end effectors	Lateral	Speed	- Position - The speed of lateral legs movement		
			Acc	- Position - The acceleration of lateral legs movement		
Inside Lower Body (#1)	Lower Body: Hips, Knees, Ankles, Ankle end effectors	Lateral (Inside)		- Position - Relative lateral closeness of legs movement		Left and right : Sum 
Downward Lower Body (#1)	Lower Body: Hips, Knees, Ankles, Ankle end effectors	Vertical: length (Down-ward)	Post	- Position - Relative downward legs movement		Left and right : Max 

Vertical Lower Body STD (#1)	Lower Body: Hips, Knees, Ankles, Ankle end effectors	Vertical: length	STD	- Position - The standard deviation of vertical legs movement		
Vertical Lower Body Max.S.A (#2)	Lower Body: Hips, Knees, Ankles, Ankle end effectors	Vertical: length	Speed	- Position - The speed of vertical legs movement		
			Acc	- Position - The Speed & acceleration of vertical legs movement		
Upward Lower Body (#1)	Lower Body: Hips, Knees, Ankles, Ankle end effectors	Vertical: length (Upward)	Post	- Position - Relative upward legs movement		Left and right : Max 

E

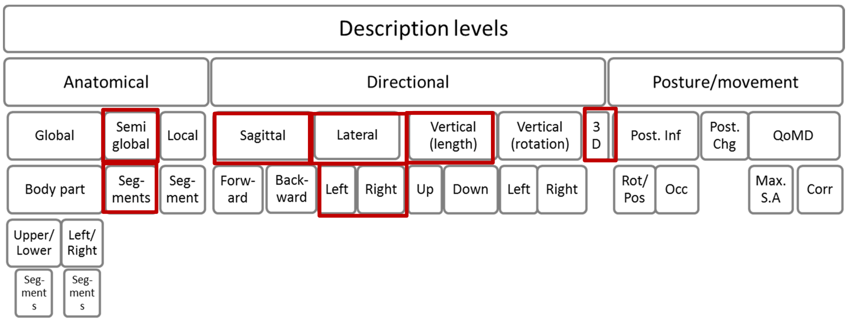
Table E.2: A summary of the 38 features describing Body Part describtion level in Anatomical dimension: Sagittal, Vertical (length) and Lateral directions of the bounding boxes that surround upper and lower body parts. Upper body parts are split into Torso+Head and Arms.

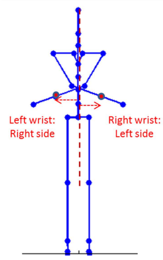
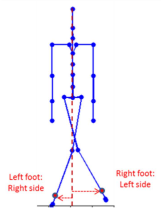
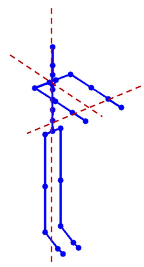
Body cues	Features
Body Part: Arms (#16)	Post-Outside-Arms Post-Inside-Arms Post-Forward-Arms Post-Backward-Arms Post-Upward-Arms Post-Downward-Arms STD-Forward-Arms STD-Backward-Arms STD-Lateral-Arms STD-Vertical-Arms Speed-Lateral-Arms Speed-Vertical-Arms Speed-Sagittal-Arms Acc-Lateral-Arms Acc-Vertical-Arms Acc-Sagittal-Arms
Body Part: Lower Body (#16)	Post-Outside-LowerBody Post-Inside-LowerBody Post-Forward-LowerBody Post-Backward-LowerBody Post-Upward-LowerBody Post-Downward-LowerBody STD-Forward-LowerBody STD-Backward-LowerBody STD-Lateral-LowerBody STD-Vertical-LowerBody Speed-Lateral-LowerBody Speed-Vertical-LowerBody Speed-Sagittal-LowerBody Acc-Lateral-LowerBody Acc-Vertical-LowerBody Acc-Sagittal-LowerBody
Body Part: Trunk (#6)	Post-Forward-TorsoHead Post-Backward-TorsoHead STD-Forward-TorsoHead STD-Backward-TorsoHead Speed-Sagittal-TorsoHead Acc-Sagittal-TorsoHead

E

E.3 Semi-Global body cues

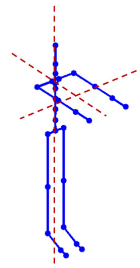
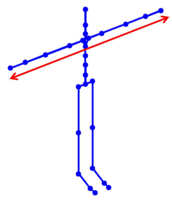
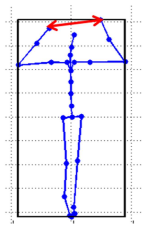
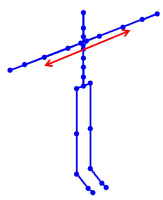
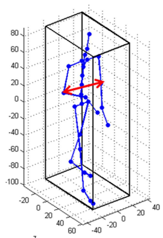
Anatomical description level: Semi Global

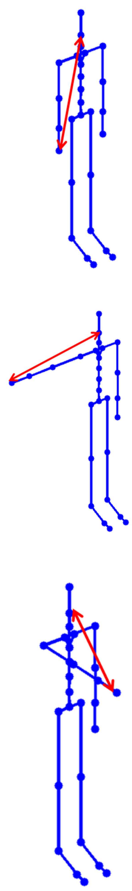
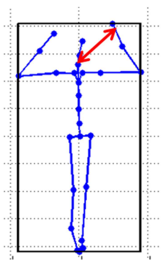


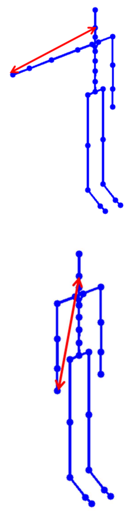
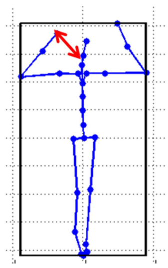
Body Cue	Anatomical Level	Directional Level	Post/ Occ/ STD/ Speed/ Acc/ Corr	Short description	Graphical representation of the Reference	Graphical representation of Feature
Hands Crossing (#1)	Wrists	Lateral	Occ	- Position - Percentage of time (percentage of frames) for which the lateral wrists positions are crossed (right wrist on the left side, left wrist on the right side)		
Feet Crossing (#1)	Feet	Lateral	Occ	- Position - Percentage of time (percentage of frames) for which the lateral feet positions are crossed (right leg on the left side, left leg on the right side)		
Hands Posture Symmetry (#1)	Wrists	ThreeD	Occ	- Position - Percentage of time (percentage of frames) for which the 3d positions of left and right wrist are approximately symmetric (in lateral, sagittal and vertical directions)		

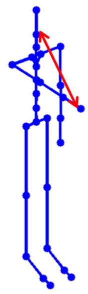
E

APPENDIX E. DETAILED DESCRIPTION OF MOTION CAPTURE FEATURES DERIVED FROM MLBNS

Symmetry Occ elbows (#1)	Elbows	ThreeD	Occ	- Position - Percentage of time (percentage of frames) for which the 3d positions of left and right elbows are approximately symmetric (in lateral, sagittal and vertical directions)		
Hands Relationship (#4)	Hands end effectors	ThreeD	Post	- Position - Relative hands relationship defined as the 3d distance between hands (end effectors)		
			STD	- Position - Standard deviation of hands relationship defined as the distance between hands (end effectors)		
			Speed	- Position - Speed of hands relationship defined as the 3d distance between hands (end effectors)		
			Acc	- Position - Acceleration of hands relationship defined as the 3d distance between hands (end effectors)		
Elbows Relationship (#4)	Elbows	ThreeD	Post	- Position - Relative elbows relationship defined as the 3d distance between elbows		
			STD	- Position - Standard deviation of elbows relationship defined as the distance		

				between elbows		
			Speed	- Position - Speed of elbows relationship defined as the 3d distance between elbows		
			Acc	Position - Acceleration of elbows relationship defined as the 3d distance between elbows		
Left Hand - Head Relationship (#4)	Head, Left Hand end effector	ThreeD	Post	- Position - Relative head-left hand relationship defined as the3d distance between left hand end effector and head.	The reference is defined as the average of maximal distance in lateral, sagittal and vertical directions. 	

			STD	- Position - Standard deviation of left hand and head relationship (the 3d distance between left hand end effector and head)		
			Speed	- Position - Speed of left hand and head relationship (the 3d distance between left hand end effector and head)		
			Acc	- Position - Acceleration of left hand end effector and head relationship (the 3d distance between left hand end effector and head)		
Right Hand – Head Relationship (#4)	Head, Right hand End effector	ThreeD	Post	- Position - Relative head-right hand relationship defined as the 3d distance between right hand end effector and head.	The reference is defined as the average of maximal distance in lateral, sagittal and vertical directions. 	

						
			STD	- Position - Standard deviation of right hand and head relationship (3d distance between right hand end effector and head)		
			Speed	- Position - Speed of right hand and head relationship (the 3d distance between left hand end effector and head)		
			Acc	Position - Acceleration of right hand and head relationship (3d distance between right hand effector and head)		
Hands Motion Correlation (#3)	Wrists	Sagittal	Corr	- Position - Pearson coefficient measure to estimate the correlation between left and right wrist movement along the sagittal direction		
		Lateral	Corr	- Position - Pearson coefficient measure to estimate the correlation between left and right wrist movement along the lateral direction		
		Vertical	Corr	- Position - Pearson coefficient measure the correlation between left and right wrist		

E

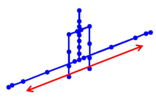
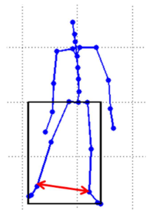
				movement along the vertical direction		
Elbows Motion Correlation (#3)	Ellbows	Sagittal	Corr	- Position - Pearson coefficient measure the correlation between left and right elbow movement along the sagittal direction		
		Lateral	Corr	- Position - Pearson coefficient measure the correlation between left and right elbow movement along the lateral direction		
		Vertical	Corr	- Position - Pearson coefficient measure the correlation between left and right elbow movement along the vertical direction		
Feet Relationship (#4)	Feet	ThreeD	Post	- Position - Relative feet relationship defined as the 3d distance between feet		
			STD	- Position - Standard deviation of feet relationship defined as the distance between feet		
			Speed	- Position - Speed of feet relationship defined as the 3d distance between feet		
			Acc	Position - Acceleration of feet relationship defined as the 3d distance between feet		

Table E.3: A summary of the features describing Semi-Global describtion level in Anatomical dimension. L and R stand respectively for Left and Right.

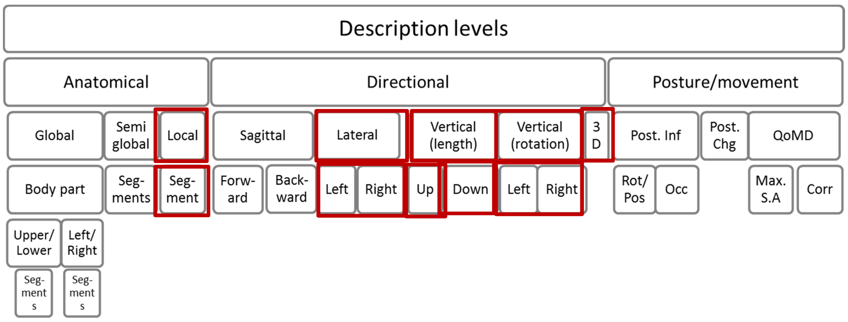
	Body cues	#	Features
Semi-Global (#30)	*Feet Relationship (LFoot.RFoot)*	#4	3D distance between feet: Post-ThreeD-Lfoot.Rfoot STD-ThreeD-Lfoot.Rfoot Speed-ThreeD-Lfoot.Rfoot Acc-ThreeD-Lfoot.Rfoot
	Hands Relationship (LHand.RHand)	#4	3D distance between the hands: Post-ThreeD-Lhand.RHand STD-ThreeD-Lhand.RHand Speed-ThreeD-Lhand.RHand Acc-ThreeD-Lhand.RHand
	Elbows Relationship (LElbow.RElbow)	#4	3D distance between the elbows: Post-ThreeD-Lelbow.Relbow STD-ThreeD-Lelbow.Relbow Speed-ThreeD-Lelbow.Relbow Acc-ThreeD-Lelbow.Relbow
	Hands Posture Symmetry	#1	The occurrence (percentage of time) of the 3D symmetry of hands positions posture: SymOcc-ThreeD-Lhand.RHand
	Elbows Posture Symmetry	#1	The occurrence (percentage of time) of the 3D symmetry of elbows positions posture: SymOcc-ThreeD-Lelbow.Relbow
	Hands Motion Symmetry	#3	Hands motion correlation in Sagittal, Vertical and Lateral directions: Corr-Lateral-Lhand.RHand Corr-Vertical-Lhand.RHand Corr-Sagittal-Lhand.RHand
	Elbows Motion Symmetry	#3	Elbows motion correlation in Sagittal, Vertical and Lateral directions: Corr-Lateral-Lelbow.Relbow Corr-Vertical-Lelbow.Relbow Corr-Sagittal-Lelbow.Relbow
	Right Hand - Head Relationship (RHand.Head)	#4	3D distance between the right hand and the head: Post-ThreeD-Rhand.Head STD-ThreeD-Rhand.Head Speed-ThreeD-Rhand.Head Acc-ThreeD-Rhand.Head
	Left Hand - Head Relationship (LHand.Head)	#4	3D distance between the left hand and the head: Post-ThreeD-Lhand.Head STD-ThreeD-Lhand.Head Speed-ThreeD-Lhand.Head Acc-ThreeD-Lhand.Head
	Arms Crossing	#1	The occurrence (percentage of time) of lateral hands crossing and legs crossing:

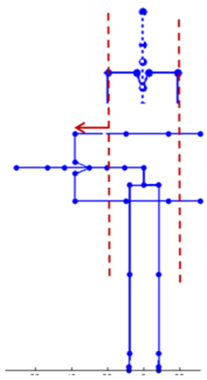
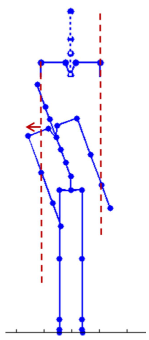
E

			CrossOcc-Lateral-Arms
	Feet Crossing	#1	The occurrence (percentage of time) of lateral feet crossing and legs crossing: CrossOcc-Lateral-Lfoot.Rfoot

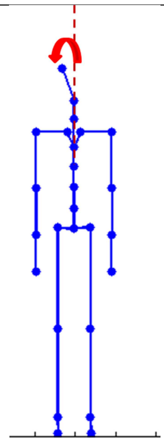
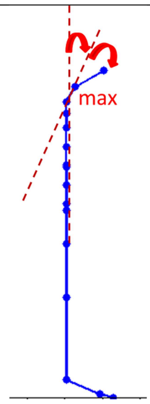
E.4 LOCAL BODY CUES

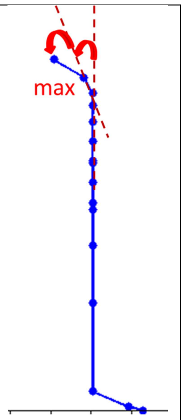
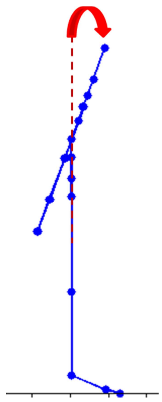
Anatomical description level: Local

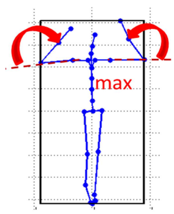
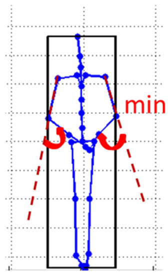


Body Cue	Anatomical Level	Directional Level	Post/ Occ/ STD/ Speed/ Acc/ Corr	Short description	Graphical representation of the Reference	Graphical representation of Feature
Lateral Torso Rotation (#4)	Torso (Shoulders)	Lateral	Post	- Position - Relative lateral torso movement		
			STD	- Position - Standard deviation of lateral torso movement		
			Speed	- Position - Speed of lateral torso movement		
			Acc	- Position - Acceleration of lateral torso movement		

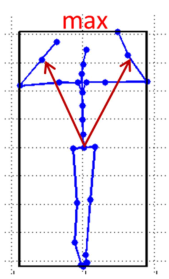
E

Lateral Head Rotation (#4)	Head	Lateral	Post	- Rotation - Lateral head movement	-	
			STD	- Rotation - Standard deviation of lateral head movement		
			Speed	- Rotation - Speed of lateral head movement		
			Acc	- Rotation - Acceleration of lateral head movement		
Head Flexion (#5)	Head, Neck	Vertical (Length)	Post	- Rotation - Downward head movement	-	

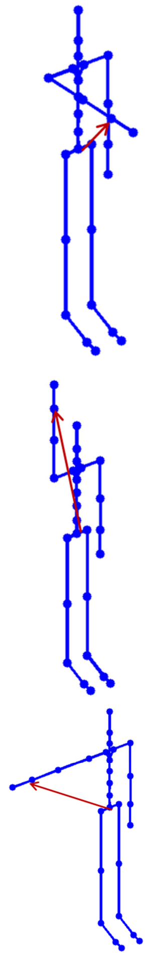
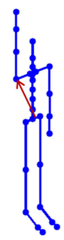
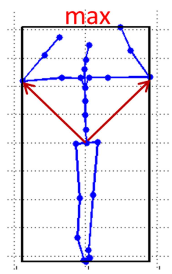
				- Upward head movement		
			STD	- Rotation - Standard deviation of head flexion movement		
			Speed	- Rotation - Speed of head flexion movement		
			Acc	- Rotation - Acceleration of head flexion movement		
Torso Flexion (#4)	Chest	Vertical (Length)	Post	- Rotation - Downward torso movement	-	

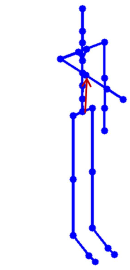
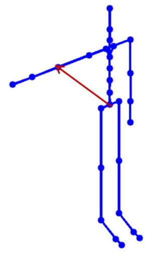
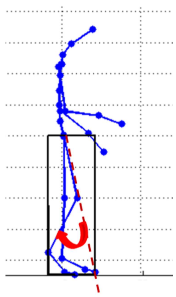
			STD	- Rotation - Standard deviation of torso flexion movement		
			Speed	- Rotation - Speed of torso flexion movement		
			Acc	- Rotation - Acceleration of torso flexion movement		
Elbows Flexion (#5)	Elbows	Vertical (Length)	Post	- Rotation - Relative angular flexion of the elbows (elbow flexion higher than pi/3)		Left and right: Max 
				- Rotation - Relative round flexion of the elbows (elbow flexion lower than pi/3)		Left and right: Min 
			STD	- Rotation - Standard deviation of elbow flexion movement		
			Speed	- Rotation - Speed of elbow flexion movement		

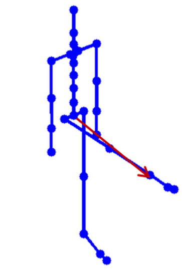
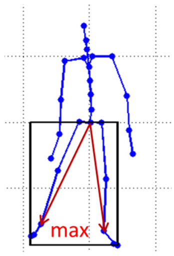
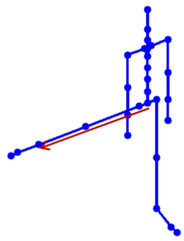
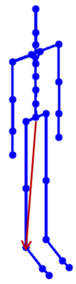
E

			Acc	- Rotation - Acceleration of elbow flexion movement		
Vertical Head Rotation (#4)	Head	Vertical (Rotation)	Post	- Rotation - Left/right head movement	-	
			STD	- Rotation - Standard deviation of left/right head movement		
			Speed	- Rotation - Speed of vertical (left/right) head movement		
			Acc	- Rotation - Acceleration of vertical (left/right) head movement		
Vertical Torso Rotation (#4)	Chest	Vertical (Rotation)	Post	- Rotation - Left/right torso movement		
			STD	- Rotation - Standard deviation of left/right torso movement		
			Speed	- Rotation - Speed of vertical (left/right) torso movement		
			Acc	- Rotation - Acceleration of vertical (left/right) torso movement		
Hands Extension (#4)		ThreeD	Post	- Position - Relative 3D hands extension according to hips center	The reference is set to the average of sagittal, vertical and lateral extension.	Left and right: Max 
			STD	- Position - Standard deviation of wrists movement defined as the distance between wrists and hips center		
			Speed	- Position - Speed of hands movement defined as the 3D distance		

E

	Wrists			between hands and hips center		
			Acc	- Position - Acceleration of wrists movement defined as the 3D distance between wrists and hips center		
Elbows Extension (#4)	Elbows	ThreeD	Post	- Position - Relative 3D elbows extension according to the hips center	The reference is set to the average of sagittal, vertical and lateral extension. 	Left and right: Max 
			STD	- Position - Standard deviation of elbows movement defined as the distance between elbows and hips center		

			Speed	- Position - Speed of elbows movement defined as the 3D distance between elbows and hips center		
			Acc	- Position - Acceleration of elbows movement defined as the 3D distance between elbows and hips center		
Knee Flexion (#4)	Knees	Vertical (Length)	Post	- Rotation - Knee flexion	-	Left and right: Mean 
			STD	- Rotation - Standard deviation of knees flexion		
			Speed	- Rotation - Speed of knees flexion		
			Acc	- Rotation - Acceleration of knees flexion		
Feet Extension (#4)		ThreeD	Post	- Position - Relative feet extension according to hips center	The reference is set to the average of sagittal, vertical and lateral extension.	

	Feet		STD	- Position - Standard deviation of feet movement defined as the distance between feet and hips center		Left and right: Max 
			Speed	- Position - Speed of feet movement defined as the distance between Ankles and hips center		
			Acc	- Position - Acceleration of feet movement defined as the distance between Ankles and hips center		

E

Table E.4: A summary of the features describing Local describtion level in Anatomical dimension. L and R stand respectively for Left and Right.

	Body cues	#	Features
Local (#46)	*Hands Extension (Hands.Hips)*	#4	3D extension of hands regarding the body center (Hips center): Post-ThreeD-Hands.Hips STD-ThreeD-Hands.Hips Speed-ThreeD-Hands.Hips Acc-ThreeD-Hands.Hips
	Elbows Extension (Elbows.Hips)	#4	3D extension of elbows regarding the body center (Hips center): Post-ThreeD-Elbows.Hips STD-ThreeD-Elbows.Hips Speed-ThreeD-Elbows.Hips Acc-ThreeD-Elbows.Hips
	Feet Extension (Feet.Hips)	#4	3D extension of feet regarding the body center (Hips center): Post-ThreeD-Feet.Hips STD-ThreeD-Feet.Hips Speed-ThreeD-Feet.Hips Acc-ThreeD-Feet.Hips
	Lateral Torso Rotation	#4	Lateral torso leaning to the left/right: Post-Lateral-Torso STD-Lateral-Torso Speed-Lateral-Torso Acc-Lateral-Torso
	Lateral Head Rotation	#4	Lateral head leaning to the left/right: Post-Lateral-Head STD-Lateral-Head Speed-Lateral-Head Acc-Lateral-Head
	Elbows Flexion	#5	Angular and round elbows flexion: Post-Upward.Flexion-Elbows Post-Downward.Flexion-Elbows STD-Flexion-Elbows Speed-Flexion-Elbows Acc-Flexion-Elbows
	Head Flexion	#5	Downward and Upward head orientation: Post-Downward.Flexion-Head Post-Upward.Flexion-Head STD-Flexion-Head Speed-Flexion-Head Acc-Flexion-Head
	Torso Flexion	#4	Collapsed torso orientation: Post-Downward.Flexion-Torso STD-Flexion-Torso Speed-Flexion-Torso Acc-Flexion-Torso
	Knee Flexion	#4	Flexion of knees: Post-Downward.Flexion-Knees

E

			STD-Flexion-Knees Speed-Flexion-Knees Acc-Flexion-Knees
	Vertical Head Rotation	#4	Left/right head rotation: Post-LRRotation-Head STD-LRRotation-Head Speed-LRRotation-Head Acc-LRRotation-Head
	Vertical Torso Rotation	#4	Left/right torso rotation: Post-LRRotation-Torso STD-LRRotation-Torso Speed-LRRotation-Torso Acc-LRRotation-Torso

E.5 Discussion

Most of the postural features of our MLBNS are defined according to a reference posture which is obtained by the maximal extension of body limbs. We refer to such postural features as “Relative” postural features. As such, these features are normalized according to the skeleton size, allowing to avoid the bias of individual size.

The set of motion capture features proposed in our work are measured along the whole sequence of movement rather than for each movement phase (e.g. one knock in knocking action, one step in walking action). The value of a postural feature measured along the whole motion sequence is set to the average of absolute and local maximum/ minimum. The final step in Figure E.1 illustrates the detection of the absolute and local maximum. For repetitive actions (such as walking, knocking and moving books from one side to another), such local maximum/ minimum may correspond to one motion phase. For instance, each local maximum of forward leg movement in walking action is probably due to one step.

The difference between right and left segments is mostly omitted in our work. For instance, left and right arm forward extensions are regrouped into one body feature referred to as forward arms extension. Different approaches can be applied for this purpose: 1) summing the measurement obtained when using the right body segments and the measurement obtained when using the left body segments, or 2) taking the largest/ lowest value of such measurements, 3) getting the average of such measurement. The first approach (summing the measurement) is often applied to combine left and right limbs extension along the lateral direction. The second approach (taking the largest/ lowest value) is often applied to combine left and right limbs extension along the sagittal direction. For instance, “taking the largest extension” approach has been applied to measure the forward arms extension. These approaches have been applied to each frame of the whole sequence. The second approach (taking the largest value) is illustrated in Figure E.1 (the second step). Such approaches (e.g. taking the largest extension, summing left and right extensions) have been also applied to combine upper and lower body movement

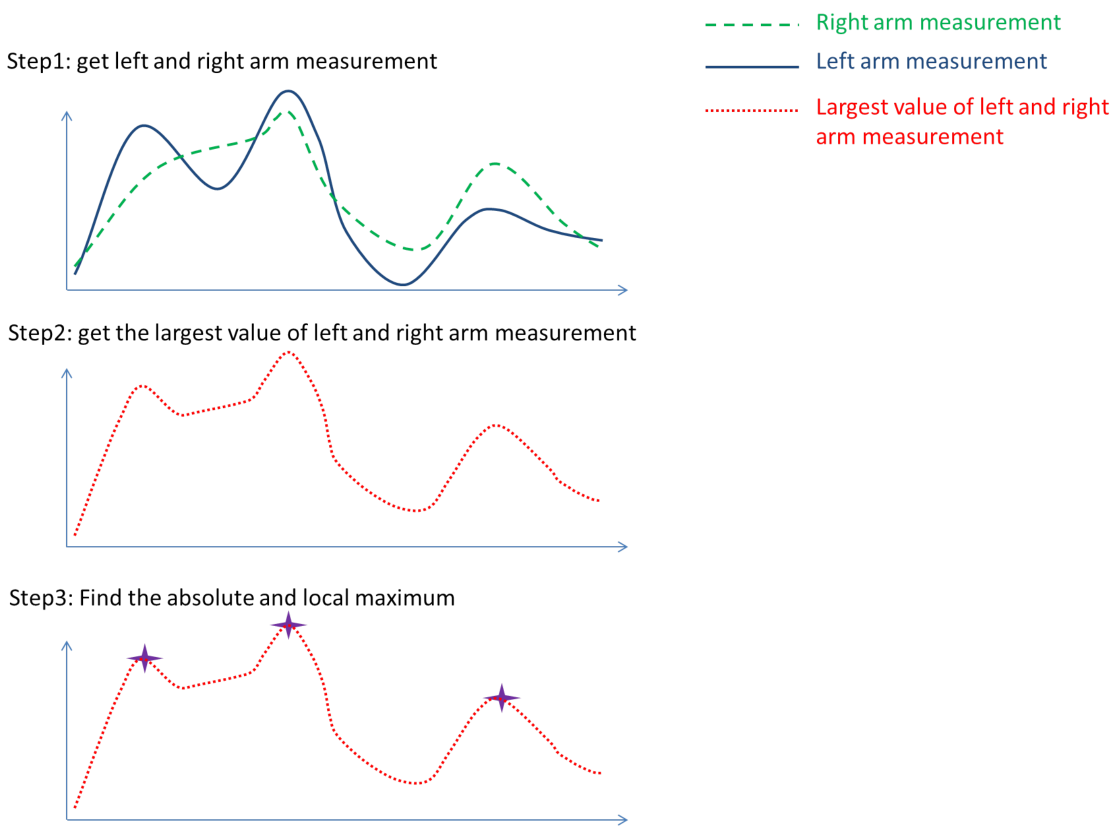


Figure E.1: Maximal arms extension measurement; the largest extension approach is applied to omit the difference between left and right measurement.

features or two levels of the same movement direction (e.g. forward and backward body movement, lateral left and right body movement).

F Scenarios used for Emilya database

F.1 Scenarios used for emotion induction in Emilya databse

Similarly to the framework followed in [Bänziger et al., 2012], the actors that participated in Emilya database recording were provided with the definition of the emotions to be portrayed as well as prototypical scenarios. As mentioned in [Bänziger and Scherer, 2010], the goal from providing the actors with that information is *to help the use of acting techniques as Stanislawsky or Acting Method*.

For each emotion, three or four scenarios were provided, but each actor had to choose only three scenarios for each emotion. The scenarios are reported in the following table (in French). The General context was "imagine that you are a student".

Table F.1: Scenarios - Emilya database (EMotional body expression In daiLY Actions)

Emotions	Short definition and scenarios
Neutral	**Definition:** Neutral emotional state, neither positive nor negative **Scenarios:** 1) "Le soir en rentrant du travail, je vais acheter mon pain à la boulangerie" 2) "Tous les matins, je croise mon voisin de palier avec qui j'échange verbalement un bonjour de politesse"
Joy	**Definition:** Feeling transported by a fabulous thing that occurred unexpectedly **Scenarios:**

	1) “Mon oncle arrive pour mon anniversaire. Il m’a demandé de regarder par la fenêtre. Son cadeau est devant la maison : la voiture que j’ai toujours rêvée d’avoir” [Scherer et al., 1991] 2) “Je reçois un coup de fil de ma/mon bien aimé(e) pour me dire qu’elle/il va enfin pouvoir venir passer quelques jours à mes côtés” 3) “J’ai gagné une somme fabuleuse au loto. Je ne m’y attendais pas du tout, j’avais acheté un billet par hasard en allant boire un café avec un ami. Je suis fou de joie lorsque je le découvre et je cours l’annoncer à mon/ma partenaire (mes parents, mes enfants)” 4) “Pour mon anniversaire, tous mes amis se sont cotisés et m’ont offert un très beau cadeau (quelque chose que je souhaitais avoir depuis longtemps mais que je ne pensais pas recevoir)” [Bänziger and Scherer, 2010]
Sadness	**Definition:** Feeling discouraged by the irrevocable loss of a person, place, or thing **Scenarios:** 1) “1) Après plusieurs années de vie en couple, mon mari/ma femme a décidé de me quitter. Les choses n’allaient plus très bien entre nous depuis déjà assez longtemps. Mais j’espérais encore que ça pourrait peut-être s’arranger. J’ai reçu les papiers du divorce ce matin. Cette fois c’est vraiment la fin de notre mariage. Je sais qu’il n’y a plus rien à faire à présent” [Bänziger and Scherer, 2010] 2) “Mon chien (ou chat) est très gravement malade. Il a déjà été opéré deux fois mais malgré cela il va de plus en plus mal. Le vétérinaire m’a expliqué ce matin qu’il n’y avait plus aucun espoir et qu’il valait mieux abréger les souffrances du pauvre animal en l’endormant. Je dois aller au cabinet du vétérinaire cet après-midi avec mon chien (chat) pour l’euthanasier” [Bänziger and Scherer, 2010]

F

	3) “Je viens de recevoir un appel me disant que ma tante préférée vient de mourir suite à l’échec de son opération médicale” (adapted from [Scherer et al., 1991]) 4) “Mon chien adoré a été opéré pour une blessure très grave. A la fin de l’opération, le vétérinaire m’annonce que l’opération a échoué, il est mort”
Anger	**Definition:** Extreme displeasure caused by someone’s stupid or hostile action **Scenarios:** 1) “Un ami qui est aussi un collègue de mon/ma partenaire (conjoint, mari, femme) m’apprend qu’une troisième personne au bureau essaie de séduire mon/ma partenaire. Cette personne est très connue pour avoir eu de nombreuses aventures avec d’autres collègues. J’ai toujours eu une assez forte antipathie pour lui/elle, mais à présent je suis carrément furieux (se) contre lui/elle. Je partage ce que je ressens avec mon ami” [Bänziger and Scherer, 2010] 2) “Durant un séjour à l’étranger j’ai sous-loué mon appartement. Au retour, je découvre que mon appartement a été abandonné dans un état lamentable par ses occupants qui n’ont, de plus, respecté aucun des engagements qu’ils avaient pris au moment où le contrat de sous-location a été signé. Une partie de mes affaires ont disparu et le loyer n’a pas été versé. Je suis absolument furieux contre ces individus irresponsables, j’exprime ce que ressens à un ami qui m’accompagne et qui a constaté l’étendue des dégâts avec moi” [Bänziger and Scherer, 2010]

	3) “Je viens de surprendre deux adolescents en train de vandaliser ma voiture. Ils ont, non seulement, forcé la portière pour voler mon auto-radio, ils ont également rayé la carrosserie, arraché l’antenne et les rétroviseurs et crevé les pneus. J’arrive à rattraper l’un des deux qui se trouve être le fils de mes voisins de palier. Je n’ai pas beaucoup de sympathie pour ces voisins qui ne cessent de se disputer avec tous les habitants de l’immeuble et passent leur temps à créer des problèmes. Je ramène l’adolescent chez ses parents et j’exprime mes sentiments sur son comportement inqualifiable sans prendre de gants” [Bänziger and Scherer, 2010] 4) “Je pensais faire une grasse matinée dimanche matin après une longue soirée, mais mon voisin commence à faire des travaux très bruyants dans sa maison à 7h du matin. Je me sens tellement énervé que je décide d’aller le gronder”
Pride	**Definition:** Feeling of triumph following a success or a personal achievement (one’s own or that of someone close) **Scenarios:** 1) “Je viens de recevoir un coup de téléphone qui m’a appris que j’ai obtenu le rôle principal dans un film (ou une pièce de théâtre). C’est mon premier grand rôle et le film (la pièce) aura certainement un très grand succès. Les candidats pour ce rôle étaient très nombreux et le réalisateur ainsi que les autres comédiens sont tous des professionnels très admirés. Je vais directement dans le salon annoncer la grande nouvelle à mon/ma partenaire (conjoint)” [Bänziger and Scherer, 2010] 2) “Un critique très sévère, mais dont l’avis est habituellement suivi, a fait un compte rendu très élogieux de ma performance d’acteur dans un rôle particulièrement difficile. Cette critique élogieuse est parue dans un grand quotidien ce matin. J’ai déjà reçu de nombreux appels de mes amis et de ma famille pour me féliciter. Mon agent a également appelé pour me dire qu’il a reçu plusieurs propositions intéressantes pour d’autres rôles” [Bänziger and Scherer, 2010]

F

	3) “Mon professeur principal me croise dans le couloir et me félicite pour le résultat de mon examen final dont la copie m’a été rendue la semaine dernière. Il m’exprime sa joie de m’avoir parmi les éléments de sa classe” 4) “Dans les deux derniers mois de mes travaux du projet de fin d’étude, mon encadrant me félicite pour le travail que j’ai fait et pour les résultats prometteurs que j’ai trouvés”
Anxiety	**Definition:** Fear of the consequences of a situation that could be unfavorable for oneself or someone close **Scenarios:** 1) “Après plusieurs années d’études, je viens de passer mes examens finaux. Ces examens m’ont semblé très difficiles. J’ai du mal à évaluer si j’ai bien réussi ou non. En attendant les résultats qui seront affichés dans quelques minutes, je commence à penser à ce qui arriverait si jamais j’échouais. Je sais que je n’aurai pas de deuxième chance” [Bänziger and Scherer, 2010] 2) “J’ai un RDV tard le soir dans un endroit isolé avec mon partenaire. Progressivement, je commence à m’imaginer que quelqu’un pourrait surgir d’un buisson. Bien que je n’entende rien bouger, je deviens de plus en plus nerveux” [Bänziger and Scherer, 2010] 3) “Je dois terminer un travail important pour la fin de la semaine et je progresse difficilement. Je crains de ne pas avoir fini à temps et je m’imagine les conséquences que cela pourrait avoir pour mon avenir professionnel. Plus je deviens nerveux, moins j’ai l’impression de progresser et plus mon inquiétude augmente. Je finis par ressentir une véritable angoisse” (adapted from [Bänziger and Scherer, 2010])

	4) “Après deux ans de travail de thèse, je n’ai pas encore trouvé des résultats prometteurs et je n’ai pas encore fini tous les objectifs imposés. Il ne me reste qu’une année pour tout finir y compris la rédaction de thèse qui prend beaucoup de temps. Je commence vraiment à me faire des soucis”
Shame	**Definition:** Self-esteem shaken by an error or clumsiness for which one feels responsible **Scenarios:** 1) “Pendant nos vacances en Turquie, dans un petit village reculé en pleine campagne, je remarque un vieil homme au physique très ingrat qui travaille dans un petit magasin. Je fais des commentaires très désobligeants (insultants) à son sujet à l’ami avec qui je suis. Je fais ces commentaires en Français en pensant que cet homme ne parle probablement que le Turc. Au moment de payer nos achats, le propriétaire du magasin remarque que nous parlons Français et appelle son employé (l’homme au physique ingrat) pour qu’il s’occupe de nous. Il s’avère qu’il a vécu en France pendant une grande partie de sa vie et qu’il parle parfaitement le Français. Je suis très gêné et j’ose à peine m’exprimer devant lui à présent” [Bänziger and Scherer, 2010] 2) “Durant une soirée, je fais des remarques très désobligeantes sur une femme qui travaille dans le même laboratoire que moi et que je trouve vraiment bête. Je raconte plusieurs anecdotes à ce sujet et je me moque ouvertement d’elle. Un peu plus tard, l’ami qui m’a invité à cette soirée me présente à l’un des hommes qui étaient présents au moment où j’ai fait ces commentaires. Il s’avère que cet homme est le mari de la femme dont je me suis moqué auparavant. Je suis très embarrassé et je ne sais plus trop quoi dire” [Bänziger and Scherer, 2010]

F

	3) “J’étais invité à une soirée. Je pensais qu’il s’agit d’une soirée de déguisement. Je suis arrivé habillé en Dracula alors que tout le monde est habillé normalement. Ils me regardent bizarrement et ils pointent le doigt vers moi. Je me sens ridicule et mon visage est de plus en plus rouge de honte.” 4) “J’ai copié des paragraphes d’un rapport de thèse d’un autre étudiant pour rédiger mon rapport. Et un jour, le professeur se rend compte et donne mon rapport comme exemple de plagiat devant toute la classe”
Panic Fear	**Definition:** Being faced with an imminent danger that threatens our survival or physical well-being **Scenarios:** 1) “Je suis en train de conduire ma voiture sur une petite route de montagne. Soudain les freins ne répondent plus du tout. Ma voiture prend de la vitesse et je ne peux plus l’arrêter, dans quelques secondes, elle va probablement quitter la route et finir dans le précipice” [Bänziger and Scherer, 2010] 2) “Je suis dans une cave où il fait très sombre en train de chercher un objet égaré par ma grand-mère. Soudain quelque chose tombe du plafond et reste accroché à mon épaule. Je suis pris de panique” [Bänziger and Scherer, 2010] 3) “Il est plus de minuit et je rentre chez moi à pied. Je suis seul pendant un moment, puis soudain je remarque qu’un homme me suit. J’entends ses pas derrière moi. J’accélère et il accélère également. Je me mets à courir et il court après moi. Je sens qu’il m’agrippe par ma veste, je vois à présent qu’il tient de l’autre main un couteau à cran d’arrêt [Bänziger and Scherer, 2010]

F

	4) “Alors que je suis dans ma banque entrain de régler des papiers avec mon conseiller, j’entends les cris de la secrétaire et en sortant dans le couloir je me trouve face à face avec un braqueur de banque ; visage caché, tenu noir et kalachnikov à la main”

F

G

Data selection for feature selection process

Table G.1: Actors (Id) whose expressive behaviors are considered as “relevant”.

	Anxiety	Pride	Joy	Sadness	Panic Fear	Shame	Anger	Neutral
SW	5, 7, 8, 9, 12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	1, 2, 4, 5, 6, 7, 8, 9, 11, 12	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12	2, 7, 8, 9, 10, 11, 12	1, 4, 6, 7, 8, 10, 11, 12	1, 2, 6, 7, 9, 11, 12	2, 3, 4, 5, 6, 7, 8, 9, 10, 11
MB	12	6, 7, 8, 10, 11, 12	1, 2, 4, 6, 8, 9, 10	1, 2, 4, 5, 6, 7, 8, 9, 10, 11	2, 6, 8, 9, 10	1, 8, 9, 11	1, 3, 6, 7, 8, 9, 11	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
WH	4, 7	1, 2, 3, 6, 8, 10, 11	2, 3, 4, 5, 6, 7, 8, 9, 10, 11	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12	1, 7, 8, 10, 12	1, 3, 4, 5, 8, 10	1, 2, 3, 5, 6, 7, 8, 11, 12	1, 2, 3, 4, 5, 7, 8, 9, 10, 11
KD	1, 2, 3, 4, 7, 8, 9, 11, 12	1, 5, 6, 7, 8, 9, 10	2, 7, 8, 9, 10, 11	1, 2, 3, 4, 6, 7, 8, 9, 10, 11	1, 5, 8, 9, 10, 11	1, 2, 4, 7, 8, 9, 11, 12	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12	2, 3, 4, 5, 7, 8, 9, 10, 11, 12
BS	1, 2, 4, 5, 6, 8	1, 2, 6, 7, 10, 11, 12	2, 10, 11	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	1, 2, 8	2, 8, 11	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12	1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12
SD	1, 2, 4, 5, 6, 8	4, 7, 10, 11, 12	2, 11	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	1, 2, 8, 9	1, 2, 5, 7, 8, 11, 12	1, 3, 5, 6, 7, 8, 9, 10, 12	1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12
Lf	2, 3, 4, 6, 8	1, 5, 6, 8, 9, 11	1, 2, 4, 6, 7, 8, 10, 11, 12	1, 2, 4, 5, 6, 7, 8, 9, 10, 11	2, 3, 8, 11	1, 8, 11	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12	2, 3, 4, 5, 7, 8, 9, 10, 11, 12
Th	7, 8, 9	1, 6, 9, 10, 11, 12	1, 5, 7, 8, 12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	2, 4, 7, 8	2, 6, 7, 8, 10, 11	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	3, 4, 5, 6, 8, 10

Table G.2: Actors (Id) whose expressive behaviors are not correctly classified by an automatic classifier (error rate of RF>0.5), but correctly recognized by human perception.

	Anxiety	Pride	Joy	Sadness	Panic Fear	Shame	Anger	Neutral
SW	7, 9, 12	6, 8, 12	1, 4, 11, 12	5, 6, 8, 12	11, 12	6	11, 12	6
MB	12	12	2, 8, 9	1, 2, 6, 10			3	1, 2, 6, 7, 10, 12
WH	4, 7	6, 8	4, 9, 11	5, 6, 8, 10	1		3, 11	5, 1
KD	1, 2, 3, 4, 7, 8, 9, 11, 12	1, 7	11	2, 3, 4, 6, 8, 10, 11			7	2, 5, 7, 9, 10, 11, 12
BS	1, 2, 4, 5, 6, 8	12	2, 11	1, 6, 7, 8, 10	1, 2	2	8, 12	2, 3, 5, 7, 9, 10, 11, 12
SD	1, 2, 4, 5, 6, 8	12	2, 11	1, 2, 3, 7, 8, 11	1	2	1, 6	1, 2, 3, 4, 7, 9, 10, 11, 12
Lf	2, 4, 6, 8		6	2, 4, 5, 6, 8, 9, 10, 11			10, 12	2, 3, 4, 7, 8, 9, 10, 11, 12
Th	7, 8, 9	12	1, 7, 12	2, 4, 6, 7, 10, 12			6	4, 5, 6, 8, 10

H Detailed characterization of emotional body expression in Emilya database

H.1 Motion capture characterization of emotions for each group of similar actions

See Figure H.1, H.2, H.3, H.4.

H.2 Perceptual characterization of expressed emotions across "similar" actions

See Figure H.5a, H.5b, H.5c, H.5d.

H.3 Motion capture characterization of emotions for each action

See Figure H.6, H.7, H.8, H.9, H.10, H.11, H.12, H.13.

H.4 Perceptual characterization of emotions for each action

See Figure H.14, H.15, H.16, H.17, H.20, H.18, H.19.

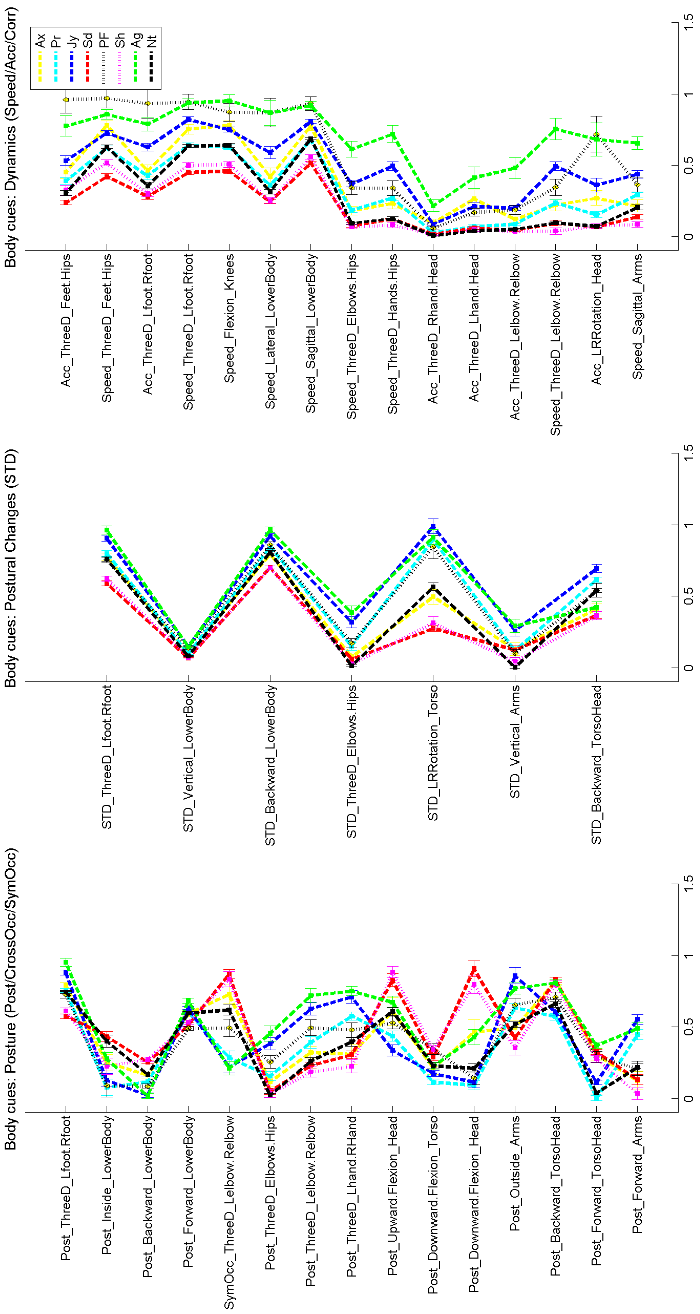


Figure H.1: Motion capture characterization of expressed emotions in walking actions (SW + WH)

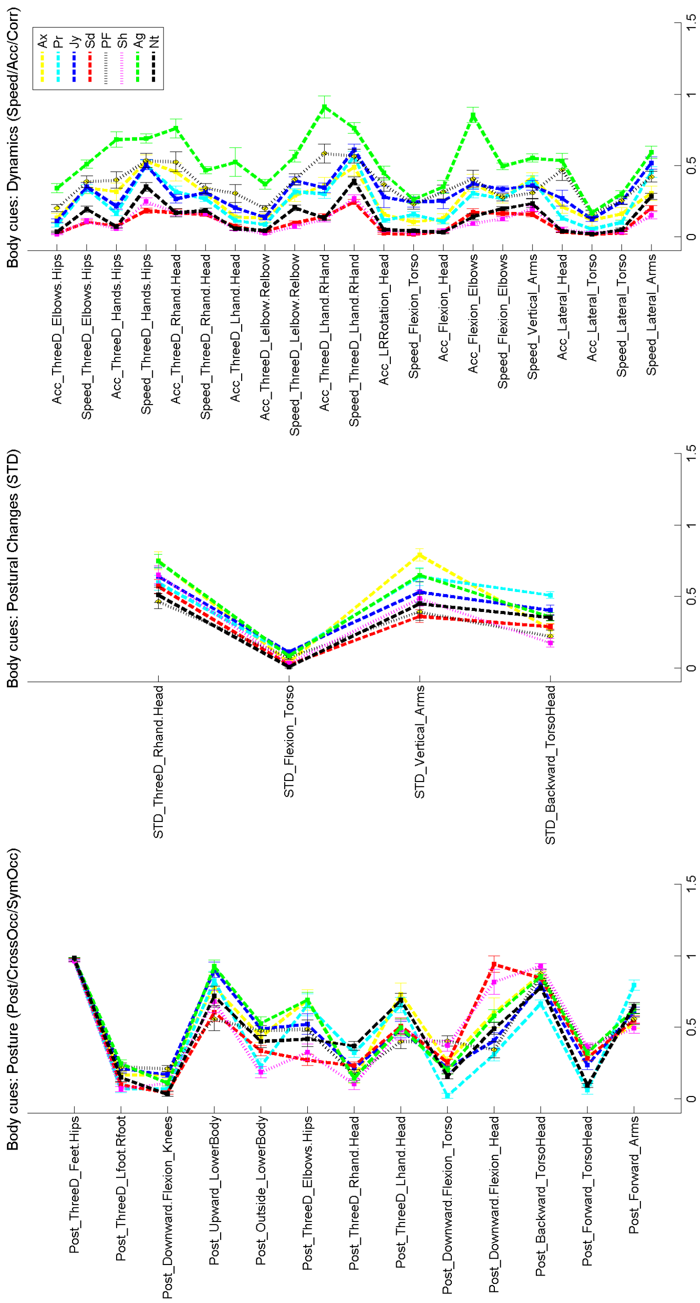


Figure H.2: Motion capture characterization of expressed emotions in Knocking and Moving books actions (KD+MB)

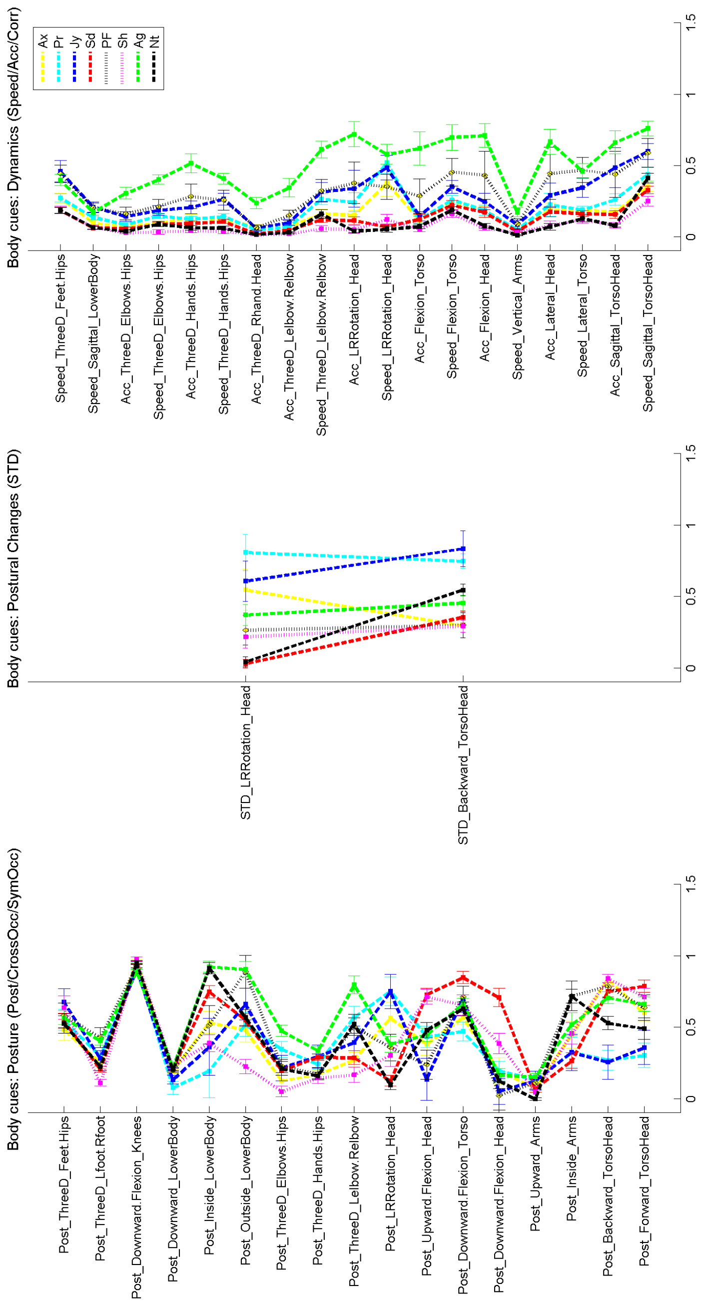


Figure H.3: Motion capture characterization of expressed emotions in Sitting Down and Being Seated actions (SD+BS)

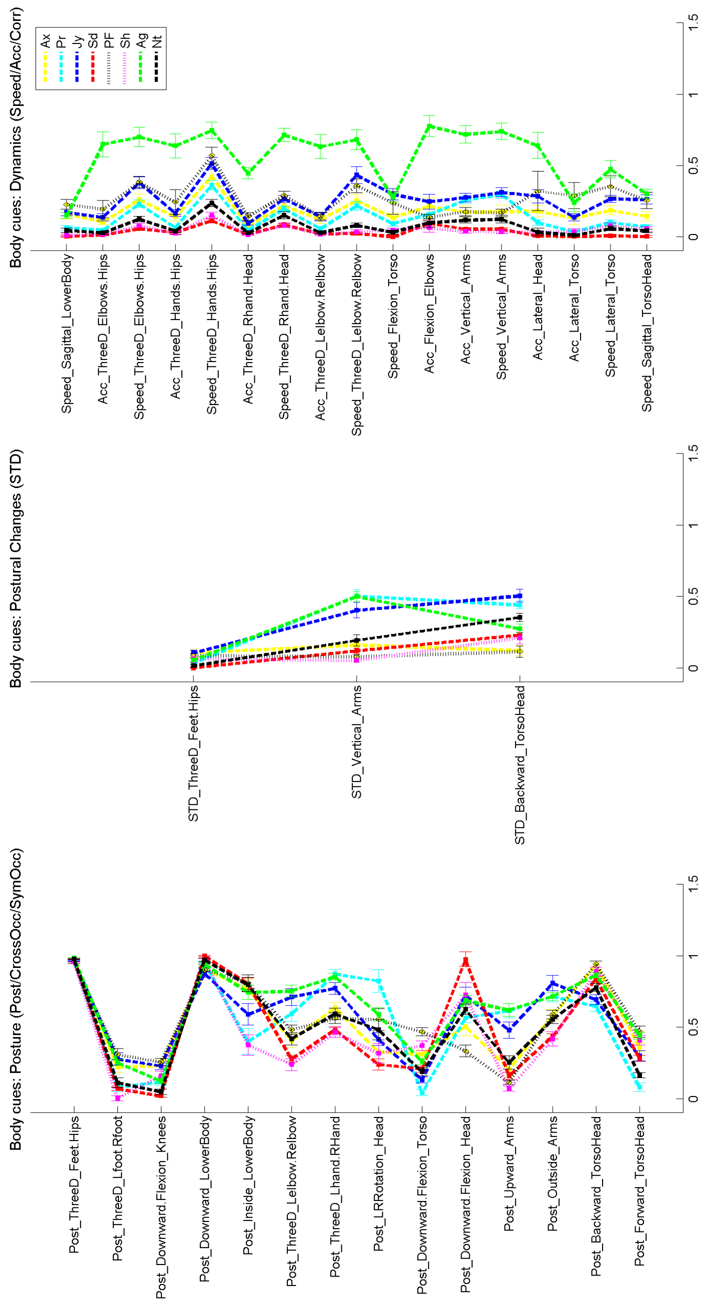
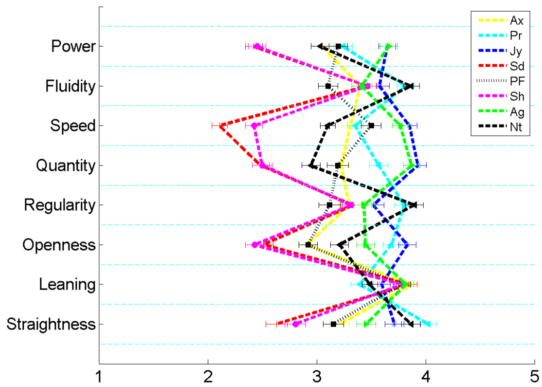


Figure H.4: Motion capture characterization of expressed emotions in Lifting and Throwing actions (Lf+Th)

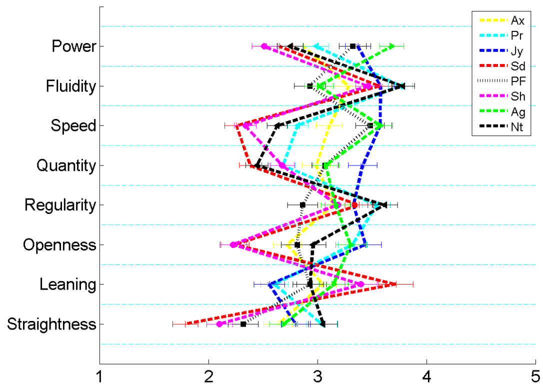
H

H.5 Random Forest Ranking of selected features

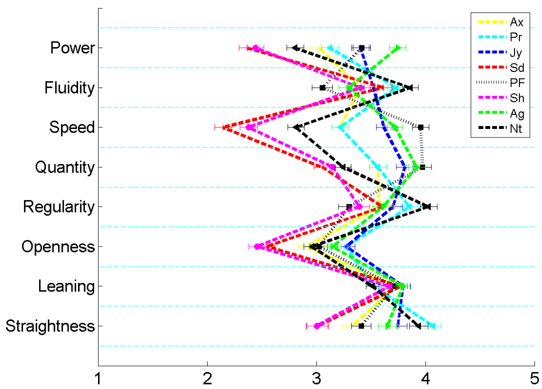
See Figures H.21, H.24, H.22, H.23, H.25.



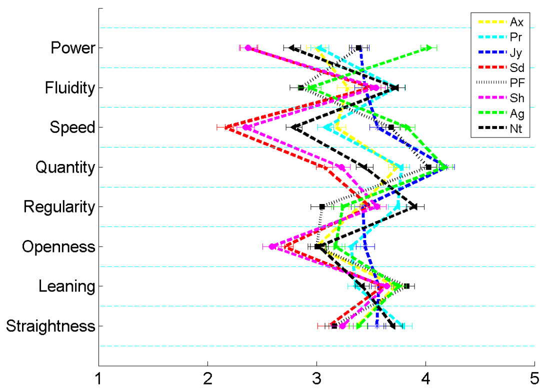
(a) Walking actions (SW and WH)



(b) Sit Down actions (SD and BS)



(c) Repetitive arms movement (MB and KD)



(d) Non-repetitive arm movement (Lf and Th)

Figure H.5: Perceptual characterization of expressed emotions across “similar” actions.

H

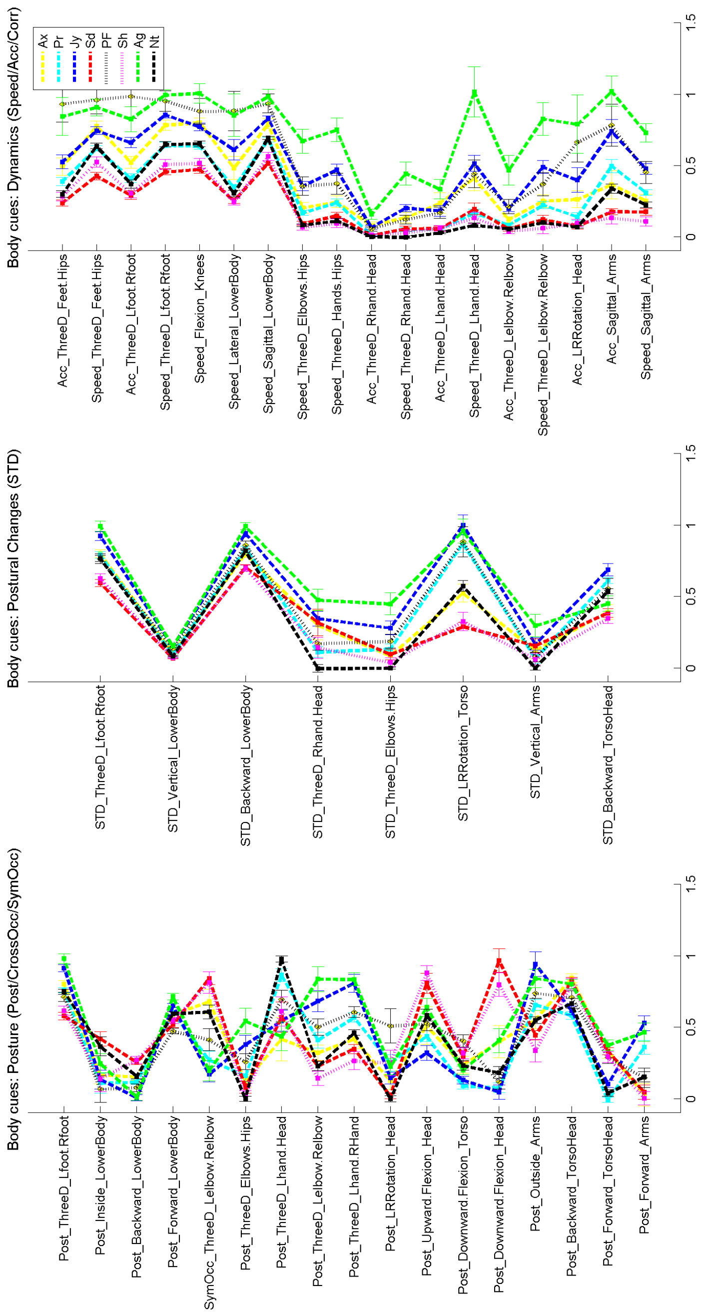


Figure H.6: Motion capture characterization of expressed emotions in simple walking actions (SW)

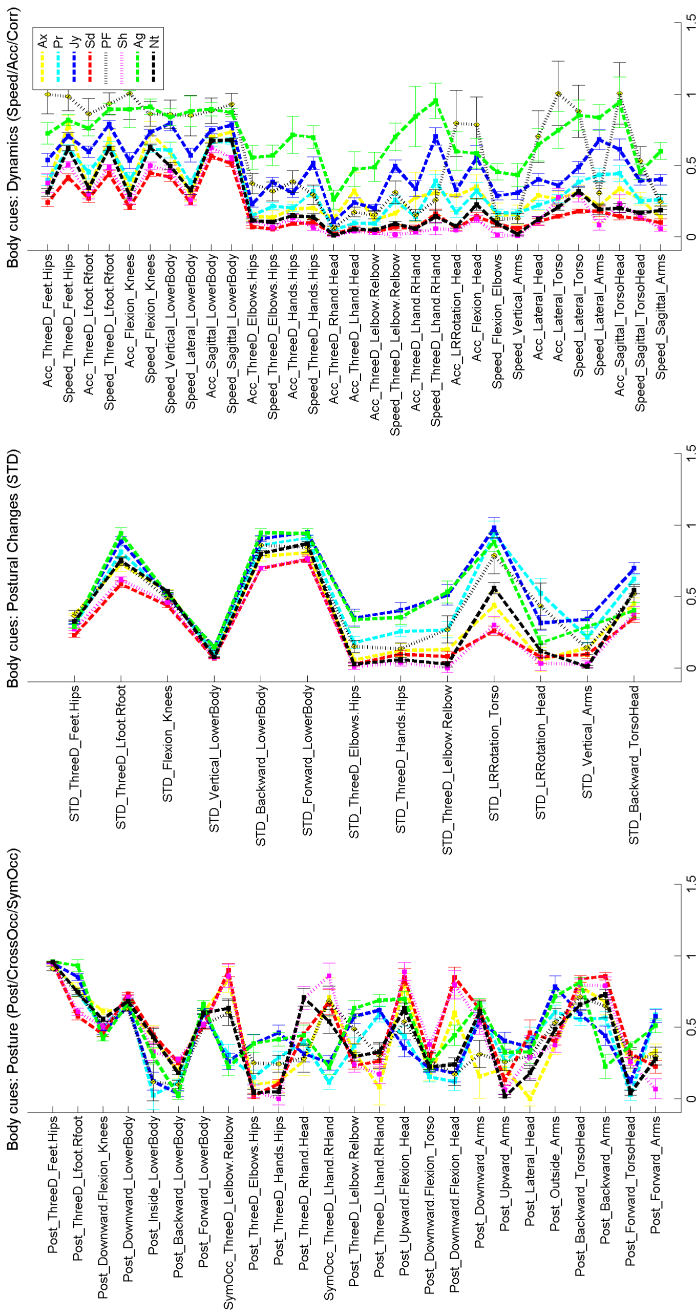


Figure H.7: Motion capture characterization of expressed emotions in walking with an object in hand actions (WH)

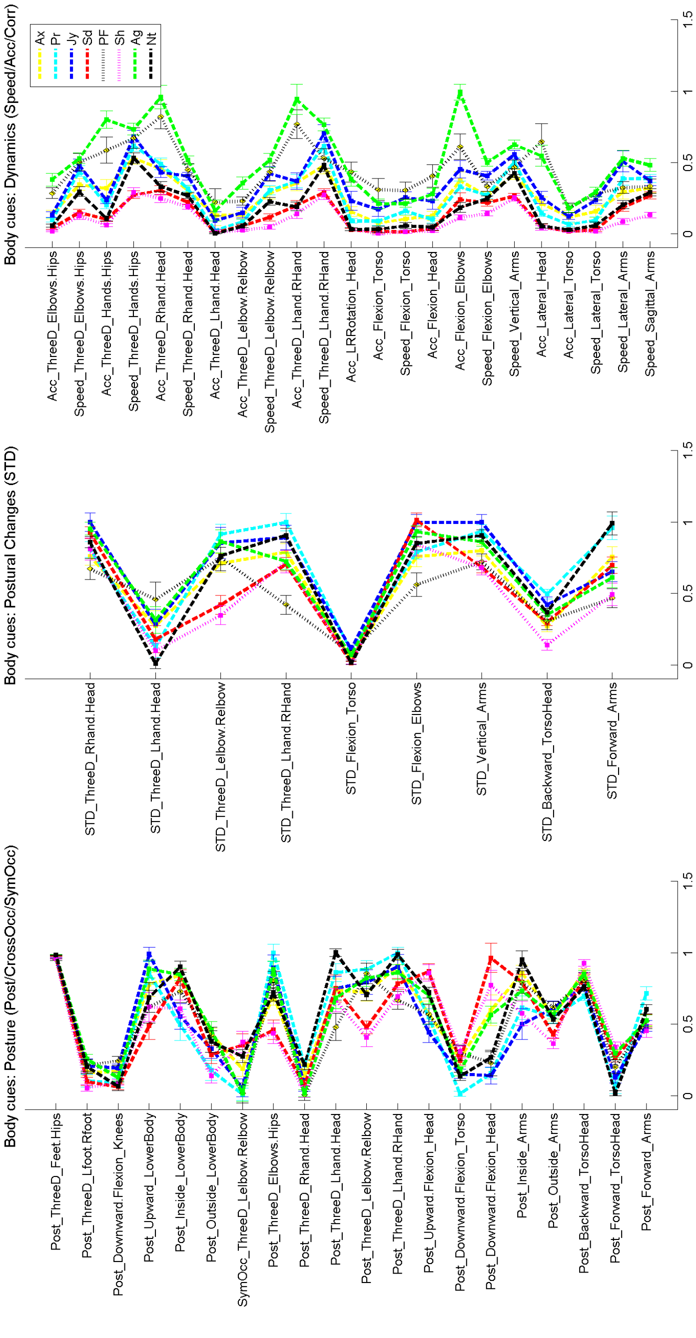


Figure H.8: Motion capture characterization of expressed emotions in Knocking action (KD)

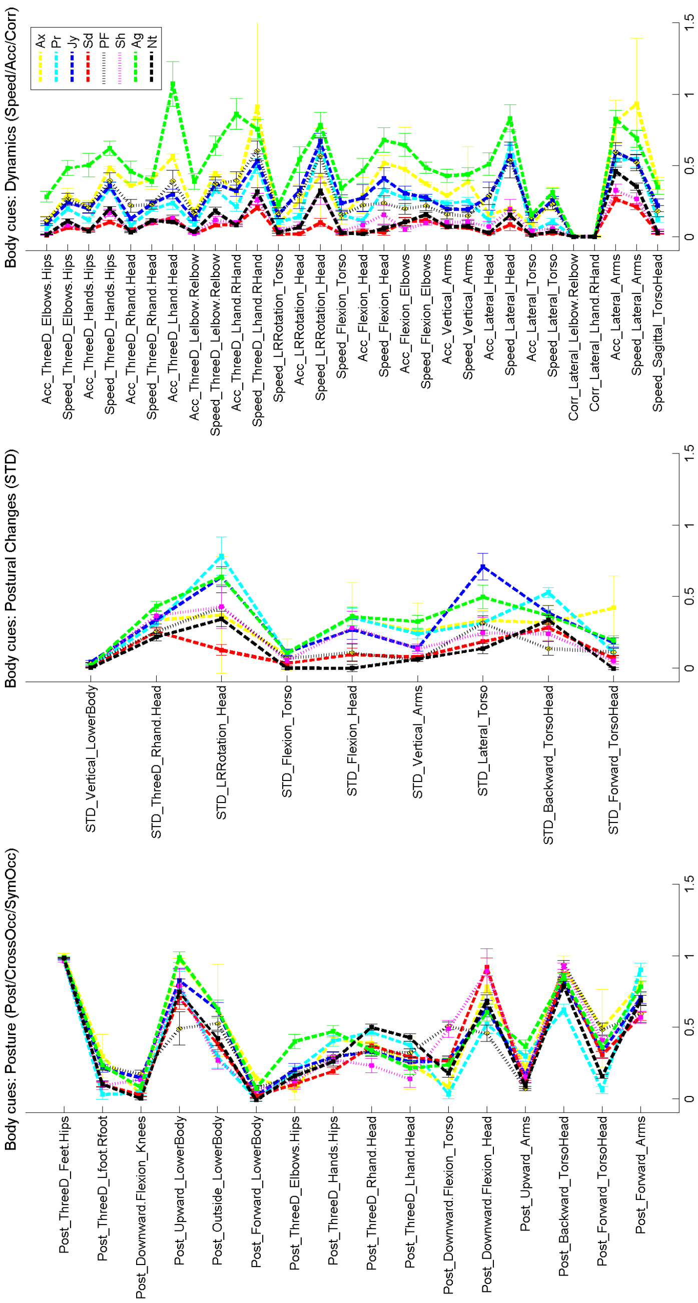


Figure H.9: Motion capture characterization of expressed emotions in Moving Books action (MB)

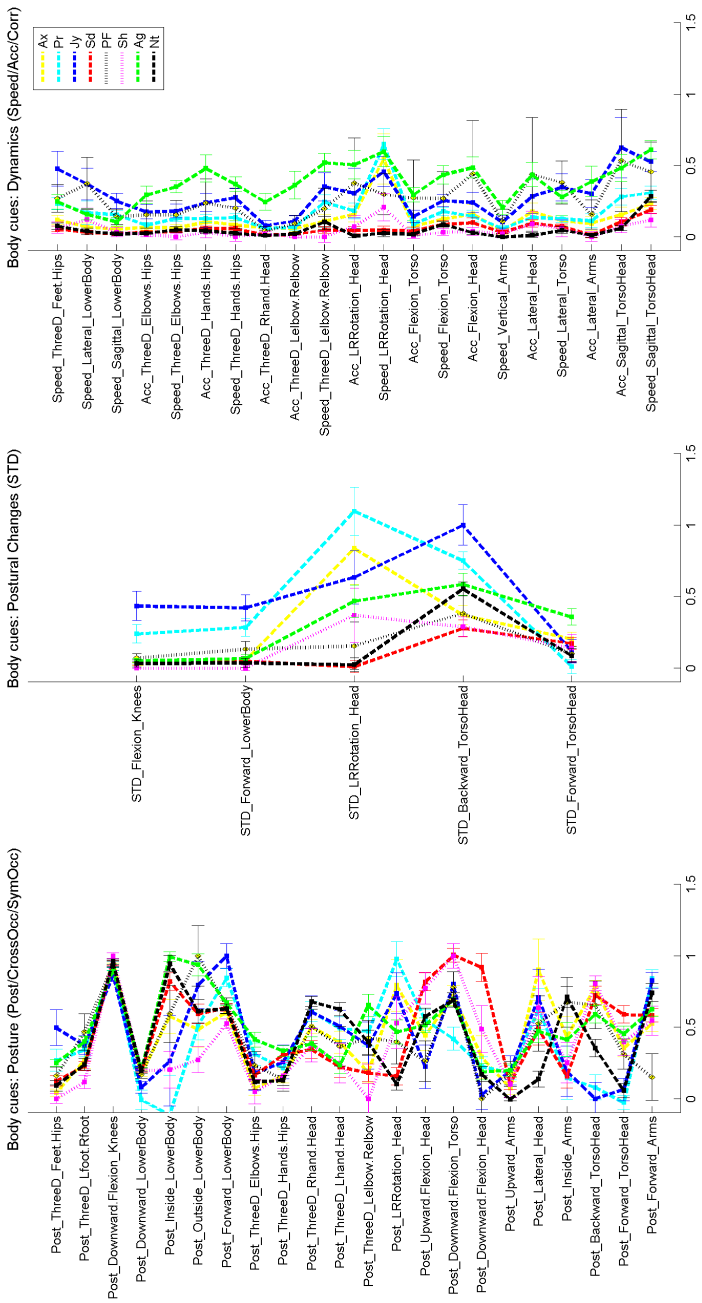


Figure H.10: Motion capture characterization of expressed emotions in Being Seated action (BS)

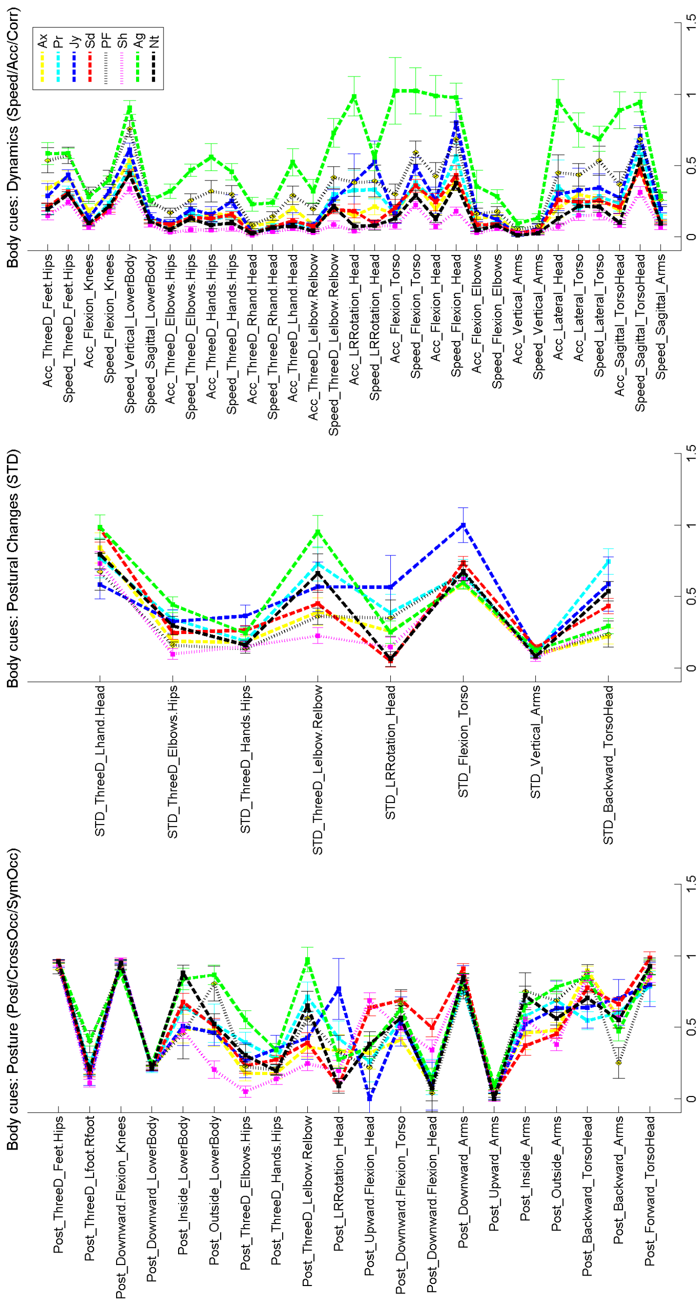


Figure H.11: Motion capture characterization of expressed emotions in Sitting Down action (SD)

H

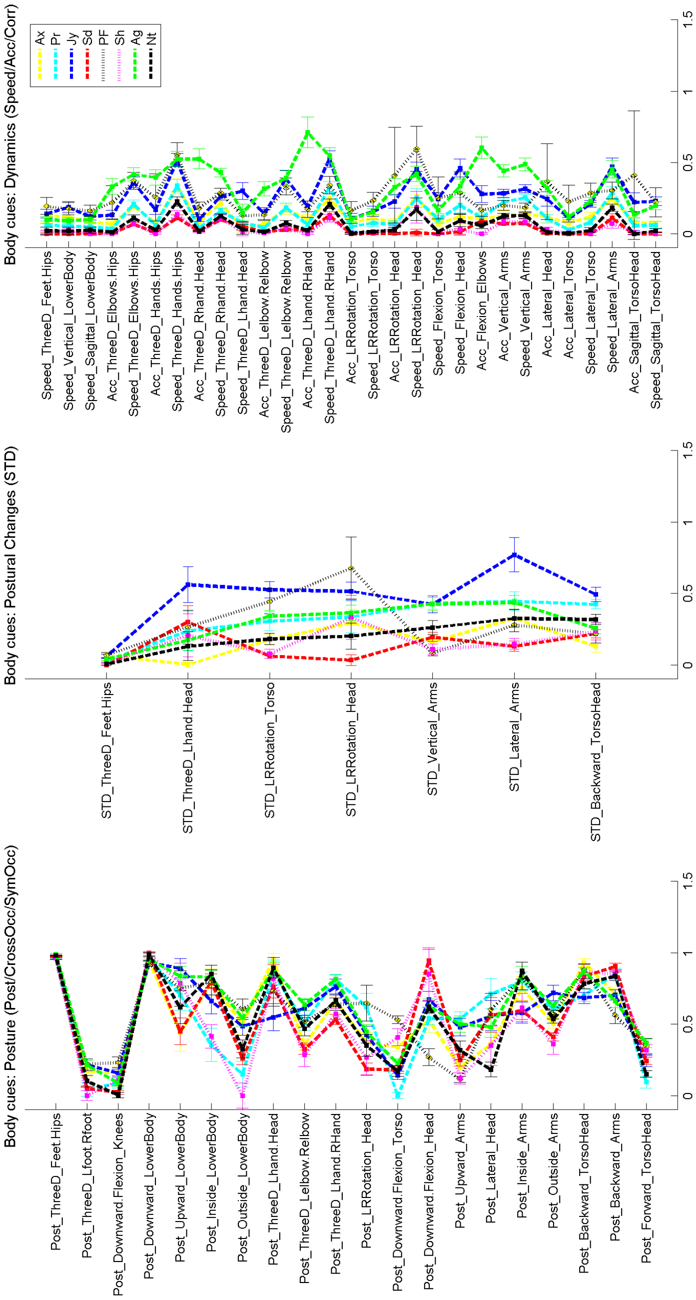


Figure H.12: Motion capture characterization of expressed emotions in Lifting action (Lf)

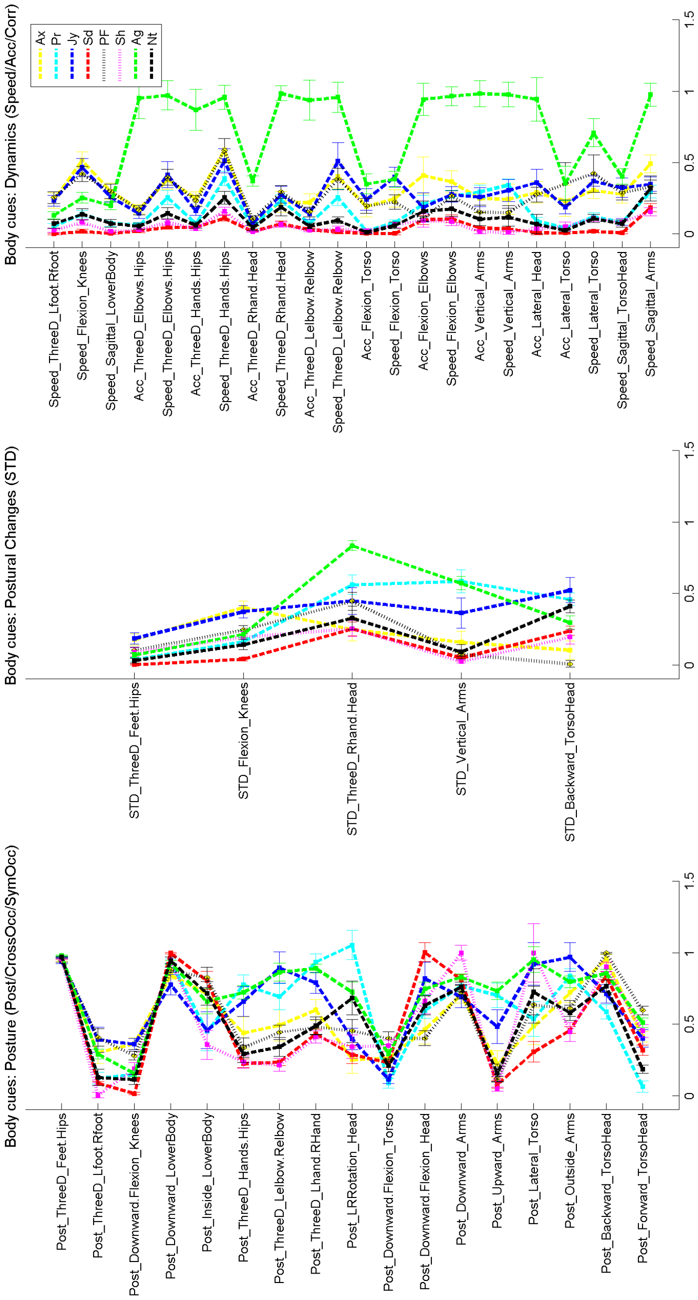


Figure H.13: Motion capture characterization of expressed emotions in Throwing action (Th)

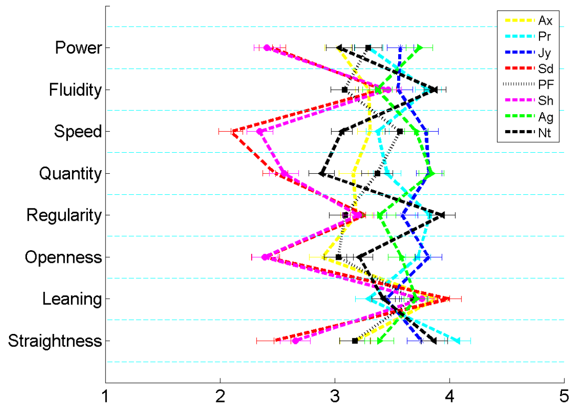


Figure H.14: Perceptual characterization of expressed emotions in simple walking actions (SW)

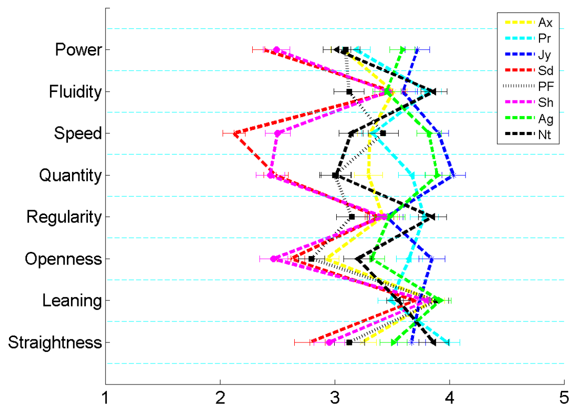


Figure H.15: Perceptual characterization of expressed emotions in walking with an object in hand actions (WH)

H

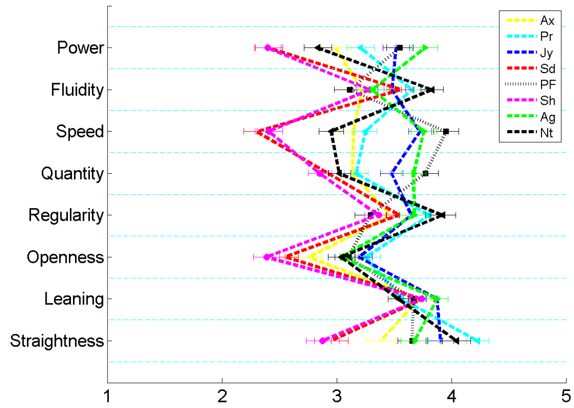


Figure H.16: Perceptual characterization of expressed emotions in Knocking action (KD)

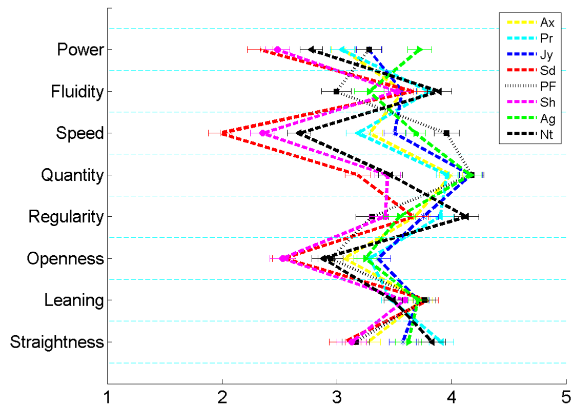


Figure H.17: Perceptual characterization of expressed emotions in Moving Books action (MB)

H

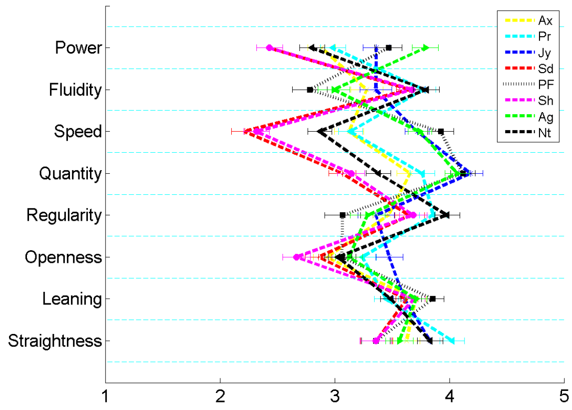


Figure H.18: Perceptual characterization of expressed emotions in Lifting action (Lf)

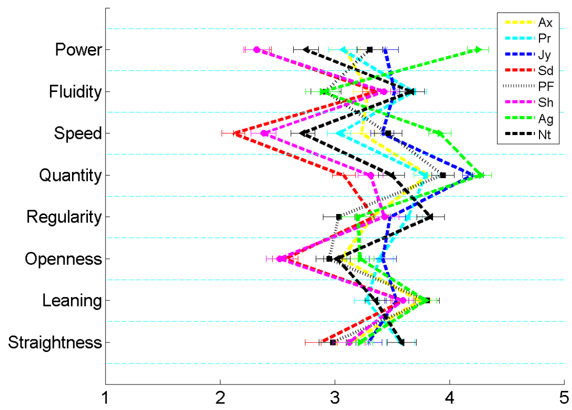


Figure H.19: Perceptual characterization of expressed emotions in Throwing action (Th)

H

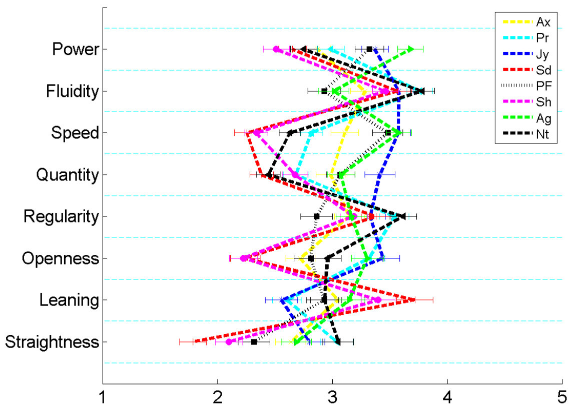


Figure H.20: Perceptual characterization of expressed emotions in Sitting Down action (SD)

H

Posture/ Movement:	Post											
Direction:	Lateral						Sagittal					
Subdirection:	Inside		Lateral		Outside		Backward			Forward		
Modality:	Arms	Lower Body	Head	Torso	Arms	Lower Body	Arms	Lower Body	Torso Head	Arms	Lower Body	Torso Head
SubModality:												
SW		25			40			17	26	30	6	1
MB						30			16	52	43	2
WH		63	52		68		56	34	14	62	21	1
KD	30	25			52	28			22	48		2
BS	10	7	47			35			4	48	21	1
SD	19	6			31	12	50		25			56
Lf	33	29	43		22	18	48		12			21
Th		34		42	23				17			6

Posture/ Movement:	Post							
Direction:	ThreeD							
Subdirection:								
Modality:	Arms						Lower Body	Lower Body
SubModality:	Elbows. Hips	Hands. Hips	Lelbow. Relbow	Lhand. Head	Lhand. RHand	Rhand. Head	Feet. Hips	Lfoot. Rfoot
SW	43		27	35	39			12
MB	55	53		40		14	42	34
WH	36	26	60		59	67	39	6
KD	39		31	14	35	46	6	40
BS	34	23	44	30		38	19	36
SD	23	36	8				52	38
Lf			46	40	24		23	8
Th		40	19		11		31	15

Posture/ Movement:	Post										SymOcc	
Direction:	Vertical Length									Vertical	ThreeD	
Subdirection:	Downward		Downward. Flexion			Upward	Upward	Upward. Flexion	Upward. Flexion		ThreeD	
Modality:	Arms	Lower Body	Head	Lower Body	Torso	Arms	Lower Body	Arms	Head	Head	Arms	
SubModality:				Knees			Lower Body	Elbows			Lelbow. Relbow	Lhand. RHand
SW			3		18				10	28	21	
MB			19	15	3	27	50					
WH	69	42	13	27	48	28			15		16	50
KD			3	42	8		44		18		41	
BS		24	2	41	6	22			18	9		
SD	39	43	11	59	40	45			42	49		
Lf		44	16	34	2	17	52			41		
Th	41	44	43	5	27	8				37		

Figure H.21: Ranking of Posture selected features for each action

Posture/ Movement:	Speed																					
Direction:	Lateral				Sagittal			ThreeD								Vertical Length						Vertical Rotation	
Subdirection:			Lateral	Lateral				ThreeD								Flexion				Vertical			
Modality:	Arms	Head	Lower Body	Torso	Arms	Lower Body	Torso Head	Arms						Lower Body		Arms	Head	Lower Body	Torso	Arms	Lower Body	Head	Torso
SubModality:								Elbows. Hips	Hands. Hips	Lelbow. Relbow	Lhand. Head	Lhand. RHand	Rhand. Head	Feet. Hips	Lfoot. Rfoot	Elbows		Knees					
SW			34		22	4		9	23	16	38		20	8	2			37					
MB	23	32		8			39	25	7	6		21	38			26	41		31	35		11	9
WH	45		33	25	38	3	23	4	20	8		37		5	2	66		19		40	58		
KD	37			13	19			23	10	17		16	49			20			15	7			
BS			42	40		37	32	5	26	17				43					28	46		8	
SD				48	22	37	4	15	28	35			29	3		46	30	5	57	41	1	14	
Lf	54			28		53	9	32	15	39	51	55	31	38			36		45	3	49	30	11
Th				4	24	12	29	18	28	14			1		36	2		9	21	7			

Figure H.22: Ranking of Speed selected features for each action

Posture/ Movement:	Acceleration																				
Direction:	Lateral			Sagittal			ThreeD								Vertical Length					Vertical Rotation	
Subdirection:															Flexion				Vertical		
Modality:	Arms	Head	Torso	Arms	Lower Body	Torso Head	Arms						Lower Body		Arms	Head	Lower Body	Torso	Arms	Head	Torso
SubModality:							Elbows. Hips	Hands. Hips	Lelbow. Relbow	Lhand. Head	Lhand. RHand	Rhand. Head	Feet. Hips	Lfoot. Rfoot	Elbows		Knees				
SW				29					7	41		24	36	11						32	
MB	22	12	37				4	13	1	17	5	18			28	47			33	10	
WH		41	24		65	54	17	53	7	57	61	46	31	30		64	44			35	
KD		11	33				5	1	9	32	24	26			4	36		45		12	
BS	31	20				12	15	39	3			45				16		14		11	
SD		20	47			16	10	13	9	17		21	44		34	18	7	32	27	2	
Lf		26	14			56	5	7	6		4	1			27				10	20	42
Th		20	30				16	22	25			3			38			39	13		

Figure H.23: Ranking of Acceleration selected features for each action

Posture/ Movement:	Postural Changes (standard deviation)																						
Direction:	Lateral		Sagittal					ThreeD								Vertical Length						Vertical Rotation	
Subdirection:			Backward		Forward											Flexion				Vertical		LRRotation	
Modality:	Arms	Torso	Lower Body	Torso Head	Arms	Lower Body	Torso Head	Arms						Lower Body	Lower Body	Arms	Head	Lower Body	Torso	Arms	Lower Body	Head	Torso
SubModality:								Elbows. Hips	Hands. Hips	Lelbow. Relbow	Lhand. Head	Lhand. RHand	Rhand. Head	Feet. Hips	Lfoot. Rfoot	Elbows		Knees					
SW			15	19				31					33		14					42	13		5
MB		48		46			44						49				45		51	20	29	24	
WH			12	18		51		11	43	49				29	22			47		32	9	55	10
KD				27	43					50	51	29	34			21			47	38			
BS				25		29	33											27				13	
SD				26				24	54	33	55								58	53		51	
Lf	25			13							47			37						19		50	35
Th				26									33	32				35		10			

Figure H.24: Ranking of Postural changes selected features for each action

Posture/ Movement:	Correlation	
Direction:	Lateral	
Subdirection:		
Modality:	Arms	
SubModality:	Lelbow. Relbow	Lhand. RHand
SW		
MB	36	54
WH		
KD		
BS		
SD		
Lf		
Th		

Figure H.25: Ranking of Correlation selected features for each action

I

Acronym

Table I.1: Acronym table

Acronym	Definition
Mocap	Motion Capture
TC	Time Code
OOB	Out Of Bag error (See section D.4, Chapter D)
CCR	Correct Classification Rate
CCR_{OOB}	Correct Classification Rate based on the OOB error returned by RF
CCR_{CV}	Correct Classification Rate based on cross-validation using a training and a test dataset
CCR_{SO}	Correct Classification Rate measured while considering the "Selected" data as the training dataset and the 'Outliers" data as the testing dataset
RC	Recall measure
RF	Random Forest approach
MLR	Multinomial Logistic Regresssion approach
TP	Temporal Profile of a motion cue
TP.Trajectory	features that describe the Temporal Profile of the 3D trajectory of a given motion cue
TP.Energy	features that related to the Temporal Profile of the Energy measure of a given motion cue
ML	Multi-Level features
FS	Feature Selection
FS-SSF	Subset of Selected Features obtained from Feature Selection approach
Int-SSF	Subset of Selected Features obtained from the Intersection of two (or more) Subsets of Selected Features
CV	Cross-Validation
VI	Variable Importance
SSF	Subset of Selected Features

Emotions:	
Ax	Anxiety
Pr	Pride
Jy	Joy
Sd	Sadness
PF	Panic Fear
Sh	Shame
Ag	Anger
Nt	Neutral
Actions:	
SW	Simple Walking
WH	Walking with an object in hands
MB	Moving Books
KD	Knocking at the door
SD	Sitting Down
BS	Being Seated
Lf	Lifting
Th	Throwing
SW+WH	"Similar" Walking actions
KD + MB	"Similar" actions involving Repetitive Arms Movement: Knocking and Moving Books actions
Lf + Th	"Similar" actions involving Non-repetitive Arm Movement: Lifting and Throwing
SD + BS	"Similar" actions involving sitting down action: Sitting Down and Being Seated
Body features acronym:	
STD	Standard deviation of body movement
Post	Postural feature (e.g. body straightness)
Occ	Occurrence of event (such as crossing arms)
Acc	Acceleration of body movement

J Publications

- Fourati, N., Pelachaud, C. (2012). Toward new expressive movement characterizations. In Motion In Games, (MIG 2012). Rennes, France.
- Fourati, N., & Pelachaud, C. (2013b). Head, shoulders and hips behaviors during turning. In 4th international workshop on Human Behavior Understanding, In conjunction with ACM Multimedia 2013 (Vol. 8212, pp. 223–234). Barcelone, Spain.
- Fourati, N., & Pelachaud, C. (2013a). A new acted emotional body behavior database. In 9th Workshop on Multimodal Corpora: Beyond Audio and Video, IVA 2013 Workshop (MMC 2013). Edinburgh, UK.
- Ochs, M., Ding, Y., Fourati, N., Chollet, M., Ravenet, B., Pecune, F., ... Pelachaud, C. (2013). Vers Des Agents Conversationnels AnimÉs Socio-Affectifs. In 25ième conférence Francophone sur L'Interaction Homme-Machine (IHM 2013) (pp. 69:69–69:78). Bordeaux, France.
- DING, Y., Huang, J., Fourati, N., Artières, T., & Pelachaud, C. (2014). Upper Body Animation Synthesis for a Laughing Character. In 14th International Conference on Intelligent Virtual Agent (IVA 2014) (pp. 164–173). Boston, MA, USA.
- Fourati, N., Huang, J., & Pelachaud, C. (2014). Dynamic stimuli visualization for experimental studies of body language. In 10th Workshop on Multimodal Corpora, In conjunction with LREC 2014 (MMC 2014) (pp. 43–46). Reykjavik, Iceland.
- Fourati, N., & Pelachaud, C. (2014b). Emilya: Emotional body expression in daily actions database. In 9th International Conference on Language Resources and Evaluation (LREC 2014) (pp. 3486–3493). Reykjavik, Iceland.
- Fourati, N., & Pelachaud, C. (2014a). Collection and characterization of emotional body behaviors. In International Workshop on Movement and Computing, (MOCO'14) (pp. 49–54). Paris, France: ACM.
- Fourati, N., & Pelachaud, C. (2015). Multi-level classification of emotional body expression. In 11th IEEE International Conference on Automatic Face and Gesture Recognition (FG2015). Ljubljana, Slovenia.
- Fourati, N., & Pelachaud, C. (2015). Relevant body cues for the classification of emotional body expression in daily actions. In sixth International Conference on Affective Computing and Intelligent Interaction (ACII2015). Xi'an,

China.

J

Bibliography

[3ds,] 3DSuit Motion Capture System,http://inertiallabs.com/3dsuit.html.

[CMU,] CMU Graphics Lab Motion Capture Database, http://mocap.cs.cmu.edu/.

[gyp,] Gypsy-7 Motion Capture System,http://www.animazoo.fr/motion-capture-systems/gypsy-7-motion-capture-system/.

[IGS,] IGS-180 Motion Capture System,http://www.animazoo.fr/motion-capture-systems/igs-180-motion-capture-system/.

[Ani,] IGS-190. Animazoo website. *http://www.animazoo.com.*

[Xse,] Xsens Technologies, http://www.xsens.com.

[Abadi et al., 2015] Abadi, M., Correa, J., Wache, J., and Yang, H. (2015). Inference of Personality Traits and Affect Schedule by Analysis of Spontaneous Reactions to Affective Videos. In *11th IEEE International Conference on Automatic Face and Gesture Recognition (FG2015)*, number i.

[Abd El Meguid and Levine, 2014] Abd El Meguid, M. and Levine, M. (2014). Fully Automated Recognition of Spontaneous Facial Expressions in Videos Using Random Forest Classifiers.

[Alaoui, 2012] Alaoui, S. F. (2012). *Dance gesture analysis and physical models based visual feedback: contribution of movement qualities to interaction*. PhD thesis, University Paris-Sud, France.

[Alaoui et al., 2012] Alaoui, S. F., Caramiaux, B., Serrano, M., and Bevilacqua, F. (2012). Movement qualities as interaction modality. In *Proceedings of the Designing Interactive Systems Conference*, DIS '12, pages 761–769, New York, NY, USA. ACM.

[Altmann et al., 2010] Altmann, A., Toloşi, L., Sander, O., and Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics (Oxford, England)*, 26(10):1340–7.

[Amelynck et al., 2012] Amelynck, D., Grachten, M., van Noorden, L., and Leman, M. (2012). Toward E-Motion-Based Music Retrieval a Study of Affective Gesture Recognition. *IEEE Transactions on Affective Computing*, 3(2):250–259.

[Anne-Helene Olivier and Cretual, 2009] Anne-Helene Olivier, Richard Kulpa, J. P. and Cretual, A. (2009). A Velocity-Curvature Space Approach for Walking Motions Analysis. *MIG 2009*, pages 104–115.

[Arikan et al., 2003] Arikan, O., Forsyth, D. a., and O'Brien, J. F. (2003). Motion synthesis from annotations. *ACM SIGGRAPH 2003 Papers on - SIGGRAPH '03*, page 402.

[Aronoff, 2006] Aronoff, J. (2006). How We Recognize Angry and Happy Emotion in People, Places, and Things. *Cross-Cultural Research*, 40(1):83–105.

[Atkinson et al., 2004] Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., and Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, 33(6):717–746.

[Aung et al., 2015] Aung, M. S. H., Kaltwang, S., Romera-paredes, B., Martinez, B., Cella, M., Valstar, M., Meng, H., Kemp, A., Shafizadeh, M., Elkins, C., Kanakam, N., Rothschild, A. D., Tyler, N., Watson, P. J., C, A. C. D., Pantic, M., and Bianchi-berthouze, N. (2015). The automatic detection of chronic pain-related expression : requirements , challenges and a multimodal dataset. *IEEE Transactions on Affective Computing*, pages 1–18.

[Aviezer et al., 2012] Aviezer, H., Trope, Y., and Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science (New York, N.Y.)*, 338(6111):1225–9.

[Ball et al., 2002] Ball, G., Mian, S., Holding, F., Allibone, R. O., Lowe, J., Ali, S., Li, G., Mccardle, S., Ellis, I. O., Creaser, C., and Rees, R. C. (2002). An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. *Bioinformatics*, 18(3):395–404.

[Banerjee, 1999] Banerjee, M. (1999). Beyond kappa: A review of interrater agreement measures. *Canadian Journal of Statistics*, 27(1):3–23.

[Bänziger et al., 2012] Bänziger, T., Mortillaro, M., and Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion (Washington, D.C.)*, 12(5):1161–79.

[Bänziger et al., 2006] Bänziger, T., Pirker, H., and Scherer, K. (2006). GEMEP-GEneva Multimodal Emotion Portrayals: A corpus for the study of multimodal emotional expressions. *Proceedings of LREC*.

[Bänziger and Scherer, 2007] Bänziger, T. and Scherer, K. R. (2007). Using Actor Portrayals to Systematically Study Multimodal Emotion Expression : The GEMEP Corpus. pages 476–487.

[Bänziger and Scherer, 2010] Bänziger, T. and Scherer, K. R. (2010). SCENARIOS - GEMEP (GEneva Multimodal Emotion Portrayals).

[Bartenieff and Lewis, 1980] Bartenieff, I. and Lewis, D. (1980). *Body Movement: Coping with the Environment*. Routledge, New York.

[Bente et al., 2001] Bente, G., Krämer, N. C., Petersen, A., and De Ruiter, J. P. (2001). Computer animated movement and person perception: methodological advances in nonverbal behavior research. *Journal of Nonverbal Behavior*, 25(3):151–166.

[Bentivogli et al., 2004] Bentivogli, L., Forner, P., Magnini, B., and Pianta, E. (2004). Revising the wordnet domains hierarchy: semantics, coverage and balancing. *Proceedings of the Workshop on Multilingual Linguistic Ressources*, pages 101–108.

[Bernhardt and Robinson, 2007] Bernhardt, D. and Robinson, P. (2007). Detecting affect from non-stylised body motions. *Affective Computing and Intelligent Interaction*, 4738:59–70.

[Birdwhistell, 1970] Birdwhistell, R. L. (1970). *Kinesics and context : Essays on body motion communication*. University of Pensylvania Press, Philadelphia, conduct an edition.

[Blum and Langley, 1997] Blum, A. and Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(97):245–271.

[Breiman, 1996] Breiman, L. (1996). Bagging predictors. *Machine learning*, 140:123–140.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

[Breiman, 2002] Breiman, L. (2002). Manual on setting up, using, and understanding random forests v3. 1. *Technical Report, Statistics Department University of California Berkeley, CA, USA*, page 29.

[Breiman et al., 1984] Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA.

[Burgoon et al., 2010] Burgoon, J. K., Guerrero, L. K., and Floyd, K. (2010). *Nonverbal communication*. Allyn & Bacon Boston, MA.

[Busso et al., 2008] Busso, C., Bulut, M., Lee, C.-C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J. N., Lee, S., and Narayanan, S. S. (2008). IEMOCAP: interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 42(4):335–359.

[Camurri et al., 2003] Camurri, A., Lagerlöf, I., and Volpe, G. (2003). Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies*, 59(1-2):213–225.

[Camurri et al., 2004] Camurri, A., Mazzarino, B., and Volpe, G. (2004). Analysis of Expressive Gesture: The EyesWeb Expressive Gesture Processing Library. *In Gesture-based Communication in Human-Computer Interaction, LNAI 2915 (Springer Verlag)*, pages 460–467.

[Cassell, 2000] Cassell, J. (2000). More Than Just Another Pretty Face : Embodied Conversational Interface Agents. In *Communication of the ACM*, pages 70–78.

[Castellano et al., 2012] Castellano, G., Mancini, M., Peters, C., and Mcowan, P. W. (2012). Correspondence Expressive Copying Behavior for Social Agents : A Perceptual Analysis. 42(3):776–783.

[Castellano et al., 2008] Castellano, G., Mortillaro, M., and Camurri, A. (2008). Automated analysis of body movement in emotionally expressive piano performances. *Music Perception*, 26(2):103–120.

[Castellano et al., 2007] Castellano, G., Villalba, S., and Camurri, A. (2007). Recognising human emotions from body movement and gesture dynamics. *ACII*, pages 71–82.

[Chi et al., 2000] Chi, D., Costa, M., Zhao, L., and Badler, N. (2000). The EMOTE model for effort and shape. In Press, A. C. M., editor, *Proceedings of the 27th annual conference on Computer graphics and interactive techniques SIGGRAPH 00*, volume ACM Comput, pages 173–182. ACM Press/Addison-Wesley Publishing Co., ACM Press.

[Clavel et al., 2011] Clavel, C., Vasilescu, I., and Devillers, L. (2011). Fiction support for realistic portrayals of fear-type emotional manifestations. *Computer Speech and Language*, 25(1):63–83.

[Coan and Gottman, 2007] Coan, J. A. and Gottman, J. M. (2007). The Specific Affect Coding System (SPAFF). In Coan, J. A. and Allen, J. J. B., editors, *Handbook of emotion elicitation and assessment.*, Series in affective science, pages 267–285. Oxford University Press.

[Coolican, 2004] Coolican, H. (2004). Research Methods & Statistics in Psychology 4th Edition. chapter Questionnaires - using psychological scales and tests in research, pages 168–199. Routledge; 4 edition.

[Coulson, 2004] Coulson, M. (2004). Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior*, 28(2):117–139.

[Cowie et al., 2011] Cowie, R., Douglas-Cowie, E., McRorie, M., Sneddon, I., Devillers, L., and Noam, A. (2011). Issues in Data Collection. In Petta, P., Pelachaud, C., and Cowie, R., editors, *Emotion-Oriented Systems: The HUMAINE Handbook*, pages 197–212.

[Crane and Gross, 2007] Crane, E. and Gross, M. (2007). Motion Capture and Emotion : Affect Detection in Whole Body Movement. *Affective computing and intelligent interaction*, pages 95–101.

[Cutler et al., 2012] Cutler, A., Cutler, D., and Stevens, J. (2012). Random Forests. In Zhang, C. and Ma, Y., editors, *Ensemble Machine Learning*, pages 157–175. Springer US.

[Dael et al., 2011] Dael, N., Mortillaro, M., and Scherer, K. R. (2011). Emotion expression in body action and posture. *Emotion*, 12(5):1085–1101.

[Dael et al., 2012] Dael, N., Mortillaro, M., and Scherer, K. R. (2012). The Body Action and Posture Coding System (BAP): Development and Reliability. *Journal of Nonverbal Behavior*, pages 97–121.

[Daems and Verfaillie, 1999] Daems, A. and Verfaillie, K. (1999). Viewpoint-dependent Priming Effects in the Perception of Human Actions and Body Postures. *Visual Cognition*, 6(6):665–693.

[Dahl and Friberg, 2007] Dahl, S. and Friberg, A. (2007). Visual Perception of Expressiveness in Musicians' Body Movements. *Music Perception*, 24(5):433–454.

[Darlow and Goldin, 2013] Darlow, A. and Goldin, G. (2013). TurkGate (Version 0.4.0) [Software]. Available from http://gideongoldin.github.com/TurkGate/.

[Darwin, 1872] Darwin, C. (1872). *The expression of emotions in man and animals*, volume 2. John Murray.

[De Gelder, 2006] De Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience*, 7(3):242–249.

[de Gelder and Van den Stock, 2011] de Gelder, B. and Van den Stock, J. (2011). The Bodily Expressive Action Stimulus Test (BEAST). Construction and Validation of a Stimulus Basis for Measuring Perception of Whole Body Expression of Emotions. *Frontiers in psychology*, 2(August):181.

[De Silva and Bianchi-Berthouze, 2004] De Silva, P. R. and Bianchi-Berthouze, N. (2004). Modeling human affective postures: An information theoretic characterization of posture features. In *Computer Animation and Virtual Worlds*, volume 15, pages 269–276.

[Dekeyser et al., 2002] Dekeyser, M., Verfaillie, K., and Vanrie, J. (2002). Creating stimuli for the study of biological-motion perception. *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc*, 34(3):375–82.

[Dell, 1977] Dell, C. (1977). *A primer for movement description using effort-shape and supplementary concepts*. Dance Notation Bureau.

[Demulier et al., 2014] Demulier, V., Bevacqua, E., and Focone, F. (2014). A Database of Full Body Virtual Interactions Annotated with Expressivity Scores. *LREC*.

[Devillers and Martin, 2012] Devillers, L. and Martin, J.-C. (2012). *Emotional Corpora: From Acquisition to Modeling*, pages 77–105. John Wiley & Sons, Inc.

[Díaz-Uriarte and Alvarez de Andrés, 2006] Díaz-Uriarte, R. and Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC bioinformatics*, 7:3.

[D'Mello and Graesser, 2009] D'Mello, S. and Graesser, A. (2009). Automatic detection of learner's affect from gross body language.

[Ekman, 1989] Ekman, P. (1989). The argument and evidence about universals in facial expressions of emotion. In Wagner, H. and Manstead, A. S. R., editors, *Handbook of social psychophysiology*, volume 58, chapter 6, pages 143–164. John Wiley & Sons Ltd.

[Ekman and Friesen, 1967] Ekman, P. and Friesen, W. V. (1967). Head and body cues in the judgment of emotion: a reformulation. *Perceptual and motor skills*, 24(3):711–724.

[Ekman and Friesen, 1969] Ekman, P. and Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1(1):49–98.

[Ekman and Friesen, 1972] Ekman, P. and Friesen, W. V. (1972). Hand Movements. *Journal of Communication*, 22(4):353–374.

[Ekman and Friesen, 1974] Ekman, P. and Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology*, 29(3):288–298.

[Ekman and Friesen, 1977] Ekman, P. and Friesen, W. V. (1977). Nonverbal behavior. In *Communication and social interaction*, pages 37–46. Grune & Stratton, New York, p.f. ostwa edition.

[Ekman and Friesen, 1978] Ekman, P. and Friesen, W. V. (1978). *Facial Action Coding System*, volume 160. Consulting Psychologists Press.

[Ennis et al., 2013] Ennis, C., Hoyet, L., Egges, A., and McDonnell, R. (2013). Emotion Capture: Emotionally Expressive Characters for Games. In *MIG '13 Proceedings of the Motion on Games*, pages 31–38.

[Eroglu Erdem et al., 2014] Eroglu Erdem, C., Turan, C., and Aydin, Z. (2014). BAUM-2: a multilingual audio-visual affective face database. *Multimedia Tools and Applications*, pages 1–31.

[Féraud and Clérot, 2002a] Féraud, R. and Clérot, F. (2002a). A methodology to explain neural network classification. *Neural Networks*, 15.

[Féraud and Clérot, 2002b] Féraud, R. and Clérot, F. (2002b). A methodology to explain neural network classification. *Neural Networks*, 15.

[Fourati and Pelachaud, 2014] Fourati, N. and Pelachaud, C. (2014). Collection and characterization of emotional body behaviors. *the International Workshop on Movement and Computing, MOCO'14*, pages 49–54.

[Freund and Schapire, 1996] Freund, Y. and Schapire, R. R. E. (1996). Experiments with a New Boosting Algorithm. In *International Conference on Machine Learning*, pages 148–156.

[Frey and Von Cranach, 1973] Frey, S. and Von Cranach, M. (1973). A method for the assessment of body movement variability. *Social Communication and Movement*, pages 389–418.

[Frijda, 1986] Frijda, N. H. (1986). *The emotions*, volume 1 of *Studies in Emotion and Social Interaction*. Cambridge University Press.

[Genuer, 2010] Genuer, R. (2010). *Forêts aléatoires: aspects théoriques, sélection de variables et applications*.

[Genuer et al., 2010] Genuer, R., Poggi, J.-M., and Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31(14):2225–2236.

[Ghattas and Ben Ishak, 2008] Ghattas, B. and Ben Ishak, A. (2008). SÉLection De Variables Pour La Classification Binaire En Grande Dimension : Comparaisons Et Application Aux DonnÉEs De Biopuces. *Journal de la Société française de statistique & Revue de statistique appliquée*, 149(3):43–66.

[Gilbert and Andrews, 1998] Gilbert, P. and Andrews, B. (1998). *Shame: Interpersonal behavior, psychopathology, and culture.*

[Glowinski et al., 2011] Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., and Scherer, K. (2011). Toward a minimal representation of affective gestures. *IEEE Transactions on Affective Computing*, 2(2):106–118.

[Graham and Jackson, 1993] Graham, P. and Jackson, R. (1993). The analysis of ordinal agreement data: beyond weighted kappa. *Journal of clinical epidemiology*, 46(9):1055–62.

[Griffin et al., 2013a] Griffin, H. J., Aung, M. S., Romera-Paredes, B., McLoughlin, C., McKeown, G., Curran, W., and Bianchi-Berthouze, N. (2013a). Laughter Type Recognition from Whole Body Motion. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 349–355.

[Griffin et al., 2013b] Griffin, H. J., Aung, M. S., Romera-Paredes, B., McLoughlin, C., McKeown, G., Curran, W., and Bianchi-Berthouze, N. (2013b). Laughter Type Recognition from Whole Body Motion. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 349–355.

[Gross et al., 2010] Gross, M. M., Crane, E. A., and Fredrickson, B. L. (2010). Methodology for Assessing Bodily Expression of Emotion. *Journal of Nonverbal Behavior*, 34(4):223–248.

[Gunes and Piccardi, 2006] Gunes, H. and Piccardi, M. (2006). A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on. IEEE*, pages 1148–1153.

[Guyon and Elisseeff, 2003] Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182.

[Guyon et al., 2002] Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1-3):389–422.

[Hager et al., 1995] Hager, J. C., Ekman, P., and Bichsel, M. (1995). Essential behavioral science of face and gesture that computer scientists need to know. In *International Workshop on Automatic Face and Gesture Recognition Preceedings*, pages 7–11. Multimedia Laboratory, University of Zürich.

[Hall and Smith, 1999] Hall, M. and Smith, L. A. (1999). Feature Selection for Machine Learning : Comparing a Correlation-based Filter Approach to the Wrapper CFS : Correlation-based Feature. In *International FLAIRS Conference*, page 5.

[Hapfelmeier and Ulm, 2013] Hapfelmeier, A. and Ulm, K. (2013). A new variable selection approach using Random Forests. *Computational Statistics & Data Analysis*, 60:50–69.

[Haugstad et al., 2006] Haugstad, G. K., Haugstad, T. S., Kirste, U. M., Leganger, S., Wojniusz, S., Klemmetsen, I., and Malt, U. F. (2006). Posture, movement patterns, and body awareness in women with chronic pelvic pain. *Journal of Psychosomatic Research*, 61(5):637–644.

[Hicheur and Berthoz, 2005] Hicheur, H. and Berthoz, a. (2005). How do humans turn? head and body movements for the steering of locomotion. *5th IEEE-RAS International Conference on Humanoid Robots, 2005.*, pages 265–270.

[Hicheur et al., 2013] Hicheur, H., Kadone, H., Grèzes, J., and Berthoz, A. (2013). Perception of emotional gaits using avatar animation of real and artificially synthesized gaits. *Humaine Association Conference on Affective Computing and Intelligent Interaction*.

[Hicheur et al., 2005] Hicheur, H., Vieilledent, S., Richardson, M. J. E., Flash, T., and Berthoz, a. (2005). Velocity and curvature in human locomotion along complex curved paths: a comparison with hand movements. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, 162(2):145–54.

[Ho, 1998] Ho, T. (1998). The random subspace method for constructing decision forests. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 20(8):832–844.

[Hothorn et al., 2006] Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15(3):651–674.

[Huang and Pelachaud, 2012] Huang, J. and Pelachaud, C. (2012). Expressive Body Animation Pipeline for Virtual Agent. In *IVA 2012,12th International Conference on Intelligent Virtual Agents*, pages 355–362.

[Hwang et al., 2006] Hwang, B. W., Kim, S., and Lee, S. W. (2006). A full-body gesture database for automatic gesture recognition. In *FGR 2006: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, volume 2006, pages 243–248.

[Imai et al., 2001] Imai, T., Moore, S. T., Raphan, T., and Cohen, B. (2001). Interaction of the body, head, and eyes during walking and turning. *Experimental Brain Research*, 136(1):1–18.

[Inza et al., 2004] Inza, I. n., Larrañaga, P., Blanco, R., and Cerrolaza, A. J. (2004). Filter versus wrapper gene selection approaches in DNA microarray domains. *Artificial Intelligence in Medicine*, 31(2):91–103.

[J., 1978] J., K. (1978). Pattern Recognition and Signal Processing. In *Chapter Feature Set Search Algorithms Sijthoff and Noordhoff, Alphen aan den Rijn*, chapter Chapter Fe, pages 41–60.

[J.A. Harrigan, 2005] J.A. Harrigan, R. R. . K. S. (2005). Proxemics, kinesics and gaze. In Oxford University Press, editor, *The new handbook of methods in nonverbal behavior research*, pages 137–198. Oxford.

[Janssen et al., 2008] Janssen, D., Schöllhorn, W. I., Lubienetzki, J., Fölling, K., Kokenge, H., and Davids, K. (2008). Recognition of emotions in gait patterns by means of artificial neural nets. *Journal of Nonverbal Behavior*, 32(2):79–92.

[Johansson, 1973] Johansson, G. (1973). biological motion. *Perception & Psychophysics*, 14(2):201–211.

[John et al., 1994] John, G., Kohavi, R., and Pfleger, K. (1994). Irrelevant Features and the Subset Selection Problem. *ICML*, pages 121–129.

[Kapoor et al., 2007] Kapoor, A., Burleson, W., and Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies*, 65(8):724–736.

[Kapur et al., 2005] Kapur, A., Kapur, A., Virji-babul, N., and Tzanetakis, G. (2005). Gesture-Based Affective Computing on Motion. pages 1–7.

[Karg et al., 2010] Karg, M., Kühnlenz, K., and Buss, M. (2010). Recognition of affect based on gait patterns. *Systems, Man, and Cybernetics*, 40(4):1050–1061.

[Karg and Samadani, 2013] Karg, M. and Samadani, A. (2013). Body movements for affective expression: a survey of automatic recognition and generation. *IEEE Transactions on Affective Computing*, 4(4):341–359.

[Khalilia et al., 2011] Khalilia, M., Chakraborty, S., and Popescu, M. (2011). Predicting disease risks from highly imbalanced data using random forest. *BMC Med Inform Decis Mak*, 11:51.

[Kleinsmith and Berthouze, 2007] Kleinsmith, A. and Berthouze, N. (2007). Recognizing Affective Dimensions from Body Posture. *Affective Computing and Intelligent Interaction*, LNCS-4738:1–12.

[Kleinsmith and Bianchi-Berthouze, 2013] Kleinsmith, A. and Bianchi-Berthouze, N. (2013). Affective Body Expression Perception and Recognition: A Survey. *IEEE Transactions on Affective Computing*, 4(1):15–33.

[Kleinsmith et al., 2011] Kleinsmith, A., Bianchi-Berthouze, N., and Steed, A. (2011). Automatic Recognition of Non-Acted Affective Postures. *IEEE Transactions on Systems Man and Cybernetics Part B Cybernetics*, 41(4):1027–1038.

[Kleinsmith et al., 2006a] Kleinsmith, A., De Silva, P. R., and Bianchi-Berthouze, N. (2006a). Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers*, 18(6):1371–1389.

[Kleinsmith et al., 2006b] Kleinsmith, A., Desilva, P., and Bianchiberthouze, N. (2006b). Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers*, 18(6):1371–1389.

[Kleinsmith et al., 2009] Kleinsmith, A., Rebai, I., Berthouze, N., and Martin, J.-C. (2009). Postural expressions of emotion in a motion captured database and in a humanoid robot. *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots - AFFINE '09*, (1):1–2.

[Kohavi and John, 1997] Kohavi, R. and John, H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(97):273–324.

[Laban, 1988] Laban, R. (1988). *The mastery of movement*. Plymouth, UK.

[Laban and Ullmann, 1966] Laban, R. V. and Ullmann, L. (1966). *The language of movement: a guidebook to choreutics*. Plays, inc.

[Levy and Duke, 2003] Levy, J. A. and Duke, M. P. (2003). The Use of Laban Movement Analysis in the Study of Personality, Emotional State and Movement Style: An Exploratory Investigation of the Veridicality of "Body Language" . *Individual Differences Research*, 1(1):39–63.

[Li et al., 2012] Li, F., Zhao, C., Ding, G., Gong, J., Liu, C., and Zhao, F. (2012). A reliable and accurate indoor localization method using phone inertial sensors. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*, page 421.

[Li et al., 2004] Li, T., Zhang, C., and Ogihara, M. (2004). A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics (Oxford, England)*, 20(15):2429–2437.

[Lunetta et al., 2004] Lunetta, K. L., Hayward, L. B., Segal, J., and Van Eerdewegh, P. (2004). Screening large-scale association study data: exploiting interactions using random forests. *BMC genetics*, 5:32.

[Ma et al., 2006] Ma, Y., Paterson, H. M., and Pollick, F. E. (2006). A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior research methods*, 38(1):134–41.

[Maldonado and Weber, 2011] Maldonado, S. and Weber, R. (2011). Embedded feature selection for support vector machines: state-of-the-art and future challenges. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 304–311.

[Maldonado et al., 2011] Maldonado, S., Weber, R., and Basak, J. (2011). Simultaneous feature selection and classification using kernel-penalized support vector machines. *Information Sciences*, 181(1):115–128.

[Markuš et al., 2014] Markuš, N., Frljak, M., Pandžić, I. S., Ahlberg, J., and Forchheimer, R. (2014). Eye pupil localization with an ensemble of randomized trees. *Pattern Recognition*, 47(2):578–587.

[McKeown et al., 2012] McKeown, G., Valstar, M., Cowie, R., Pantic, M., and Schroder, M. (2012). The SEMAINE Database: Annotated Multimodal Records of Emotionally Colored Conversations between a Person and a Limited Agent. *IEEE Transactions on Affective Computing*, 3(1):5–17.

[McNeill, 2005] McNeill, D. (2005). Gesture and Thought. *Continuum*, (18):University of Chicago Press.

[Mehrabian, 1972] Mehrabian, A. (1972). *Nonverbal Communication*. Aldine: Atherton., Chicago.

[Meijer, 1989] Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, 13(4):247–268.

[Metallinou et al., 2010] Metallinou, A., Lee, C., and Busso, C. (2010). The USC CreativeIT database: A multimodal database of theatrical improvisation. *Workshop on Multimodal Corpora, LREC*.

[Mickelborough et al., 2000] Mickelborough, J., van der Linden, M. L., Richards, J., and Ennos, a. R. (2000). Validity and reliability of a kinematic protocol for determining foot contact events. *Gait & posture*, 11(1):32–7.

[Milders et al., 2003] Milders, M., Crawford, J. R., Lamb, A., and Simpson, S. A. (2003). Differential deficits in expression recognition in gene-carriers and patients with Huntington's disease. Technical Report 11, Department of Psychology, University of Aberdeen, Aberdeen AB24 2UB, Scotland, UK. m.milders@abdn.ac.uk.

[Montepare et al., 1999] Montepare, J., Koff, E., Zaitchik, D., and Albert, M. (1999). The use of body movements and gestures as cues to emotions in younger and older adults. *Journal of Nonverbal Behavior*, 23(2):133–152.

[Montepare et al., 1987] Montepare, J. M., Goldstein, S. B., and Clausen, A. (1987). The identification of emotions from gait information. *Journal of Nonverbal Behavior*, 11(1):33–42.

[Mota and Picard, 2003] Mota, S. and Picard, R. W. (2003). Automated Posture Analysis for Detecting Learner's Interest Level. *2003 Conference on Computer Vision and Pattern Recognition Workshop*, 5:49–49.

[Multon, 2013] Multon, F. (2013). Sensing human walking: Algorithms and techniques for extracting and modeling locomotion. pages 177–197.

[MVN BIOMECH system. Xsens website,] MVN BIOMECH system. Xsens website. http://www.xsens.com/.

[Narsky and Porter, 2014a] Narsky, I. and Porter, F. C. (2014a). Decision trees. In *Statistical Analysis Techniques in Particle Physics*, pages 307–329. Wiley-VCH Verlag GmbH and Co. KGaA.

[Narsky and Porter, 2014b] Narsky, I. and Porter, F. C. (2014b). Methods for variable ranking and selection. In *Statistical Analysis Techniques in Particle Physics*, pages 385–415. Wiley-VCH Verlag GmbH & Co. KGaA.

[Neill and Caswell, 1993] Neill, S. and Caswell, C. (1993). *Body Language for Competent Teachers*. Taylor & Francis.

[Niewiadomski et al., 2011] Niewiadomski, R., Hyniewska, S. J., and Pelachaud, C. (2011). Constraint-Based Model for Synthesis of Multimodal Sequential Expressions of Emotions. *IEEE Transactions on Affective Computing*, 2(3):134–146.

[Niewiadomski et al., 2013] Niewiadomski, R., Mancini, M., and Baur, T. (2013). MMLI: Multimodal multiperson corpus of laughter in interaction. *In proceeding of: 4th international workshop on Human Behavior Understanding, In conjunction with ACM Multimedia 2013, At Barcelona, Spain*, 8212:pp 184–195.

[Olivier and Cretual, 2007] Olivier, A.-H. and Cretual, A. (2007). Velocity/curvature relations along a single turn in human locomotion. *Neuroscience letters*, 412(2):148–53.

[Olivier et al., 2009] Olivier, A.-H., Kulpa, R., Pettre, J., and Armel, C. (2009). A Velocity-Curvature Space Approach for Walking Motions Analysis. *MIG 2009*, pages 104–115.

[Oshiro et al., 2012] Oshiro, T. M., Perez, P. S., and Baranauskas, J. A. (2012). How many trees in a random forest? In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 7376 LNAI, pages 154–168.

[Pang et al., 2012] Pang, H., George, S. L., Hui, K., and Tong, T. (2012). Gene selection using iterative feature elimination random forests for survival outcomes. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM*, 9(5):1422–31.

[Park et al., 2004] Park, H., Park, J.-I., Kim, U.-M., and Woo, W. (2004). Emotion recognition from dance image sequences using contour approximation. In Fred, A., Caelli, T., Duin, R., Campilho, A., and de Ridder, D., editors, *Structural, Syntactic, and Statistical Pattern Recognition*, volume 3138 of *Lecture Notes in Computer Science*, pages 547–555. Springer Berlin Heidelberg.

[Patterson et al., 2001] Patterson, H. M., Pollick, F. E., and Sanford, A. J. (2001). The Role of Velocity in Affect Discrimination. *Dimension Contemporary German Arts And Letters*, pages 756–761.

[Pollick et al., 2001] Pollick, F. E., Paterson, H. M., Bruderlin, A., and Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, 82(2):B51–B61.

[Prinzie and Van den Poel, 2008] Prinzie, A. and Van den Poel, D. (2008). Random Forests for multiclass classification: Random MultiNomial Logit. *Expert Systems with Applications*, 34(3):1721–1732.

[Quinlan, 1986] Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1):81–106.

[Quinlan, 1993] Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[Rakotomamonjy, 2003] Rakotomamonjy, A. (2003). Variable Selection Using SVM-based Criteria. *Journal ofMachine Learning Research*, 3:1357–1370.

[Roetenberg et al., 2009] Roetenberg, D., Luinge, H., and Slycke, P. (2009). Xsens MVN: Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors. *Xsens Technologies Technical Whitepaper*.

[Roether et al., 2009] Roether, C. L., Omlor, L., Christensen, A., and Giese, M. A. (2009). Critical features for the perception of emotion from gait. *Journal of Vision*, 9(6):1–32.

[Rose et al., 1998] Rose, C., Cohen, M. F., and Bodenheimer, B. (1998). Verbs and adverbs: Multidimensional motion interpolation. *IEEE Computer Graphics and Applications*, 18(5):32–40.

[Russell, 1994] Russell, J. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological bulletin*, 115(1):102–41.

[Russell, 1980] Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.

[Saeys et al., 2008] Saeys, Y., Abeel, T., and Peer, Y. (2008). Robust Feature Selection Using Ensemble Feature Selection Techniques. In Daelemans, W., Goethals, B., and Morik, K., editors, *European conference on Machine Learning and Knowledge Discovery in Databases (ECML/PKDD)*, pages 313–325. Springer-Verlag.

[Saeys et al., 2007] Saeys, Y., Inza, I. n., and Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics.

[Salarian et al., 2009] Salarian, A., Zampieri, C., Horak, F. B., Carlson-Kuhta, P., Nutt, J. G., and Aminian, K. (2009). Analyzing 180 degrees turns using an inertial system reveals early signs of progression of Parkinson's disease. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2009:224–7.

[Salkind, 2010] Salkind, N. J. (2010). *Encyclopedia of Research Design*. SAGE Publications.

[Savva et al., 2012] Savva, N., Scarinzi, A., and Bianchi-Berthouze, N. (2012). Continuous Recognition of Player's Affective Body Expression as Dynamic Quality of Aesthetic Experience. *IEEE Transactions on Computational Intelligence and AI in Games*, 4(3):199–212.

[Scherer et al., 1991] Scherer, K. R., Banse, R., Wallbott, H. G., and Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, 15(2):123–148.

[Scherer and Ellgring, 2007] Scherer, K. R. and Ellgring, H. (2007). Multimodal expression of emotion: affect programs or componential appraisal patterns? *Emotion Washington Dc*, 7(1):158–171.

[Sigal et al., 2010] Sigal, L., Balan, A. O., and Black, M. J. (2010). HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International Journal of Computer Vision*, 87(1-2):4–27.

[Sreenivasa et al., 2008] Sreenivasa, M. N., Frissen, I., Souman, J. L., and Ernst, M. O. (2008). Walking along curved paths of different angles: the relationship between head and trunk turning. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, 191(3):313–20.

[Strapparava and Valitutti, 2004] Strapparava, C. and Valitutti, A. (2004). WordNet-Affect: an affective extension of WordNet. *Proceedings of the 4th International Conference on Language Resources and Evaluation*, pages 1083–1086.

[Strobl et al., 2008] Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. (2008). Conditional variable importance for random forests. *BMC bioinformatics*, 9:307.

[Strobl et al., 2007] Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC bioinformatics*, 8:25.

[Strobl et al., 2009] Strobl, C., Malley, J., and Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods*, 14(4):323–348.

[Svetnik et al., 2004] Svetnik, V., Liaw, A., Tong, C., and Wang, T. (2004). Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. *Multiple Classifier Systems*, pages 334–343.

[Tilmanne, 2013] Tilmanne, J. (2013). *Data-driven Stylistic Humanlike Walk Synthesis*. PhD thesis.

[Tilmanne and Dutoit, 2010a] Tilmanne, J. and Dutoit, T. (2010a). Expressive gait synthesis using pca and gaussian modeling. In Boulic, R., Chrysanthou, Y., and Komura, T., editors, *Motion in Games*, volume 6459 of *Lecture Notes in Computer Science*, pages 363–374. Springer Berlin Heidelberg.

[Tilmanne and Dutoit, 2010b] Tilmanne, J. and Dutoit, T. (2010b). Expressive Gait Synthesis Using PCA and Gaussian Modeling. pages 363–374.

[Tilmanne and Dutoit, 2011] Tilmanne, J. and Dutoit, T. (2011). Continuous Control of Style through Linear Interpolation in Hidden Markov Model Based Stylistic Walk Synthesis. pages 232–236.

[Tilmanne et al., 2009] Tilmanne, J., Sebbe, R., and Dutoit, T. (2009). A Database for Stylistic Human Gait Modeling and Synthesis. *in Proceedings of the eNTERFACE'08 Workshop on Multimodal Interfaces*, pages 91–94.

[Tolosi and Lengauer, 2011] Tolosi, L. and Lengauer, T. (2011). Classification with correlated features: unreliability of feature ranking and solutions. *Bioinformatics (Oxford, England)*, 27(14):1986–1994.

[Tong and Granat, 1999] Tong, K. and Granat, M. H. (1999). A practical gait analysis system using gyroscopes. *Medical engineering & physics*, 21(2):87–94.

[Tracy and Robins, 2007] Tracy, J. L. and Robins, R. W. (2007). The prototypical pride expression: development of a nonverbal behavior coding system. *Emotion Washington Dc*, 7(4):789–801.

[Tracy and Robins, 2008] Tracy, J. L. and Robins, R. W. (2008). The nonverbal expression of pride: evidence for cross-cultural recognition. *Journal of Personality and Social Psychology*, 94(3):516–530.

[Urbain et al., 2010] Urbain, J., Bevacqua, E., Dutoit, T., Moinet, A., Niewiadomski, R., Pelachaud, C., Picart, B., Tilmanne, J., and Wagner, J. (2010). The avlaughtercycle database. In Chair), N. C. C., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., and Tapias, D., editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

[Van Den Stock et al., 2007] Van Den Stock, J., Righart, R., and De Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion Washington Dc*, 7(3):487–494.

[van der Laan, 2006] van der Laan, M. J. (2006). Statistical Inference for Variable Importance. *The International Journal of Biostatistics*, 2(1).

[Velloso et al., 2013] Velloso, E., Bulling, A., and Gellersen, H. (2013). AutoBAP: Automatic coding of body action and posture units from wearable sensors. In *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, pages 135–140.

[Verikas et al., 2011] Verikas, a., Gelzinis, a., and Bacauskiene, M. (2011). Mining data with random forests: A survey and results of new tests. *Pattern Recognition*, 44(2):330–349.

[Vieilledent et al., 2001] Vieilledent, S., Kerlirzin, Y., Dalbera, S., and Berthoz, a. (2001). Relationship between velocity and curvature of a human locomotor trajectory. *Neuroscience letters*, 305(1):65–9.

[Visser et al., 2007] Visser, J. E., Voermans, N. C., Oude Nijhuis, L. B., van der Eijk, M., Nijk, R., Munneke, M., and Bloem, B. R. (2007). Quantification of trunk rotations during turning and walking in Parkinson's disease. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 118(7):1602–6.

[Volkova et al., 2014a] Volkova, E., de la Rosa, S., Bülthoff, H. H., and Mohler, B. (2014a). The MPI Emotional Body Expressions Database for Narrative Scenarios. *PLoS ONE*, 9(12):e113647.

[Volkova et al., 2014b] Volkova, E. P., Mohler, B. J., Dodds, T. J., Tesch, J., and Bülthoff, H. H. (2014b). Emotion categorization of body expressions in narrative scenarios. *Frontiers in Psychology*, 5(JUN).

[von Laban and Ullmann, 1966] von Laban, R. and Ullmann, L. (1966). *Choreutics*. Macdonald & Evans.

[Wallbott, 1998] Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, 28(6):879–896.

[Wallbott and Scherer, 1986] Wallbott, H. G. and Scherer, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, 51(4):690–699.

[Winters, 2005] Winters, A. (2005). Perceptions of Body Posture and Emotion : A Question of Methodology. *Psychology*, 3(2):9.

[Zacharatos et al., 2014] Zacharatos, H., Gatzoulis, C., and Chrysanthou, Y. (2014). Automatic emotion recognition based on body movement analysis: A survey. *Computer Graphics and Applications, IEEE*, 34(6):35–45.

J

Classification and characterization of emotional body expression in daily actions

Nesrine Fourati

RESUME : Ce travail de thèse peut être résumé en quatre étapes principales. Premièrement, nous avons proposé un système d'annotation multi-niveaux pour décrire le mouvement corporel expressif dans différentes actions. Deuxièmement, nous avons enregistré une base de données de l'expression corporelle des émotions dans des actions quotidiennes. Cette base de données constitue un large corpus de comportements expressifs considérant l'expression de 8 émotions dans 7 actions quotidiennes, combinant à la fois les données audio-visuelle et les données de capture de mouvement et donnant lieu à plus que 8000 séquences de mouvement expressifs. Troisièmement, nous avons exploré la classification des émotions en se basant sur notre système d'annotation multi-niveaux. L'approche des forêts aléatoires est utilisée pour cette fin. L'utilisation des forêts aléatoires dans notre travail a un double objectif : 1) la fiabilité du modèle de classification, et 2) la possibilité de sélectionner un sous-ensemble de paramètres pertinents en se basant sur la mesure d'importance retournée par le modèle. Nous avons aussi comparé la classification automatique des émotions avec la perception humaine des émotions exprimées dans différentes actions. Finalement, nous avons extrait les paramètres les plus pertinents qui retiennent l'expressivité du mouvement en se basant sur la mesure d'importance retournée par le modèle des forêts aléatoires. Nous avons utilisé ce sous-ensemble de paramètres pour explorer la caractérisation de l'expression corporelle des émotions dans différentes actions quotidiennes. Un modèle d'arbre de décision a été utilisé pour cette fin.

MOTS-CLEFS : Capture de mouvement, Expression corporelle, Classification, paramètres corporels, Sélection des paramètres

ABSTRACT : The work conducted in this thesis can be summarized into four main steps. Firstly, we proposed a multi-level body movement notation system that allows the description of expressive body movement across various body actions. Secondly, we collected a new database of emotional body expression in daily actions. This database constitutes a large repository of bodily expression of emotions including the expression of 8 emotions in 7 actions, combining video and motion capture recordings and resulting in more than 8000 sequences of expressive behaviors. Thirdly, we explored the classification of emotions based on our multi-level body movement notation system. Random Forest approach is used for this purpose. The advantage of using Random Forest approach in our work is double-fold : 1) reliability of the classification model and 2) possibility to select a subset of relevant features based on their relevance measures. We also compared the automatic classification of emotions with human perception of emotions expressed in different actions. Finally, we extracted the most relevant features that capture the expressive content of the motion based on the relevance measure of features returned by the Random Forest model. We used this subset of features to explore the characterization of emotional body expression across different actions. A Decision Tree model was used for this purpose.

KEY-WORDS : Motion capture, Bodily expression, Classification, Body cues, Feature selection