Automatic Recognition and Generation of Affective Movements

by

Ali-Akbar Samadani

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2014

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Body movements are an important non-verbal communication medium through which affective states of the demonstrator can be discerned. For machines, the capability to recognize affective expressions of their users and generate appropriate actuated responses with recognizable affective content has the potential to improve their life-like attributes and to create an engaging, entertaining, and empathic human-machine interaction.

This thesis develops approaches to systematically identify movement features most salient to affective expressions and to exploit these features to design computational models for automatic recognition and generation of affective movements. The proposed approaches enable 1) identifying which features of movement convey affective expressions, 2) the automatic recognition of affective expressions from movements, 3) understanding the impact of kinematic embodiment on the perception of affective movements, and 4) adapting pre-defined motion paths in order to "overlay" specific affective content.

Statistical learning and stochastic modeling approaches are leveraged, extended, and adapted to derive a concise representation of the movements that isolates movement features salient to affective expressions and enables efficient and accurate affective movement recognition and generation. In particular, the thesis presents two new approaches to fixed-length affective movement representation based on 1) functional feature transformation, and 2) stochastic feature transformation (Fisher scores). The resulting representations are then exploited for recognition of affective expressions in movements and for salient movement feature identification. For functional representation, the thesis adapts dimensionality reduction techniques (namely, principal component analysis (PCA), Fisher discriminant analysis, Isomap) for functional datasets and applies the resulting reduction techniques to extract a minimal set of features along which affect-specific movements are best separable. Furthermore, the centroids of affect-specific clusters of movements in the resulting functional PCA subspace along with the inverse mapping of functional PCA are used to generate prototypical movements for each affective expression.

The functional discriminative modeling is however limited to cases where affect-specific movements also have similar kinematic trajectories and does not address the interpersonal and stochastic variations inherent to bodily expression of affect. To account for these variations, the thesis presents a novel affective movement representation in terms of stochastically-transformed features referred to as Fisher scores. The Fisher scores are derived from affect-specific hidden Markov model encoding of the movements and exploited to discriminate between different affective expressions using a support vector machine (SVM) classification. Furthermore, the thesis presents a new approach for systematic identification of a minimal set of movement features most salient to discriminating between different

affective expressions. The salient features are identified by mapping Fisher scores to a low-dimensional subspace where dependencies between the movements and their affective labels are maximized. This is done by maximizing Hilbert Schmidt independence criterion between the Fisher score representation of movements and their affective labels. The resulting subspace forms a suitable basis for affective movement recognition using nearest neighbour classification and retains the high recognition rates achieved by SVM classification in the Fisher score space. The dimensions of the subspace form a minimal set of salient features and are used to explore the movement kinematic and dynamic cues that connote affective expressions.

Furthermore, the thesis proposes the use of movement notation systems from the dance community (specifically, the Laban system) for abstract coding and computational analysis of movement. A quantification approach for Laban Effort and Shape is proposed and used to develop a new computational model for affective movement generation. Using the Laban Effort and Shape components, the proposed generation approach searches a labeled dataset for movements that are kinematically similar to a desired motion path and convey a target emotion. A hidden Markov model of the identified movements is obtained and used with the desired motion path in the Viterbi state estimation. The estimated state sequence is then used to generate a novel movement that is a version of the desired motion path, modulated to convey the target emotion. Various affective human movement corpora are used to evaluate and demonstrate the efficacy of the developed approaches for the automatic recognition and generation of affective expressions in movements.

Finally, the thesis assesses the human perception of affective movements and the impact of display embodiment and the observer's gender on the affective movement perception via user studies in which participants rate the expressivity of synthetically-generated and human-generated affective movements animated on anthropomorphic and non-anthropomorphic embodiments. The user studies show that the human perception of affective movements is mainly shaped by intended emotions, and that the display embodiment and the observer's gender can significantly impact the perception of affective movements.

Acknowledgements

First, my special gratitude goes to my advisors, Prof. Rob Gorbet and Prof. Dana Kulić. Thanks for the opportunity to do research in the emerging and exciting field of affective computing. Thanks for your continuous support and encouragements along the way, for reading all my draft papers and reports, and for your detailed feedbacks on them. You have always colored the black and white sketch of ideas in my head with your valuable inputs. Without your guidance, this work would have not been possible.

During my PhD, I was lucky to engage in different collaborations between the adaptive systems lab and industry partners, which has enriched my PhD experience. Thanks for these opportunities, Prof. Kulić! You have also been and are a great mentor and career advisor for which I will always be grateful.

I want to thank all the past and present members of the adaptive system lab for their warm company and the interesting research discussions in our group meetings, and the more interesting discussions about existential crisis of graduate students! I also want to thank them for being the first volunteers in the user-studies I conducted during my PhD.

The availability of rich movement data was pivotal to the success of my PhD research. I would like to thank Prof. Nadia Bianchi-Berthouze and Dr. Andrea Kleinsmith for sharing their affective movement data with me.

I would like to thank Prof. Sarah Jane Burton for all her help and insights to Laban movement analysis and for her patience with me through designing and collecting hand-arm movements and completing Laban annotation questionnaires. I also want to express my gratitude to Dr. Eric Kubica for his support and help during the initial phase of my PhD.

My deepest thanks goes to my parents and siblings for their continuous support, unconditional love, and warm encouragements across the globe. I also want to thank my parents-in-law for being geographically close and for all their support and encouragements.

Last and foremost, I wish to thank Khadige Abboud for all her love and warmth, her beautiful company, the afternoon runs and yoga exercises, and her exceptional espresso beverages all that made the pressure of work bearable.

Dedication

To my parents

Table of Contents

List of Tables xii

List of Figures xiv

Nomenclature xvi

1 Introduction 1

1.1 Objectives . . . 2

1.2 Contributions . . . 3

1.3 Thesis Organization . . . 6

2 Related Work 7

2.1 Affect: Definition, Expression, and Representation . . . 8

2.2 Movement Notation Systems . . . 10

2.2.1 Discussion . . . 13

2.3 Affective Movement Perception . . . 14

2.3.1 Full Body Movements . . . 15

2.3.2 Gait . . . 16

2.3.3 Upper Body Movements . . . 16

2.3.4 Gender-specific Differences in Affective Movement Perception . . . 17

2.3.5 Embodiment-specific Differences in Affective Movement Perception 18

2.3.6 Summary . . . 19
2.4 Automatic Affective Movement Recognition . . . 19
2.4.1 Full-body Movements . . . 20
2.4.2 Upper-body Movements . . . 21
2.4.3 Summary . . . 22
2.5 Automatic Affective Movement Generation . . . 23
2.5.1 Summary . . . 25

3 Affective Movement Recognition and Generation Based on Functional Movement Representation 27

3.1 Functional Affective Movement Recognition . . . 28
3.1.1 Basis Function Expansion . . . 29
3.1.2 Functional Dimensionality Reduction . . . 30
3.1.2.1 Functional Principal Component Analysis (FPCA) . . . 30
3.1.2.2 Functional FDA (FFDA) . . . 33
3.1.2.3 Functional Supervised PCA Based on HSIC (FSPCA) . . . 35
3.1.2.4 Functional Isomap (F-Isomap) . . . 36
3.1.3 Experiments . . . 38
3.1.4 Results . . . 40
3.1.5 Discussion . . . 41
3.2 Functional Affective Movement Generation . . . 46
3.2.1 Landmark Alignment and Length Normalization . . . 47
3.2.2 Prototypical Movement Generation . . . 48
3.2.3 Experiments . . . 48
3.2.4 Results . . . 53
3.2.4.1 The Effect of Generation Source . . . 55
3.2.4.2 The Effect of Intended Emotion . . . 56
3.2.5 Discussion . . . 56

3.3 The Effect of Embodiment and Gender on the Perception of Affective Movements . . . 57
3.3.1 Experiments . . . 57
3.3.2 Results . . . 60
3.3.2.1 The Effect of Intended Emotion . . . 60
3.3.2.2 The Effect of Embodiment . . . 62
3.3.2.3 The Effect of Affect Representation Model . . . 63
3.3.2.4 Discussion . . . 64
3.3.3 The Effect of Gender . . . 66
3.3.3.1 Discussion . . . 72
3.4 Summary . . . 72

4 Affective Movement Recognition Based on Generative and Discriminative Stochastic Dynamic Models 75
4.1 Hybrid Generative-Discriminative Affective Movement Modeling . . . 77
4.1.1 HMM-based Movement Modeling . . . 78
4.1.2 Fisher Score Representation . . . 80
4.1.3 Discriminative Embedding of Fisher Scores . . . 82
4.2 Experiments . . . 82
4.2.1 Datasets . . . 82
4.2.1.1 Full-body Affective Movement Dataset . . . 83
4.2.1.2 Hand-arm Affective Movement Dataset . . . 83
4.2.2 HMM Initialization and Model Selection . . . 86
4.2.3 Fisher Scores and SVM Classification . . . 88
4.2.4 Discriminative Lower-dimensional Embedding . . . 88
4.3 Results . . . 88
4.3.1 Full-body Dataset . . . 89
4.3.2 Hand-arm Dataset . . . 90

4.4 Discussion . . . 91
4.4.1 Interpersonal Affective Movement Recognition . . . 93
4.4.2 Comparison with other Recognition Studies . . . 94
4.4.3 Salient Affective Movement Features . . . 95
4.4.3.1 Full-body Movements . . . 96
4.4.3.2 Hand-arm Movements . . . 99
4.4.3.3 Discussion . . . 107
4.5 Summary . . . 108

5 Affective Movement Generation Using Laban Effort and Shape and Hidden Markov Models 110

5.1 Laban Effort and Shape Quantification . . . 112
5.1.1 Weight Effort . . . 112
5.1.2 Time Effort . . . 113
5.1.3 Space Effort . . . 114
5.1.4 Flow Effort . . . 114
5.1.5 Shape Directional . . . 115
5.1.6 Experiments . . . 115
5.1.6.1 CMA Annotation . . . 115
5.1.7 Results and Discussion . . . 117
5.2 Full-Body Movement Generation Methodology . . . 123
5.2.1 Laban Effort and Shape Components for Full-body Movements . . . 123
5.2.2 Regularized Multinomial Logistic Regression . . . 124
5.2.3 HMM-based Movement Modeling . . . 125
5.2.4 The Proposed Generation Approach . . . 126
5.3 Experimental Setup . . . 128
5.3.1 RMLR Parameter Selection . . . 130
5.3.2 HMM Initialization and Model Selection . . . 130

5.3.3 Affective Movement Validation . . . 131
5.3.3.1 Objective Validation . . . 131
5.3.3.2 Subjective Validation . . . 131
5.4 Results . . . 133
5.4.1 Objective Validation . . . 135
5.4.2 Subjective Validation . . . 135
5.5 Discussion . . . 139
5.6 Summary . . . 141

6 Conclusions **144**
6.1 Concluding Remarks and Future Work . . . 147

APPENDICES **151**

A Summary Tables for the Literature Review in Chapter 2 **152**
A.1 Affective Movement Perception . . . 152
A.2 Affective Movement Recognition . . . 154

References **158**

List of Tables

2.1 Laban Effort components adapted from [1] 13
2.2 Laban Shape components [2]. 14

3.1 Leave-One-Out cross validation training and testing recognition rates for functional DR techniques applied on the affective hand movements 41
3.2 Leave-One-Out cross validation training and testing recognition rates for functional DR techniques applied on the affective full-body movements . . 41
3.3 Null hypotheses in the generation source user study 54
3.4 ANOVA results in generation source user study 55
3.5 Descriptive keywords in the embodiment user study 59
3.6 Null hypotheses in the embodiment user study 61
3.7 ANOVA results for the embodiment user study 62
3.8 Confusion matrix showing percentage (%)* of emotion rating for different affective movements. 64
3.9 Correlation (%) between the arousal ratings and peak velocity (first column) and peak acceleration (second column). 65
3.10 Null hypotheses in the gender user study 67
3.11 ANOVA results for the gender user study (Male participants) 68
3.12 ANOVA results for the gender user study (Female participants) 68
3.13 Confusion matrix for the mender user study 71

4.1 Hand-arm motion paths* 86

4.2 Recognition rates for the full-body data using hybrid generative-discriminative modeling . . . 89
4.3 Confusion matrix (%) for the 10-fold cross-validated 14NN classification in the 3D sPCA subspace for the full-body dataset . . . 89
4.4 Recognition rates for the hand-arm data using hybrid generative-discriminative modeling . . . 91
4.5 Confusion matrix (%) for the 10-fold cross-validated 2NN classification in 6D sPCA subspace for the hand-arm dataset . . . 91
4.6 Affective recognition rates for the full-body dataset . . . 95
4.7 Highly weighted entries of the discriminative class-specific CFS's in the sPCA transformations . . . 103

5.1 Correlations (%) between CMA-annotated and quantified Laban effort components along with their p-values. "*" indicates a significant correlation. . . . 119
5.2 Discriminative Laban components for each pair of emotions . . . 134
5.3 Confusion matrix (%) for the automatic recognition of the generated movements . . . 135

A.1 A selective list of affective movement perception studies. . . . 152
A.2 Recognition of affective expressions from body movements. . . . 154

List of Figures

3.1 Screen shots of an animated hand movement. 39

3.2 Affective hand movement embeddings in the 2D functional subspaces . . . 40

3.3 Affective full-body movement embedding for a) training data, b) testing data in the resulting 2D subspaces. 42

3.4 Perturbation plots from functional dimensionality reduction 45

3.5 Hylozoic series [3] 49

3.6 An example of landmark alignment and length normalization 50

3.7 FPCA 2D embedding of affective human hand movements. 51

3.8 The human-like hand model used to display the generated and original movements. 52

3.9 The schematic representation of the cicumplex model used in the questionnaire. 53

3.10 Average ratings in the generation source user study 54

3.11 Display embodiments in the embodiment user study 58

3.12 Average ratings in the embodiment user study 61

3.13 Average ratings in the gender user study 70

4.1 Schematic of the proposed hybrid generative-discriminative affective movement modeling approach 79

4.2 Pseudocode for the validation procedure. 84

4.3 Movement exemplars from the full-body dataset 85

4.4 Marker set used for collecting hand-arm movements. 87

4.5 sPCA lower-dimensional embedding of the affective full-body movements. . 90
4.6 Heat maps showing the maximum weights for Fisher scores derived with respect to the covariances of the class-specific HMMs in the sPCA dimensions for the full-body dataset. 97
4.7 Salient affective full-body movement features 98
4.8 Heat maps showing the maximum weights for Fisher scores derived with respect to the Gaussian mixture covariance matrices of the class-specific HMMs in the sPCA dimensions for the hand-arm dataset. 101
4.9 Salient affective hand-arm movement features 104

5.1 Average quantified Weight and Time for the affective hand-arm movements 120
5.2 Quantified Shape Directional for different movements. The dashed line illustrates a suitable threshold value. 122
5.3 Schematic of the proposed generation approach. 127
5.4 Pseudo-code for the proposed generation approach. 129
5.5 Average participants' ratings of synthetically-generated affective movements 136
5.6 The confusion matrix of participants' ratings of the synthetically-generated affective movements 137
5.7 A heat map of pair-wise differences between participants' ratings of target emotions and other emotions 138
5.8 An illustration of velocity changes for a sadness to anger conversion 141

Nomenclature

In this thesis, Greek and Latin letters are used for parameters and data, respectively. Furthermore, lower and uppercase letters are used to denote scalars, lowercase bold-face are used to denote vectors, and uppercase bold-face are used to denote matrices. Time-series observations and functions are also denoted by lowercase letters, however the accompanied text makes it clear when a lowercase letter refers to a scalar, a time-series observation, or a function. For instance, $x(t)$ and x denote a function and a time-series observation, respectively. x_t denotes a discrete time-series observation sampled at time t.

α^k	The mass coefficient of the k^{th} body part used in quantifying the Laban Effort components
$\bar{x}(t)$	The point-wise average of functions across n replications
$\ddot{x}_{t_i}$	The second derivative of the motion trajectory x at time t_i
$\dot{x}_{t_i}$	The first derivative of the motion trajectory x at time t_i
$\gamma_t(i)$	The probability that an HMM $\mathbf{\Lambda}$ is in state i (S_i) at time t given a movement x: $P(q_t = S_i \mid x, \mathbf{\Lambda})$
$\gamma_t(i,j)$	The probability of the j^{th} multivariate Gaussian distribution associated with state i of an HMM $\mathbf{\Lambda}$ being active at time t for a given movement x

$\int x(t)y(t)dt$	The inner-product of two functions $x(t)$ and $y(t)$.
κ	The curvature of a motion path
λ	A Lagrange multiplier
$\|.\|_p$	The norm p
$\mathbf{\Lambda}$	An HMM
$\mathbf{\Lambda}_k$	A class-specific HMM encoding movements from the k^{th} class
$\mu_{\mathbf{i}}$	The mean of the Gaussian output of the i^{th} state of HMM_{NN}
$\mathbf{A}$	The state transition matrix of an HMM
$\mathbf{a}_{t_i}^k$	The Cartesian acceleration of the k^{th} body part at time t_i
$\mathbf{c}_{ji}$	A vector of BFE coefficients for a time-series observation x_i
$\mathbf{e}$	A column vector of $\mathbf{1}$'s
$\mathbf{H}$	A centering matrix
$\mathbf{I}_n$	An identity matrix of size n
$\mathbf{L}$	The set of all Laban Effort and Shape components
$\mathbf{L}'_k$	The set of affectively discriminative Laban components for k^{th} class
$\mathbf{L}_k$	The set of affectively non-discriminative Laban components for k^{th} class
$\mathbf{L}_{e_t}$	The set of affectively non-discriminative Laban components for an emotion e_t
$\mathbf{N}$	A matrix representing the pair-wise geodesic neighbourhood relationship in Isomap

$\mathbf{O}$	The modulated version of a desired motion path conveying a target emotion
$\mathbf{p}_{t_i}^k$	The Cartesian position of the k^{th} body part at time t_i
$\mathbf{x}_j^{\mathbf{L}}$	The Laban Effort and Shape representation of the j^{th} movement from the set $\mathcal{D}$
$\mathcal{D}$	A dataset of labeled training movements
$\mathcal{D}^{\mathbf{L}}$	The Laban Effort and Shape representation of the set $\mathcal{D}$
$\mathcal{D}^{\mathbf{L}_{e_t}}$	The representation of a dataset $\mathcal{D}$ in terms of affectively non-discriminative Laban components $\mathbf{L}_{e_t}$
$\mathcal{E}$	A set of affective labels
$\mathcal{N}$	A Gaussian distribution
$\mathfrak{B}$	The distribution of observations for the states of an HMM
$\mathfrak{G}$	A neighbourhood graph in Isomap
$\otimes$	The cross-covariance operator
Φ	A functional vector of length G whose elements are the basis functions $\varphi_g(t)$ [4]
π_i	The prior or initial state probability of state i
$\pi_{i,k}$	The initial state probability of state i of an HMM $\mathbf{\Lambda}_k$
$\mathbf{v}_{t_i}^k$	The Cartesian velocity of the k^{th} body part at time t_i
θ	The parameters of the posterior function (inverse logit) in multinomial logistic regression
$\top$	The transpose operator

$\varphi_g(t)$	g^{th} basis function used in BFE of time-series observations
$\|\mathbf{t}\|$	The length of a vector of discrete sampling points $\mathbf{t}$
$\zeta_t(i,l)$	The probability of an HMM $\mathbf{\Lambda}$ being in states i and l at times t and $t+1$ for a given movement x: $P(q_t = S_i, q_{t+1} = S_l \mid x, \mathbf{\Lambda})$
a_{il}	The transition probability from state i to state l of an HMM: $P(q_{t+1} = S_l \mid q_t = S_i)$
d_{spline}	The degree of a B-spline
$E^k_{t_i}$	The kinetic energy of the k^{th} body part at time t_i
e_t	A target emotion
$f_{ji,t}$	The value of the j^{th} time-series feature of movement x_i at time t
G	The total number of basis functions used in BFE
K	The number of classes to which datapoints in a dataset belong
L	A sequence of two or more connected points in the neighbourhood graph in Isomap
M	Number of mixtures of Gaussian per state in an HMM
m	The number of time-series features (joint angles or Cartesian positions) describing a movement observation
N	The number of HMM states
n	The number of movement samples in a dataset $\mathcal{D}$
n_d	The number of copies of the desired motion path x_d encoded in HMM_{NN}

n_k	The number of movement samples in the k^{th} class
n_L	The number of Laban Effort and Shape components
O_t	A test movement observation
q_t	The actual HMM state at time t
S	The most likely HMM_{NN}'s state sequence for a desired motion path x_d
T_d	The length of a desired motion path x_d
T_i	The length of a sampled affective movement observation x_i
$v_{t_i}^k$	The speed of the k^{th} body part at time t_i
$w(t)$	A functional transformation produced by a functional dimensionality reduction technique
x_d	A desired motion path
$x_d^{\mathbf{L}_{e_t}}$	The representation of a desired motion path x_d in terms of affectively non-discriminative Laban components for a target emotion e_t
x_i	A sampled affective movement observation of length T_i (the movement is defined by T_i sequential observations, where each observation represents Cartesian positions of body joints at a particular time, $t \in [1, T_i]$)
1NN	one-nearest neighbour
2D	Two-dimensional
3D	Three-dimensional

kNN	k-nearest neighbour
ANN	Artificial neural network
ANOVA	Analysis of variance
BAP	Body action and posture
BFE	Basis function expansion
CFS	Covariance Fisher score
CMA	Certified motion analyst
CRBM	Conditional restricted Boltzmann machine
DR	Dimensionality reduction
F-Isomap	Functional Isomap
FACS	Face action coding system
FDA	Fisher discriminant analysis
FFDA	Functional Fisher discriminant analysis
FPCA	Functional principal component analysis
FSPCA	Functional supervised principal component analysis
GOC	Goodness of clustering metric
GP	Gaussian process
GRBF	Gaussian radial basis function
HCI	Human-computer interaction
HMM	Hidden Markov model

HMM_{NN}	The HMM encoding of the nearest neighbours of a desired motion path (x_d)
HRI	Human-robot interaction
HSIC	Hilbert Schmidt independence criterion
LASSO	Least absolute shrinkage and selection operator
LDA	Linear discriminant analysis
LMA	Laban movement analysis
LOOCV	Leave-one-out cross validation
LOSOCV	Leave-one-subject-out cross validation
LPF	Low pass filter
MDS	Multidimensional scaling
ML	Maximum likelihood
MLP	Multi-layer perceptron
MSE	Mean square error
NN	Nearest neighbours
$\text{NN}(x_d)$	the nearest neighbours of a desired motion path x_d
PAD	Pleasure-arousal-dominance dimensional model of emotion
PC	Principal component
PCA	Principal component analysis
RKHS	Reproducing kernel Hilbert space

RMLR	Regularized multinomial logistic regression
sPCA	Supervised principal component analysis
SVD	Singular value decomposition
SVM	Support vector machine
T-pose	A body posture in which the arms are raised to shoulder level and extended to the sides
Tr	A matrix trace

Chapter 1

Introduction

Research in psychology and cognitive science indicates that human affective states regulate cognitive processes such as decision making, creativity, attention, action planning, and social interaction [5]. Humans are also adept at perceiving the affective states of others, as manifested through different modalities including body movements [6]. Studies in psychology suggest that body movements are an important non-verbal communication medium and might carry information about human affective state not observed from facial or verbal cues [7, 8]. Body movements are a primary communication medium especially when affective expressions are estimated from a distance, verbal communication is not available, facial expressions are disguised as a result of social editing, or an emotion is better expressed via body movements (e.g., nervousness). Humans associate different body movements and postures with distinct affective states and are able to identify the feelings and comprehend the expression encoded in a displayed movement even when the demonstrator intends to conceal the expression [8].

Although human affect has been studied for over a century and a half by psychologists [9, 10, 11], it was only in the 1990s that research into computational analysis of affect attracted the interest of engineers and computer scientists [12]. As a result, the last two decades have seen a growing number of interactive machines such as social robots [13], kinetic sculptures [3, 14], and animated agents [15]. If we are to realize the full potential of these machines, it is important to equip them with affective communication capabilities in order to enable an engaging, entertaining, and empathic human-machine interaction. To this end, computational models are emerging that aim to accurately recognize users' expressions and generate appropriate responses through different modalities including movements.

The majority of the existing works on computational affective movement analysis are tailored to a specific set of motion paths and a limited set of emotions, and often are not readily adaptable to other motion paths and emotions. Furthermore, it is not yet known what movement features are salient to affective expression, and the selection of these features for automatic affective movement recognition and generation is usually done in an ad-hoc manner [8]. The computational affective movement analysis literature also mainly focuses on full-body movements of anthropomorphic kinematic embodiments, and less attention has been directed toward affective expression via movements of isolated body parts and non-anthropomorphic embodiments. In addition, perceptual validation of the computational models has been lacking in most cases, despite the subjective nature of affective movement perception [8].

This thesis is motivated by the development of a series of interactive sculptures called the *Hylozoic series* [3], and develops a new framework for computational analysis of affective movements consisting of: 1) a systematic identification of movement features salient to affective expressions, 2) automatic recognition of affective expressions from movements, 3) automatic modulation of a desired motion path to convey a target affective expression, 4) experimental validation of the human perception of affective movements and the impact of display embodiments and the observer's gender on the perception of affective movements. The computational analysis in this thesis is motivated by the hypothesis that there exist low-dimensional spaces agnostic to movement kinematics where automatic and accurate affective movement recognition and generation can be performed. The thesis proposes a systematic approach for identifying these discriminative and generative subspaces via statistical learning and stochastic modeling of affective movement observations.

1.1 Objectives

The following objectives and research questions are addressed in this thesis:

1. Develop computational models for affective movement recognition.
    a) What are the discriminative features of affective movements?
    b) What is a suitable recognition model to discriminate between different affective movements?
    c) How can the identified discriminative features and the recognition method be validated?

2. Develop a computational model for modulating a desired motion path to convey a target emotion.

    a) What is a suitable representation of movements that captures both kinematic and affective characteristics?

    b) How can the representation be quantified and used for generating novel motion paths?

    c) How can the expressivity of the generated movements be verified?

3. Assess the human perception of affective movements displayed on different kinematic embodiments via perceptual user studies.

    a) How well are the intended emotions in the movements perceived and what are the emotions that are easily communicated through body movements?

    b) What influence do the kinematic embodiment and its physical appearance have on the perception of affective movements?

    c) What is the impact of the observer gender on the perception of affective movements?

1.2 Contributions

The thesis develops a systematic framework for recognition and generation of affective movements displayed on anthropomorphic and non-anthropomorphic structures. The thesis makes the following contributions:

A. *Affect recognition and generation based on functional movement representation and dimensionality reduction* (**Chapter:** 3).

   Affective movements are captured in terms of time-series signals describing the motion of many different body joints, resulting in high-dimensional datasets. Furthermore, there are significant temporal variations caused by phase and/or length differences even between kinematically and/or affectively similar movements. This thesis hypothesizes that there exists a low-dimensional space where movements with different affective expressions are best separable, and aims to identify such a low-dimensional space using dimensionality reduction techniques. However, to enable

application of dimensionality reduction techniques to affective movements, a fixed-length representation of the movements is first required. To this end, the thesis proposes a systematic approach for converting variable-length movement observations into fixed-length observations defined as a weighted linear combination of a number of basis functions (i.e., basis function expansion). Subsequently, the thesis adapts various dimensionality reduction techniques to enable application to the functional movement representation and exploits the resulting reduction techniques to identify movement features along which affective movements are best separable. The identified features span low-dimensional spaces where affective movement recognition is then performed. In particular, Fisher discriminant analysis (FDA), supervised principal component analysis (sPCA), and Isomap are adapted and their performance along with that of the functional principal component analysis (FPCA) [4] in producing a discriminative mapping is evaluated using cross-validation tests and nearest neighbour classification. Therefore, the proposed functional feature transformation and dimensionality reduction techniques enable application of discriminative learning techniques to variable-length time-series observations. In addition, the thesis proposes an approach for generating prototypical affective movements that uses the resulting inverse FPCA transformations to reconstructs high-dimensional joint trajectories for centroids of affect-specific clusters of movements in the resulting FPCA space. The expressivity of the generated prototypical movements is verified in a perceptual user study.

B. *Affect recognition based on hybrid generative-discriminative movement modeling* (**Chapter:** 4).

An affective expression can be communicated through a number of kinematically different movements. The functional movement analysis (Contribution A.) is limited to within-class affective movements that are also similar in their kinematic motion path. Furthermore, the functional analysis requires movement landmarks to be identified and then temporally aligned, which can be a very tedious and expensive task for large datasets. To this end, the thesis develops an alternative fixed-length movement representation based on a stochastic feature transformation using hidden Markov models (HMM)s and represents each movement in terms of the gradients of the likelihood of the resulting models with respect to the model parameters (Fisher score representation). The resulting representation is used in a support vector machine classification whose performance in discriminating between different affective movements is evaluated using cross-validation tests. Furthermore, the thesis presents a new systematic approach for automatic identification of movement features most salient to affective expressions. The proposed approach maps Fisher scores to a lower-dimensional

space where dependencies between movements and their affective labels are maximized. This mapping is done by maximizing Hilbert Schmidt independence criterion between Fisher scores and their affective labels. The resulting mapping highlights a minimal set of features most salient to affective expressions. The identified features are represented in terms of parameters of the learned models and often lack any semantic meaning when assessed against the human perception. The thesis presents a high-level interpretation of these features in terms of movement kinematics and dynamics and sheds light on movement cues salient to different affective expressions. Therefore, the Fisher score representation supersedes the functional representation approach for affective movement recognition as it enables a more generic affective movement recognition regardless of interpersonal, kinematic, and temporal variabilities inherent in body movements.

C. *Affective movement generation based on Laban movement analysis and generative movement modeling* (**Chapter:** 5).

The application of the Laban movement analysis is proposed for computational generation of affective movements. Quantifications for the Laban Effort and Shape components are developed and validated in collaboration with a certified motion analyst (CMA). The quantified Laban components provide a compact description of movements that captures both expressive and kinematic characteristics. The quantified components are validated against those annotated by the CMA and exploited in the development of a new automatic approach for affective movement generation. The developed generation approach makes use of two movement abstractions: 1) Laban Effort and Shape components and 2) hidden Markov modeling, and modifies an arbitrary motion path to display a target affective expression. This approach uses available affective movement datasets, and for a desired motion path and a target emotion, searches the dataset for movements that are kinematically-similar to the desired motion path and belong to the target emotion class. An HMM for the identified movements is learned, and the most likely state sequence of the learned HMM for the desired motion path is obtained using the Viterbi algorithm. The resulting state sequence is used to generate a modulated version of the desired motion path encoding the target emotion. The expressivity of the synthetically-generated movements is evaluated objectively using a validated automatic recognition model from contribution B. and subjectively with a user study in which participants rated the perceived affective content observed in the movements.

D. *Affective movement perception* (**Chapters:** 3, 5).

To better understand the human perception of affective movements, perceptual user

studies have been conducted in which participants rate perceived affective expressions from videos of animated affective movements. These user studies address the third objective of the thesis and offer insights into the perception of different affective expressions from movements and the impact of display embodiment and an observer's gender on the perception of affective movements. The user studies are also used to subjectively verify the expressivity of the synthetically-generated movements in Contributions A. and C.

1.3 Thesis Organization

This thesis is organized as follows. Chapter 2 provides relevant background information on the definition, expression, and representation of affect and movement notation systems, followed by a literature review on affective movement perception, automatic recognition, and generation. Chapter 3 presents an approach for recognizing and generating affective movements based on a functional representation of joint trajectories mapped to lower-dimensional subspaces derived using dimensionality reduction techniques. Chapter 3 also presents a user study evaluating the human perception of synthetically-generated affective hand movements displayed on anthropomorphic and non-anthropomorphic embodiments. An affective movement recognition approach based on hybrid generative-discriminative modelling of affective movements is presented in Chapter 4. Chapter 5 describes a quantification for the Laban movement notation, and presents an affective movement generation approach based on the quantified Laban notation and a stochastic movement abstraction. Chapter 6 presents the conclusions and future directions for the work on computational analysis of affective movements.

Chapter 2

Related Work

■ *An earlier version of the text in this chapter has appeared in the following publications*: [8, 16, 17].

Since the formation of the affective computing field by pioneers such as Rosalind Picard [12], the research on computational analysis of affective phenomena has been of growing interest among engineers and computer scientists. As facial expressions and verbal communication dominate during face-to-face interaction, they are the modalities that have been predominantly studied in affective computing [18, 19]. However, there is evidence that body movements also communicate affect (see Section 2.3), might convey expressions not detectable in facial or verbal modalities, and are less prone to social editing. Furthermore, kinematic movements are an important modality for non-verbal and non-facial machines, and can be used for affective and empathic communication with their users, which in turn helps to create a more engaging human-machine interaction [20].

In this chapter, relevant background information on the definition, expression, and representation of affect and movement notation systems is presented, followed by an overview of the affective movement perception literature to motivate and support the computational analysis of affective movements. Finally, a literature review on the automatic recognition and generation of affective movement is presented.

2.1 Affect: Definition, Expression, and Representation

Affective phenomena (broadly termed affect [21]) include emotions, feelings, moods, attitudes, temperament, affective dispositions, and interpersonal stances [11, 22, 23, 24]. A categorization of these terms based on event focus, intrinsic and transactional appraisal, synchronization, rapidity of change, behavioural impact, intensity, and duration is provided in [23, 25, 24] (e.g., emotions change rapidly, are short-term and intense, whereas mood covers a longer time-span and changes more slowly).

Among the affective phenomena, emotions are those most widely studied in human-machine interaction [22, p.4]. Scherer defines *emotion* as "an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism" [23]. The five subsystems are: cognitive (for evaluation), physiological (for regulation), motivational (for preparation of an action), subjective feeling, and motor expression (for communication) [23]. Theories of emotion expression are often based on facial expressions; fewer studies have been conducted to investigate the extent to which existing theories predict and explain bodily expressions of emotion [26, 27, 28, 29]. Effective emotional communication requires both the ability to encode and send one's own emotion in an appropriate and comprehensible manner, and the ability to receive and decode the emotions of others [22]. For virtual agents and robots, these two skills refer to automatic generation and recognition of emotional expressions, respectively. These two modules can be integrated into higher-level computational models which cover the generation of appropriate emotions, and relations to cognitive and motivational subsystems [30].

The terms affect and emotion have been used inconsistently, often interchangeably, in the field of human-machine interaction (which can be subdivided to human-computer interaction (HCI) and human-robot interaction (HRI)), and both terms can be found in studies on automatic recognition and generation of affective movements. In this thesis, the term *affective expression* is used to refer to the affective phenomena communicated via movements, and movements that are communicating affective expressions are referred to as *affective movements*. In the review of related work in this chapter, the term emotion is used when studies explicitly address emotions.

Humans can control their affective expressions to a certain extent (e.g., using display rules to achieve a social goal [31, p.72]). This gives rise to the possibility of a difference between the internal experience and external expression. This possible discrepancy

has been widely studied and affective expressions are divided into either spontaneous or strategic ones [32, 31, 6]. Spontaneous expressions are involuntary and their content is non-propositional, whereas strategic or symbolic expressions are goal-oriented and their content is propositional [32, 31]. It is important to note that an observed expression may not necessarily be the observable manifestation of an internal state, but rather, displayed to achieve a social goal. The majority of the studies in computational analysis of affective expressions do not consider differences between the internal and observed expressions and assume a single affective attribute for each movement stimulus; the same approach is followed in this thesis.

The automatic recognition and generation of affective movements are relatively unexplored in comparison to studies on other modalities such as facial expression and physiology [7]. Body movements are advantageous over other modalities 1) for perception from a distance, because bodily expressions are more easily visible from a distance than subtle changes in the face [33, 34, 35], 2) for analyzing types of expressions that are less susceptible to social editing, because people are often less aware of their body than their facial expressions [36, 37], and 3) for conveying affective states which are more easily expressed through movement (e.g., intense positive or negative emotions [38]).

Affective expressions can be represented using a set of distinct categorical labels, or a dimensional model. *Categorical labels* describe affective expressions based on their linguistic use in daily life. Different sets of categorical labels can be chosen depending on the study. Most frequently, happiness, sadness, fear, and anger are included, a subset of the basic emotions [39]. A *basic emotion* is defined by a set of neural correlates in the brain, a specific set of bodily expressions, and a motivational component for action tendencies [39]. A popular set of basic emotions is proposed by Paul Ekman and contains anger, happiness, sadness, surprise, disgust, and fear [40, 41].

Dimensional models represent an affective state as a point on a continuum spanned by a set of independent dimensions. A popular example is the two-dimensional circumplex model, where similar affective expressions are arranged adjacent to one another on a circle, and dissimilar affective expressions are arranged to lie opposite to each other on the circle [42]. The dimensions of circumplex space are *arousal* and *valence*. Another common model applied in affective computing is the PAD-model, with the dimensions *pleasure* (or *valence*), *arousal*, and *dominance* [43]. Arousal corresponds to the level of activation, mental alertness, and physical activity. Dominance represents the amount of control over others and the surroundings versus feeling controlled by external circumstances, and valence ranges from negative (unpleasant) to positive (pleasant) expressions. Categorical labels can be mapped to the continuous PAD space (e.g., happiness, amusement, and contentment are related to high pleasure, whereas anger, fear, and sadness are related to low

pleasure [44, 45, 46, 47]). A dimensional representation may relate more to the underlying physiological changes [27] and Barrett suggests that categorical labels are a result of subjective categorization of emotions using conceptual knowledge (e.g., similar to color perception [48]).

In this thesis both categorical and dimensional emotion models are used in the conducted perceptual studies, and in the computational analysis of affective movements.

2.2 Movement Notation Systems

The work on affect and its bodily manifestation in psychology and the dance community provides valuable insights that can advance the research in affective computing. In particular, movement notation systems provide a rich tool to capture both kinematic characteristics and expressive qualities of movements in a more compact representation compared to high-dimensional joint trajectories. As FACS[1] pushed the research on computational analysis of facial expressions forward, movement notation systems could also advance the computational analysis of affective movements by providing an objective and systematic movement representation.

Burgoon *et al.* divide movement notation systems into functional and structural ones [6]. *Functional systems* describe the communicative function of a displayed movement using verbal labels. The Ekman and Friesen formulation of kinesic[2] behaviours into five categories (emblems, illustrators, affective displays, regulators, and manipulators) is an example of a functional notation system [51].

- *Emblems*: culturally shaped body movements (e.g., thumb up).
- *Illustrators*: actions accompanied with speech to augment the verbalized message (e.g., pointing at an example).
- *Affect displays*: distinctive bodily expression of different emotion categories.
- *Regulators*: body movements maintaining the flow of verbal conversation (e.g., head nods).

[1] Face action coding system (FACS) provides a comprehensive set of action units that can be used to objectively describe any type of facial movement [49].

[2] Birdwhistell referred to non-verbal communicative body movements as *kinesics* [50].

- *Adaptors*: body movements learned and practiced to satisfy personal needs (self-adaptors; occur frequently in private, e.g., head scratching), body language practiced during interpersonal contact such as giving, taking, attacking or being attacked (alter-adaptors; e.g., crossed arm on the chest as a protective posture), movements learned to perform instrumental activities (e.g., driving a car).

Affective movements are classified under the affect display category of the kinesics behaviours [51], however, other categories of body movements might be involved in the expression of emotion (e.g., self-adaptors can be perceived as a sign of discomfort or anxiety) as these categories are intentionally allowed to overlap.

Structural systems are primarily concerned with the question of what a movement looks like and provide sufficient structural and expressive details for movement replication [6], and are therefore more appropriate for computational affective movement analysis than functional systems. In the following, a review of structural systems that have been applied for computational analysis of affective movements is presented.

Inspired by linguistic notation systems, Birdwhistell proposed a structural movement notation system that parallels phonemic transcription in linguistics [50]. Birdwhistell introduced *kine* (the smallest perceivable body motion, e.g., raising eye brows), *kineme* (a group of movements with a same social meaning, e.g., one nod, two nods, three nods), and *kinemorphs* (a combination of kinemes forming a gesture) followed by *kinemorphic* classes and complex kinemorphic constructs, which are analogous to sentences and paragraphs in linguistics. Birdwhistell used motion qualifiers and action modifiers that define: 1) the degree of muscular tension involved in executing a movement, 2) the duration of the movement, and 3) the range of the movement. The *kinegraph* is introduced as a tool for notating individual kines and their direction at different body sections. The Birdwhistell system is capable of micro analysis of body movements as its kines capture barely perceivable body motion ranging from 1/50 seconds to 3 seconds in duration [50]. Birdwhistell emphasizes the importance of context for inferring the meaning of an observed movement. A Birdwhistell-inspired annotation was used to extract semantic areas in emoticons[3] for automatic recognition of their expressed emotions [52].

Delsarte [53] classifies emotion as a form of expression in gestures and divides the body into zones within which mental, moral, and vital components are defined. He identifies nine laws that contribute to the meaning of a movement: altitude, force, motion (expansion, contraction), sequence, direction, form, velocity, reaction, and extension. The Delsarte system has been used for automatic generation of affective full-body [54] as well as hand

[3]An emoticon is a string of symbols used in text communication to express users' emotions [52].

and arm [55] movements. In [55], participants' perception of a set of Delsarte-generated hand and arm movements displayed on an animated agent was shown to be consistent with the Delsarte model prediction.

Recently, Dael *et al.* proposed BAP (body action and posture), a structural notation system for a systematic description of temporal and spatial characteristics of bodily expression of emotions. Analogous to FACS, BAP introduces 141 behavioural categories for coding action, posture, and function of an observed body movement [26]. BAP segments body movements into localized units in time and describes them at three levels: anatomical (articulation of different body parts), form (direction and orientation of the movements), and functional (behavioural classes categorized in kinesics emblems, illustrators, and manipulators). BAP anatomical and form variables are Boolean (0 for absence and 1 for presence), while functional variables are ordinal (1 for very subtle and 5 for very pronounced). BAP was developed using the GEMEP corpus of emotion portrayals. Since the movements are captured from the knees upwards in GEMEP, the current version of BAP does not code whole body postures and leg movements. BAP also does not code dynamic movement characteristics such as velocity, acceleration, and energy. BAP reliability has been demonstrated by assessing intercoder agreement (two coders) on occurrence, temporal precision, and segmentation of posture and action units [26]. There is only a single report on the application of BAP for computational analysis of affective movements at the time of writing this thesis, in which BAP behavioural categories are employed for recognition of 12 affective states encoded in 120 movements demonstrated by 10 actors [56]. Recently, AutoBAP has been proposed for automatic annotation of posture and action units based on BAP anatomical and form (and not functional) coding guidelines [57].

The Laban system is a prominent example of a structural movement notation system, which was developed for writing and analyzing both the structure and expressivity of movements in dance choreography [58, 59]. The Laban system has four major components: Body, Effort, Shape, and Space. Body indicates the involved body parts, and the sequence of their involvement in the movement. Space defines where the movement is happening, and the motion directions of the body and body parts. Shape characterizes the bodily form, and its changes in space. Effort describes the inner attitude toward the use of energy. In other words, Body and Space describe *What one does through a movement*, whereas Effort and Shape describe *How the movement is performed* [1].

Effort and Shape are the most relevant Laban major components for the study of affective movements. Bartenieff presents Effort and Shape as a complete system for the objective study of movements, from behavioural and expressive perspectives [1]. Effort has four bipolar semantic components: Weight, Time, Space, Flow (see Table 2.1), and Shape has three components: Shape Flow, Directional, and Shaping/Carving (see Table

Table 2.1: Laban Effort components adapted from [1]

	Extremes	Example
Space: Attention to surroundings	Direct	Pointing to a particular spot
	Indirect	Waving away bugs
Weight: Sense of the impact of one's movement	Light	Dabbing paint on a canvas
	Strong	Punching
Time: Sense of urgency	Sustained	Stroking a pet
	Sudden	Swatting a fly
Flow: Attitude toward bodily tension and control	Free	Waving wildly
	Bound	Carefully carrying a cup of hot liquid

2.2). Computational Laban analysis has been carried out for movement recognition (e.g., [60]), and generation (e.g., [61]), and to relate Laban components to low-level movement features e.g., velocity and acceleration [62, 63], and different affective expressions [64].

2.2.1 Discussion

For automatic affective movement recognition and generation, there is a need for consistent and quantitative description of movements, leading to a preference for structural notation systems that provide a fixed number of distinct movement descriptors such as the Laban system. However, despite their proven suitability for movement coding, except for BAP, the structural notation systems do not explicitly provide quantitative measures, which is perhaps the main barrier to their application in computational movement analysis. In addition, the extensive attention to microanalysis (e.g., Birdwhistell system [50]), and the need for special training (e.g., Laban system) hamper their adoption in affective computing. Furthermore, some structural notation systems require the coder to infer the meaning or function of an observed movement (e.g., Delsarte [53]). The correspondence between movements and affective expressions is not transcultural and transcontextual, and there may be idiosyncratic, gender-specific, or age-specific differences in affective movements [65]. Such movement/affective expression discrepancies result in a drawback for the notation systems that code the meaning or function of an observed movement. Moreover, the amount and intensity of an affective expression is important for computational analysis; hence, the preference for structural notation systems that code such information (e.g., the

Table 2.2: Laban Shape components [2].

	Elements	Example
Shape Flow: is self-referential and defines readjustments of the whole body for internal physical comfort.	Growing	Self-to-self communication, stretching to yawn
	Shrinking	Exhaling with a sigh
Directional: is goal-oriented and defines the pathway to connect or bridge to a person, object, or location in space.	Arc-like	Swinging the arm forward to shake hands
	Spoke-like	Pressing a button
Shaping/Carving: is process-oriented and is the three dimensional "sculpting" of body oriented to creating or experiencing volume in interaction with the environment.	Molding, Contouring, or Accommodating	Cradling a baby

Laban system).

Among the above structural systems, the Laban system is the most popular notation system [8] and enables 1) consistent movement representation in terms of a fixed number of distinct descriptors and 2) studying encoded emotions and their intensity in addition to kinematic characteristics of the movements. However, the Laban descriptors are qualitative and to enable their application in computational analysis, they need to be first quantified.

In collaboration with a certified motion analyst (CMA), this thesis develops and validates a quantification for Laban Effort and Shape descriptors, which is then used to design an approach for automatic affective movement generation (Chapter 5).

2.3 Affective Movement Perception

To pave the path toward fully interactive machines capable of affective communication with their users, it is important to develop a sufficient understanding of how humans perceive and express affect via body movements. Perceptual user studies can help to develop a better understanding of affective movement perception, and offer insights into whether movements

of full-body or isolated body parts are capable of communicating affect and the movement kinematics and dynamics salient to each affective expression. Many researchers have tested the human perception of affective expressions through dance and body movements, using a variety of stimuli including both full-light and point-light videos [66, 67, 68, 69], and reported above-chance recognition rates.

This section presents an overview of the affective movement perception literature to motivate the computational analysis for affective movement recognition and generation sought in this thesis.

2.3.1 Full Body Movements

Boone and Cunningham tested the ability of children (4 to 8 years) and adults (17 to 22 years) to recognize affect encoded in expressive dance movements [66]. The recognition rate was above chance for both children and adults. They report the following six movement cues used by participants for the perception of affective dance movements: changes in tempo (anger), directional changes in face and torso (anger), frequency of arms up (happiness), duration of arms away from torso (happiness), muscle tension (fear), the duration of time leaning forward (sadness). Furthermore, sadness was recognized most accurately. In a similar study, Camurri and colleagues [67] tested human movement perception in four emotional categories (anger, fear, grief, and happiness) conveyed through the same dance movement. They suggest that the duration of the movement, quantity of the movement (the amount of observed movement relative to the velocity and movement energy represented as a rough approximation of the physical momentum), and contraction index (measured as the amount of body contraction/expansion) play key roles in the perception of affect from the dance movement.

Recognition rates of 63% (for point-light display) and 88% (for full-light display) were reported for dance movements conveying fear, anger, grief, joy, surprise, and disgust [68]. Using Laban notation, Weight and Time Effort characteristics of dance movements were suggested as the key discriminative features for the perception of encoded emotions (happiness, sadness, anger, fear) [70]. Four basic emotions of happiness, sadness, anger, and fear were recognized from body postures depicted in a set of gray-scale photographs, with fearful body expressions being the hardest to recognize [71]. Coulson investigated the perception of communicative postures of a mannequin figure conveying anger, disgust, fear, happiness, sadness, and surprise [72], and reports that the static body postures constitute a reliable medium for communicating emotions, especially those of anger, sadness, and happiness. He also argues that the less well recognized emotions (fear, disgust, and surprise)

are emergency responses; hence, their perception depends on the perceived movement velocity and form, which is not depicted in the static postures. The perception of affective expressions (anger, sadness, happiness, fear, disgust) encoded in static and dynamic full-body communicative movements demonstrated using full-light and point-light displays was evaluated in a user study and an above chance recognition was observed [69]. The perception accuracy was higher for exaggerated emotions and dynamic full-light displays, with anger and disgust being the least recognized expressions.

2.3.2 Gait

Studies from psychology indicate that affective states can be expressed during walking and recognized above chance level by human observers (e.g., videos [73, 74], animations [75, 76]). These studies show that affective states modulate kinematics and dynamics such as walking speed, shoulder and elbow range of motion, and head orientation [73, 77, 74, 76]. For instance, sadness and depression are characterized by reduced walking speed, arm swing, and vertical head movements as well as slumped postures and larger lateral body sway [77]. Using the PAD emotion model, human gait observers recognize differences in arousal better than differences in pleasure [75].

2.3.3 Upper Body Movements

There are fewer works on the perception of affect from individual body parts in isolation from the rest of the body. There are situations where only individual body parts are observable due to occlusion of the rest of the body (e.g., the head and hands movements in a video call). Furthermore, there is interest in displaying affective movements on embodiments that, due to kinematic constraints, are incapable of full-body movements and rather display only isolated limb movements (e.g., [78, 79]). Therefore, it is important to explore expression and perception of affect through individual body parts.

Ekman and Friesen suggest that head orientation is an indicator of gross affective state (i.e., positive vs. negative) as well as intensity of emotion [80]. Busso *et al.* conducted a user study using the PAD model to evaluate the perception of affect from head movements during affective speech [81]. They report that head movements corresponding to different affective states (anger, happiness, sadness, neutral) are characterized by distinct motion activation, range, and velocity.

There is evidence that hand and arm movements are most significant for distinguishing between affective states [82, 83, 84]. In an early study by Carmicheal *et al.* [85], affec-

tive hand and arm gestures performed by an actor were correctly recognized (e.g., hand and arm gestures for prayer, fear, anxiety). Different affective states conveyed with hand and arm movements are recognized above chance level (full-light videos of communicative [85, 86, 87] and abstract [88] hand movements, and point-light animation of functional hand movements (drinking and knocking) [29]). Velocity, acceleration, and finger motion range are frequently reported as important hand and arm movement features for distinguishing different affective states [88, 29]. For instance, in [88], happy movements were characterized by indirect arm trajectories, angry movements were forceful and fast, whereas sad movements were slow and weak. Perceived arousal was found to be correlated with velocity, acceleration, and jerk (rate of change of acceleration) of the arm movement [29]. Affective states are also recognized above chance level during sign language communication, even for observers who do not understand the language being signed [86, 87].

2.3.4 Gender-specific Differences in Affective Movement Perception

The effect of gender on the perception of bodily expression of affect is largely unexplored. Differences in the perception of affective movements may exist due to the gender of the demonstrator or observer. In general, reports on gender differences in the perception of affective state mainly focus on facial expressions. There are various and sometimes contradictory findings on the abilities of men and women to decode facial expressions [89]. Women perceive affective states through facial expressions more accurately than men [90, 91, 92]. Men are found to be superior in recognizing angry facial expressions, e.g., [93, 94, 95], whereas women are found to be better at perceiving happy facial expressions, e.g., [96, 97] and sad facial expressions, e.g., [95]. However, in a recent study on decoding affective movements, male observers outperformed female observers in recognizing happy movements, whereas the female observers were better at recognizing angry and neutral knocking movements [98]. In another study, no significant gender differences in the perception of affective hand and arm movements were observed [85].

Other studies investigate the role of the demonstrator's gender in the perception of affective movements. In a user study, participants tended to apply social stereotypes to infer the gender of a point-light display throwing a ball with different emotions. Angry movements were judged to be demonstrated by men and sad movements were more likely to be attributed to women [99]. Similarly, the perception of fearful gait is facilitated if the walker is female [100], due to kinematic similarities between fearful gait and natural female gait. Significant differences in perception of emotion from the same static postures of *Venus* and *Apolo* (postures with different arm positions) are reported in [101].

2.3.5 Embodiment-specific Differences in Affective Movement Perception

The embodiment of an artificial agent expressing an affective state can be physical (robotic agent) or virtual (animated agent). Physical and virtual embodiments can be further subdivided into anthropomorphic (human-like kinematics and appearance, e.g., physical anthropomorphic [102], virtual anthropomorphic [103]) and non-anthropomorphic (non-human-like kinematics and appearance, e.g., physical non-anthropomorphic [104], virtual non-anthropomorphic [105]).

It is well known that humans can perceive affective states from non-anthropomorphic demonstrators. For example, humans can perceive life-like affective states from movements of abstract geometrical shapes [106]. However, there are conflicting reports on the role of embodiment in perception. In some studies, the perception of affective movements was not influenced by non-anthropomorphic appearance [105] and kinematics [104] of demonstrators, whereas other studies have shown that non-anthropomorphic kinematics and appearance of display embodiment may influence the affective movement perception [107, 108]. For instance, Takahashi *et al.* investigate the perception of affective states for a teddy bear robot and a black-suited teddy bear robot performing the same expressive motion [79]. Their results show that the perception of anger and disgust depends on the appearance of the robot (regular versus black-suited), whereas the perception of joy, sadness, surprise, and fear does not depend on the appearance.

Movement features salient to the perception of affect from animated geometric agents include absolute velocity, relative angle, relative velocity, relative heading, relative vorticity, absolute vorticity, and relative distance [109]. The arousal component of perceived affective movements displayed on embodiments with non-anthropomorphic kinematics [104, 110] is found to be correlated with the velocity and acceleration of the movements. In [110], the valence component of perceived affective movements, displayed by an interactive device with non-anthropomorphic kinematics, is related to the smoothness of the movements.

There are also conflicting reports on differences between perceiving affective states from physical embodiments (e.g., robot) and virtual embodiments depicted in video or animation. Some studies report no significant difference [111, 112], whereas significant differences between the perception of affective movements from videos of an actor and his animated replicates are reported in [103].

2.3.6 Summary

The psychology literature indicates that affective expression can be discerned from body posture and movements. The majority of psychological and behavioural studies on the affective movement perception examine categorical emotions. The categories happy, sad, and angry are better perceived from movements than categories such as fear and disgust [73, 72]. The perceptual studies largely focus on full body movements and less attention has been directed to affective expression via the movements of individual body parts. There are reports on multiple and sometimes conflicting features which might be salient to the affective movement perception. Mainly, velocity, acceleration and jerk are suggested as the key contributing movement features in conveying affect [8]. For instance, depression is associated with slow movements and elation is characterized by fast and expansive movements [84].

There are also reports on the role of gender in the perception and demonstration of affective movements, which emphasize the importance of considering gender in studying affective movements. For computational affective movement analysis, to remove (or control for) the potential influence of gender, databases should contain a balanced number of male and female demonstrators, and the reliability of the databases should be evaluated with both male and female observers.

Findings from perceptual studies on human-machine interaction suggest that the embodiment of affective displays may influence their interaction with human users. Therefore, further systematic exploration is warranted to identify kinematics and appearances that might impede the human-machine interaction, especially the perception of affective movements displayed by the machine. Table A.1 summarizes a selection of the perceptual studies discussed above.

2.4 Automatic Affective Movement Recognition

When observing a movement, in addition to "what is the movement", humans are also adept at interpreting "how is that movement performed", which may lead to ascribing an affective attribute to that movement [82, 113]. For instance, when observing a knocking movement, we use functional cues to understand the activity being performed (knocking), and expressive cues to interpret the feeling and intention underlying the movement (e.g., knocking angrily). Eliminating expressive cues from movement recognition can yield an incomplete picture and result in misinterpretation of movements. Therefore, it is important to develop automatic recognition models that can infer affective expressions encoded in

movements. Such recognition models would enable machines to better understand the activities of their users and create a more engaging human-machine interaction.

Studies on automatic recognition of affective expression from movements vary in terms of the number of affective categories and dimensions, the number of demonstrators, the type of movements,and the extent of kinematic variability in the movements (movements with fixed and specified trajectories, compared to a variety of unconstrained movements). Furthermore, the developed computational models are either person-specific or interpersonal. In *person-specific* models, both testing and training movements come from the same individual, whereas in *interpersonal* models, testing and training movements come from different individuals. In this section, a literature review on automatic recognition of affective expression from full-body and upper-body movements is presented.

2.4.1 Full-body Movements

Studies on automatic recognition of affect from gait dominate the literature of automatic recognition of affective movements. An artificial neural network (ANN) was used in [114] to recognize affective expressions (neutral, happy, sad, angry) from gait using kinetic features (measured by a force plate), kinematic features (joint angle trajectories), and angular velocities of the arm, hip, knee, and ankle, and an interpersonal (22 individuals) recognition rate of around 80% was reported. PCA, kernel PCA, linear discriminant analysis (LDA), and general discriminant analysis were used to extract a set of features to improve automatic recognition of anger, sadness, happiness, and neutral expressions from gait using ANN, naive Bayes, and support vector machine (SVM) classifiers [75]. An interpersonal (13 individuals) recognition rate of 69% was achieved.

In [115], a model based on non-linear mixture discriminant analysis was used to discriminate between full-body postures intended to convey sadness, happiness, anger, and fear (communicative postures). The postures were defined using a set of features describing distances between different body joints and the labeled with the most frequent observer labels from three cultures (American, Japanese, and Sri Lankan). A separate model was trained for each culture, and recognition rates of 78% (American model), 88% (Sri Lankan model), and 90% (Japanese model) were achieved. Kleinsmith and Bianchi-Berthouze report testing recognition rates ranging between 79% to 81% for correct recognition of affective posture attributes along the arousal, valence, potency, and avoidance dimensions [116]. In their study, the dimensional attributes for a posture were assigned as the observers' average ratings of the affective dimensions for that posture. Five different classification approaches (logistic regression, decision trees, naive bayes, artificial neural network (ANN),

and support vector machine (SVM)) were evaluated for automatic recognition of full-body communicative movements intended to convey anger, sadness, joy, and fear [117]. Mean marker velocity and acceleration values and standard deviations of the marker position, velocity and acceleration were manually selected as the movement features and ANN and SVM were reported as the most accurate classifiers with a person-specific recognition rate as high as 92%.

A model based on recurrent neural networks was proposed for automatic recognition of affective expressions (high and low intensity negative emotions, happiness, concentration) from unconstrained abstract movements of 9 players during a Wii tennis game and person-dependent and leave-one-subject-out recognition rates of 61% and 55% were achieved, respectively [118]. Another study reports a rate of 66% recognizing five types of laughter (hilarious, social, awkward, fake, non-laughter) from unconstrained body movements demonstrated by 9 subjects [119]. In another study [120], a model based on constrained body gestures achieved the highest rate (67%), recognizing 8 emotions (anger, despair, interest, irritation, joy, pleasure, pride, sadness) demonstrated by 10 subjects.

There are also reports on automatic recognition of affect from *artistic* movements. Camurri *et al.* used decision trees to distinguish between a dancer's choreographed movement expressing anger, fear, grief, and joy, and report a recognition rate of 40% on test data [121]. Non-choreographed dance movements of 10 dancers expressing happiness, surprise, anger, and sadness were recognized at a rate if 73% using a multi-layer perceptron (MLP) classification driven by movement features extracted using singular value decomposition (SVD) [122].

2.4.2 Upper-body Movements

Fewer works focus on automatic recognition of affective expressions from movements of isolated body parts. For functional upper-body movements, Bernhardt and Robinson report a 50% recognition rate of neutral, happy, angry, and sad knocking movements demonstrated by 30 individuals, using support vector machines (SVMs) with a set of dynamic movement features [123]. They improved the recognition rate to 81% by accounting for idiosyncratic effects. Boredom, confusion, delight, flow, and frustration were detected at a rate of 39% from the torso movements while sitting (captured from pressure exerted on a chair) of 28 participants using a set of classifiers including Bayesian, SVM, decision trees, artificial neural network (ANN), and kNN [124]. In another study on functional movements, a model based on decision trees was proposed in [125] for recognizing emotional expressions (sad, serene, over-expressive) from functional bodily movements of a pianist, and a recognition rate of 58% was reported.

Gunes and Piccardi [35] studied the temporal/phase synchrony of affective face and unconstrained *communicative* upper-body expressions for the automatic recognition of affect. They considered 12 expressions: anxiety, boredom, uncertainty, puzzlement, neutral/positive/negative surprise, anger, disgust, fear, happiness, and sadness. Different classification approaches including SVM, decision trees, ANN, and Adaboost were applied and the best interpersonal (23 individuals) recognition rate (77%) was obtained using Adaboost with decision trees. In [88], an abstract arm movement with a specified kinematic trajectory was used by 10 dancers to express happiness, anger, and sadness. Using maximum velocity, maximum acceleration, and directness of movement, an interpersonal recognition rate of 78% was achieved.

Recognition of affective states from observing only head movements is investigated in [126, 127, 128, 129]. Using only a small set of possible movement types (e.g., nodding, shaking, tilting, no movement, leaning forward and backward) and the way they are performed (modulation), results in person-dependent recognition rates above chance level. Person-independent recognition rates above chance level are reported when combining head movements with arm gestures or observing only arm gestures and torso movements, e.g., [130, 35].

2.4.3 Summary

The affective movement recognition literature reports on various computational models capable of recognizing affective expressions conveyed via different communicative movements (e.g., [117, 131, 35, 130]), artistic movements (e.g., [122, 121]), abstract movements (e.g., [132, 88]), and functional movements (e.g., [75, 114, 123]). The majority of the studies consider a discrete emotion set and only a few consider affective dimensions (typically, arousal and valence). Table A.2 summarizes the studies on automatic recognition of affective movements.

There are a number of challenges in automatic affective movement recognition ranging from inter-individual to kinematic and stochastic differences in movements. Affective movements can be highly variable in terms of intensity, timing, and range of movement, even when the same demonstrator repeats a single movement multiple times (viz., stochastic variabilities). In addition to the inter-individual and stochastic differences, an affective expression can be communicated through a number of kinematically different movements, and there also exist kinematically similar movements that convey distinct affective expressions. Person-specific recognition models have been developed to overcome limitations associated with inter-individual variabilities and are shown to surpass interpersonal mod-

els in recognizing affective movements (e.g., [133, 75, 114]). To address kinematic variabilities, computational models are designed for a specific kinematic movement (e.g., gait [75, 114], knocking [123], an abstract arm raising movement [88], or a choreographed dance movement [121]). However, models capable of recognizing emotions from a wide range of kinematic movements are favoured over movement-specific models. Furthermore, despite the diverse literature on affective movement recognition, the movement features critical to affect recognition are not yet precisely known, and the selection of features for affective movement recognition is usually done in an ad-hoc manner.

To enable inter-individual affective movement recognition robust to kinematic and stochastic variabilities, features most salient to discriminating between different affective expressions need to be identified and exploited in developing a computational model for affective movement recognition.

This thesis proposes a movement representation based on basis function expansion and identifies a minimal set of functional features most salient to discriminating between different affective expressions using dimensionality reduction techniques adapted for functional datasets (Section 3). The affective movement recognition is then performed in a space spanned by the identified functional features using nearest neighbour classification. Despite its suitability for modeling variable-length affective movements, the proposed functional modeling approach is sensitive to kinematic, stochastic, phase variations inherent in human movements acquired under naturalistic settings. Subsequently, the thesis develops a more generic approach that can handle temporal, stochastic, and inter-individual variabilities, as well as kinematically dissimilar within-class affective movements, in a systematic manner, using Fisher score movement representation and support vector machine classification (Section 4).

2.5 Automatic Affective Movement Generation

Affective movements are commonly generated by designers and animators using techniques such as keyframing [134]. These techniques, however, are time consuming and require trained animators and designers. Furthermore, in human-machine interaction scenarios, the expert-based generation techniques might limit the extent and duration of online interaction as the manual generation and control of affective movements require direct involvement of a human expert for extended periods of time. Therefore, it is important to develop computational models for automatic generation of affective movements. Such computational models enable the generation of automatic affective responses and allow machines to engage in a long-term empathic interaction with their users.

In general, there are two main approaches for automatic movement generation: 1) rule-based, and 2) example-based. Rule-based approaches propose a set of motion generation/modification rules that are mainly driven by an intuitive mapping between movement features and different affective expressions (or motion styles). On the other hand, example-based approaches generate a desired affective movement by concatenating or blending movement exemplars from a labeled dataset of movements.

Several examples of rule-based approaches can be found in the literature. A mapping from Laban Effort and Shape to upper-body movement kinematics is proposed in [61]. Correlations between Laban components and discrete emotions are used to devise a mapping from discrete emotions to joint angles of a humanoid robot [135]. In another study, a set of motion modification rules adjust the velocity, the range of motion, and the basic posture for a given movement based on specified levels of the arousal and valence affective dimensions [136]. A set of expressivity parameters derived from the psychology literature (overall activation, spatial extent, temporal extent, fluidity, power, repetition) are used to develop rules that select a gesture type and modify it to adopt a desired style (e.g., abrupt) for an embodied conversational agent [137].

Example-based approaches range from concatenation of labeled motion segments (e.g., [138]) to sophisticated generative models learned from exemplar movements (e.g., [139]). Human movements vary in length and are intrinsically high-dimensional, and learning in high-dimensional motion space suffers from the limitations associated with the "curse of dimensionality". Different approaches are reported that make use of a low-dimensional representation of affective movements along with an inverse mapping to generate new human movements. For instance, movements can be represented as a weighted linear combination of a set of basis functions (functional representation). The human gait is generated by interpolating or extrapolating Fourier expansions of actual human movements corresponding to different locomotion types (walk and run) subject to kinematic constraints (step size, speed, and hip position) [140]. Then, a gait is imbued with a target emotion via superposition of the Fourier expansion of the gait with a Fourier characteristic function for the target affective gait. A limitation of the functional modeling approaches for affective movement generation is that they require movements to be kinematically similar with aligned landmarks, which is a difficult constraint to attain under natural conditions.

Other example-based approaches learn a mapping between interpolation parameters and the human joint space to generate different motion trajectories. Mukai and Kuriyama [141] employ a Gaussian process regression approach to learn such a mapping. For each action (e.g., arm reaching), a set of control parameters are manually selected (e.g., turning angle and height of the target for the reaching movement) and a mapping between the parameter space and high-dimensional joint space is learnt for that action. The resulting

action-specific mapping is used to generate different styles of the action (e.g., reaching to different target positions). Movement styles were annotated in terms of Laban Weight, Time, and Flow Effort for the movements available in a labeled dataset [142], and for each pair of kinematically similar movements, a new movement is generated via space-time interpolation. Then, a mapping between resulting motion styles and the interpolation parameters is learnt. For a test movement and a target style, a pair of movements in the labeled dataset are selected, whose blend is kinematically similar to the test movement and provides the best approximation of the target style [142]. The studies [141, 142] do not explicitly address affective movement generation.

Another class of example-based approaches derives a stochastic generative modeling of the movements. Among stochastic generative approaches, those that derive a mapping from a low-dimensional hidden space to the movement space are popular in human motion analysis. Style machines, a parametric hidden Markov modeling approach, provide a stochastic modeling of movements parametrized by a style variable to be used to generate movements with a desired style [143]. Style machines were tested with a set of similar movements performed in different styles (biped locomotion, and choreographed ballet dance), and were not explicitly tested to generate affective movements [143]. A Gaussian process (GP) dynamical model was used to learn motion styles as low-dimensional continuous latent variables and their mapping to movement joint space. Using the learned model, a new movement is generated by adjusting the latent variable to adopt a desired style [144]. GP-based models are often limited to a small number of aligned and similar movements, and are computationally expensive to learn, which hinder their application for large scale motion learning and generation. Taylor *et al.* [139] propose a parametric hierarchical modeling of human movements composed of conditional restricted Boltzmann machines (CRBM) for exact inference and generating stylistic human movements. Higher level layers representing high-level latent factors can be sequentially added to the model as needed, to capture high-level structures in the movements, resulting in a deep belief network. In an extension to [139], a set of context/style variables (discrete or continuous) is added to the CRBM to generate different motion styles, synthesize transitions between stylized motions as well as interpolating and extrapolating training movements [145]. However, learning CRBMs is challenging due the large number of parameters and metaparameters that need to be set [146].

2.5.1 Summary

The literature on automatic stylistic movement generation can be divided into two main categories of rule-based and example-based approaches, and only a few studies explicitly

address conveying emotion via movements. The success of rule-based approaches depends on *a priori* knowledge of postural and motion cues salient to different affective expressions and how these cues are incorporated into the rules in terms of quantitative variables that can be tuned to generate a desired affective movement. A drawback of the rule-based approaches is that they are designed for a specific kinematic embodiment and a set of movements, and often are not readily adaptable to other embodiments and movements. While this limitation also exists for the example-based approaches, they do not rely on any deterministic rules and are capable of generating novel movement trajectories provided that adequate movement exemplars are available. Therefore, the success of the example-based approaches depends on the richness of the labeled dataset used in these approaches.

In general, the same challenges as those in automatic recognition of affective movements (kinematic, temporal, and stochastic variabilities) also hold for both the rule-based and example-based affective movement generation. Furthermore, there are inter-individual differences in affective movements that specifically influence example-based approaches that use labeled dataset of movement from different demonstrators. Therefore, an automatic generation model should suppress the irrelevant sources of variations and exploit movement features salient to affective expression in modulating a desired motion path to convey a target emotion. In addition, the model should be capable of generating affective movements in a computationally efficient manner.

In this thesis, a new example-based affective movement generation approach is presented (Section 5). Using a labeled dataset of affective movements, the proposed approach exploits movements kinematically similar to a desired motion path and belonging to a target emotion class (nearest neighbours) to generate a modulated version of the desired motion path, that conveys the target emotion. The Laban Effort and Shape notation is used to identify the nearest neighbours for the desired motion path. The nearest neighbours are encoded in an HMM and the most likely state sequence of the resulting HMM for the desired motion path is used to generate a modulated version of the desired motion path.

Chapter 3

Affective Movement Recognition and Generation Based on Functional Movement Representation

■ *An earlier version of the text in this chapter has appeared in the following publications*: [16, 17, 105, 147].

In this chapter, an affective movement recognition approach based on functional movement representation and dimensionality reduction (DR) is first presented. The approach represents human movements in terms of temporal functions using B-spline basis function expansion, and adapts DR techniques (Fisher discriminant analysis (FDA), supervised principal component analysis (sPCA), and Isomap) to enable application to the functional data. The functional DR techniques along with functional PCA are applied on an affective full-body movement dataset and an affective hand movement dataset, and their performance is evaluated using leave-one-out cross validation with one-nearest neighbour classification in the corresponding low-dimensional subspaces. The results show that functional sPCA outperforms the other DR techniques in terms of classification accuracy and time resource requirements.

Next, an approach for generating prototypical affective movements based on functional movement representation and functional principal component analysis (FPCA) is presented. A low-dimensional encoding of affective movements is obtained using FPCA, and the centroids of emotion-specific clusters of movements in the resulting functional sub-

space along with the FPCA inverse mapping are used to generate prototypical movements for each emotion class. The proposed approach is applied to the affective hand movement dataset, and a user study is conducted to evaluate the expressivity of the generated prototypical hand movements in comparison with those displayed by a demonstrator.

Finally, this chapter presents a user study conducted to evaluate the impact of display embodiment and an observer's gender on the perception of the generated affective hand movements. To this end, the generated affective movements were animated on a human-like and a frond-like hand model, and their expressivity was rated by male and female observers. In summary, this chapter presents:

- An approach for affective movement modeling and recognition based on functional movement representation and functional dimensionality reduction.
- An approach for generating prototypical affective movements based on functional movement representation and functional principal component analysis, and a user study to evaluate the expressivity of generated movements in comparison with those displayed by a demonstrator.
- A user study to evaluate the impact of display embodiment and an observer's gender on the perception of the affective movements.

3.1 Functional Affective Movement Recognition

A sampled affective movement x_i of length T_i is described by a collection of m joint positions. As a result, the movement x_i at a time instant $t \in \{1, 2, \ldots, T_i\}$ is defined by

$$x_{i,t} = \{f_{1i,t}, f_{2i,t}, \ldots, f_{mi,t}\}, \tag{3.1}$$

where $f_{ji,t}$ is the value of the j^{th} time-series feature of movement x_i at time t and there are a total of m features.

The variable length of affective movements impedes the application of conventional DR techniques as these techniques can only be applied to fixed-length data. In this section, a systematic approach for fixed-length vectorial movement representation based on basis function expansion (BFE) is presented.

After transforming sequential observations into basis function space, DR techniques need to be adapted to enable application to the resulting functional datasets. Ramsay

introduces functional principal component analysis (FPCA), an approach for applying PCA to functional datasets [4]. In this section, an approach similar to [4] is employed to adapt Fisher discriminant analysis (FDA) [148], supervised principal component analysis (sPCA) [149], and Isomap [150] for application to functional datasets. The resulting functional DR techniques along with FPCA are then used to obtain a lower-dimensional embedding of affective movements. The discriminative performance of the functional representation is evaluated using leave-one-out cross validation (LOOCV) with one-nearest neighbour (1NN) classification in the resulting reduced spaces.

3.1.1 Basis Function Expansion

Basis function expansion (BFE) is used to represent sequential observations as temporal functions computed at every time instant t. Using BFE, time-series features are represented as a weighted linear combination of a set of basis functions $\varphi_g(t), g = 1, \ldots, G$, where G is the total number of basis functions. Considering a multivariate time-series observation x_i with $t = 1, 2, \ldots, T_i$, the BFE for its j^{th} time-series feature (f_{ji}) is denoted as $f_{ji}(t)$ and expressed as

$$f_{ji}(t) = \sum_{g=1}^{G} c_{jig}\varphi_g(t) \rightarrow f_{ji}(t) = \mathbf{c}_{ji}^{\top}\Phi. \tag{3.2}$$

Φ is a functional vector of length G containing basis functions $\varphi_g(t)$, and $\mathbf{c}_{ji}$ is a vector of basis function coefficients for the j^{th} feature of the time-series observation x_i. Therefore, the high-dimensional time-series features are now represented within the framework of the basis function coefficient vector $\mathbf{c}$. The BFE coefficients of individual features are then concatenated in a single vector to represent the multivariate time-series observation x_i. This forms a dataset $\mathcal{D} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$, with $\mathbf{x}_i \in \mathbb{R}^h$ ($h = m \cdot G$) being a vectorial representation of the multivariate time-series movement sequence x_i. Therefore, the variable-length time-series movement observations are now represented as fixed-length observations.

This thesis uses B-splines as the basis functions as they are best suited for representing observations that lack any strong cyclic variations [4]. B-splines are defined as piecewise polynomials with smoothly connected polynomial pieces, and they provide two types of independent controls: 1) a set of fixed control points (knots) and 2) spline degree. Individual B-splines are constructed recursively using the Cox-DeBoor recursion formula [151]. For a B-spline of degree d_{spline}, there are $d_{spline} + 1$ polynomials of degree d_{spline} joined at d_{spline} inner knots. Except at the boundaries, each B-spline overlaps with $2d_{spline}$ polynomial pieces and is positive over an interval of $d_{spline} + 2$ knots and zero elsewhere. The degree

of the B-spline used depends on the desired order of derivative. For instance, if we require a smooth acceleration curve, then B-splines of degree 3 and above should be used.

After choosing the type of basis function, least squares regression regularized by a roughness penalty is used to obtain an optimal set of coefficients. In this thesis, the square of the second derivative of the curve at each time instant t is used as the roughness penalty [4]. Regularizing the least squares criterion with the roughness penalty helps to avoid overfitting and increases curve differentiability.

3.1.2 Functional Dimensionality Reduction

In this section, the functional formulation of PCA proposed by Ramsay [4] is fist reviewed. An extension of this methodology is then proposed to adapt Fisher discriminant analysis, supervised PCA, and Isomap to enable application to functional datasets.

3.1.2.1 Functional Principal Component Analysis (FPCA)

Principal component analysis (PCA) identifies main modes of variation in data of dimensionality d by projecting it into a lower-dimensional space formed by p principal components (PCs; directions of maximum variation in the data), where $p < d$ [152]. Suppose $\mathbf{X} \in \mathbb{R}^{d \times n}$ is a matrix whose columns represent n centered multivariate data points $\mathbf{x} \in \mathbb{R}^d$, and let the first PC defines a weight vector, $\mathbf{w}_1$ for which the linear combination values $\mathbf{w}_1^\top \mathbf{X}$ have the largest variance. The weights $\mathbf{w}_1$ are obtained by solving the following optimization problem

$$\begin{aligned} &\max_{\mathbf{w}_1} \mathbf{w}_1^\top \mathbf{X}\mathbf{X}^\top \mathbf{w}_1 \\ &\text{s.t. } \mathbf{w}_1^\top \mathbf{w}_1 = \|\mathbf{w}_1\|_2^2 = 1. \end{aligned} \tag{3.3}$$

Subsequent $p-1$ PCs are obtained by solving the corresponding optimization similar to 3.3, subject to additional orthogonality constraints. For instance, for the r^{th} PC, there are $r-1$ orthogonality constraints: $\mathbf{w}_q^\top \mathbf{w}_r = 0,\ q < r$.

Alternatively, PCA can be defined as finding the eigenvectors corresponding to the p largest eigenvalues of the sample covariance matrix $\mathbf{S} = \mathbf{X}\mathbf{X}^\top$. To this end, the following generalized eigenvalue decomposition problem is solved

$$\mathbf{S}\mathbf{W} = \lambda \mathbf{W}, \tag{3.4}$$

where $\mathbf{W}$ is the weight matrix of size $d \times p$ whose columns are the eigenvectors of $\mathbf{S}$ with λ being the corresponding eigenvalues. The eigenvectors are the principal components of $\mathbf{X}$. The weighted linear combination of the columns of $\mathbf{X}$, with the weights coming from the first principal component ($\mathbf{w}_1^\top \mathbf{X}$) has the largest variance amongst all linear combinations of the columns of $\mathbf{X}$.

Ramsay introduced functional PCA (FPCA) to enable application of PCA on functional data. In FPCA, the PCs are a set of orthonormal eigenfunctions expressed as a weighted linear combination of basis functions. For functional observations $x_i(t)$, the variation in the dataset is approximated using the bivariate covariance function v.

$$v(s,t) = \frac{1}{n-1} \sum_i (x_i(s) - \bar{x}(s))(x_i(t) - \bar{x}(t)), \tag{3.5}$$

where $\bar{x}(t)$ is the mean of the functional observations in the dataset

$$\bar{x}(t) = \frac{1}{n} \sum_i x_i(t). \tag{3.6}$$

In order to obtain FPCA principal component weight function $w_1(t)$, the variance of the FPCA embedding (FPCA scores) along $w_1(t)$ should be maximized

$$\begin{aligned} &\max_{w_1(t)} \sum_i (S_{i1})^2 \\ &\text{s.t.} \int w_1^2(s)ds = \|w_1\|_2^2 = 1, \end{aligned} \tag{3.7}$$

where S_{i1} is the FPCA embedding for a functional observation $x_i(t)$ along $w_1(t)$ and is derived using

$$S_{i1} = \int w_1(s) x_i(s) ds. \tag{3.8}$$

Similarly, the functional principal component $w_2(t)$ can be obtained using Equation 3.7, introducing an additional orthogonality constraint on eigenfunctions

$$\int w_1(s) w_2(s) ds = 0 \tag{3.9}$$

This procedure is continued to obtain p functional PCs.

Similar to the multivariate case, the FPCA can be defined as a generalized eigenvalue decomposition problem. In FPCA, the eigenequation to be solved is

$$\int v(s,t)w_r(s)ds = \lambda w_r(t). \tag{3.10}$$

where $w_r(t)$ is the r^{th} eigenfunction and $r = 1, 2, \ldots, p$. Suppose our n time-series observations are represented in terms of weighted linear combination of G basis functions as described in Section 3.1.1; the i^{th} time-series observation: $x_i = \mathbf{c}_i^\top \Phi$. Using BFE, the r^{th} eigenfunction (PC) of FPCA, $w_r(t)$, is expressed as

$$w_r(t) = \sum_{g=1}^{G} b_{rg}\varphi_g(t) \rightarrow w_r(t) = \Phi^\top \mathbf{b}_r, \tag{3.11}$$

where Φ is a functional vector of length G containing basis functions $\varphi_g(t)$, and $\mathbf{b}_r$ represents coefficients corresponding to successive basis functions $\varphi_g(t)$. Let $\mathbf{C}$ be a matrix of size $G \times n$ containing the coefficients for BFE of n time-series observations. Each column of $\mathbf{C}$ represents BFE coefficients for a single time-series observation x_i. The covariance function in Equation (3.5) can be re-written as

$$v(s,t) = (n-1)^{-1}\Phi^\top(s)\mathbf{C}\mathbf{C}^\top\Phi(t). \tag{3.12}$$

Hence, the corresponding eigenequation (3.10) can be expressed as

$$\int (n-1)^{-1}\Phi^\top(s)\mathbf{C}\mathbf{C}^\top\Phi(t)\Phi^\top(t)\mathbf{b}_r dt = \lambda\Phi^\top(s)\mathbf{b}_r. \tag{3.13}$$

$\mathbf{b}_r$ is a vector carrying the coefficients of the eigenfunction $w_r(t)$. This equation holds for all the temporal arguments s, and therefore can be reduced to

$$(n-1)^{-1}\mathbf{C}\mathbf{C}^\top\mathbf{M}\mathbf{b}_r = \lambda\mathbf{b}_r, \tag{3.14}$$

where $\mathbf{M} = \int \Phi(t)\Phi^\top(t)dt$ is an order G symmetric matrix, and furthermore, the constraint $\|w_r(t)\|_2^2 = 1$ is equivalent to

$$\|w_r(t)\|_2^2 \equiv \mathbf{b}_r^\top\mathbf{M}\mathbf{b}_r = 1. \tag{3.15}$$

If we define, $\mathbf{u}_r = \mathbf{M}^{1/2}\mathbf{b}_r$, we can rewrite the FPCA eigenequation (3.14) as

$$\begin{aligned} &(n-1)^{-1}\mathbf{M}^{1/2}\mathbf{C}\mathbf{C}^\top\mathbf{M}^{1/2}\mathbf{u}_r = \lambda\mathbf{u}_r \\ &\text{s.t. } \mathbf{u}_r^\top\mathbf{u}_r = 1 \end{aligned} \tag{3.16}$$

Equation (3.16) is a generalized eigenvalue problem and can be solved for $\mathbf{u}_r$. The resulting $\mathbf{u}_r$ can be used to compute $\mathbf{b}_r$, which in turn is used to compute the eigenfunction $w_r(t)$ using Equation (3.11).

A complete definition of a function would require specifying its value at every possible time t, which will result in an arbitrary large dimensional definition for the function. The function values at different t's is equivalent to the variables in the case of multivariate observations. However, unlike the multivariate PCA, where the maximum number of PCs is equal to the dimensionality of the dataset, the maximum number of PCs in FPCA is $n-1$ provided that the n functional observations $x_i(t)$ are linearly independent and $|\mathbf{t}| \gg n$, which is often the case when dealing with functional observations. In the case of a concatenated multivariate time-series dataset obtained as explained in Section 3.1.1, the eigenfunction corresponding to each individual time-series feature can be obtained by breaking down the resulting eigenfunctions from FPCA into pieces of length equal G.

3.1.2.2 Functional FDA (FFDA)

For multivariate datasets of dimensionality d, Fisher discriminant analysis [148] projects a K-class dataset into a $K-1$ dimensional space in an attempt to maximize the distance between projected means and minimize the variance of each projected class. The FDA objective function in a K-class multivariate problem can be formulated as

$$\max_{\mathbf{W}} \frac{Tr(\mathbf{W}^\top \mathbf{S}_b \mathbf{W})}{Tr(\mathbf{W}^\top \mathbf{S}_w \mathbf{W})} \equiv \begin{cases} \max_W \mathrm{Tr}(\mathbf{W}^\top \mathbf{S}_b \mathbf{W}) \\ s.t.\ \mathrm{Tr}(\mathbf{W}^\top \mathbf{S}_w \mathbf{W}) = 1, \end{cases} \tag{3.17}$$

where $\mathbf{W}$ is a weight matrix of size $d \times K-1$, and $\mathbf{S}_b$ and $\mathbf{S}_w$ are the between-class and within-class covariances, respectively. Introducing a Lagrange multiplier λ and setting the first derivative of the Lagrange function with respect to $\mathbf{W}$ to zero results in a generalized eigenvalue problem

$$\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{W} = \lambda \mathbf{W}. \tag{3.18}$$

Therefore, $\mathbf{W}$ is a matrix of eigenvectors associated with the eigenvalues of $\mathbf{S}_w^{-1}\mathbf{S}_b$. Here, a similar approach to FPCA is employed to adapt FDA for sequential observations estimated using BFE. Assume that there are n multivariate time-series movements x_i belonging to K classes and each movement x_i is defined over a temporal interval $t = 1, 2, \ldots, T_i$. As before, these movements are estimated as temporal functions using G basis functions. The within-class covariance for the k^{th} class of the functional observations is computed as

$$v_k(s,t) = (n_k - 1)^{-1} \Phi^\top(s) \mathbf{C}_k \mathbf{C}_k^\top \Phi(t), \tag{3.19}$$

where n_k is the number of movements belonging to the k^{th} class, Φ is a functional vector of length G containing basis functions $\varphi_g(t)$, and $\mathbf{C}_k$ is a matrix of size $G \times n$ carrying the coefficients corresponding to the basis functions $\varphi_g(t)$ of the movements in class k. Similar to conventional FDA, the between-class covariance ($v_b(s,t)$) can be estimated by subtracting the within-class covariances ($v_w(s,t)$) from the total covariance ($v_t(s,t)$). The total covariance is computed using Equation (3.5). Now, if we solve for the FDA optimization problem by introducing a Lagrange multiplier λ, we get the below eigenequation for the functional observations

$$\int v_b(s,t)w_r(t)dt = \lambda_r \int v_w(s,t)w_r(t)dt, \tag{3.20}$$

where $w_r(t) = \Phi^\top(t)\mathbf{b}_r$ is the r^{th} eigenfunction ($r = 1, 2, \ldots, K-1$) associated with the eigenvalue λ_r. Using the expressions obtained for the within-class covariance and the between-class covariance, Equation (3.20) can be written for eigenfunction $w_r(t)$ as follows

$$\int z_b \Phi^\top(s)\mathbf{C}_b\mathbf{C}_b^\top\Phi(t)\Phi^\top(t)\mathbf{b}_r dt = \lambda \int z_w \Phi^\top(s)\mathbf{C}_w\mathbf{C}_w^\top\Phi(t)\Phi^\top(t)\mathbf{b}_r dt, \tag{3.21}$$

which can be reduced to

$$z_b\mathbf{C}_b\mathbf{C}_b^\top\mathbf{M}\mathbf{b}_r = \lambda z_w\mathbf{C}_w\mathbf{C}_w^\top\mathbf{M}\mathbf{b}_r, \quad \mathbf{M} = \int \Phi(t)\Phi^\top(t)dt. \tag{3.22}$$

z_b and z_w are the normalizing terms for between-class and within-class covariances. Similar to FPCA, the corresponding constraint to the eigenequation (3.22) is

$$\|w_r(t)\|_2^2 \equiv \mathbf{b}_r^\top\mathbf{M}\mathbf{b}_r = 1. \tag{3.23}$$

Let $\mathbf{u}_r = \mathbf{M}^{1/2}\mathbf{b}_r$, the eigenequation (3.22) can be written as

$$\begin{aligned} &z_b\mathbf{C}_b\mathbf{C}_b^\top\mathbf{M}^{1/2}\mathbf{u}_r = \lambda z_w\mathbf{C}_w\mathbf{C}_w^\top\mathbf{M}^{1/2}\mathbf{u}_r, \\ &\left(z_w\mathbf{C}_w\mathbf{C}_w^\top\mathbf{M}^{1/2}\right)^{-1}\left(z_b^{-1}\mathbf{C}_b\mathbf{C}_b^\top\mathbf{M}^{1/2}\right)\mathbf{u}_r = \lambda\mathbf{u}_r. \end{aligned} \tag{3.24}$$

The above generalized eigenvalue problem can be solved for $\mathbf{u}_r$, which is used to compute the FFDA functional feature transformations $w_r(t)$. Subsequently these feature transformations will be used to embed the movements into the lower dimensional embedding of FFDA using Equation (3.8).

3.1.2.3 Functional Supervised PCA Based on HSIC (FSPCA)

In cases where there are non-linear relations between two random variables, non-linear dependency measures need to be used to explore the correlation between these variables. Gretton *et al.* [153] propose the use of the square of the Hilbert Schmidt norm of the cross-covariance operator between reproducing kernel Hilbert spaces (RKHS)s, as a measure of independence between the variables. Suppose that $\mathcal{F}$ and $\mathcal{G}$ are separable RKHSs containing all real-valued, bounded, and continuous functions mapping $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}$ to $\mathbb{R}$, respectively. Then, for each $\mathbf{x} \in \mathcal{X}$, there exists a $\varphi(\mathbf{x}) \in \mathcal{F}$ such that $\mathbf{K}(\mathbf{x}, \mathbf{x}') = \langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle_{\mathcal{F}}$ is a unique positive definite kernel associated with $\mathcal{F}$. It can be shown that there exists a linear cross-covariance operator between $\mathcal{F}$ and $\mathcal{G}$ that maps elements in $\mathcal{F}$ to elements of $\mathcal{G}$; $C_{xy} : \mathcal{F} \to \mathcal{G}$.

$$C_{xy} := \mathbf{E}_{xy}[(\varphi(\mathbf{x}) - \mathbf{E}_x[\varphi(\mathbf{x})]) \otimes (\psi(\mathbf{y}) - \mathbf{E}_y[\psi(\mathbf{y})])], \tag{3.25}$$

where $\otimes$ is the tensor product.

The Hilbert Schmidt independence criterion (HSIC) between $\mathcal{X}$ and $\mathcal{Y}$ is defined as the square of the Hilbert Schmidt norm of the cross-covariance operator between their RKHSs. In terms of kernel functions $\mathbf{K}$ and $\mathbf{B}$ associated with $\mathcal{F}$ and $\mathcal{G}$, respectively, the HSIC measure is defined as

$$\begin{aligned} \mathrm{HSIC}(\mathcal{P}_{\mathcal{XY}}, \mathcal{F}, \mathcal{G}) = \mathbf{E}_{\mathbf{xx'yy'}}[\mathbf{K}(\mathbf{x}, \mathbf{x}')\mathbf{B}(\mathbf{y}, \mathbf{y}')] \dots \\ + \mathbf{E}_{\mathbf{xx'}}[\mathbf{K}(\mathbf{x}, \mathbf{x}')]\mathbf{E}_{\mathbf{yy'}}[\mathbf{B}(\mathbf{y}, \mathbf{y}')] \dots \\ - 2\mathbf{E}_{\mathbf{xy}}[\mathbf{E}_{\mathbf{x'}}[\mathbf{K}(\mathbf{x}, \mathbf{x}')]\mathbf{E}_{\mathbf{y'}}[\mathbf{B}(\mathbf{y}, \mathbf{y}')]], \end{aligned} \tag{3.26}$$

where $\mathbf{E}_{\mathbf{xx'yy'}}$ is the expectation over independent pairs $(\mathbf{x}, \mathbf{y})$ and $(\mathbf{x}', \mathbf{y}')$ drawn from distribution $\mathcal{P}_{\mathcal{XY}}$. For n independent observations $Z = \{(\mathbf{x_1}, \mathbf{y_1}), \dots, (\mathbf{x_n}, \mathbf{y_n})\} \subseteq \mathcal{X} \times \mathcal{Y}$ drawn from the distribution $\mathcal{P}_{\mathcal{XY}}$, the empirical estimation of HSIC between $\mathcal{X}$ and $\mathcal{Y}$ is

$$\mathrm{HSIC}(Z, \mathcal{F}, \mathcal{G}) = \frac{1}{(n-1)^2}\mathrm{Tr}(\mathbf{KHBH}), \ \mathbf{H} = \mathbf{I}_n - \frac{1}{n}\mathbf{e}\mathbf{e}^\top. \tag{3.27}$$

$\mathbf{H}$ is a constant matrix used to centralize $\mathbf{K}$ and $\mathbf{B}$, $\mathbf{e}$ is a column vector of $\mathbf{1}$'s, and $\mathbf{I}_n$ is an identity matrix of size n. It can be shown that the HSIC of two independent variables is zero. In this thesis, linear and Gaussian radial basis function (GRBF) kernels are used.

Suppose $\mathbf{X}$ is a matrix of size $d \times n$ carrying a set of multivariate datapoints in its columns and $\mathbf{y}$ is a vector carrying the datapoint labels. The HSIC-based supervised PCA [149] is a nonlinear dimensionality reduction technique aiming to find a set of feature transformations (PCs) $\mathbf{w}_r$ along which $\mathrm{Tr}(\mathbf{KHBH})$ is maximized, i.e., that it finds $\mathbf{w}_r$'s

such that the $\mathbf{w}_r^\top \mathbf{X}$ is highly dependent on $\mathbf{y}$ ($r = 1, 2, \ldots, p$, where p is the desired dimensionality of the reduced subspace). In the following, supervised PCA is formulated with a linear kernel on $\mathbf{X}$ and any type of kernel on $\mathbf{y}$.

$$\begin{aligned} &\mathbf{w}_r^\top X \xrightarrow{\textit{linear kernel}} \mathbf{K} = \mathbf{X}^\top \mathbf{w}_r \mathbf{w}_r^\top \mathbf{X} \\ &\mathbf{y} \xrightarrow{\textit{any kernel}} \mathbf{B} \\ &\max_{\mathbf{w_r}} \Big(\mathrm{Tr}(\mathbf{w}_r^\top \mathbf{XHBHX}^\top) \mathbf{w}_r \Big) \\ &\text{s.t: } \mathbf{w}_r^\top \mathbf{w}_r = 1. \end{aligned} \tag{3.28}$$

Similar to classical PCA, the constraint $\mathbf{w}_r^\top \mathbf{w}_r = 1$ is introduced to make the optimization problem well-posed. It can be shown that the solution to the maximization problem (3.28) are the eigenvectors corresponding to the largest eigenvalues of $\mathbf{XHBHX}^\top$ [149]. A special case of the above formulation is when $\mathbf{B} = \mathbf{I}$, which results in the conventional PCA.

Here, a modification of supervised PCA is presented that allows application of this nonlinear DR technique to the functional dataset. A modification of Equation (3.28) to accommodate the basis function representation of the observations is as follows

$$\begin{aligned} &\int \mathbf{b}_r^\top \Phi(t) \Phi(t)^\top \mathbf{C} \, dt \xrightarrow{\textit{linear kernel}} \int \mathbf{C}^\top \Phi(t) \Phi(t)^\top \mathbf{b}_r \mathbf{b}_r^\top \Phi(t) \Phi(t)^\top \mathbf{C} \, dt \\ &\mathbf{y} \xrightarrow{\textit{any kernel}} \mathbf{B} \\ &\max_{\mathbf{u}_r} \Big(\mathrm{Tr}(\mathbf{u}_r^\top \mathbf{D} \mathbf{u}_r) \Big), \\ &\text{s.t. } \mathbf{u}_r \mathbf{u}_r^\top = 1 \\ &\text{where } \mathbf{D} = \mathbf{M}^{1/2} \mathbf{CHBHC}^\top \mathbf{M}^{1/2}, \\ &\mathbf{u}_r = \mathbf{M}^{1/2} \mathbf{b}_r \text{ and } \mathbf{M} = \int \Phi(t) \Phi(t)^\top \, dt. \end{aligned} \tag{3.29}$$

The above maximization can be formulated as a generalized eigenequation by introducing a Lagrange multiplier λ. Then, the resulting eigenequation can be solved for $\mathbf{u}_r$, which will be used to obtain $w_r(t)$. Finally, the FSPCA embedding can be obtained using Equation (3.8).

3.1.2.4 Functional Isomap (F-Isomap)

Isomap is a non-linear extension of the Multidimensional Scaling (MDS) [154] dimensionality reduction technique, which performs MDS on the geodesic space of a non-linear data

manifold, preserving the pairwise geodesic distances between datapoints in the reduced subspace [150]. Isomap embedding is performed in three steps: 1) finding neighbours of each point (e.g., k nearest neighbours) and constructing a neighbouring graph $\mathfrak{G}$, 2) computing the geodesic pairwise distance between all the points, and 3) embedding the data with MDS based on the geodesic distances between the datapoints. To enable the application of Isomap to sequential observations, functional estimations of these observations are first obtained and then the k-nearest neighbours for each observation are found. Next, the geodesic distance between two sequential observations is computed as the shortest path between the two observations in the neighbourhood graph, which passes through neighbour observations

$$d_{\mathbf{N}}(a,b) = \min_{L} \sum_{i=\{1,\dots,l-1\}} d_{\mathfrak{G}}(x_i, x_{i+1}). \tag{3.30}$$

In Equation (3.30), L includes two or more connected movements in the neighbourhood graph, with $x_1 = a$ and $x_l = b$. The x_i and x_{i+1} are the nearest neighbours. l is the number of movements in the geodesic path between a and b, including a and b, and $d_{\mathfrak{G}}(x_i, x_{i+1})$ is the graph Euclidean distance between x_i and x_{i+1}. Next, matrix $\mathbf{N}$ is formed with entries corresponding to the square of pairwise geodesic distances $d_{\mathbf{N}}$. Eigenvalue decomposition is performed on $\mathbf{N}$ to obtain the eigenfunctions $w_r(t) = \Phi^\top(t)\mathbf{b}_r$ corresponding to the top p eigenvalues. The functional Isomap lower-dimensional embedding is computed using the top eigenfunctions as follows:

$$S_{ir} = \sqrt{\lambda_r} w_r(t_i), \tag{3.31}$$

where S_i is the lower-dimensional embedding of x_i, S_{ir} is the r^{th} dimension of S_i, $w_r(t_i)$ is the i^{th} discrete approximation of the r^{th} eigenfunction $w_r(t)$ with λ_r being the corresponding eigenvalue for $w_r(t)$.

Unlike other DR techniques in this section, Isomap embedding does not provide a parametric transformation that can be used for transforming previously-unseen high-dimensional observations into the Isomap low-dimensional space. In [155], a non-parametric estimation of the Isomap low-dimensional transformation is introduced to test the Isomap embedding. It is shown that the Isomap embedding for a test observation O_t denoted as S_t can be approximated as

$$S_t = \frac{1}{\lambda_r} \sum_i w_r(t_i) \tilde{\mathbf{K}}(O_t, x_i), \tag{3.32}$$

where $(\lambda_r, w_r(t))$ are the eigenvalue-eigenfunction pairs obtained from performing eigenvalue decomposition on the neighbourhood matrix $\mathbf{N}$. $\tilde{\mathbf{K}}(a,b)$ is a kernel function that produces the neighbourhood matrix $\mathbf{N}$ for Isomap embedding and is defined as

$$\tilde{\mathbf{K}}(a,b) = -\frac{1}{2}\Big(d^2_{\mathbf{N}}(a,b) - \mathbf{E}_{O_t}\Big[d^2_{\mathbf{N}}(O_t,b)\Big] - \mathbf{E}_{\acute{\mathbf{X}}}\Big[d^2_{\mathbf{N}}\Big(a,\acute{\mathbf{X}}\Big)\Big] + \mathbf{E}_{O_t,\acute{\mathbf{X}}}\Big[d^2_{\mathbf{N}}\Big(O_t,\acute{\mathbf{X}}\Big)\Big]\Big), \tag{3.33}$$

where O_t is the test observation and $\acute{\mathbf{X}}$ is the set of training observations (landmarks). In the case of sequential observations, the eigenfunction $w_r(t)$ is approximated using BFE as $\Phi^\top \mathbf{b}_r$. Considering $\acute{\mathbf{X}} = \{x_1, x_2, \ldots, x_i\}_{i=1}^{n_l}$ as a subset of sequential observations (a total number of n_l landmarks), the Isomap embedding for the test observation O_t in the r^{th} eigenfunction $w_r(t)$ is computed as

$$S_t = \frac{1}{2\sqrt{\lambda_r}} \sum_i \Big(b_r \Phi(t_i)\Big)\Big(\mathbf{E}_{\acute{\mathbf{X}}}\Big[d^2_{\mathbf{N}}\Big(\acute{\mathbf{X}}, x_i\Big)\Big] - d^2_{\mathbf{N}}(x_i, O_t)\Big), \tag{3.34}$$

where $\mathbf{E}_{\acute{\mathbf{X}}}$ is the average over the training observations. Detailed discussion and proof for Isomap testing can be found in [155]. Here, Equation (3.32) is used to test the F-Isomap embedding.

3.1.3 Experiments

Two affective movement datasets are used to evaluate and compare the discriminative performance of the functional dimensionality reduction techniques: 1) a dataset consisting of a single hand movement performed by a single demonstrator [105], and 2) a larger dataset consisting of a variety of full-body movements performed by multiple demonstrators [115].

The hand movement dataset considers one movement type, closing and opening the hand, mainly involving phalangeal and carpo-metacarpal joint movements. Three sets of movements of 10 trials were collected, where each set conveys a different emotion. Three different emotions were considered: sadness, happiness, and anger. For each emotion, 5 trials were performed on the right hand and 5 on the left hand. A demonstrator performed the hand movements while wearing a Dataglove (ShapeHand from Measurand [156]). The Cartesian coordinates for the root joint (wrist) and three joints A, B and C along each finger (Figure 3.1) were collected at 84 Hz.

Next, a challenging dataset of full-body affective movements was used to further assess the discriminative and computational qualities of the functional DR techniques. The full-body dataset was collected using a Vicon 612 motion capture system and contains 183 acted full-body affective movements obtained from thirteen demonstrators who freely expressed movements conveying anger, happiness, fear, and sadness with no kinematic constraints

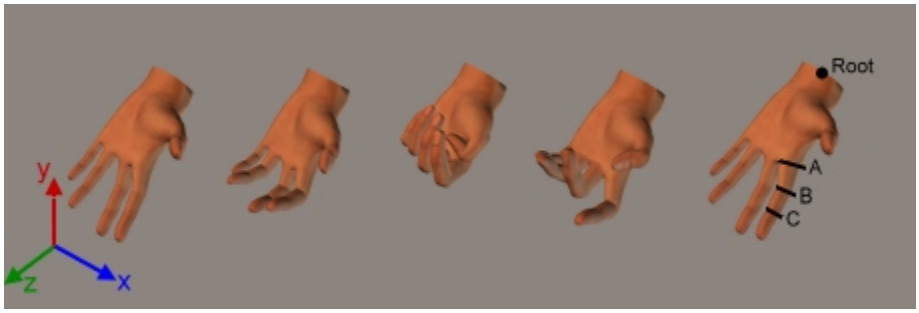

Figure 3.1: Screen shots of an animated hand movement. Local Euler angles were collected for the wrist (root) and three joints (A, B, and C) along each finger. Joints in each finger are named as shown on the index finger of the far right hand (A: proximal joint, B: intermediate joint, C: distal joint).

[115]. There are 32 markers attached to bodily landmarks[1] and their 3D (3 dimensional) Cartesian coordinates are collected using 8 motion capture cameras at a rate of 120Hz[2]. There are 46 sad, 47 happy, 49 fearful, and 41 angry movements in the full-body dataset. Each movement starts from a known pose, the T-pose (arms raised to shoulder level and extended to the sides).

For both datasets, the affective movements are preprocessed through BFE before the application of functional DR techniques. Two hundred B-splines of 4^{th} degree are chosen to represent the affective movement time series. BFE is performed using MATLAB code provided in [157]. Next, FPCA, FFDA, FSPCA, and F-Isomap are applied to obtain discriminative lower-dimensional embeddings of the transformed affective movements (i.e., functional estimation of the movements). For the FSPCA, two types of kernels are applied to the movement labels; a linear kernel and the Gaussian radial basis function (GRBF) kernel. MATLAB code provided in [158] is modified to generate the F-Isomap embedding. For FPCA, FFDA and FSPCA, Equation (3.8) is used to obtain the lower-dimensional embedding for test observations. For F-Isomap, lower-dimensional embeddings of test observations are computed using Equation (3.34). The performance of the functional DR

[1]The markers are placed on the following bodily landmarks: left front head, right frond head, left back head, right back head, top chest, center chest, left front waist, right front waist, left back waist, right back waist, top of spine, middle of back, left outer metatarsal, right outer metatarsal, left toe, right toe, left shoulder, right shoulder, left outer elbow, right outer elbow, left hand, right hand, left wrist inner near thumb, right wrist inner near thumb, left wrist outer opposite thumb, right wrist outer opposite thumb, left knee, right knee, left ankle, right ankle, left heel, right heel.

[2]Some movements in the full-body dataset are recorded at 250Hz, which were down-sampled to 120Hz in our study to keep the sampling rate consistent for all the movements.

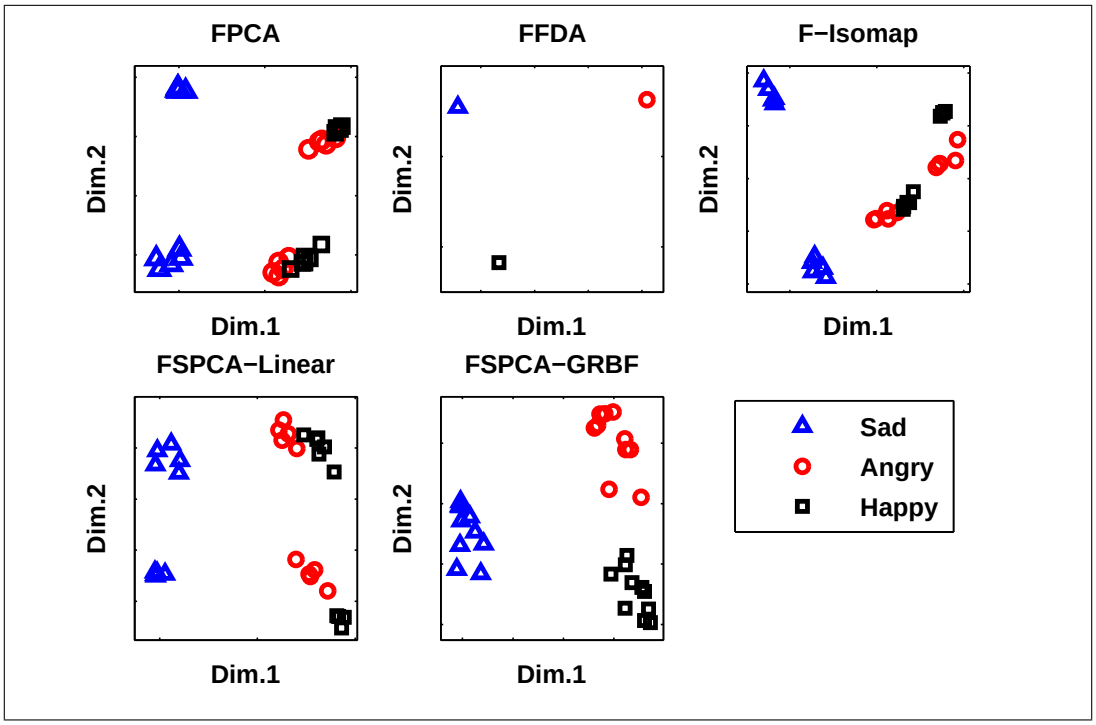

Figure 3.2: Affective hand movement embedding in the resulting 2D subspaces. For FPCA, F-Isomap, and FSPCA-Linear, movements from the same emotional category performed on left and right hands are clustered separately in the resulting embedding.

techniques in discriminating between affective movements is examined with leave-one-out cross validation (LOOCV) using one-nearest-neighbour (1NN) classification.

3.1.4 Results

The two-dimensional embeddings of all the affective hand movements obtained by FPCA, FFDA, FSPCA and F-Isomap are shown in Figure 3.2. The LOOCV training and testing recognition rates for different functional DR techniques along with their training time are shown in Table 3.1.

Table 3.2 shows the LOOCV training and testing recognition rates for the full-body dataset obtained using the 1NN classifier in the resulting reduced spaces. 3D subspaces of the functional DR techniques are used for the full-body movements to compute LOOCV error due to their discriminative advantage over 2D subspaces. The 2D embedding of

Table 3.1: Leave-One-Out cross validation training and testing recognition rates for functional DR techniques applied on the affective hand movements

	Training recognition rate (%)	**Testing recognition rate (%)**	**Elapsed training time (sec)**
FPCA	87	67	0.07
FFDA	100	30	474.20
F-Isomap	96	57	0.13
FSPCA-Linear	97	97	0.09
FSPCA-GBRF	99	93	0.09

training and testing full-body movements are shown in Figure 3.3 to illustrate the ability of the functional DR techniques to discriminatively embed the high-dimensional affective movements in a low dimensional space.

Table 3.2: Leave-One-Out cross validation training and testing recognition rates for functional DR techniques applied on the affective full-body movements

	Training recognition rate (%)	**Testing recognition rate (%)**	**Elapsed training time (sec)**
FPCA	44	43	0.70
FFDA	100	37	564.59
F-Isomap	47	44	1.81
FSPCA-Linear	44	44	1.68
FSPCA-GBRF	**59**	**54**	**1.71**

3.1.5 Discussion

FFDA separates class-specific training observations in distinct clusters (nearly a single point; FFDA embedding in Figures 3.2 and 3.3). However, it fails to accurately separate the test observations (low testing recognition rates of 30% and 37% for the hand and full-body movements, Tables 3.1 and 3.2). The FFDA overfitting effect observed for the hand movements (Figure 3.2) and for the full-body movements (Figure 3.3) is due to the small number of high-dimensional observations used (smaller number of observations than the

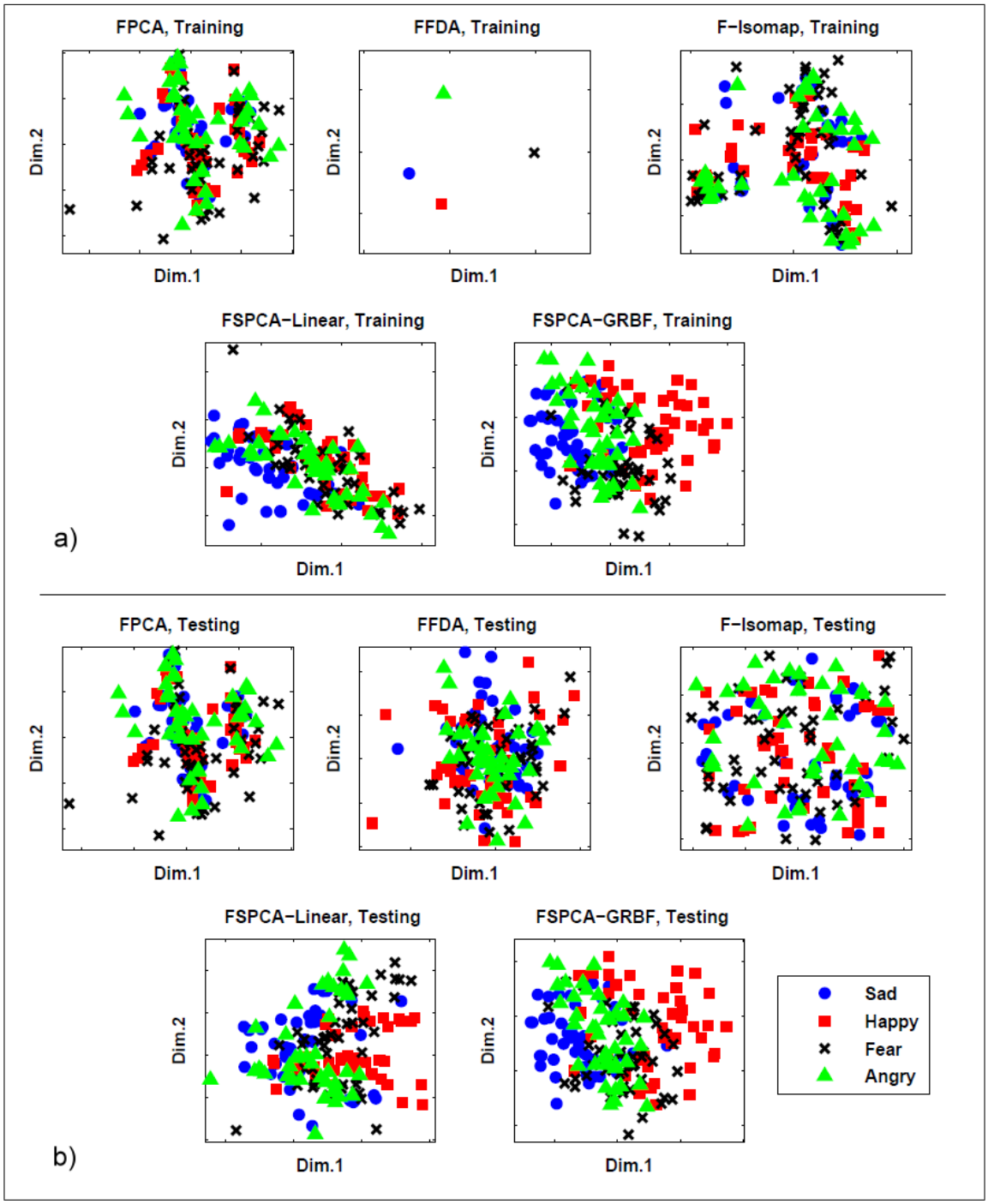

Figure 3.3: Affective full-body movement embedding for a) training data, b) testing data in the resulting 2D subspaces.

dimensionality of the observation), and confirms earlier findings that FDA performs poorly on high dimensional problems when few training points are available [159].

Separated clusters of hand movements belonging to the same affective class are clearly observable in the reduced subspaces of FPCA, F-Isomap, and FSPCA (Figure 3.1). These distinct clusters correspond to the movements performed on the left and right hands. Furthermore, in the FPCA-Linear and F-Isomap embeddings, angry and happy hand movements overlap to some extent, while sad hand movements form distinct clusters (Figure 3.1).

F-Isomap achieves good performance only for the training hand movements (96%, Table 3.1). As Isomap does not provide a parametric transformation that can be used for evaluating the generalizability of the resulting reduced subspace to out-of-sample movements, an approximation of Isomap out-of-sample embedding proposed in [155] is adapted here to test the generalizability of the F-Isomap to unseen movements. This approximation might be the reason for the poor testing performance of the F-Isomap (57%, Table 3.1). Furthermore, the performance of Isomap deteriorates if the datapoints belong to disjoint underlying manifolds, which might be the case here [160].

For the full-body dataset, among the functional DR techniques, FSPCA-GRBF embedding shows dense and more distinct within-class clusters of movements in the resulting low-dimensional space. By visual inspection, it is easy to associate different subintervals of dimensions of the FSPCA-GRBF subspace to distinct affective movements. For instance, lower values of the first dimension of the FSPCA-GRBF subspace are occupied by sad movements, whereas happy movements are distributed along the higher values of the first dimension. FSPCA-GRBF results in the highest LOOCV training recognition rate (59%, Table 3.2) and testing recognition rates (54%, Table 3.2). FPCA, F-Isomap, and FSPCA-Linear embeddings show a large overlapping between full-body movements from different classes, resulting in poor discrimination between the training and testing affective movements (Table 3.2).

As discussed in Section 2.4, in the affective movement recognition literature, the automatic interpersonal emotion recognition rates range from 40% to 70% [123, 121, 124, 75], depending on the number of intended emotions, number of demonstrators, and the amount of within-class kinematic variations in the movements. Using apex postures from 108 of the movements in the full-body dataset used here, Kleinsmith *et al.* [115] tested human perception of the intended emotions. The overall recognition rate was 55% with the least recognized postures being fearful ones (50% recognition rate) and the most recognized ones being the sad postures (63% recognition rate). The FSPCA-GRBF DR applied on the full-body dataset achieves overall training recognition rate of 59% (Sad: 66%, Happy:

64%, Fearful: 51%, Angry: 56%) and testing recognition rate of 53.6% (Sad: 61%, Happy: 62%, Fearful: 45%, Angry: 46%), which are comparable to human recognition rates on the same dataset as reported in Kliensmith *et al.* perceptual study [115].

If we consider the extent to which each class in the reduced subspaces is spread as a measure of quality of the embeddings of the functional DR techniques, by visual inspection, one can argue that FSPCA-GRBF kernel results in the most compact embedding of the classes for both hand and full-body movements. In the case of FFDA, despite the compact embedding of each class to a single point, as discussed above, poor embedding of the test observations is observed due to overfitting. The compact embeddings of the high-dimensional movements facilitate the interpretation of the reduced subspace dimensions, as distinct subintervals of these dimensions can be associated to distinct affective movement classes (i.e., in FSPCA-GRBF embedding shown in Figure 3.2, sad movements are uniquely characterized by lower values of dimension 1).

The resulting functional transformations can be further explored to identify salient characteristics associated with different emotion classes. To this end, transformations obtained by FPCA, FFDA, and FSPCA for the hand dataset are plotted as perturbations of the overall mean of the feature; $\mu(t) \pm \alpha_p w(t)$, where $\mu(t)$ is the functional feature mean across the movements, α_p is the perturbing constant, and $w(t)$ is the functional transformation corresponding to that feature. Figure 3.4 shows examples of perturbation plots for the hand movements resulting from FPCA, FFDA, and FSPCA techniques: Z-trajectory of joint A of the thumb corresponding to the second dimension of the reduced subspaces and Y-trajectory of joint C of the middle finger corresponding to the first dimension of the reduced subspaces. The perturbation plots for other hand coordinates can be obtained similarly. These perturbation plots help to evaluate the importance of different movement functional features in constructing the discriminative reduced spaces either as a whole or over subintervals. For the Y-trajectory of joint C of the middle finger, the perturbation plots for FPCA, and FSPCA with linear and Gaussian RBF kernels are quite similar while being different from the one corresponding to FFDA embedding (the right column in Figure 3.4). According to these perturbation plots, for the Y-trajectory of joint C of the middle finger, the functional feature variations at the beginning and the end of the movements play an important role in producing the FPCA and FSPCA subspaces. If a functional feature has little effect in producing the discriminative subspaces, it appears as overlap of the functional feature mean with its positive and negative functional transformation perturbations as is the case in the FFDA perturbation plot for the Y-trajectory of joint C of the middle finger. Therefore, the contribution of the Y-trajectory of joint C of the middle finger in constructing FFDA embedding is not significant. In both the FPCA and FSPCA subspaces, sad movements are embedded along the lower values of the

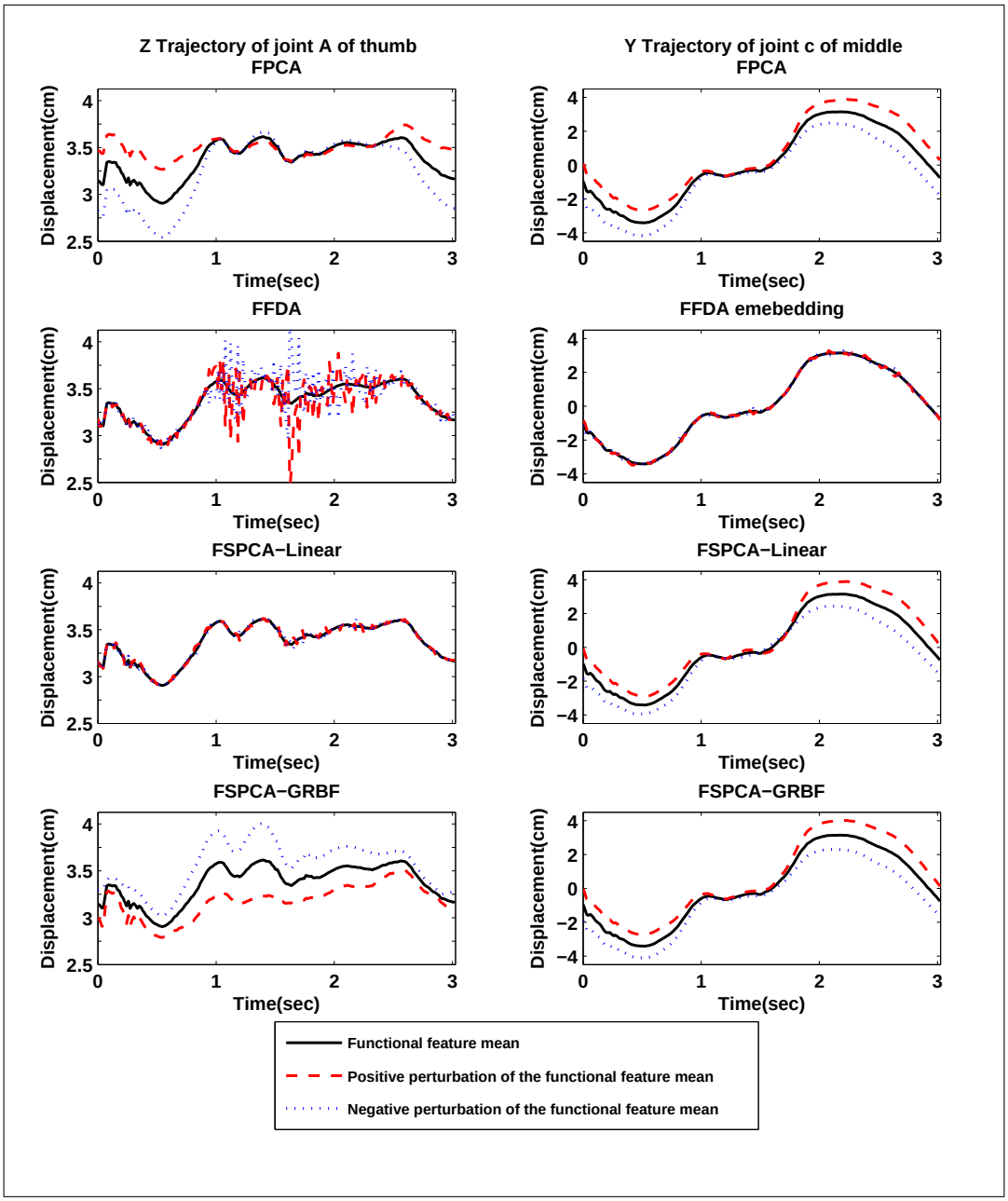

Figure 3.4: Perturbation plots corresponding to the reduced subspaces for the hand movements produced by different functional DR techniques: Z-trajectory of joint A of the thumb corresponding to the second dimension of the reduced subspaces (left column), Y-trajectory of joint C of the middle finger corresponding to the first dimension of the reduced subspaces (right column).

first dimension, while happy and angry movements are characterized by higher values of the first dimension (Figure 3.2); hence similar perturbation plots for the first dimension of the FPCA and FSPCA subspaces are obtained. Differences between FPCA and FSPCA embeddings occur along the second dimension.

An example of the perturbation plots for the Z-trajectory of joint A of the thumb corresponding to the second dimension of the reduced subspaces produced by FPCA, FFDA, and FSPCA is shown in the left column of Figure 3.4. For the Z-trajectory of joint A of the thumb, the perturbation plot for FPCA demonstrates that the functional feature variations at the beginning and the end of the movements play an important role in producing the FPCA embedding. For the FFDA embedding, the functional feature transformation introduces a highly variable trend of weights starting at one-third of the movement and attenuating toward the end (FFDA perturbation plot in the left column of Figure 3.4). This demonstrates the FFDA search for a direction in high-dimensional movement space that maximally separates different movement classes, while forming compact classes along that direction through weighting individual basis functions. The contribution of the Z-trajectory of joint A of the thumb in constructing the FSPCA-linear embedding is not significant. The perturbation plot for the FSPCA-GRBF shows that the entire Z-trajectory of joint A of the thumb plays an important role in constructing the discriminative FSPCA-GRBF embedding. Therefore, FPCA and the two variations of FSPCA DR techniques result in a different functional feature transformation for discriminative embedding of the affective movements. The superiority of the FSPCA techniques over the FPCA in the discriminative analysis (Table 3.1) is likely due to the fact that FSPCA benefits from movement labels in constructing the discriminative lower-dimensional subspace.

Among the functional DR techniques covered here, FFDA is the most computationally expensive and this is due to the requirement for computing the overall covariance as well as individual class covariances. The least computationally expensive is FPCA followed by FSPCA. The computational complexity of the F-Isomap algorithm depends on the computation of pairwise geodesic distances [150].

3.2 Functional Affective Movement Generation

In this section, an approach for generating prototypical movements for different emotion classes in a dataset is presented. First, movements in the dataset are preprocessed through landmark alignment and length normalization. The transformed movements are then represented in terms of functional features, and functional principal component analysis (FPCA)

is used to identify features representing most of the variance within the dataset. The centroids of within-class clusters of movements in a space spanned by the resulting functional principal components along with the inverse FPCA mapping are used to generate class-specific prototypical movements.

The affective hand movement dataset described in Section 3.1.3 is used to demonstrate the performance of the proposed functional affective movement generation approach. To confirm that synthetically-generated movements are perceived in a similar way to human-generated movements, a user study was conducted. Participants were asked to rate the affective expression perceived from human-generated and synthetically-generated affective movements.

3.2.1 Landmark Alignment and Length Normalization

Human movements are highly variable in terms of amplitude, phase, and length, even when the same demonstrator repeats a single movement multiple times. To normalize the data with respect to time while aligning the relative temporal locations of key points in the movements (the landmarks), landmark alignment and length normalization is performed. Here, landmark alignment and length normalization are done in two sequential steps: 1) within-class landmark alignment and length normalization, and 2) between-class length normalization.

The within-class landmark alignment and length normalization are performed simultaneously at the level of individual time series features, using piece-wise linear re-sampling. Movement landmarks are selected as the start, end, and extrema locations of local joint Euler angle trajectories. To improve accuracy, the landmarks are selected manually for every time series feature across the within-class movements. Following landmark selection, the rest of the pre-processing (landmark alignment, length normalization and basis function expansion), feature extraction and subsequent affective movement generation are all automated processes. After selecting the landmarks, an exemplar from each class is selected as the reference movement and the length and landmarks of other movements from the same class are normalized and aligned to those of the reference movement using piecewise linear re-sampling.

To enable the application of FPCA, between-class length normalization is performed by converting the previously transformed movements (fixed-length and aligned within-class movements) from different classes to fixed-length (equal to the length of the shortest within-class transformed movement) vectors using linear re-sampling. After the between-class movements are length normalized, their time-series features are decomposed into

temporal functions computed at every time instant t using basis function expansion (BFE) as described in Section 3.1.1.

3.2.2 Prototypical Movement Generation

For generating the prototypical movements, FPCA is applied on the preprocessed movements and the movements are embedded into the resulting FPCA space. The centroids of the within-class clusters of movements in the FPCA space are used as the representatives of different emotion classes. The inverse FPCA functional feature transformations (i.e., top eigenfunctions) are then used to reconstruct the high-dimensional functional features for the representatives using Equation 3.4. Next, the reconstructed functional features of each class representative are linearly re-sampled to the average length of the original movements in the corresponding emotion class. The resulting functional features are the local Euler angle trajectories for different body joints of the prototypical movement for that emotion class.

3.2.3 Experiments

The affective hand movement dataset described in Section 3.1.3 is used to demonstrate the performance of the generation approach. The hand is an important medium for communicative gestures [82], [161]. Furthermore, the proposed generation approach is originally motivated by the development of a series of architectural responsive environments, called the *Hylozoic Series* [78] (Figure 3.5), and the hand closely resembles the motion style and structure of the moving components of these environments.

The angry and sad movements in the affective hand movement dataset are characterized by high and low tempo, respectively. The happy movements are characterized by fast tempo as well as multiple changes in velocity and direction. The affective hand movements (their local Euler angle trajectories) are first preprocessed through landmark alignment and length normalization as described in Section 3.2.2. An example of length normalization and landmark alignment is shown in Figure 3.6. Basis function expansion is then carried out before the application of FPCA, as described in Section 3.1.1. Here, the number of B-splines and their degree are chosen empirically to be 150 and 4, respectively. The basis function expansion resulted in the R^2 statistic above 92% for all the time-series features. In addition, the goodness of fit for each individual time-series feature is visually monitored using plots of the original time-series, their BFE fit, and their corresponding residual plots to ensure the quality of the fit over all subintervals.

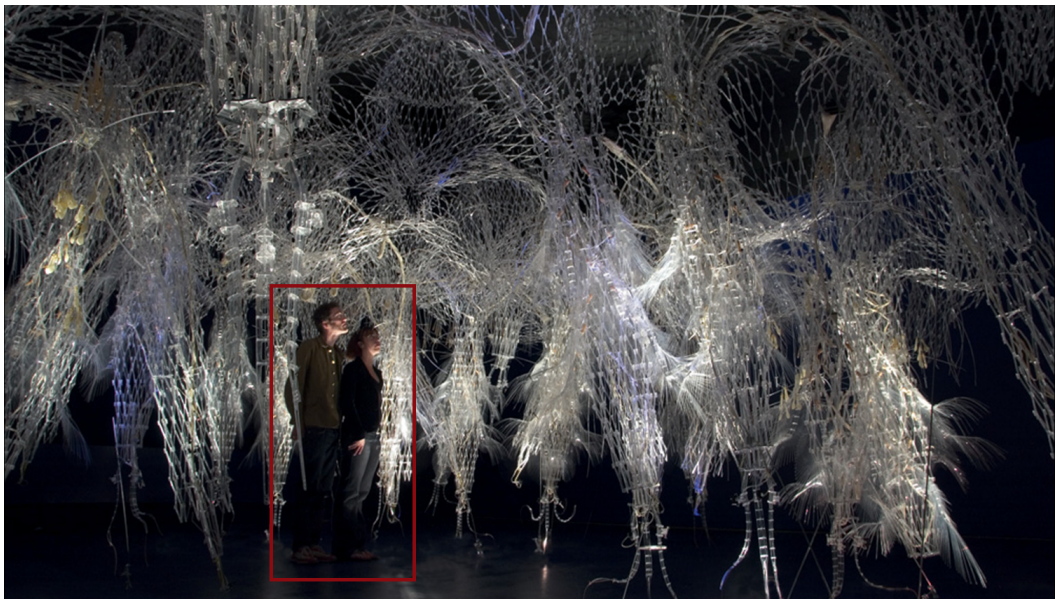

Figure 3.5: Two visitors highlighted with a red outlined rectangle immersed in Hylozoic series, a responsive architectural geotextile environment [3]. The Hylozoic series uses massively repeating components, microprocessors, sensors and actuators to create decentralized responsive systems capable of subtle motions giving the impression that the environments are 'sensitive' and may even have affective states. Knowledge of key contributing movement features in conveying affect and methods for automatic affective movement generation would enable these interactive environments to display movements that convey recognizable affective expressions and engage in meaningful interaction with humans. Reprinted with permission.

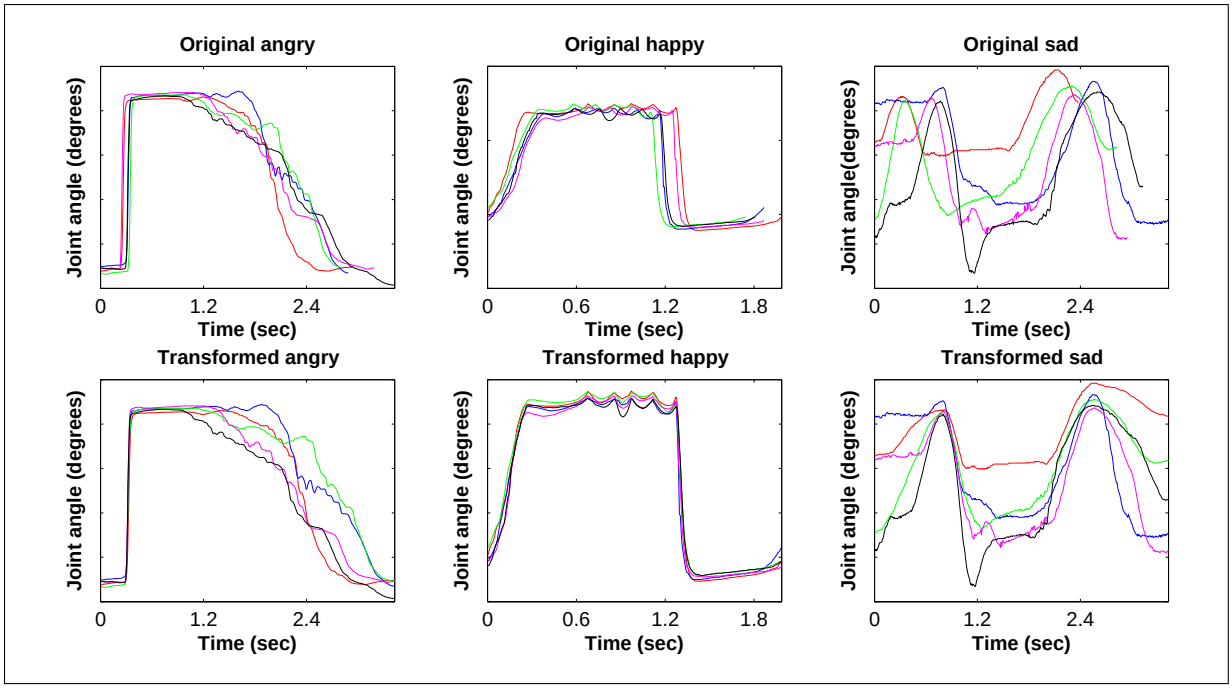

Figure 3.6: An example of landmark alignment and length normalization for time-series features representing local Euler angle trajectories for: X-rotation of joint B of the little finger in angry movements (right column), X-rotation of index's joint A in happy movements (middle column), Z-rotation of wrist in sad movements (left column).

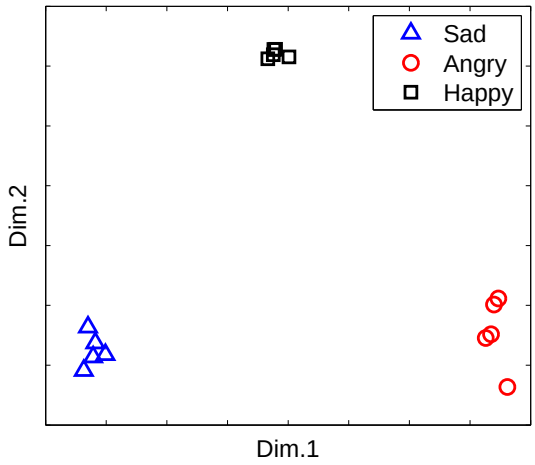

Figure 3.7: FPCA 2D embedding of affective human hand movements.

Next, the main modes of variation in the resulting transformed movements are found using FPCA, and the corresponding two-dimensional (2D) subspace is obtained (Figure 3.7).

As can be seen in Figure 3.7, there is a clear separation between different classes of affective hand movements in the FPCA 2D embedding, as within-class movements are embedded near one another, forming dense clusters distant from clusters corresponding to other classes of the affective hand movements. Therefore, any clustering algorithm (e.g., k-means) can find the 3 clusters each containing instances of a class of affective movement. Using the centroids of these clusters, the high dimensional joint trajectories of the prototypical movements for different emotion classes are constructed using the approach described in Section 3.2.2. FPCA 3D embedding is used for movement generation, as movement generation from FPCA 2D space compromised some fine movement details. Videos of the animated embodiments using the original and regenerated movements are available at [162].

Since human interpretation of displayed affective movements can be very subjective, subjective measures need to be employed to assess the perception of generated affective movements in comparison with those displayed by the demonstrator. For this purpose, a user study was conducted in which participants rated the expressivity of the synthetically-generated and human-generated movements animated on a human-like hand model (Figure 3.8). Therefore, there are two movement generation sources (human-generated movements (Original) and synthetically-generated movements (Generated)), and three intended emo-

Figure 3.8: The human-like hand model used to display the generated and original movements.

tions (happiness, sadness, and anger) in the user study.

Each questionnaire session included 6 animations: 3 showing the original movements and the other 3 animations showing the generated movements. The animations were shown to the participants in a randomized order. Each animation was accompanied by two questions. The first question asked the participants to rate on a Likert scale the extent to which each of the six basic Ekman emotions was conveyed in the animation, with 1 being "not conveyed at all" and 6 being "strongly conveyed". Offering participants the choice of six emotions gives a more accurate picture of recognition rate, since it does not artificially constrain the responses and shows whether emotions are unambiguously recognized. In the second question, participants were asked to rate the arousal and valence components of the emotion perceived for each movement, using a 7-point scale. A description of the arousal and valence dimensions of emotion was provided, along with a schematic representation of circumplex model of emotion (Figure 3.9). Low intensity-high intensity and unpleasant-pleasant were the adjective pairs displayed at the extremes of arousal and valence scales, respectively, to further guide the participants in evaluating the arousal and valence components.

A total of 22 participants (26.1 years $\pm$ 5.8 years, 12 male and 10 female) completed the user study. Participants were healthy adults and recruited from among students at the University of Waterloo. They were provided with detailed information on the study and the procedure to complete the computer-based questionnaire. All the questionnaire sessions took place at the same location and were administered by the same researcher to ensure a uniform experience for all the participants. The study received ethics approval from the Office of Research Ethics, University of Waterloo, and a consent form was signed

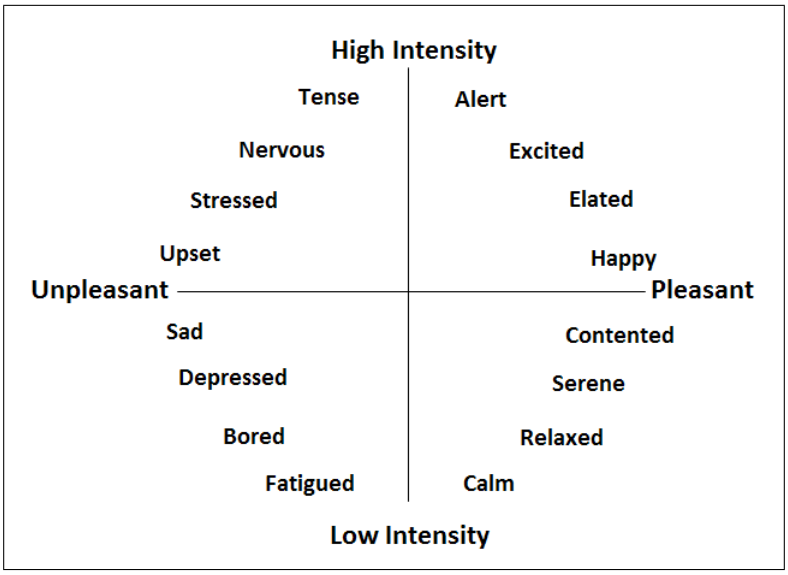

Figure 3.9: The schematic representation of the cicumplex model used in the questionnaire.

electronically by each participant prior to the start of the questionnaire.

3.2.4 Results

The main and interaction effects of the generation source and intended emotion on the perception of original and generated affective movements are evaluated using participants' responses in the user study[3].

The order of the affective movement animations seen by the participants was randomized. Using one-way ANOVA, we tested if viewing multiple animations caused any habituation by considering the presentation order as an independent variable. The effect of presentation order on the participants' evaluations of the animated movements was not found to be significant at $p < 0.05$, indicating that habituation was not a significant effect.

Bar charts of the ratings of anger, happiness, sadness, arousal, and valence for each generation source and intended emotion combination, averaged over the participants, are shown in Figure 3.10. Five two-way repeated measure ANOVAs are performed, each testing the main and interaction effects of the generation source (original and generated) and intended emotions (anger, happiness, and sadness) on the participants' ratings of anger,

[3]The SPSS statistical software package [163] is used to analyse the user study results.

Table 3.3: Null hypotheses tested in the repeated measure ANOVAs performed to evaluate the effects of the generation source and intended emotion on the participants' perception; $i = \{Anger, Happiness, Sadness, Arousal, Valence\}$.

$H_0^A(1,i)$:	*The means of the participants' ratings of i for original and generated movements are equal.*
$H_0^A(2,i)$:	*The means of the participants' ratings of i for different intended emotions are equal.*
$H_0^A(3,i)$:	*Generation sources and intended emotions are independent and no interaction effect between the two is present on the ratings of i.*

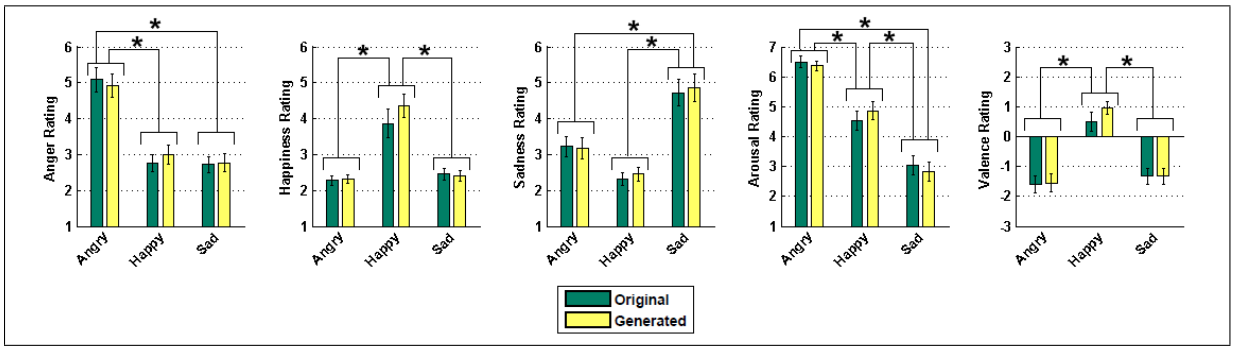

Figure 3.10: Average participants' ratings (mean ± SE) for the original and generated affective movements displayed on the human-like structure. From left to right, ratings for: anger, happiness, sadness, arousal, valence. "*" sign indicates a significant pair-wise difference between the levels of intended emotion. SE: standard error.

happiness, sadness, arousal, and valence. Table 3.3 shows the null hypotheses tested in each repeated measure ANOVA.

The main and interaction effects of the independent variables (generation source and intended emotion) on the participants' ratings of the dependent variables (ratings of anger, happiness, sadness, arousal, and valence) are considered significant at $p < 0.05$. Table 3.4 shows the resulting F-statistics, p-values, and effect sizes (η^2). In addition, a post-hoc analysis consisting of paired t-tests with Bonferroni correction is performed to identify the significant differences from the ANOVA tests.

Table 3.4: F-statistics, p-values, and effect size (η^2) results from repeated measure ANOVAs for the main and interaction effects of generation source and intended emotion on participants' ratings of anger, happiness, sadness, arousal, and valence. Greenhouse-Geisser correction is used when sphericity assumption is violated. "*" sign indicates a significant difference.

	$i:$	Anger	Happiness	Sadness	Arousal	Valence
Generation source	$F(1,21) =$	0.023	2.333	0.427	0.007	0.805
$\left(H_0^A(1,i)\right)$	$p =$	0.882	0.142	0.520	0.934	0.380
	$\eta^2 =$	0.000	0.004	0.001	0.000	0.003
Intended Emotion	$F(2,42) =$	31.191	27.038	28.030	49.009	45.015
$\left(H_0^A(2,i)\right)$	$p =$	0.000*	0.000*	0.000*	0.000*	0.000*
	$\eta^2 =$	0.399	0.360	0.358	0.559	0.398
Generation source ×	$F(2,42) =$	0.461	2.398	0.114	1.123	0.543
Intended emotion	$p =$	0.634	0.121	0.797	0.335	0.585
$\left(H_0^A(3,i)\right)$	$\eta^2 =$	0.003	0.007	0.001	0.004	0.004

3.2.4.1 The Effect of Generation Source

To analyze the effect sizes, we apply Cohen's recommended magnitudes for effect sizes, with a η^2 of 0.01, 0.06, and 0.14 indicating small, medium, and large effect sizes, respectively [164]. According to the ANOVA results, the interaction between the generation source and intended emotion is not significant and the corresponding effect sizes are small ($\eta^2 <= 0.01$) in all cases (Table 3.4); hence, we can not reject $H_0^A(3,i)$.

The observed effect of the generation source and the paired differences in the participants' ratings of anger, happiness, sadness, arousal, and valence with respect to the generation source are not significant as the corresponding p-values are larger than the Bonferroni-corrected significance level of 0.025. Therefore, we retain the first set of null hypotheses ($H_0^A(1,i)$; equal mean ratings of movements from different generation sources for all affective labels). Furthermore, the observed effects of the generation source are of a small size ($\eta^2 < 0.004$); hence, the generation source is responsible for only a very small portion of variability ($< 0.4\%$ of the overall variability (effect + error)) in the participants' perception of the movements.

Therefore, the small differences in the perception of the original versus generated movements are likely due to chance and the generation source has very little impact on the participants' perception of the affective movements. This indicates that the movements

generated with the proposed functional approach are perceived very similarly to the original movements displayed by the human demonstrator.

3.2.4.2 The Effect of Intended Emotion

The intended emotion has a significant effect on the participants' perception of the affective movements in all cases (rejecting $H_0^A(2, i)$ for all i's). The detected main effect of the intended emotion is large ($\eta^2 > 0.14$) in all cases (Table 3.4), indicating that the intended emotion is a critical factor in evaluating the expression in different animations.

Using the follow-up post-hoc analyses, with respect to the intended emotions, there are significant pairwise differences between the participants' ratings of the intended emotions (anger, happiness, and sadness) at a Bonferroni corrected significance level ($p < 0.017$). For instance, anger ratings for the angry movements are significantly higher than the anger ratings for the sad and happy movements (anger rating in Figure 3.10). All the pairwise differences in the arousal ratings in light of the intended emotions are significant at $p < 0.017$ (arousal rating in Figure 3.10). For the valence rating, happy movements are significantly different from either the sad or angry movements. The observed pairwise differences between the valence rating of the sad and angry movements are not significant (valence rating in Figure 3.10), which might be due to the similarity in their valence characteristics (negative valence).

3.2.5 Discussion

The results of the ANOVA tests show that all the medium to large main effects can be detected with our sample size (22 participants). When no significant effect was observed, the size of the effect was extremely small. The extremely small effects are likely due to chance and hence, the generation source has very little influence on the participants' perception of the affective movements. On the other hand, the significant effects detected in the user study indicate that human perception of the affective movements is strongly influenced by the intended emotion. Therefore, the proposed approach is capable of generating prototypical movements for different affective categories that are as expressive as their counterparts displayed by the human demonstrator.

However, there are limitations to the proposed generation approach. This generation approach is only capable of generating a prototypical movement for a collection of within-class affective movements that are also kinematically similar. Furthermore, the functional

generation approach is sensitive to mis-alignment of spatial landmarks and length variations between the within-class kinematically-similar training movements and mandates a preprocessing step where landmarks are identified and aligned followed by length normalization. Landmark alignment and length normalization can be a tedious task for large datasets, which in turn will impede the application of the proposed approach for large scale movement generation problems. Therefore, other approaches are warranted to generate new movements that are kinematically different from the training movements and convey a target emotion, and do not require landmark alignment and length normalization. The automatic generation approach presented in Chapter 5 aims to address these limitations.

3.3 The Effect of Embodiment and Gender on the Perception of Affective Movements

Identifying human capabilities in perceiving affective expressions is essential for developing interactive machines that can engage with their human users. In order to ensure that the behaviour of the interactive machines is perceived as intended, any embodiment- or gender-specific differences in the perception of affective expressions are an important design consideration. This section presents a user study conducted to evaluate the role of the display embodiment and an observer's gender on the perception of affective movements.

3.3.1 Experiments

The generated affective hand movements from Section 3.2.3 were displayed on a human-like and a frond-like hand model 3.11. Therefore, there are two embodiments in the user study (human-like and frond-like), three intended emotions (happiness, sadness, and anger), and the gender is the between-subject factor. It should be emphasized that the movements used in this experiment are the synthetically-generated ones only. Therefore, by fixing the generation source, we will investigate the role of intended-emotion and display embodiment in the perception of synthetically-generated movements displayed on the human-like and the frond-like embodiments.

The same participants as those reported in Section 3.2.3 completed the user study (22 participants: 26.1 years $\pm$ 5.8 years, 12 male and 10 female). They were provided with detailed information on the study and the procedure to complete the computer-based questionnaire. All the questionnaire sessions took place at the same location and were

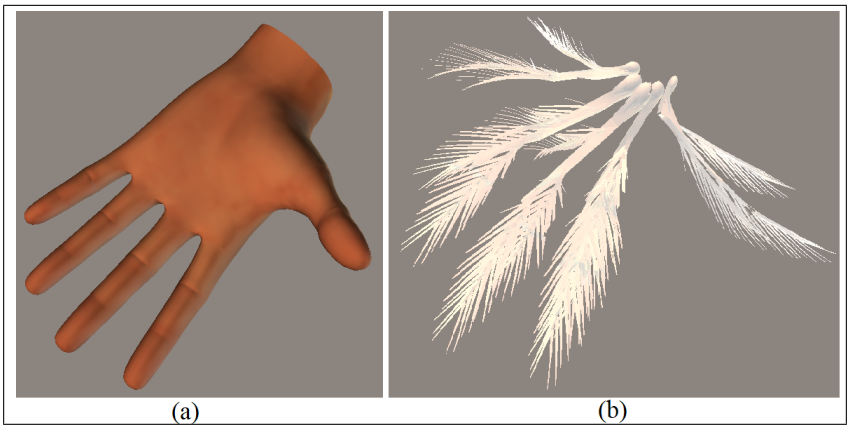

Figure 3.11: Display embodiments used to display affective movements. a) anthropomorphic (human-like) hand model, b) non-anthropomorphic frond-like hand model. These animated embodiments are produced using Poser (version 8, Smith Micro Inc.).

administered by the same researcher to ensure a uniform experience for all the participants. The study received ethics approval from the Office of Research Ethics, University of Waterloo, and a consent form was signed electronically by each participant prior to the start of the questionnaire.

Each questionnaire session included 6 animations: 3 showing the generated movements on the human-like embodiment and the other 3 animations showing the generated movements displayed on the frond-like embodiment. The following naming format is used to refer to the animations in the rest of this section: "(embodiment: human-like, frond-like)_(intended emotion: angry, happy, sad)" (e.g., "Frond_Happy" represents the happy movement displayed on the frond-like embodiment).

The animations were shown to the participants in a randomized order. Each video was accompanied by three questions. The first question was a multiple-selection question asking participants to select among a list of keywords those that most closely described the animated embodiment in the video. The list of keywords and average user responses are shown in Table 3.5. The second and third questions asked the participants to rate the six basic Ekman emotions and arousal and valence in the same way as the user study presented in Section 3.2.3.

Table 3.5: The percentage (%) of descriptive keywords selected for different combinations of the intended emotion and embodiment. The descriptive keywords selected over 50% of the time are highlighted.

Embodiment	Hand			Frond		
Intended emotion	Sad	Happy	Angry	Sad	Happy	Angry
Human-like	**73**	**64**	**64**	14	14	18
Robot-like	18	18	36	18	23	14
Organic	36	36	23	50	41	**55**
Artificial	27	27	36	23	32	27
Hand-like	**73**	**77**	**68**	23	18	36
Plant-like	0	0	0	**73**	**59**	**64**
Represents biological motion	**64**	**64**	50	**55**	50	50
Represents mechanical motion	5	18	27	14	27	23
Friendly	32	23	23	41	23	5
Unfriendly	5	0	45	5	23	50
Cartoon-like	9	18	9	18	14	14
Pleasant	5	27	0	45	18	9
Unpleasant	23	5	**59**	0	9	41
Aggressive	5	18	**64**	0	41	41
Gentle	**59**	18	0	**77**	27	18
None	0	0	0	0	0	0

3.3.2 Results

The effect of presentation order on the participants' perception of the animated embodiments was not found to be significant at $p < 0.05$ using one-way ANOVA, indicating that habituation was not a significant effect.

The participants' perception of animated embodiments was assessed using the set of descriptive keywords listed in Table 3.5. As expected, the human-like embodiment was frequently described as human-like (67%) and hand-like (73%), regardless of the intended emotion. The frond-like embodiment was described as plant-like in 65% of the responses. Both the human-like and frond-like embodiments were perceived to be representing biological motion in 59% and 52% of the responses, respectively. Notably, the frond-like embodiment displaying the sad movement was frequently described as gentle (77%), friendly (41%), and pleasant (45%), while these descriptors were not commonly associated with the human-like embodiment. The human-like embodiment displaying angry movements was perceived as unpleasant (59%) and aggressive (64%), whereas these descriptors were less frequently attributed to the angry movement displayed on the frond-like embodiment. The participants' ratings of affective movements displayed on the human-like and frond-like embodiments are shown in Figure 3.12.

In order to test whether the physical appearance of the embodiment influences the participants' ratings of the animated affective movements, five two-way repeated measure ANOVAs are performed with independent variables being the embodiment and intended emotion. Table 3.6 shows the null hypotheses tested in each repeated measure ANOVA. The resulting F-statistics, p-values, and effect sizes (η^2) are reported in Table 3.7. The ANOVA results are considered significant at $p < 0.05$. Paired t-test comparisons with Bonferroni correction are carried out to further assess the significant effects reported in Table 3.7 at different levels of the independent variables.

3.3.2.1 The Effect of Intended Emotion

The interaction between the embodiment and intended emotion is significant in the participants' ratings of happiness, sadness, and valence, with relatively large sizes (Table 3.6); hence, rejecting $H_0^B(3, happiness)$, $H_0^B(3, sadness)$, and $H_0^B(3, valence)$. According to the ANOVA tests, similarly to Section 3.2.4, the intended emotion has a main effect on the participants' perception of the affective movements and demonstrates large effect sizes (Table 3.7). Therefore, we reject the first set of hypotheses in Table 3.6 ($H_0^B(1, i)$) that states the mean participants' ratings of different levels of intended emotions are equal. For the arousal rating, significant pairwise differences in ratings of different affective movements

Table 3.6: Null hypotheses tested in the repeated measure ANOVAs performed to evaluate the role of embodiment and intended emotion on the participants' perception; $i = \{Anger, Happiness, Sadness, Arousal, Valence\}$.

$H_0^B(1,i)$:	*The means of the participants' ratings of i for different intended emotions are equal.*
$H_0^B(2,i)$:	*The means of the participants' ratings of i for different embodiments are equal.*
$H_0^B(3,i)$:	*Embodiment and intended emotions are independent and no interaction effect between the two is present on the ratings of i.*

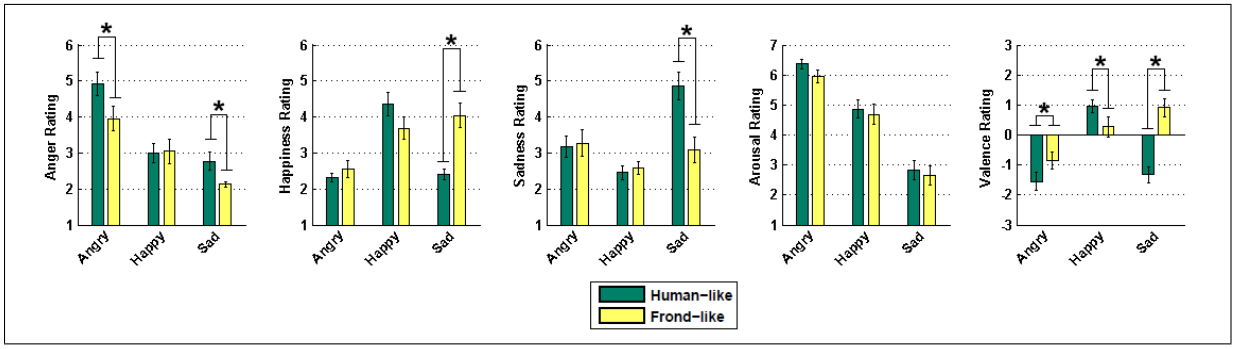

Figure 3.12: Average participants' ratings (mean ± SE) for the affective movements displayed on the human-like and frond-like embodiments. From left to right, ratings for: anger, happiness, sadness, arousal, valence. "*" indicates a significant pairwise difference between the levels of the embodiments.

Table 3.7: F-statistics, p-values, and effect sizes (η^2) from two-way repeated measure ANOVAs each testing the main and interaction effects of embodiment and intended emotion on participants' ratings of anger, happiness, sadness, arousal, and valence. Greenhouse-Geisser correction is used when sphericity assumption is violated. "*" sign indicates a significant difference.

	i :	Anger	Happiness	Sadness	Arousal	Valence
Intended Emotion	$F(2,42)$ =	24.406	23.710	9.453	46.660	19.638
$\left(H_0^B(1,i)\right)$	p =	0.000*	0.000*	0.000*	0.000*	0.000*
	η^2 =	0.283	0.198	0.138	0.532	0.201
Embodiment	$F(1,21)$ =	6.081	5.502	12.066	1.569	17.121
$\left(H_0^B(2,i)\right)$	p =	0.022*	0.029*	0.002*	0.224	0.000*
	η^2 =	0.027	0.019	0.026	0.004	0.050
Embodiment ×	$F(2,42)$ =	2.448	10.450	10.564	0.196	14.310
Intended emotion	p =	0.099	0.001*	0.001*	0.767	0.000*
$\left(H_0^B(3,i)\right)$	η^2 =	0.017	0.107	0.077	0.001	0.128

are observed regardless of the embodiment. It should be noted that the movements used in this experiment are the synthetically-generated ones and the effect of intended emotion and display embodiment on the perception of affective movements is evaluated, unlike the experiment in Section 3.2.4 in which the affective movements were either human-generated or synthetically-generated. In the case of anger, happiness, sadness, and valence ratings for the movements displayed on the human-like embodiment, pairwise differences between different levels of the intended emotion are present. However, due to the interaction effect between the embodiment and intended emotion, pairwise differences at different levels of the intended emotion demonstrate a different trend for the frond-like structure (Figure 3.12). These differences are discussed in the next section.

3.3.2.2 The Effect of Embodiment

Embodiment has a significant main effect on the participants' ratings of anger, sadness, and valence at the Bonferroni corrected level $p < 0.025$, rejecting the null hypotheses $H_0^B(2, anger)$, $H_0^B(2, sadness)$, and $H_0^B(2, valence)$ in Table 3.6. Furthermore, the effect of embodiment on the ratings of happiness is found marginal at $p < 0.025$. The observed sizes of the embodiment effects on anger, happiness, sadness, and valence ratings are medium (Table 3.7). Post-hoc paired t-tests between the ratings of the movements dis-

played on different embodiments were performed to identify significant differences detected in ANOVA tests.

The most significant effect of the embodiment is due to differences in the perception of sad movements. The sad movement displayed on the frond-like embodiment was perceived as a happy movement and conveying positive valence (happiness and valence ratings in Figure 3.12). Therefore, there is no significant difference between the valence ratings for the happy and sad movements displayed on the frond-like embodiment (valence rating in Figure 3.12), which is not the case when the display embodiment is human-like. This is in-line with the participants' description of the Frond_Sad animation as pleasant (45%), friendly (41%) and gentle (77%). Furthermore, sad movements displayed on the frond-like embodiment are not rated as conveying sad; no significant difference between the sadness rating of this movement and other movements is observed regardless of embodiment.

3.3.2.3 The Effect of Affect Representation Model

This thesis next investigates how different representations of affect (discrete versus continuous) influence participants' responses. Participants were able to distinguish differing levels of arousal in all three classes of affective movements used here (high arousal values for anger, medium arousal values for happiness, and low arousal values for sad movements (arousal ratings in Figures 3.12). These arousal ratings are consistent regardless of embodiment.

For valence, the average ratings show that the angry movements are correctly perceived as conveying negative valence regardless of embodiment (Figure 3.12). Participants' evaluations of valence for the happy movements displayed on the human-like embodiment (Figure 3.12) are positive and consistent with the valence property of happy emotions [165], but high variance is observed for the frond-like embodiment (valence rating in Figure 3.12). For the sad movement displayed on the human-like embodiment, the valence component is rated negatively, which is consistent with the valence property of the sad emotion [165]. However, for the frond-like embodiment displaying the sad movement, the valence rating has a positive average value.

When the display embodiment is human-like, the intended emotion is rated significantly higher than the other emotions communicated in the dataset (anger, happiness, and sadness ratings in Figure 3.12). These rating differences are all found to be significant at $p < 0.017$ using paired t-tests with Bonferroni correction.

Next, the perception for the emotions not communicated in the movement dataset (surprise, fear, and disgust) is considered. For this purpose, we generated a confusion

matrix for the perceived Ekman emotions as shown in Table 3.8. For the confusion matrix, an emotion is considered recognized if it is rated 3 or above on the Likert scale. Note that this recognition cut-off is applied only for illustrative purposes in Table 3.8, and all the analysis in Section 3.2.4 is done on the full scale of ratings obtained in the questionnaire study. When angry movements, both original and generated, are displayed on the human–like embodiment, in addition to being correctly perceived as angry, they are also frequently perceived as fear and surprise. The frond-like embodiment displaying angry movement was perceived as fear and surprise more frequently than anger. The happy movements are perceived both as happiness and as surprise regardless of the generation source and embodiment. For the human-like embodiment, the sad movement is clearly and uniquely recognized. The sad movement executed on the frond-like embodiment is misperceived as happiness.

Table 3.8: Confusion matrix showing percentage (%)* of emotion rating for different affective movements.

	Perceived emotions					
	Anger	Happiness	Sadness	Surprise	Fear	Disgust
Hand_Angry	77%	9%	41%	50%	73%	45%
Hand_Happy	23%	55%	14%	55%	9%	14%
Hand_Sad	18%	14%	77%	23%	32%	23%
Hand_Angry	82%	5%	32%	73%	77%	32%
Hand_Happy	32%	77%	14%	64%	9%	23%
Frond_Angry	55%	14%	36%	73%	77%	36%
Frond_Happy	27%	55%	23%	50%	27%	5%
Frond_Sad	0%	55%	27%	27%	9%	0%

∗ Since participants were asked to rate all 6 Ekman emotions on a 6 point Likert scale, there are cases where an expressive movement was rated 3 or above for more than one emotion. On the other hand, there are cases in which anger, happiness and sadness were all rated below 3. This is why none of the emotion ratings add up to 100% in the confusion matrix.

3.3.2.4 Discussion

The user study shows that display embodiment influences the perception of affective movements as significant differences were observed in the perception of the same affective movement displayed on different embodiments. The effect of embodiment on participants' ratings found in this study contradicts previous findings [104], [105] that reject the role of

Table 3.9: Correlation (%) between the arousal ratings and peak velocity (first column) and peak acceleration (second column).

	Peak velocity	Peak acceleration
Human-like	74%	75%
Frond-like	70%	70%

embodiment on affective movement perception. The animated frond-like embodiment here is different than the one used in a previous study [105]. In [105], the frond-like embodiment was frequently described as aggressive, unfriendly, and unpleasant, while here the frond-like embodiment is described as gentle, friendly, and pleasant when displaying the sad movement. This difference in embodiment perception may be the reason for the differences in the participants' assessment of the affective movements displayed on the frond-like embodiment in [105] as compared to the one reported in this section. One possible hypothesis is that it may be more difficult to perceive negative valence emotions (anger, sadness) on a embodiment which is associated with positive emotional characteristics such as gentleness.

In addition, the results of the user study indicate that the arousal ratings of the affective movements are consistent and comparable with the arousal properties of the corresponding emotions reported in [165], regardless of embodiment. The average valence ratings for the animated human-like embodiment are also consistent and similar to [165], whereas inconsistencies are observed with the frond-like embodiment.

When considering ratings for all six Ekman emotions, there is some confusion between the intended emotions and non-intended emotions as highlighted in Table 3.8. One interpretation is that the Ekman emotions that are easily mis-perceived (e.g., anger, fear, and surprise) have similar arousal-valence characteristics. Based on the results obtained in this study, it appears that the circumplex model of emotion is a more reliable tool for measurement of the emotion for movement stimuli, as it is not clear that participants are capable of distinguishing discrete emotion categories with similar arousal-valence characteristics when observing affective movements alone.

We also computed correlations between the arousal ratings and peak velocity and peak acceleration characteristics of the affective movements (Table 3.9). The resulting correlations corroborate with previous works [29] and [104], which have found that the level of perceived arousal is correlated to the velocity and acceleration of the movements.

3.3.3 The Effect of Gender

The role of gender on the participants' perception of affective movement was also studied using the user study reported in Section 3.3. In particular, the following questions were investigated:

1. Did the gender of the observers have an influence on the perception of affective hand movements?

2. Did the intended-emotion and display embodiment have a different impact on male or female observers?

To investigate the effect of gender and its interaction with the embodiment and intended-emotion on the participants' ratings of the affective movements, a three-way repeated measure ANOVA can be used with gender as a between-subject variable, and display structure and intended-emotion as within-subject variables. However, the interpretation of the significant effects from a three-way repeated measure ANOVA is difficult due to the large number of variables and their main and interaction effects (7 main and interaction effects). Furthermore, a larger sample size would be needed to detect significant effects of all the variables presented in the study.

To reduce the number of effects and simplify the analysis, we instead performed two sets of two-way repeated measure ANOVAs (each set contains five ANOVA tests) to assess the main and interaction effects of the intended-emotion and embodiment on the ratings of anger, happiness, sadness, arousal, and valence by the male participants (set 1) and female participants (set 2). This way, we have reduced the number of variables to two within-subject variables: intended-emotion and embodiment. Table 3.10 shows the null hypotheses tested in each repeated measure ANOVA. Tables 3.11 and 3.12 show the resulting F-statistics, p-values, and effect sizes (η^2) for male and female participants, respectively.

According to the ANOVA results in Table 3.11, there is a significant interaction between embodiment and intended-emotion in the male participants' ratings of anger, happiness, sadness, and valence; hence rejecting $H_0^{male}(3, i)$ for $i = \{anger, happiness, sadness, valence\}$. However, no significant interaction between the intended-emotion and embodiment in the female participants' ratings was observed (Table 3.12); hence, retaining $H_0^{female}(3, i)$'s.

There are also differences in the main effects of the intended-emotion and embodiment on the male and female participants' perception. The intended-emotion was found to significantly influence the ratings of both the male and female participants in all the cases in this study except for the male sadness ratings. The embodiment has a significant

main effect on the female participants' ratings in all the cases at $p < 0.05$ except for the happiness ratings (rejecting $H_0^{female}(2, i)$ for $i = \{anger, sadness, arousal, valence\}$), whereas the effect of embodiment on the male participants' ratings was not found significant at $p < 0.05$.

Bar charts of average ratings of anger, happiness, sadness, arousal, and valence by male and female participants are shown in Figure 3.13. Paired t-tests are performed between the pairs of the male and female participants' ratings of the affective movements displayed on the hand–like and frond–like embodiments and significant pair-wise differences are shown using "*" in Figure 3.13. Table 3.13 shows the confusion matrix of the perception of the intended-emotions. For the confusion matrix, an emotion is considered recognized if it is rated 3 or above on the Likert scale.

As can be seen in Table 3.13, the perception of anger by female participants was significantly affected by the embodiment as the angry movement displayed on the frond-like embodiment was less frequently recognized as conveying anger in comparison with the angry movement displayed on the human-like embodiment (female anger rating of the angry movements in Figure 3.13). The male participants equally attributed high arousal and negative valence to the angry movement and correctly recognized angry movement regardless of the embodiment (male anger, arousal, and valence ratings of the angry movement in Figure 3.13). However, the female participants associated a lower-level of arousal and less-negative valence to the frond-like embodiment displaying the angry movement (female arousal and valence ratings of the angry movement in Figure 3.13). The better performance of the male participants in recognizing angry movements is congruent with [93, 94] suggesting that men are more accurate in recognizing angry expressions.

The happy movement displayed on the human-like embodiment is correctly recognized as conveying happiness and positive valence by both male and female participants, whereas

Table 3.10: Null hypotheses tested in the repeated measure ANOVAs for the participants' ratings; $i = \{Anger, Happiness, Sadness, Arousal, Valence\}$, $G = \{male, female\}$.

$H_0^G(1, i)$:	*The means of the G participants' ratings of i for different intended-emotions are equal.*
$H_0^G(2, i)$:	*The means of the G participants' ratings of i for different embodiments are equal.*
$H_0^G(3, i)$:	*Embodiment and intended-emotions are independent and no interaction effect between the two is present in the G participants' ratings of i.*

Table 3.11: F-statistics, p-values, and effect size (η^2) results from two-way repeated measure ANOVAs each testing the main and interaction effects of embodiment and intended-emotion on male participants' ratings of anger, happiness, sadness, arousal, and valence. There are 12 male participants. Greenhouse-Geisser correction is used when sphericity assumption is violated. "*" sign indicates a significant effect. Bonferroni adjustment was made for multiple comparisons.

	$i:$	Anger	Happiness	Sadness	Arousal	Valence
Intended–Emotion	$F(2,22) =$	15.006	20.749	2.127	18.947	9.981
$\left(H_0^{male}(1,i)\right)$	$p =$	0.000*	0.000*	0.143	0.000*	0.001*
	$\eta^2 =$	0.315	0.244	0.089	0.503	0.162
Embodiment	$F(1,11) =$	0.014	4.068	3.000	1.232	2.129
$\left(H_0^{male}(2,i)\right)$	$p =$	0.908	0.069	0.111	0.291	0.172
	$\eta^2 =$	0.000	0.027	0.014	0.004	0.012
Embodiment ×	$F(2,22) =$	5.421	7.871	8.406	1.329	17.488
Intended-emotion	$p =$	0.012*	0.003*	0.002*	0.285	0.000*
$\left(H_0^{male}(3,i)\right)$	$\eta^2 =$	0.071	0.166	0.103	0.008	0.228

Table 3.12: F-statistics, p-values, and effect size (η^2) results from two-way repeated measure ANOVAs each testing the main and interaction effects of embodiment and intended-emotion on the female participants' ratings of anger, happiness, sadness, arousal, and valence. There are 10 female participants. Greenhouse-Geisser correction is used when sphericity assumption is violated. "*" sign indicates a significant effect. Bonferroni adjustment was made for multiple comparisons.

	$i:$	Anger	Happiness	Sadness	Arousal	Valence
Intended–Emotion	$F(2,18) =$	8.825	7.676	14.333	33.081	15.221
$\left(H_0^{female}(1,i)\right)$	$p =$	0.002*	0.004*	0.000*	0.000*	0.000*
	$\eta^2 =$	0.254	0.174	0.230	0.612	0.311
Embodiment	$F(1,9) =$	16.308	1.385	11.000	11.184	38.383
$\left(H_0^{female}(2,i)\right)$	$p =$	0.002*	0.269	0.009*	0.009*	0.000*
	$\eta^2 =$	0.114	0.009	0.042	0.047	0.122
Embodiment ×	$F(2,18) =$	1.619	2.739	3.508	0.360	2.521
Intended-emotion	$p =$	0.226	0.092	0.084	0.703	0.108
$\left(H_0^{female}(3,i)\right)$	$\eta^2 =$	0.019	0.050	0.064	0.003	0.060

the frond-like embodiment displaying happiness is less frequently recognized as happy. The male participants frequently misperceived the happy movement displayed on the frond-like embodiment as conveying anger, which might be the reason for the slightly negative valence attributed to the Frond_Happy movement by the male participants. Frond_Happy movement is correctly recognized by the female participants. Although there is a significant difference between the average arousal ratings of the Frond_Happy and Human_Happy movements by the female participants, these average ratings are relatively high for both embodiments.

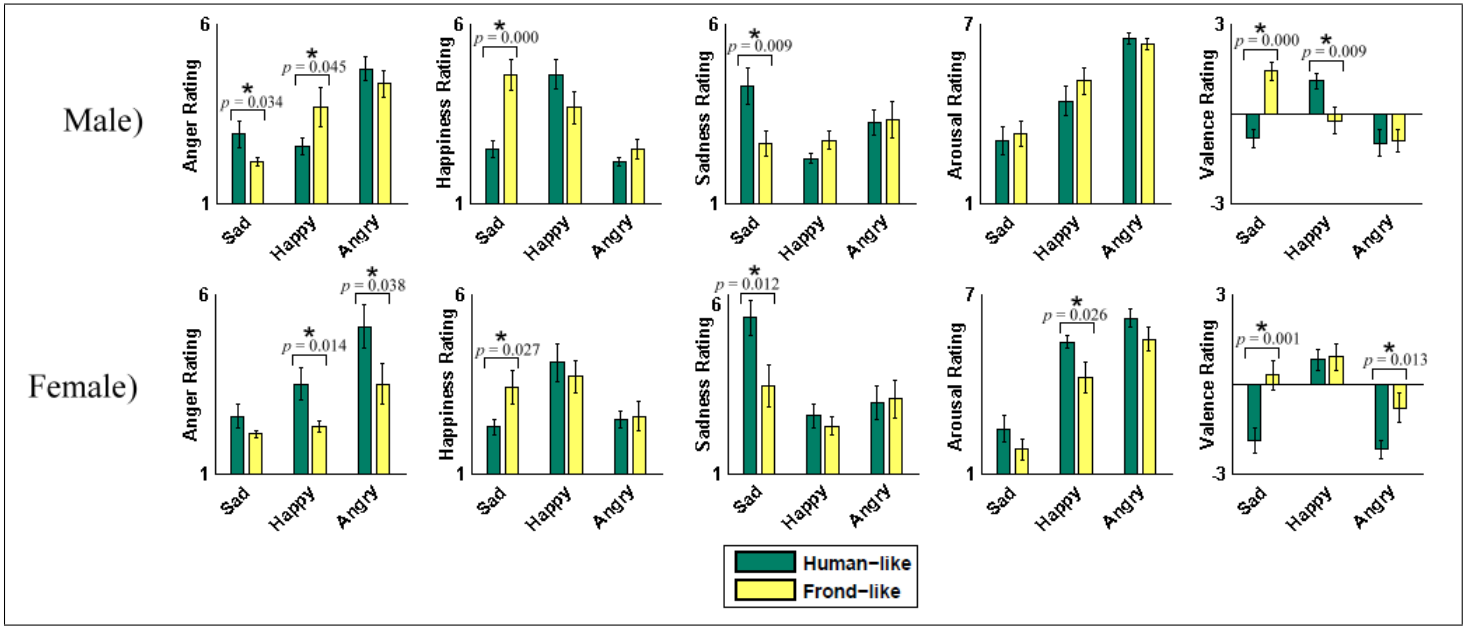

Figure 3.13: Average ratings (mean ± SE) for the affective movements displayed on the human-like and frond-like embodiments by 12 male and 10 female participants. From left to right, ratings for: anger, happiness, sadness, arousal, valence. Significant pair-wise differences between the ratings of an intended-emotion displayed on different embodiments are indicated using "*" sign and their p-values are reported.

The relatively higher accuracy of the female participants in recognizing happy movements in comparison to the male participants is similar to the reports in [166, 96] suggesting that women are more tuned to experiencing positive expressions.

Both the male and female participants correctly rated the sad movement displayed on the human-like embodiment as sad, with low arousal and negative valence attributes, while the Frond_Sad movement is less frequently recognized as sad. The Frond_Sad movement is frequently perceived as conveying happiness and positive valence, especially by the male participants. Overall, both male and female participants correctly recognized differing levels of arousal from the affective movements, while women rate the perceived valence more accurately, which is consistent with [167].

Table 3.13: Confusion matrix showing percentage (%)* of anger, happiness, and sadness ratings for different affective movements by the 12 male and 10 female participants. The recognition rates greater than 50% are highlighted.

	Perceived emotions		
	Anger	Happiness	Sadness
Hand_Angry (male)	**92%**	0%	33%
Frond_Angry (male)	**75%**	8%	33%
Hand_Angry (female)	**70%**	10%	30%
Frond_Angry (female)	30%	20%	40%
Hand_Happy (male)	17%	**83%**	0%
Frond_Happy (male)	50%	50%	25%
Hand_Happy (female)	50%	**70%**	30%
Frond_Happy (female)	0%	**60%**	20%
Hand_Sad (male)	25%	17%	**58%**
Frond_Sad (male)	0%	**67%**	17%
Hand_Sad (female)	20%	10%	**90%**
Frond_Sad (female)	0%	40%	40%

∗ There are cases where an affective movement was rated 3 or above for more than one emotion. On the other hand, there are cases in which anger, happiness and sadness were all rated below 3. This is why none of the emotion ratings add up to 100% in the confusion matrix.

3.3.3.1 Discussion

In this user study, gender-specific differences in the perception of affective hand movements displayed on human-like and frond-like embodiments were investigated. It was found that the gender significantly influenced the perception of the affective movements in many cases. Furthermore, cases were observed in which the impact of the intended-emotion and embodiment on the participants' perception of the affective movements varied between male and female participants (e.g., anger ratings for Frond_Angry movement). The male participants perceived angry movements more accurately than the female participants regardless of embodiment, whereas the female participants performed better in recognizing happy movements. Both male and female participants frequently misperceived sad movements displayed on the frond-like embodiment as conveying a positive expression.

In general, male and female participants exhibited a more similar affective movement perception when the embodiment was human-like (Figure 3.8 and Table 3.13). Such embodiment-specific effects on the perception of female and male participants merit further investigation and would potentially motivate the use of more human-like embodiments for communicating affect during human-machine interaction to ensure consistent perception.

3.4 Summary

In this chapter, an approach based on functional movement representation and dimensionality reduction for affective movement recognition and generation was presented (Section 3.1). A fixed-length representation of affective movements using B-spline basis function expansion was obtained, and functional versions of the DR techniques, FPCA, FFDA, FSPCA and FIsomap, were applied on the BFE representation of the affective movements. One affective full-body dataset and one affective hand dataset were used to evaluate the discriminative performance of the functional DR techniques. Considering testing error (from leave-one-out cross validation using one-nearest-neighbour classification) for the datasets and elapsed training time as assessment criteria, the FSPCA outperformed other functional DR techniques. Furthermore, for the full-body dataset, considering the large number of freely expressed affective movements displayed by 13 different demonstrators, the FSPCA technique shows promising performance when compared with other perceptual and automatic interpersonal affective movement recognition studies [123, 121, 124, 75, 115].

The presented movement recognition approach is particularly useful since it uses a minimal set of systematically obtained feature transformations (dimensions spanning the lower-dimensional subspaces), rather than affective movement recognition in the original

high-dimensional joint trajectory space, which most likely contains many redundant and irrelevant features to the recognition task. Furthermore, BFE is an efficient way to represent the high-dimensional and variable-length sequential observations as functions estimated by a weighted linear combination of a fixed number of basis functions, which satisfies the discriminative DR techniques' requirement for fixed-length vectorial representation of the sequential observations. BFE can also be regarded as an intermediate dimensionality reduction step as it produces a smoother down-sampled version of the original temporal observations.

For affective movement generation (Section 3.2), a transformation based on piece-wise linear re-sampling was applied to produce a set of fixed-length within-class movement sequences with aligned landmarks. The transformed movements were then represented using B-spline BFE, and FPCA was used to obtain a lower-dimensional embedding of the BFE representation of the movements. Prototypical affective movements were then generated from the centroids of emotion-specific clusters of movements in the FPCA space by inverse FPCA mapping to the high-dimensional joint trajectory space. The expressivity of the generated prototypical movements in comparison with their human-generated counterparts were subjectively verified using a user study (Section 3.2.4).

This chapter also investigated the role of embodiment on the perception of affective hand movements via a user study (Section 3.3). The generated hand movements from Section 3.2 were animated on a human-like and a frond-like embodiment and their expressivity was rated by the participants in the user study. Participants were able to perceive, above chance level, the affective content in the movements from both the human-like and frond-like embodiments. However, the accuracy of perception of the affective expressions varied between the embodiments. Participants perceived the expressions encoded in the hand movements as intended when the movements were displayed on the human-like embodiment. On the other hand, the sad movements displayed on the frond-like embodiment were perceived as happy movements. Furthermore, the frond-like embodiment displaying sad movements was frequently described as pleasant, friendly, and gentle, which also indicates the influence of the embodiment on the perception of affective movements.

We have also conducted an experiment to investigate the potential role of the observer's gender in the perception of affective hand movements displayed on different embodiments (Section 3.3.3). We found that the observer's gender significantly influences their perception of affective movements in many cases and that the impact of the display embodiment and intended-emotion on the perception of the affective movements differs between male and female observers. For instance, male observers correctly recognized angry movements regardless of the embodiment, whereas female observers associated less arousal and less-negative valence to the frond-like embodiment displaying angry movements.

Overall, the significant effects detected in the user study indicate the importance of the intended emotion, embodiment, and observer's gender (ordered according to the corresponding affect sizes; hence their importance) in conveying affective expressions in the human-machine interaction field. To the best of our knowledge, there has been no report on the combined role of the observer's gender and embodiment on the perception of affective movements.

Despite its suitability for discriminative analysis of affective movements, the functional movement representation is sensitive to within-class stochastic, temporal, and kinematic variations inherent in body movement. To address these limitations, an approach based on hybrid generative-discriminative modeling of movements is proposed in Chapter 4. Furthermore, affective movement generation based on functional movement modeling is limited to within-class kinematically similar movements and requires a tedious land-mark alignment and length normalization. To overcome these limitations, a more generic approach for affective movement generation is presented in Chapter 5 that is capable of modulating an arbitrary motion path to overlay a target emotion.

Chapter 4

Affective Movement Recognition Based on Generative and Discriminative Stochastic Dynamic Models

■ *An earlier version of the text in this chapter has appeared in the following publication*: [168].

As discussed in section 2.4, a major challenge in designing automatic models for affective movement recognition is the large amount of variability in human movement. Ideally, a recognition model should be able to recognize affective expressions in the presence of kinematic (motion trajectory), interpersonal, and stochastic variabilities inherent in body movements.

In general, the approaches to movement modeling can be divided into temporal (e.g., functional dimensionality reduction as discussed in Chapter 3), and feature-based (e.g., classical multivariate dimensionality reduction). Temporal approaches account for movement dynamics by analyzing the time-series movement features, whereas feature-based approaches require a movement representation in terms of discrete features (e.g., key poses or statistical quantities such as extrema of the motion pattern). The continuous hidden

Markov model (HMM) is a temporal approach that drives a generative[1] model for sequences of time-series observations. HMM represents sequential time-series observations (e.g., multiple trials of a movement) as a stochastic process encoding both temporal and stochastic variabilities as well as spatial characteristics (characteristic key-frames) of the observations. Recognition based on HMM is usually done by training class-specific HMMs and using the forward algorithm [169] to compute the likelihood of a movement being generated by each class-specific HMM. The movement is recognized as belonging to the class with the maximum likelihood (ML classification).

As summarized in Table A.2, only a few studies report using HMMs for affective movement recognition. Using HMMs, 79% of affective upper-body movements (happy, joyful, frustrated, and angry) demonstrated by 20 children playing a computer game were correctly recognized [130]. In another study [35], an interpersonal (23 individuals) recognition rate of 12.6% was obtained for affective upper-body movements (anger, anxiety, boredom, disgust, fear, happiness, neutral/positive/negative surprise, uncertainty, puzzlement, and sadness) using HMM-based recognition. The low recognition rate in [35] could be due to the use of PCA extracted features rather than original movement features for HMM training and testing. The application of HMM for affective movement segmentation is also reported in [35, 170].

For recognition tasks, discriminative approaches are known to be superior to generative ones [171]. However, in order to apply discriminative approaches to variable-length sequential observations such as affective movements, a fixed-length representation of these observations is needed. A fixed-length representation of sequential observations can be obtained using a fixed number of meta-features (derived features intended to describe movement time-series behaviour such as average velocity [132]), interpolation techniques (re-sampling the movements to a fixed length; e.g., [172]), basis function expansion (Chapter 3), or by defining a stochastic feature transformation (e.g., [173]).

Meta-features are usually selected in an ad hoc manner and may not fully capture important dynamic or kinematic movement features contributing to affective expression. For re-sampling and BFE approaches, movement landmarks need to be identified and then temporally aligned, which can be a very tedious and expensive task for large datasets. In contract, stochastic generative techniques such as HMM are robust to the movement variations caused by phase and/or length differences; hence, eliminating the need for length normalization and landmark alignment. Jaakkola and Haussler [171] proposed the Fisher

[1]Discriminative approaches are specialized in defining the boundaries between dissimilar movements to perform recognition, whereas generative approaches model the movement class itself, to enable both movement recognition and regeneration.

score representation, a stochastic feature transformation to obtain a fixed length representation of variable-length sequential observations. Discriminative analysis based on Fisher scores has been reported to result in recognition rates in many fields including speech analysis [173], bioinformatics [171], document classification [174], object recognition [175], facial recognition [176], and sign language recognition [177].

By exploiting stochastically-transformed time-series features, this chapter presents a framework for discriminative analysis of affective movements consisting of: 1) a new recognition approach based on a hybrid generative-discriminative time-series modeling, 2) an approach for automatically extracting features most salient to discriminating between different affective expressions, 3) a low-dimensional embedding of the movements based on the salient features, where affective movement recognition is optimized.

The proposed framework has three main blocks that operate in series: 1) stochastic modeling of class-specific movements based on HMMs with a mixture of Gaussian outputs, 2) movement representation in terms of stochastically-transformed features highlighting differences in generative processes of various observations (HMM-based Fisher scores), and support vector machine classification in the resulting Fisher score space, 3) salient feature extraction in the Fisher score space using supervised principal component analysis (sPCA) and the Hilbert Schmidt independence criterion. The salient features consist of features maximally correlated with the classification labels and encompassing the main modes of variation in the time series data.

This chapter also presents an experimental evaluation of the proposed framework in identifying salient movement features and affective movement recognition using two datasets with movement exemplars for various affective expressions displayed by different demonstrators.

4.1 Hybrid Generative-Discriminative Affective Movement Modeling

As discussed above, the main objective is to develop an affective movement recognition model robust to kinematic, interpersonal, and stochastic variations in body movements. To this end, a hybrid generative-discriminative movement modeling approach is proposed. Class-specific movements are encoded using separate HMMs, which are subsequently used to derive a Fisher score representation of the movements. Next, SVM classification is performed in the Fisher score space and the resulting recognition rate is compared with HMM-based maximum-likelihood (ML) recognition to confirm the discriminative quality

of the Fisher score representation and demonstrate the benefits of combining the generative HMM and the discriminative SVM to enhance recognition. SVM is used as the discriminative model for its reliability in high-dimensional classification tasks [178, 179], robustness to over-fitting, and computational efficiency as its complexity does not depend on the dimensionality of the feature space. Another advantage of the SVM classifier is that it enables direct incorporation of kernels.

Another objective of this chapter is to automatically identify a minimal set of movement features most salient to discriminating between different affective expressions. For this purpose, sPCA [149] is applied in the Fisher score space to obtain a set of discriminative transformations that maximize dependency between the movements and their affective labels. The resulting sPCA transformations span a lower-dimensional subspace suitable for discriminative analysis and visualizing high-dimensional movements. kNN classification is performed in the sPCA subspace to evaluate the discriminative quality of the sPCA transformations. A high recognition rate in the sPCA subspace indicates that the sPCA transformations represent a minimal set of discriminative movement features. The sPCA transformations are further analyzed to identify movement features salient to different affective expressions. A schematic of the proposed approach is shown in Figure 4.1.

4.1.1 HMM-based Movement Modeling

HMM is a generative technique that models a sequential observation as a stochastic process whose dynamics are described by a discrete hidden state variable. The hidden state varies between N hidden state values based on a state transition matrix $\mathbf{A}$ of size $N \times N$. The observation variables (outputs) are described by a vector of size m. The distribution of the observations for each hidden state is modeled as a mixture of M multivariate Gaussians and is denoted as $\mathfrak{B}$. Furthermore, there is an initial state probability $\pi_i|_{i=1}^{N}$ for each hidden state. Therefore, an HMM model $\mathbf{\Lambda}$ consists of $\mathbf{\Lambda}(\mathbf{A}, \mathfrak{B}, \pi)$.

Efficient algorithms exist for estimating the model parameters $\mathbf{A}, \mathfrak{B}, \text{and } \pi$ (e.g., the Baum-Welch algorithm, an expectation-maximization algorithm), evaluating the likelihood that a new observation sequence was generated from the model (e.g., the forward algorithm), and estimating the most probable state sequence (Viterbi algorithm). A detailed review of HMMs can be found in [169].

In this work, an HMM with a mixture of Gaussian outputs is trained to encode movements from each affective category (class-specific HMM). A single affective category can contain both kinematically similar and kinematically dissimilar movements. Kinematically similar movements in an affective category are clustered in one output mixture. Therefore,

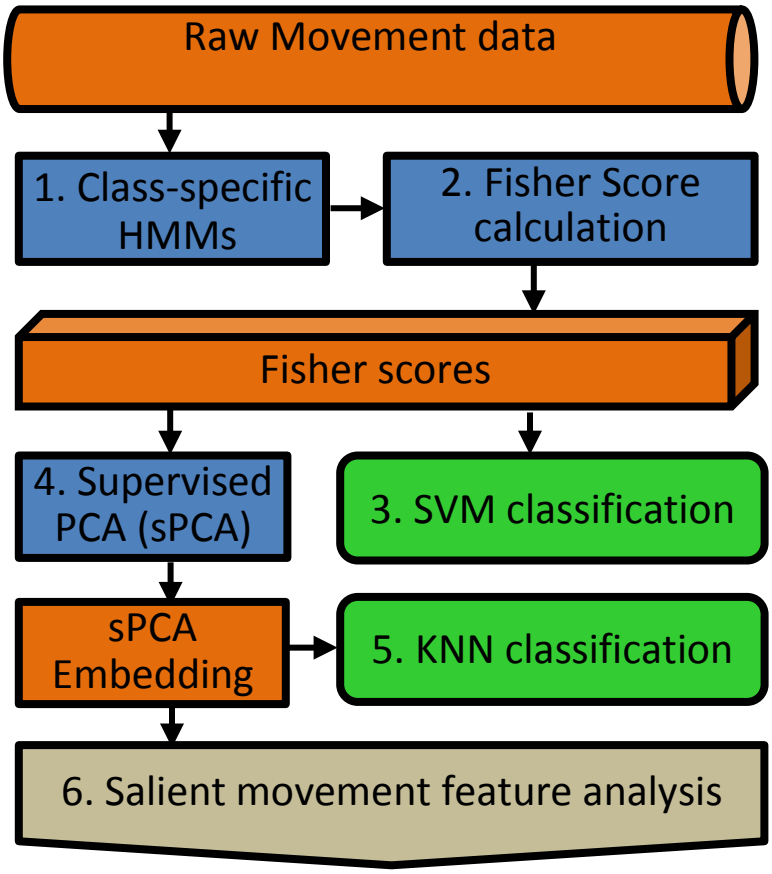

Figure 4.1: Schematic of the proposed approach. 1) Class-specific raw movement observations are encoded in separate HMM models, 2) Using the resulting class-specific HMM models, a Fisher score representation of the movements is obtained, 3) Affective movement recognition in the Fisher score space using SVM classification, 4) A discriminative lower-dimensional embedding of the Fisher scores is derived using sPCA, 5) Affective movement recognition in the resulting sPCA space using kNN classification, 6) Analysing salient discriminative movements features spanning the sPCA subspace.

the mixture of Gaussian outputs allow for encoding kinematically dissimilar movements from an affective category in a single HMM.

4.1.2 Fisher Score Representation

Jaakkola and Haussler proposed incorporating generative models in designing discriminative models, hence benefiting from the strengths of both models [171]. In particular, they exploit a stochastic generative encoding of variable-length sequential observations to transform them into fixed-length observations. The fixed-length observations are referred to as Fisher scores[2] and defined in terms of the gradient of the log-likelihood with respect to parameters of a generative stochastic model $\mathbf{\Lambda}$

$$\mathbf{F}(x) = \nabla_{\mathbf{\Lambda}} \log P(x|\mathbf{\Lambda}), \tag{4.1}$$

where x and $\mathbf{F}(x)$ represent a multivariate sequential observation of length T (e.g., an affective movement) and its Fisher score representation, respectively, $\mathbf{\Lambda}$ is the generative model (represented by a vector of generative model parameters), and $\log P(x|\mathbf{\Lambda})$ is the log-likelihood of x with respect to parameters of the generative model $\mathbf{\Lambda}$. A Fisher score indicates the contribution of each generative model parameter to the process of generating a particular sequence.

In this thesis, HMM-based Fisher scores are used to enable the application of SVM for affective movement recognition. Suppose we have a fully-connected HMM[3] $\mathbf{\Lambda}_k$, encoding movements from the k^{th} affective category. For a time-series movement x of length T in this affective category, the observation probability for the i^{th} HMM state is defined as

$$\begin{aligned} b_i(x_t) &= \sum_{j=1}^{M} w_{ij} \mathcal{N}(x_t, \mu_{ij}, \Sigma_{ij}) = \\ &\sum_{j=1}^{M} w_{ij} \frac{1}{(2\pi)^{m/2} |\Sigma_{ij}|^{1/2}} e^{-1/2(x_t-\mu_{ij})^T \Sigma_{ij}^{-1} (x_t-\mu_{ij})}, \end{aligned} \tag{4.2}$$

[2]Note that the Fisher scores are different than the classical Fisher separability criterion used in dimensionality reduction and feature selection [180]. Fisher score representation is a stochastic mapping technique to transform variable-length sequential observations to fixed-length observations.

[3]A fully-connected or ergodic HMM is an HMM in which every state of the model can be reached in one step from every other state of the model [169]. As a result, the fully-connected HMM has a full transition matrix.

where x_t is the observation vector (movement) at time t, and w_{ij}, μ_{ij}, and Σ_{ij} are the weight, mean, and covariance of the j^{th} multivariate Gaussian distribution of size m at state i. We denote the Fisher scores for the movement x derived based on the fully connected HMM $\mathbf{\Lambda}_k$ as $\mathbf{F}(x)_k$. $\mathbf{F}(x)_k$ consists of gradients of log-likelihood with respect to initial state probabilities, transition probabilities, and weights, means and covariances of the mixtures of Gaussian outputs (Equation 4.2)

$$
\begin{aligned}
&\mathbf{F}(x)_k = [\nabla\pi_{i,k}\ \nabla a_{il,k}\ \nabla w_{ij,k}\ \nabla\mu_{ij,k}\ \nabla\Sigma_{ij,k}],\ \text{where}\\
&\nabla\pi_{i,k} = \frac{\gamma_0(i)}{\pi_{i,k}},\\
&\nabla a_{il,k} = \sum_{t=1}^{T}\frac{\zeta_t(i,l)}{a_{il,k}},\\
&\nabla w_{ij,k} = \sum_{t=1}^{T}\frac{\gamma_t(i,j)}{w_{ij,k}},\\
&\nabla\mu_{ij,k} = \sum_{t=1}^{T}\gamma_t(i,j)(x_t-\mu_{ij,k})^T\Sigma_{ij,k}^{-1},\\
&\nabla\Sigma_{ij,k} = \sum_{t=1}^{T}\gamma_t(i,j)\big(-\Sigma_{ij,k}^{-1}-\Sigma_{ij,k}^{-T}(x_t-\mu_{ij,k})(x_t-\mu_{ij,k})^T\Sigma_{ij,k}^{-T}\big).
\end{aligned}
\tag{4.3}
$$

In Equation 4.3, $i = l = 1,\ldots,N$ states and $j = 1,\ldots,M$ mixtures per state in $\mathbf{\Lambda}_k$. $\gamma_0(i)$ is the probability of being in state i at time $t = 0$ given the movement x and HMM $\mathbf{\Lambda}_k$, a_{il} is the transition probability from state i to state l, $\zeta_t(i,l)$ is the probability of being at states i and l at times t and $t+1$, respectively, and $\gamma_t(i,j)$ is the probability of the j^{th} multivariate Gaussian distribution associated with state i being active at time t. $\pi_{i,k}$ is the initial state probability of state i in $\mathbf{\Lambda}_k$.

The dimensionality of the Fisher scores derived with respect to $\mathbf{\Lambda}_k$ are: $\nabla\pi_{i,k}|_{N\times 1}$, $\nabla a_{il,k}|_{N\times N}$, $\nabla w_{ij,k}|_{N\times M}$, $\nabla\mu_{ij,k}|_{m\times N\times M}$, $\nabla\Sigma_{ij,k}|_{m\times m\times N\times M}$, where m is the number of time-series features in each movement, and N and M are the number of states and mixtures per state in $\mathbf{\Lambda}_k$, respectively.

In this thesis, the class-specific HMMs, $\mathbf{\Lambda}_k$, are used to obtain Fisher scores for each movement as follows: Fisher scores for a movement x are computed with respect to each $\mathbf{\Lambda}_k$ ($\mathbf{F}(x)_k|_{k=1}^{K}$, K classes) separately, and then the resulting class-specific Fisher scores are concatenated in a vector $\mathbf{F}(x)$ to represent that movement

$$\mathbf{F}(x) = [\mathbf{F}(x)_1\ \mathbf{F}(x)_2 \ldots \mathbf{F}(x)_K]. \tag{4.4}$$

4.1.3 Discriminative Embedding of Fisher Scores

Despite their discriminative advantage, Fisher scores are very high-dimensional, which impedes the identification and interpretation of intrinsic movement features salient to affective expression. In order to extract a minimal set of salient movement features, we apply sPCA (described in Section 3.1.2.3) in the Fisher score space. sPCA is a supervised dimensionality reduction technique that identifies a minimal set of dimensions encompassing the main modes of variation in the movements and along which the distribution of the movements is highly correlated with their respective affective labels [149], while eliminating noisy or redundant information in the high-dimensional Fisher score space. sPCA extracts these dimensions by maximizing the correlation between the movements and their affective labels using the Hilbert Schmidt independence criterion [153].

The discriminative quality of the resulting sPCA subspace is evaluated using kNN classification. We use a simpler classifier (kNN) in the sPCA subspace rather than a more sophisticated one such as SVM in order to evaluate the discriminative quality of the resulting subspace. In other words, we aim to identify a discriminative sPCA subspace where a simple classifier such as kNN results in a high recognition rate comparable with SVM recognition rates in the high-dimensional Fisher score space. These discriminative sPCA transformations are then further explored to identify movement features salient to affective expressions.

4.2 Experiments

A 10-fold stratified cross-validation setup as described in Figure 4.2 is used to test the efficacy of the proposed approach. The 10-fold stratified cross-validation is used for its reliability in estimating the accuracy of a classifier [181, 182].

4.2.1 Datasets

As discussed earlier, an affective expression can be communicated through a number of kinematically different movements, and there might exist kinematically similar movements that convey distinct affective expressions. Furthermore, there are interpersonal

and stochastic differences in affective movement expressions. To assess the performance of the proposed affective movement recognition approach, datasets that provide exemplars representing kinematic, interpersonal, and stochastic variabilities are needed. In this chapter two such datasets are used: 1) an affective full-body movement dataset, and 2) an affective hand-arm movement dataset.

4.2.1.1 Full-body Affective Movement Dataset

The full-body dataset was collected using a Vicon 612 motion capture system and contains 183 acted full-body affective movements obtained from thirteen demonstrators who freely expressed movements conveying anger, happiness, fear, and sadness with no kinematic constraints [115]. There are 32 markers attached to bodily landmarks[4] and their 3 dimensional (3D) Cartesian coordinates are collected using 8 motion capture cameras at a rate of 120Hz[5]. There are 46 sad, 47 happy, 49 fearful, and 41 angry movements in the full-body dataset. Each movement starts from a known pose, the T-pose (arms raised to shoulder level and extended to the side). The full-body movements range from 1.39 seconds (167 frames) to 9.17 seconds (1100 frames) in duration. Due to interpersonal (and possibly idiosyncratic) differences as well as the lack of kinematic constraints on the movements, there is a wide range of within-class variabilities (Figure 4.3a and 4.3b). There are also between-class kinematic similarities (Figure 4.3b and 4.3c).

4.2.1.2 Hand-arm Affective Movement Dataset

The second dataset consists of hand-arm movements. This dataset was used because there is evidence that these movements are most significant for distinguishing between affective expressions [82], [83], [84]. Furthermore, there are situations where only individual body parts are observable due to occlusion of the rest of the body (e.g., head and hand movement in a video call). There is also interest in displaying affective movements on embodiments that, due to kinematic constraints, are incapable of full-body movements and rather display only isolated limb movements (e.g., [3]).

[4]The markers are placed on the following bodily landmarks: left front head, right frond head, left back head, right back head, top chest, center chest, left front waist, right front waist, left back waist, right back waist, top of spine, middle of back, left outer metatarsal, right outer metatarsal, left toe, right toe, left shoulder, right shoulder, left outer elbow, right outer elbow, left hand, right hand, left wrist inner near thumb, right wrist inner near thumb, left wrist outer opposite thumb, right wrist outer opposite thumb, left knee, right knee, left ankle, right ankle, left heel, right heel.

[5]Some movements in the full-body dataset are recorded at 250Hz, which were down-sampled to 120Hz in our study to keep the sampling rate consistent for all the movements.

Validation Procedure

Inputs: A dataset $\mathcal{D}$ of movement from K affective expression class.
10-fold stratified cross validation

Data_k: movements from the k^{th} class.
$\text{Data}_k \xrightarrow{\textbf{Permute}} \text{Data}_k^p$
$\text{Data}_k^p \xrightarrow{\textbf{Split into 10 parts}} \text{Data}_{k,1}^p, \text{Data}_{k,2}^p, \ldots, \text{Data}_{k,10}^p$

for $j = 1, \ldots, 10$
- o $\text{Test}_{k,j} \leftarrow \text{Data}_{k,j}^p$
- o $\text{Train}_{k,j} \leftarrow \text{Data}_k^p \setminus \text{Data}_{k,j}^p$

HMM Training
- o HMM Initialization to encode $\text{Train}_{k,j}$
- o Train $\text{HMM}_{k,j}$ using $\text{Train}_{k,j}$
- o ML classification of $\text{Test}_{k,j}$: ML_j

Compute Fisher scores using trained $\text{HMM}_{k,j}$
- o $\mathbf{F}(tr)_j \leftarrow \text{Train}_{k,j}$
- o $\mathbf{F}(te)_j \leftarrow \text{Test}_{k,j}$

SVM Classification in Fisher score space
- o SVM training using $\mathbf{F}(tr)_j$
- o SVM testing on $\mathbf{F}(te)_j$: SVM_j

SPCA embedding
- o Compute SPCA transformations using $\mathbf{F}(tr)_j \rightarrow \text{SPCA}_j$
- o Embed $\mathbf{F}(tr)_j$ and $\mathbf{F}(te)_j$ using SPCA_j
- o kNN classification on $\text{SPCA}(\mathbf{F}(te)_j)$: $k\text{NN}_j$

end

ML-Rate $\leftarrow$ mean(**ML**)
SVM-Rate $\leftarrow$ mean(**SVM**)
kNN-Rate $\leftarrow$ mean(***k*NN**)

Figure 4.2: Pseudocode for the validation procedure.

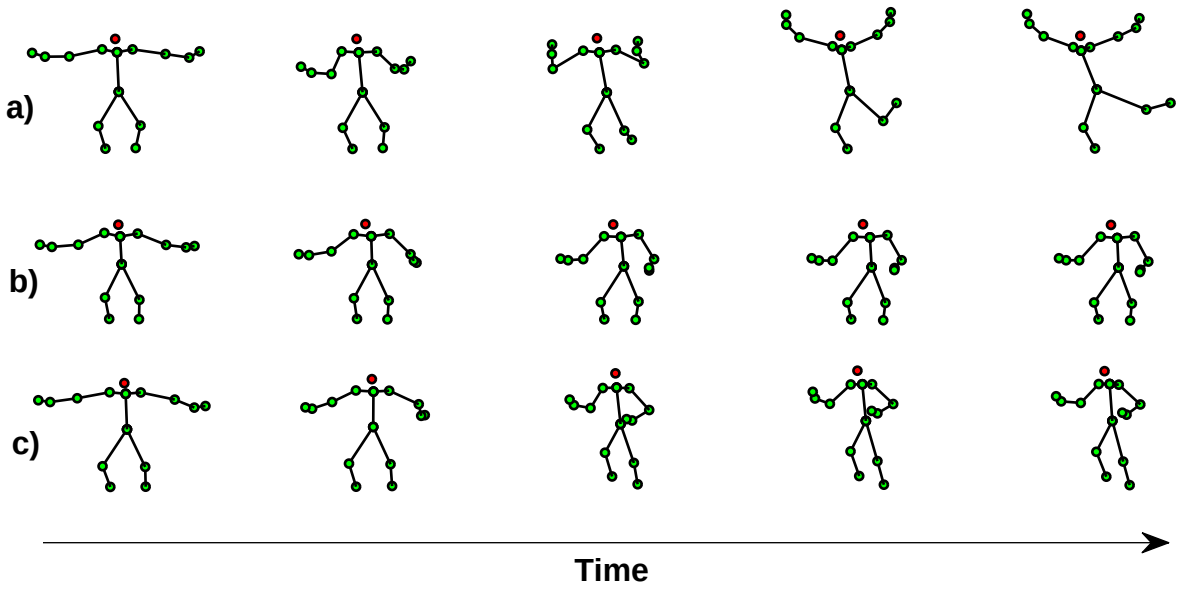

Figure 4.3: Movement exemplars from the full-body dataset: (a) and (b) are two happy movements, which are kinematically different, whereas the happy movement in (b) is similar to the fear movement in (c).

The hand-arm dataset was collected at the university of Waterloo using a Motion-Analysis[6] optical motion capture system equipped with 8 cameras (providing 640 x 480 full resolution with sub-millimeter spatial accuracy) at a sampling rate of 200Hz and the marker set shown in Figure 4.4. We designed six hand-arm motion paths in consultation with a certified movement analyst. Each motion path was used by a professional actor to convey the six basic Ekman emotions (i.e., anger, happiness, sadness, fear, disgust, and surprise [41]) one at a time (each demonstrated movement, in its entirety, was intended to convey a single, specific emotion). Table 4.1 describes the hand-arm motion paths. For each pair of motion path and emotion, 5 trials were recorded. The hand-arm movements range from 1.23 seconds (249 frames) to 25.49 seconds (5099 frames) in duration. A few movement exemplars from each emotion class (4 sad, 5 happy, 2 angry, 4 fearful, 6 surprised, and 8 disgusted) were excluded from analysis due to poor quality of the recorded trajectories (large number of markers missing, occluded, or flipped). In total, 26 sad, 25 happy, 28 angry, 26 fearful, 24 surprised, and 22 disgusted movements were considered in the analysis. Therefore, the collected hand-arm movements provide exemplars for the full set of basic Ekman emotions, which include kinematically dissimilar/affectively similar and kinematically similar/affectively different movements. Both the full-body and hand-arm

[6]Motion Analysis Corporation, www.motionanalysis.com

Table 4.1: Hand-arm motion paths*

Motion path	Description
A	From Self, right arm down along side to forward mid level, reaching to take something, palm up; palm remains open.
B	Similar motion as A, but in the reverse direction.
C	Right palm open on upper chest, opening and extending right arm fully directly in front of the right shoulder at forward mid-level, with the palm facing left, perpendicular to the floor similar to the hand-shake position.
D	Similar motion as C, but in the reverse direction.
E	Right arm is extended forward at mid-level with open palm facing down and the hand parallel to the floor. Moving backward ending with the open palm facing forward near the right shoulder.
F	Similar motion as E, but in the reverse direction.

$*$ During these movements, the wrist and finger orientations remain constant.

datasets were manually segmented. In the proposed approach a single inference is made for each movement.

4.2.2 HMM Initialization and Model Selection

In order to reduce variability due to differences in demonstrators' physical builds (e.g., height), a movement normalization is applied prior to HMM training. The normalization helps to reduce the risk of local optimization of HMM parameters by reducing the variation in trajectories (amplitude) between subjects and/or movement types. For this purpose, each movement observation is normalized by the length of its demonstrator's hand and arm measured from his/her right shoulder to the right wrist. Since all the movements in the full-body and hand-arm datasets start from a known pose (full-body dataset: T-pose, hand-arm dataset: known starting poses reported in Table 4.1), the hidden state sequence for the HMMs always starts at state 1; hence, the hidden state priors are defined as: $\pi_i = 1$ for $i = 1$, and $\pi_i = 0$, otherwise.

To initialize the training process, the movement observations are divided into N (num-

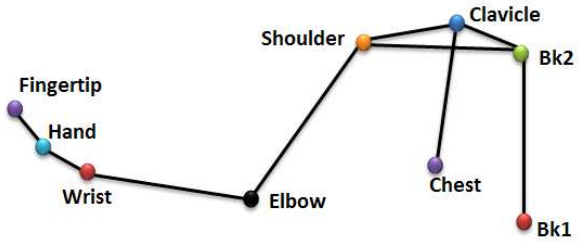

Figure 4.4: Marker set used for collecting hand-arm movements.

ber of hidden states) equal segments, and k-means clustering is run within each segment to identify M clusters (M mixtures). Then, the means and covariances of the identified clusters within each segment (hidden state) are used as the initial means and covariances of the output mixture of Gaussians associated with that hidden state. Full covariances for the mixtures of Gaussians are used to account for potential correlations between body joint trajectories. In each segment, the mixture weight is initialized as the ratio of the number of data points in the cluster to the total number of data points in that segment.

Constraints can be placed to limit the allowable transitions between states. For the full-body dataset, initially, we allowed full transition matrices (i.e., ergodic HMM). However, after training, the transition matrices are reduced to either a left-to-right transition matrix with forward jumps, or a transition matrix with some loops (or backward transitions) to accommodate cyclic patterns in the movements (e.g., two jumps, three walking steps). For the hand-arm dataset, the transition matrix is constrained to left-to-right, since the motion paths (Table 4.1) are direct movements with defined start and stop positions.

The best configuration of the HMMs in terms of number of states and mixtures is selected to optimize 10-fold cross validated testing movement recognition. In each fold, HMMs are built using the training set, and their performance is measured as the percentage of correctly recognized testing movements. This is done by computing the likelihood that each testing movement is generated by the trained HMMs, $\mathbf{\Lambda}_k$, assuming that there are K HMMs for K affective movement classes. The new observation x belongs to the model with the highest likelihood ($\arg \max_{1 \leq k \leq K} P(x|\mathbf{\Lambda}_k)$).

The choice of HMM parameters, particularly the number of mixtures, is based on the dataset, and therefore must be experimentally selected. The described approach in this chapter is therefore suitable for generalization, since the number of mixtures will be adapted to what is in the data, rather than requiring manual tuning.

4.2.3 Fisher Scores and SVM Classification

After training the appropriate class-specific HMMs, the Fisher score representation of the movements is obtained with respect to the class-specific HMMs as described in Section 4.1.2. We eliminate Fisher scores with respect to HMM priors since the starting state for all the movements in both datasets is the first state.

Following the 10-fold cross validation setup, the testing data Fisher scores are classified using linear SVM classifiers trained on the training data Fisher scores from each cross validation fold as illustrated in Figure 4.2. The SVM recognition performance is used to demonstrate the efficacy of the proposed approach for affective movement recognition. The SVM classification is done using the LIBSVM library [183].

4.2.4 Discriminative Lower-dimensional Embedding

The goal of the lower-dimensional embedding step is to identify a minimal set of salient features with enough discriminative power to maintain the cross-validated SVM recognition rates achieved in the higher-dimensional Fisher score space. Following the 10-fold cross validation setup, both training and testing movements are embedded using sPCA transformations derived using only the training movements in each fold. The discriminative quality of the sPCA lower-dimensional space is then evaluated using kNN classification of the testing movements in each cross validation fold. Two types of kernels are tested on affective labels of the movements in sPCA: linear and Gaussian radial basis function (GRBF) kernels. For the GRBF kernel, the proper σ value is selected from a range of 0.25 to 25 with jumps of 0.25 using cross-validation to minimize the testing errors. Furthermore, the cross-validated kNN classification in the resulting sPCA subspaces is done for different subspace dimensionalities and nearest neighbours, and the subspace with the highest kNN testing recognition rate is selected for salient feature analysis.

4.3 Results

In this section, the results of the experimental evaluation of the hybrid generative-discriminative modeling approach are presented for: 1) the full-body dataset (Section 4.3.1) and 2) the hand-arm dataset (Section 4.3.2).

4.3.1 Full-body Dataset

The best HMM configuration for the full-body dataset is found to be seven states and four mixtures per state, which results in a testing maximum likelihood (ML) recognition rate of 74% ± 3.2%. To ensure that HMM training is complete and the resulting models accurately encode various movements in each affective category, training data are also classified using the forward algorithm. The average ML recognition rate on the training data is 100% for the full-body dataset, indicating accurate modeling of the training movements by the resulting HMMs. Therefore, the HMM configuration with seven states and four mixtures per state is selected as the appropriate HMM configuration for further analysis.

Table 4.2 shows the recognition rates achieved for the full-body movements at different stages of the movement modeling in the proposed approach. The table also includes the F_1-measure for the testing movements. For the full-body dataset, the 14NN classification in the 3D subspace produced by sPCA with GRBF kernel ($\sigma = 0.25$) results in the highest 10-fold cross validated recognition rate (77% ± 3.1%) among sPCA subspaces. Table 4.3 shows the confusion matrix for the 10-fold cross-validated kNN classification of the full-body movements in the 3D sPCA subspace.

Table 4.2: Cross validated recognition rates (%) and testing F_1-measures for the full-body dataset

Technique	Train	Test	F_1-measure
HMM + ML	100	74±3.2	0.75
FS + SVM	100	76±3.1	0.80
FS + sPCA + 14NN	91%±2.1%	77±3.1	0.81

Table 4.3: Confusion matrix (%) for the 10-fold cross-validated 14NN classification in the 3D sPCA subspace for the full-body dataset

	Sadness	Happiness	Fear	Anger
Sadness	**85**	2	7	7
Happiness	2	**73**	12	12
Fear	10	0	**90**	0
Anger	6	13	17	**64**

Figure 4.5 shows the sPCA subspaces spanned by: 1) 1^{st} and 2^{nd} sPCA dimensions, 2) 1^{st} and 3^{rd} SPCA dimensions, and 3) 2^{nd} and 3^{rd} sPCA dimensions for the full-body dataset as the 3D sPCA was found the most discriminative for the full-body movements.

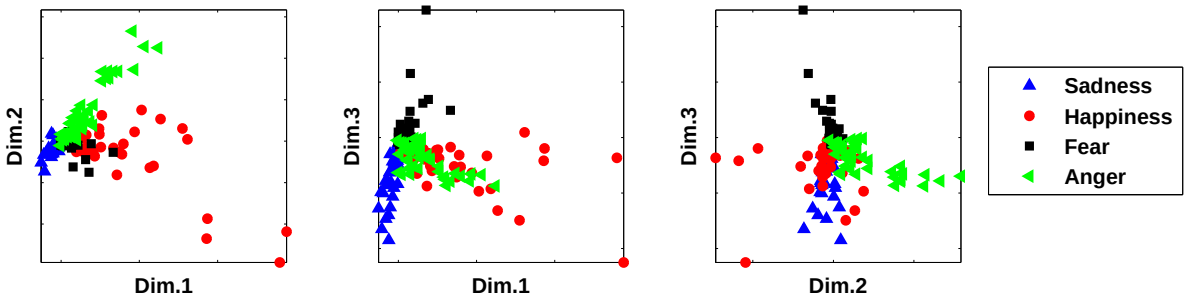

Figure 4.5: sPCA lower-dimensional embedding of the affective full-body movements.

4.3.2 Hand-arm Dataset

For the hand-arm dataset, an HMM configuration with five states and four mixtures per state achieves the highest testing ML recognition rate (79% ± 3.3%). There are three movement types in the hand-arm dataset that are used in both forward and reverse directions (making the six motion paths in Table 4.1). Motion paths A, C, and F start at distinct positions and end at a same position (with the right arm extended in front of the body at the chest level). Therefore, the starting positions of A, C, and F, in addition to the starting position of B, D, and E (the end position of A, C, and F) form four distinct starting positions for the movements in the dataset; hence, the choice of four mixtures enables capturing the distinct motion paths in the hand-arm dataset.

The trained HMMs are used to classify the training hand-arm movements and an average ML recognition rate of 100% is achieved. The 100% ML recognition of the training movements confirms an accurate encoding of the training movements by the resulting HMMs. Therefore, a configuration with five states and four mixtures per state is selected for the hand-arm movements as the appropriate class-specific HMM configuration for further analysis. Table 4.4 shows the recognition rates achieved at different stages of the hand-arm movement modeling in the proposed approach. The table also includes the F_1-measure for the testing movements. For the hand-arm dataset, the 2NN classification in a 6D sPCA-GRBF ($\sigma = 0.5$) subspace results in the highest recognition rate (96% ± 1.6%) among sPCA subspaces. Table 4.5 shows the confusion matrix for the 10-fold cross-validated kNN classification of the hand-arm movements in the 6D sPCA subspace.

Table 4.4: Cross validated recognition rates (%) and testing F_1-measures for the hand-arm dataset

Technique	Train	Test	F_1-measure
HMM + ML	100	79±3.3	0.84
FS + SVM	100	97±1.4	0.99
FS + sPCA + 2NN	100	96±1.6	0.98

4.4 Discussion

The SVM recognition rates and F_1-measures in the Fisher score spaces exceed the ML recognition rates for both the full-body and the hand-arm datasets (Tables 4.2 and 4.4, respectively). The high SVM recognition rates and F_1-measures in the Fisher score spaces illustrate the discriminative quality of the Fisher scores and confirm earlier findings that discriminative analysis based on Fisher scores results in improved recognition rates as compared to generative models used to derive the Fisher scores [173, 171, 174, 175, 176, 177]. Affective movement recognition based on HMM-derived Fisher scores exploits the differences between the generative processes for different affective movements (modeled in terms of the gradient of log likelihood with respect to the parameters of the class-specific HMMs), and results in a higher recognition rate than the ML classification that makes use of the differences in likelihoods only (posteriors of the movements given the class-specific HMM models).

The proposed approach is robust to kinematic differences and can handle multiple kinematic expressions within the same affective category via mixture of Gaussians. In such a model, kinematically dissimilar movements populate distinct Gaussians in the mixture

Table 4.5: Confusion matrix (%) for the 10-fold cross-validated 2NN classification in 6D sPCA subspace for the hand-arm dataset

	Sadness	Happiness	Fear	Anger	Surprise	Disgust
Sadness	**96**	0	0	0	0	4
Happiness	0	**96**	4	0	0	0
Fear	0	0	**100**	0	0	0
Anger	0	0	0	**100**	0	0
Surprise	0	4	0	4	**92**	0
Disgust	0	0	5	0	0	**95**

at each hidden state.

For both datasets used in this study, the recognition rate and F_1-measure in the sPCA subspaces are comparable to the SVM recognition rates in the Fisher score spaces, which demonstrates the discriminative and generalizability (i.e., ability to discriminate between unseen test movements) qualities of the resulting low-dimensional sPCA subspace. The HMM-derived Fisher scores highlight the movements' important kinematic and dynamic features, and sPCA identifies the most discriminative features in the high-dimensional Fisher score space. Therefore, high recognition rates comparable to SVM recognition rates in the Fisher score space are achieved in the sPCA subspace spanned by only a few dimensions. Another appealing property of the resulting sPCA embeddings is that the subspace dimensions form a minimal set of features salient for affective movement recognition, which can be used to explore correspondences between low-level and high-level movement features.

3D and 6D sPCA subspaces resulted in the highest testing recognition rates for the full-body and hand-arm datasets, respectively. Provided that the affective expressions are perceivable as intended, the resulting 3D sPCA subspace for the full-body movements might correspond to the pleasure-arousal-dominance (PAD) model in which the four intended emotions (anger, sadness, happiness, and fear) in the full-body dataset are characterized with distinct levels of pleasure, dominance, and arousal [43]. On the other hand, the hand-arm dataset has exemplars for all the six basic Ekman emotions, which is probably the reason we need a higher dimensional subspace (6D) as compared to the full-body dataset in order to discriminate between the hand-arm movements. For instance, surprise is shown to be distinguishable from other basic emotions along the unpredictability dimension [184], and to distinguish disgust, the avoidance dimension might be needed [185].

The intended affective expressions are recognized above chance level in both datasets (Tables 4.3 and 4.5). In the full-body dataset, mainly the angry and happy movements are confused with the other movements. For the angry movements, the highest confusion is between fearful (17%) and happy movements (13%). Similarly, for happy movements, the highest confusion is between fearful (12%) and angry movements (12%). Considering the circumplex model of emotion [165], the observed confusions seem to be related to the similarities between the affective expressions along the arousal and valence dimensions. For instance, anger and fear are both high arousal and negative valence expressions.

For the hand-arm dataset, most of the confusion occurs in the recognition of surprised movements (Table 4.5), as they are confused with happy (4%) and angry movements (4%). Happiness, anger, and surprise expressions are similar in the arousal dimension of the circumplex model of emotion (high-arousal). The observed confusions in the affective

movement recognition and the possible link to the dimensional models of emotion merit further investigation to identify whether distinct basic emotions can be correctly recognized from movements or if only varying levels of affective dimensions (e.g., arousal and valence) are accurately recognizable from movement.

We also compared the recognition performance in our study to that achievable by human observers for the full-body dataset. Note that the full-body movement labels are the actor labels. Kleinsmith *et al.* [115] tested human perception of the intended emotions from the most expressive postures of 108 of movements in the full-body dataset (referred to as apex postures in [186]). The overall recognition rate was 54.7% with the least recognized postures being fearful ones (49.4 % recognition rate), and the most recognized being the sad postures (63.4 % recognition rate). Therefore, the proposed recognition approach is superior to human observers in decoding affective expressions. It should be emphasized that this comparison is made to illustrate the discriminative quality of the proposed recognition approach and that the perceptual study is done using apex frames from the movements and not the whole movements. Nonetheless, the achieved recognition rates using the proposed approach are promising.

Similar to other data-driven approaches, the success of the proposed approach relies heavily on the training data. For accurate recognition, the training data should provide exemplars covering a wide range of movements and emotions with kinematic, stochastic, and interpersonal variabilities. Furthermore, the proposed approach is a supervised technique that requires movement labels as an input.

4.4.1 Interpersonal Affective Movement Recognition

There are person-specific differences in bodily expression of affect. Such differences, along with kinematic and stochastic ones, make the task of automatic recognition of affective movements challenging. To further evaluate the performance of the proposed approach for person-independent affective movement recognition, leave-one-subject-out cross validation (LOSOCV) is performed for the full-body dataset, which includes affective movements from 13 demonstrators. In each fold of LOSOCV, a subject is left out (testing subject) and the models are trained using the remaining subjects (training subjects). The best HMM configuration is found to be an HMM with seven states and four mixtures per state that optimizes the LOSOCV testing recognition rate. The following LOSOCV balanced testing

recognition rates[7] were achieved using the proposed approach: HMM + ML classification (70% ± 4.0%), FS + SVM (71% ± 3.9%), and FS + sPCA + 14NN (72% ± 3.9%).

The resulting LOSOCV recognition rates are slightly lower than the 10-fold cross validated recognition rates (Table 4.2). Visual inspection of the full-body affective movements reveals idiosyncratic differences in the encoding of the expressions by different subjects. The lower LOSOCV recognition rates indicate that the performance of the proposed approach depends on the availability of movement exemplars which cover the kinematic and dynamic space for the particular affective expressions (i.e., a rich training set). Nevertheless, the LOSOCV recognition rates are comparable with those of 10-fold cross validation (Table 4.2), which demonstrates the capability of the proposed approach to generalize to affective movements performed by unseen subjects, as long as sufficient training data is available.

4.4.2 Comparison with other Recognition Studies

As reviewed in Section 2.4 (and summarized in Table A.2), there are many reports on the automatic recognition of affective movements, with interpersonal recognition rates ranging from 30% to 81%. Considering 13 demonstrators and the high degree of within-class kinematic and dynamic variabilities in the full-body dataset, the recognition rate in the present study demonstrates the suitability of the HMM-based movement modeling and Fisher score representation for discriminative analysis of affective movements. However, direct comparison between the performance of our approach and other affective movement recognition models in the literature is not possible due to the fact that different datasets were used in these studies.

Most closely related to our work are the works by Bianchi-Berthouze *et al.* on automatic recognition of apex (most expressive) postures extracted from the movements in the full-body dateset [115, 116, 188, 189]. In [115], a model based on non-linear mixture discriminant analysis was proposed, which uses a set of low-level features describing distances between different body joints and the most frequent labeling of observers from three cultures (American, Japanese, and Sri Lankan) as input for training. A separate model was trained for each culture, and recognition rates of 78% (American model), 88% (Sri Lankan model), and 90% (Japanese model) were achieved. It should be noted that these are training recognition rates as all the postures were used to train the culture-specific

[7]Balanced accuracy is used to overcome the imbalance in the number of affective movements by each subject in each affective category in the full-body dataset and is computed as the arithmetic mean of specificity and sensitivity measures [187].

models and the prediction accuracy was not evaluated for test postures. De Silva *et al.* [188] apply a similar approach as [115] for affective posture recognition using the apex postures from the movements in the full-body dataset and report a recognition rate of 56%. In another study [189], a recognition model was designed using 182 postures and achieved 60% recognition rate on the training set manipulated by adding noise. Kleinsmith and Bianchi-Berthouze report testing recognition rates ranging between 79% to 81% for correct recognition of affective posture attributes along the arousal, valence, potency, and avoidance dimensions [116]. In their study, the dimensional attributes for a posture were assigned as the observers' average ratings of the affective dimensions for that posture.

For further comparison, the functional recognition approach described in Section 3.1 is tested on the full-body dataset using 10-fold cross validation, and a recognition rate of 54% $\pm$ 3.7% in the 3D sPCA subspace using 14NN classifier is found. Table 4.6 summarizes the recognition rates achieved using the full-body dataset (as a posture or a movement dataset) in different studies.

Table 4.6: Affective recognition rates for the full-body dataset

		Recognition rates (%)	
Study	**Stimuli**	**Training**	**Testing**
De Silva *et al.* [188]	Postures	84	56
Kleinsmith *et al.* [115]	Postures	78-90	-
Kleinsmith *et al.* [189]	Postures	79	60
Samadani *et al.* [17]	Movements	61	54
Current approach	Movements	91	77

The current study investigates affective expression through movement stimuli rather than postures. The achieved cross-validated and LOSOCV recognition rates are high and surpass those achieved on the posture dataset. It also outperforms the functional recognition approach reported in Chapter 3. The resulting high-recognition rates in our work might indicate the importance of considering both kinematic and dynamics features for discriminating between different affective categories. The critical role of movement dynamics is frequently reported in the literature, e.g., [72, 69, 76].

4.4.3 Salient Affective Movement Features

The resulting principal components from the sPCA reduction are further inspected to identify salient movement features for recognizing affective expressions. The Fisher scores

with relatively large weights in the resulting sPCA dimensions correspond to the movement features most salient to affective movement recognition. For both the full-body and hand-arm datasets, the covariance Fisher scores (CFS) are associated with the highest weights in the sPCA transformations derived using the entire dataset for training. The covariances provide a rich source of information about kinematic variabilities in the datasets in terms of variances and covariances between different pairs of body joints, which also reflect dynamic variabilities (e.g., velocity) of the movements; hence, their importance for discriminative analysis.

For each class-specific HMM, the maximum sPCA weight for each CFS at each mixture of the HMM states is identified. The CFS's maximum weights are then normalized and shown using heat maps to visualize the relative importance of the class-specific HMM states and mixtures along the sPCA dimensions for discriminative analysis. The salient feature analysis for full-body movements and hand-arm movements are presented in Sections 4.4.3.1 and 4.4.3.2, respectively.

4.4.3.1 Full-body Movements

Figure 4.6 shows the heat map highlighting the contribution of the class-specific HMM states and their corresponding mixtures in discriminating between affective full-body movements.

Considering the left-to-right nature of the class-specific HMMs, the importance of the end states (illustrated by higher weights in Figure 4.6) indicates the important role of the full-body postures toward the end of the movement interval for discriminating between the full-body affective movements. To visualize these key full-body postures, means of the HMM mixtures with highly weighted CFS's are shown in Figure 4.7.

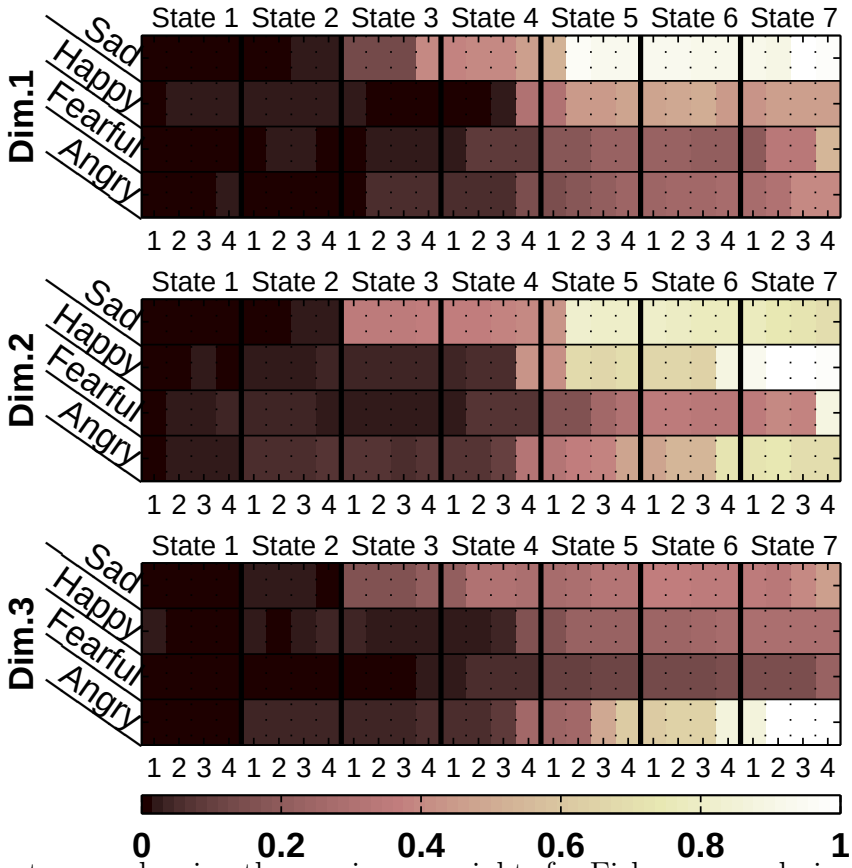

Figure 4.6: Heat maps showing the maximum weights for Fisher scores derived with respect to the covariances of the class-specific HMMs in the sPCA dimensions for the full-body dataset.

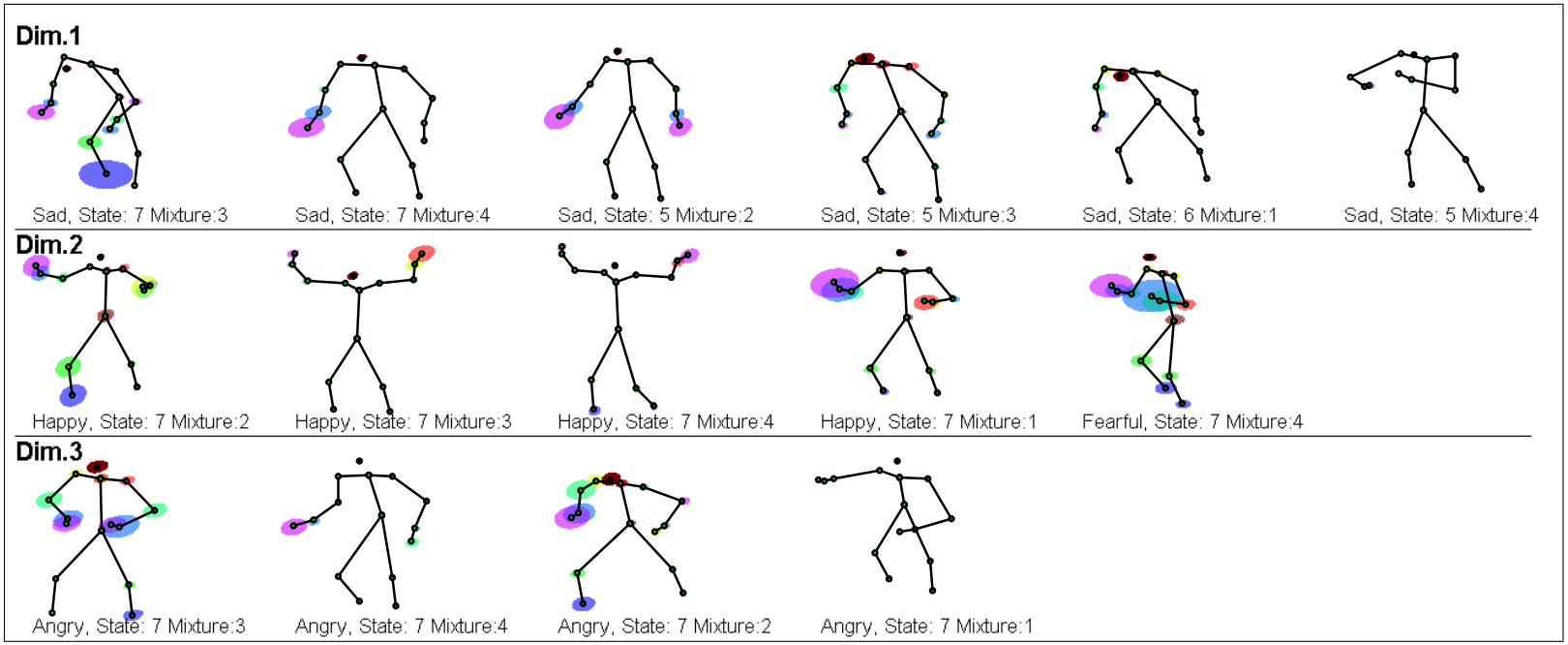

Figure 4.7: The full-body postures corresponding to the top weighted CFS's in the sPCA dimensions. The top weighted CFS's are identified using the heat maps in Figure 4.6. These postures are the means of the corresponding output mixtures. Joint variances are also shown and color coded using spheres whose sizes are proportional to the variances. The postures are labeled and ordered according to the sPCA weights of their corresponding CFS's.

One general observation from Figure 4.7 is that the upper body plays an important role in affective expression for the affective movement exemplars in the full-body dataset. To further investigate the importance of individual body parts for affective full-body movement recognition, we examined the highly weighted CFS's and found that for all the affective categories, the variances of the left and right hand are highly weighted. This indicates the importance of hand postures and their dynamic profile for discriminating between different affective categories. In addition, the head variance of the angry movements is highly weighted which indicates the importance of head position and movement in expressing anger. These observations are congruent with reports on the importance of the upper body limbs and head movements in affective expressions [190].

Figure 4.7 also shows characteristic postures, particularly of the upper body and hands, for different affective categories. As can be seen in Figure 4.7 (Dim. 1), the sad movements are generally characterized by a drooped head and shoulders, an overall slumped posture, and a protective crouch posture. Similar attributes for sad body language are reported in [77, 72, 191, 82]. In Figure 4.7 (Dim. 2), the happy movements are characterized by open and expansive postures, upward hand and arm movements, and advancing movements, congruent with the reports in [66, 72, 191], whereas the salient fearful posture seems to show an alert, retreating, and cowering movement similar to [66, 72]. The angry movements are characterized by arms akimbo postures, angled forward postures, and clenching fist, similar to the reports on the bodily expression of anger in [82, 72]. Furthermore, these key postures are similar to those manually selected as the most expressive postures in the full-body dataset [115].

Next, we estimated the average state durations for the ergodic class-specific HMMs as: $d = \frac{1}{N-1} \sum_{i=2}^{N} 1/(1-a_{ii})$, where a_{ii} is self transition probability for state i [169]. The initial state was excluded from the calculation since it mainly encodes the known starting T-pose position. We found that the sad movements have the longest average state durations (2.11 seconds) followed by fearful (1.99 seconds) and happy (1.17 seconds) movements, respectively. Angry movements have the shortest average state durations (0.87 seconds). The longer state durations for the sad category suggest slow motions for sad movements, which are generally characterized as lingering in time, whereas happy and angry movements are characterized by fast motion [7, 67]. Fearful movements also have a long average state duration, which might be due to the freezing postures common to this category [67].

4.4.3.2 Hand-arm Movements

For the hand-arm dataset, classification in the 6D sPCA subspace results in the highest 10-fold cross-validated testing recognition rate (96% ± 1.6%). The dimensions of the re-

sulting 6D sPCA subspace are further explored to identify movement features salient to affective movement recognition. Similar to the full-body dataset, the covariance Fisher scores (CFS)'s have the highest weights in the resulting sPCA transformations for the hand-arm dataset. Therefore, to explore the movement features most salient to the good recognition performance of the proposed approach for the hand-arm movements, we analyzed the highly weighted CFS's in the sPCA transformations.

The heat maps shown in Figure 4.8 visualize the relative importance of the class-specific HMM states and mixtures along the sPCA transformations. The first observation from Figure 4.8 is that the end states and their mixtures are weighted higher than the early states in the resulting sPCA transformations, indicating the importance of the movement features toward the end of the movements for discriminating between different affective categories. This occurs even though both the starting and ending positions are specified to the demonstrator. However, we observe much greater variability in the end pose compared to the starting pose.

It should be emphasized that in the hand-arm dataset, the same set of prescribed hand and arm motion trajectories are used to convey each of the six basic Ekman emotions. As a result, there are no (or very limited) postural variabilities between the affective hand-arm movements, and therefore no emotion-specific characteristic postures. However, the affective hand-arm movements do vary along dynamic dimensions (e.g., velocity). These dynamic variations are important for discriminating between the hand-arm affective movements, and are accounted for by the position variances and covariances of different body joints in the mixtures of Gaussian outputs and by varying state durations in the class-specific HMMs.

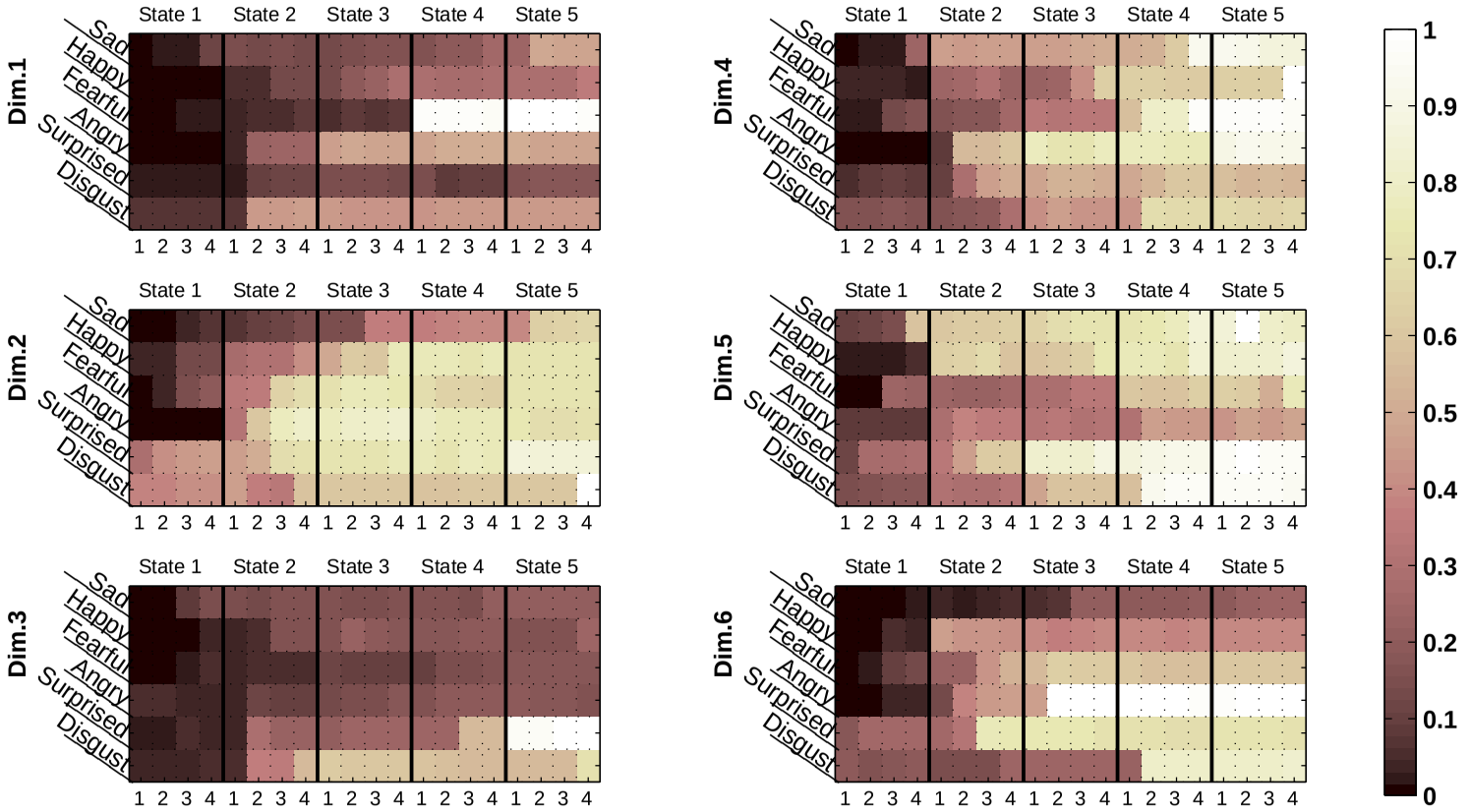

Figure 4.8: Heat maps showing the maximum weights for Fisher scores derived with respect to the Gaussian mixture covariance matrices of the class-specific HMMs in the sPCA dimensions for the hand-arm dataset.

To further explore the movement features that are useful for discriminating between different affective hand-arm classes, we examined the highly weighted CFS's to identify their highly-weighted entries (variances and covariances). In particular, we have selected the entries of the covariance matrices whose weights are greater than a threshold (the threshold was set to 70% of the maximum weight in the sPCA transformations) and reported them in Table 4.7. It is important to note that a high weight for a variance (or a covariance) indicates only that the variance (or the covariance) is salient for affective-class discrimination; the actual value of the discriminative variance (or covariance) may range from low to high and doesn't itself determine its relevance to discrimination.

Figure 4.9 shows the means of Gaussian mixtures whose CFS's are highly weighted in the resulting 6D sPCA subspace. As can be seen in the figure, there are no discriminating postures in this dataset, as the movement trajectories are constrained to be the same for all affective classes. Figure 4.9 also shows the highly-weighted variances reported in Table 4.7, using red spheres. The sizes of the spheres are proportional to the magnitudes of the corresponding variances. As a result, the highly weighted variances that have very small magnitudes are shown using very small spheres (e,g., var(Shoulder) in Disgust(S:5 M:2)).

Furthermore, the highly weighted covariances reported in Table 4.7 are shown using green lines whose thickness is proportional to absolute magnitudes of the corresponding covariances. The highly weighted covariances are normalized by the maximum covariance in their corresponding mixtures and scaled for illustration. The scaling is done to overcome the graphic tool's minimum line thickness and the resulting inability to display very small covariances. Even with the scaling, there are two highly weighted covariances that are too small to be visible using line thickness alone. These covariances, cov(Fingertip-Bk1) and cov(Shoulder-Bk1) for the fearful class, are shown as thin red lines in Figure 4.9.

Table 4.7: Highly weighted entries of the discriminative class-specific CFS's in the sPCA transformations

Emotions					
Sad	Happy	Angry	Fearful	Surprised	Disgust
var(Shoulder)	cov(Shoulder-Bk1)	var(Elbow)	var(Hand)	var(Shoulder)	var(Shoulder)
var(Hand)	cov(Fingertip-Bk1)	var(Fingertip)	cov(Shoulder-Bk1)	var(Chest)	var(Chest)
var(Bk1)	cov(Elbow-Hand)	cov(Hand- Fingertip)	cov(Fingertip-Bk1)	var(Hand)	var(Hand)
	var(Bk1)	var(Wrist)	cov(Hand-Fingertip)	cov(Elbow-Fingertip)	var(Fingertip)
		var(Clavicle)	var(Elbow)	cov(Elbow-Hand)	var(Bk1)
		var(Chest)	var(Fingertip)	var(Bk1)	
		var(Bk2)	var(Bk1)		

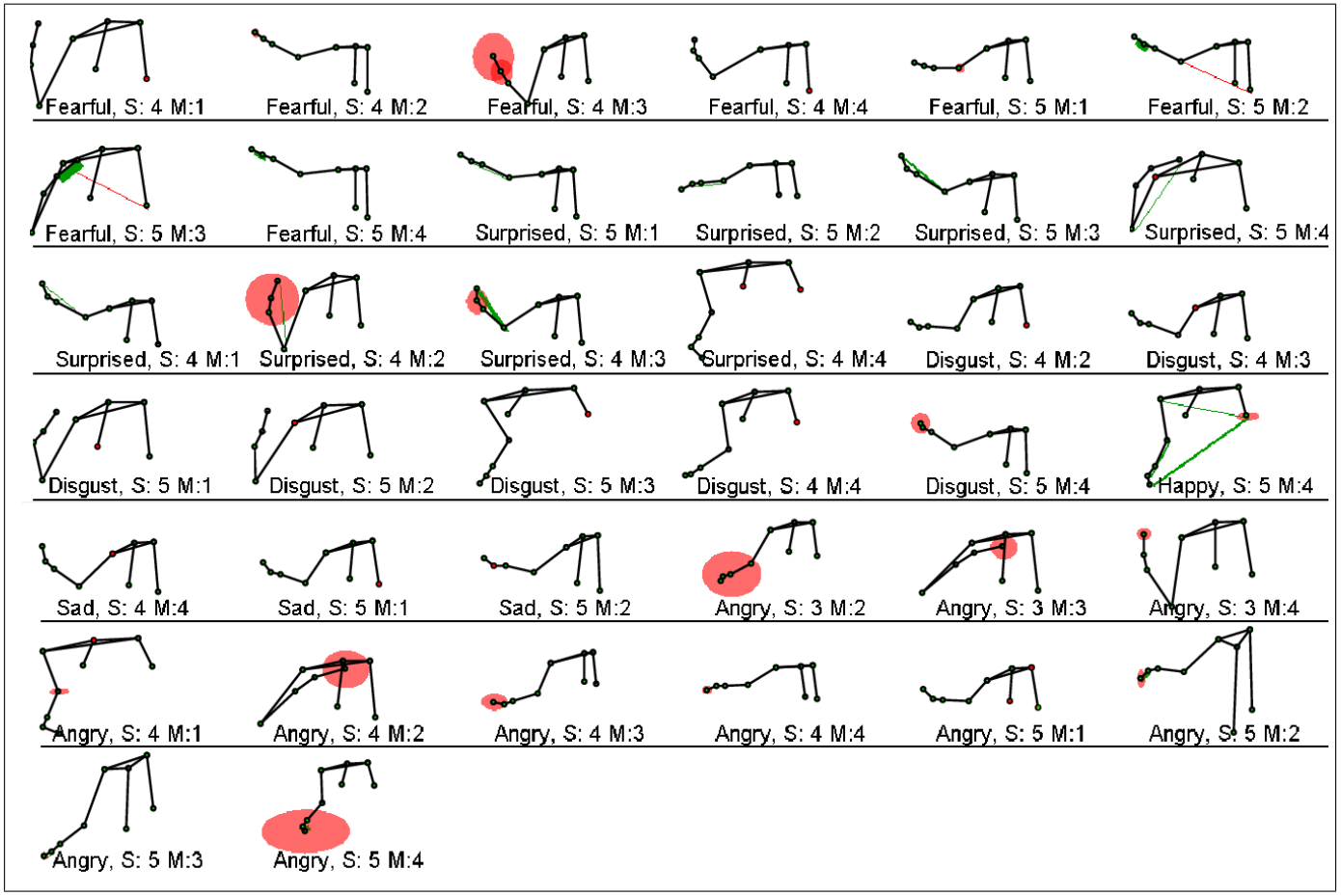

Figure 4.9: The means of mixtures whose CFS's are highly weighted in the sPCA transformations. The highly weighted variances in the CFS's are shown using spheres whose sizes are proportional to the magnitude of the corresponding variances. Highly weighted normalized covariances are shown using green lines whose thickness is proportional to the magnitude of the corresponding covariances. The highly weighted covariances that are too small to appear are shown with thin red lines (Fearful(S:5 M:2) and Fearful(S:5 M:3)). S:State, and M: Mixture. See Figure 4.4 for marker locations.

The average state durations for the left-to-right class-specific HMMs for the hand-arm movements are computed by excluding the first state (mainly encoding the known starting poses) and the last state (the resting poses). The average state durations are: sad (1.54 seconds), happy (0.53 seconds), angry (0.74 seconds), fearful (1.88 seconds), surprised (0.35 seconds), and disgusted (1.07 seconds). As can be seen from these state durations, the surprised movements have the shortest state duration followed by happy and angry movements. Surprised, happy, and angry movements are high arousal movements, which are generally characterised by high velocity and frequent tempo changes [84, 74, 110, 67, 29]. On the other hand, the fearful movements have the longest state durations, followed by sad and disgusted movements. Fearful, sad and disgusted movements are all low-dominance emotions and we found that they are generally characterized by slower movements. Furthermore, the frequent freezing postures common to the fearful class [67] also contribute to the long state durations.

Using Table 4.7, Figure 4.9, and the average state durations, the significance of the highly weighted variances for different affective categories is discussed in Section 4.4.3.2.1, followed by a discussion on the highly weighted covariances (Section 4.4.3.2.2).

4.4.3.2.1 Variance Analysis

A high variance can be indicative of: 1) high velocity movements, 2) jittery movements, or 3) lack of pose consistency between trials. It is possible to differentiate between velocity and the latter two causes by considering also the state duration, i.e., a short state duration and high variance indicates high velocity, while a long state duration and high variance indicates jittery movements or a lack of pose consistency between trials. Conversely, a low variance and long state duration indicates low velocity, smooth movements.

For the sad movements, variances of the shoulder, hand, and the lower back (Bk1 in Figure 4.4) are the entries with the highest weights in the sPCA transformations (Table 4.7); hence, the most discriminative entries for the sad movements. As can be seen in Figure 4.9, these variances are very low in magnitude. The low-magnitude variances and the long average state duration for the sad movements indicate smooth and slow trajectories for this class of affective hand-arm movements. The slow and smooth nature of the sad movements is congruent with previous reports [84], [7].

For the happy movements, the variance of the lower back is highly weighted (Table 4.7), and is low in magnitude (Figure 4.9 Happy(S:5 M:4)). The low position variance of the lower back coupled with short state durations indicates high pose consistency, i.e., a rigid pose for the back during the happy movements.

For the angry movements, the state durations are short while the variances of different upper body joints are highly weighted (Table 4.7). The variances at the upper body joints (lower back, clavicle, and chest) are low in magnitude, indicating a rigid pose of the upper body. On the other hand, the highly weighted variances at the wrist, fingertip, and elbow have high magnitude (Figure 4.9). The large variances at the wrist, fingertip, and elbow considered with the short average state duration of the angry movements indicate the high-velocity nature of the angry movements for the hand and arm [67].

For the fearful movements, the variances of hand, elbow, fingertip, and lower back are highly weighted (Table 4.7). The hand, elbow, and fingertip variances are large in magnitude (Figure 4.9). These large variances, when considered with the long average state duration for the fearful class, indicate jittery (also called "jerky" in the literature) movements [66, 67].

For the surprised movements, the variances of the hand, shoulder, chest, and lower back are highly weighted (Table 4.7). The variances of the shoulder, chest, and lower back are low in magnitude. On the other hand, from Figure 4.9, we can see that the variances at the hand for the surprised movements are large in magnitude. The surprised movements also had the shortest average state duration, which along with the large hand variances indicate the high velocity profile of this class and may suggest a sudden nature for the surprised movements.

For the disgust movements, the variances of the chest, shoulder, hand, fingertip, and lower back are highly weighted (Table 4.7). As can be seen in Figure 4.9, the variances at the shoulder, chest, hand, and lower back are low in magnitude, whereas there is a considerable variance at the fingertip for the disgust movements (Figure 4.9 Disgust(S:5 M:4)). The disgust movements have a medium average state duration relative to other affective classes. Therefore, the large fingertip variance does not imply a high velocity profile for the disgust movements, rather the large fingertip variance might be caused by low pose consistency between trials for the disgusted movements.

4.4.3.2.2 Covariance Analysis

In general, it is more difficult to interpret the highly weighted covariances as the measure involves two joints and their relative movements. A large-magnitude covariance between a pair of joints indicates that the positioning of the two joints is consistent across the movements (i.e., the two joints have synchronized movements). On the other hand, a low-magnitude covariance between a pair of joints could indicate out-of-phase, random movements, or no movements for the joints. There are highly weighted covariances for

the happy, angry, fearful, and surprised classes. The following discussion is one possible interpretation of the highly-weighted covariances, which is supported by the reports in the literature on movement features salient to discrimination between different affective classes (e.g., [191, 66, 67]).

For the happy hand-arm movements, the covariance of the lower back and shoulder, covariance of the lower back and fingertip, and covariance of the elbow and hand are the entries with the highest weights (Table 4.7). These covariances have large magnitudes (Figure 4.9 Happy(S:5 M:4)). These covariances could indicate the elevating posture and overall upward and synchronized movement of the upper body parts common to the happy movements [7, 191, 66]. Furthermore, the covariance between the shoulder and lower back is specific to the happy movements, which again might be caused by the upward upper body movements characterized by raising shoulders.

There is a highly weighted covariance between the hand and fingertip for the angry class (Table 4.7). This covariance appears to be large in magnitude as demonstrated in (Figure 4.9 Angry(S:5 M:4) and Angry(S:5 M:1)), which might indicate a rigid hand posture during the expression of anger.

For the fearful movements, the covariances between the shoulder and lower back, between the fingertip and lower back, and between the hand and fingertip are highly weighted (Table 4.7). From Figure 4.9, the highly weighted covariances (cov(Shoulder-Bk1) and cov(Fingertip-Bk1)) are very low in magnitude (shown by thin red lines), which might be due to the jittery nature of the fearful movements causing unaligned movement of distal body parts (e.g., lower back and fingertip) in terms of direction and speed (i.e., velocity). On the other hand, the highly weighted covariance between the hand and fingertip is high in magnitude (Figure 4.9 Fearful(S:5 M:2), Fearful(S:5 M:3), and Fearful(S:5 M:4)), indicating a rigid hand shape for the expression of fear.

For the surprised movements, there are highly weighted covariances between the elbow and fingertip, and elbow and hand (Table 4.7). These covariances appear to be large (see for instance Figure 4.9 Surprised(S:4 M:3), Surprised(S:5 M:3), and Surprised(S:5 M:2)). This could imply a rigid arm and hand posture during the surprised movements.

4.4.3.3 Discussion

The above analysis demonstrates the utility of the hybrid generative-discriminative abstraction presented in this chapter for the automatic identification of salient movement features. The proposed approach can be used to identify both characteristic postures and characteristic dynamic features for different affective expressions. As can be seen from

the analysis, dynamic features such as the velocity and the degree of co-ordination between body parts play an important role in conveying affect, particularly when the posture trajectory is constrained (hand-arm movements). Posture trajectory constraints are an important consideration not just for laboratory datasets such as the one considered here, but also for functional movements (e.g., gait, sign language), where trajectory generation is constrained by the functional task and only trajectory modulation can be used to convey affect. The importance of movement dynamics in addition to postural characteristics for affective expressions is also reported in the literature (e.g., [72, 69, 76]).

4.5 Summary

This chapter presented a new approach for the recognition and discriminative abstraction of affective movements based on hybrid generative-discriminative movement modeling. Class-specific HMMs with a mixture of Gaussian outputs (generative modeling) are used to encode movement exemplars for different affective expressions. The HMM is robust to temporal variabilities, which may be caused by stochastic and interpersonal differences in body movements, and directly encodes observed joint trajectories without the need for costly and ad-hoc landmark alignment and/or length-normalization.

The class-specific HMMs are then used to derive a fixed-length representation of the movements in terms of Fisher scores. Subsequently, the resulting HMM-based Fisher score representation is used to optimize the discrimination between affective movements using an SVM classification. The SVM classification in the Fisher score space resulted in improved 10-fold cross-validated testing recognition rates of 76% and 97% for the full-body and hand-arm datasets, respectively.

This chapter also presented an approach for the automatic identification of salient movement features based on hybrid generative-discriminative movement abstraction, namely, sPCA abstraction of HMM-based Fisher scores. The resulting sPCA abstraction is tailored toward discriminating between different affective expressions and enables a computationally efficient discriminative analysis as compared with that performed in the high-dimensional Fisher score space. 10-fold cross-validated testing recognition rates of 77% and 96% were achieved by using simple kNN classifiers for the sPCA abstraction of the full-body and hand-arm datasets, respectively. The resulting recognition rates in the sPCA subspaces are comparable with the SVM recognition rates in the Fisher score spaces, despite the significant dimensionality reduction. Therefore, the proposed approach identifies a low-dimensional discriminative subspace where a simple classifier (kNN) results in high recognition.

The sPCA transformations are particularly important as they represent movement features salient for discriminative analysis; hence, no assumptions about the importance of specific movement features for discriminating between affective movements are needed. Further analysis of the sPCA dimensions revealed class-specific characteristic postures and dynamic features congruent with reports from perceptual user studies on the body language most relevant to different affective expressions. Therefore, the recognition of affective movements is best served by considering both postural and dynamic movement features.

Chapter 5

Affective Movement Generation Using Laban Effort and Shape and Hidden Markov Models

■*An earlier version of the text in this chapter has appeared in a published article* [192] *and a manuscript under review* [193].

The ability to express affect via movements has the potential to improve the perceived intelligence and empathic attributes of autonomous machines (e.g., social robots [13] and kinetic sculptures [3]) and create a more engaging human-machine interaction. Therefore, it is important to equip machines with computational models capable of modulating machines' actuated responses to manifest target affective expressions. In this chapter, one such model for automatic affective movement generation is presented.

As discussed in Section 2.5, the large amount of kinematic, temporal, interpersonal, and stochastic variabilities in human movements pose challenges to automatic generation of affective movements. Therefore, an automatic generation model should be robust to these variabilities and incorporate movement features salient to a target emotion in generating movements to strengthen the expression of the target emotion. Furthermore, the model should provide a computationally efficient process to generate affective movements. The affective movement generation approach proposed in this chapter aims to fulfil these requirements. In particular, the research question tackled here is: how can we modulate a desired motion path to overlay a target emotion?

When dealing with high-dimensional affective movements, it is important to derive an abstract representation of the movements that captures both kinematic and affective characteristics. To this end, the proposed approach derives two movement abstractions using Effort and Shape components of Laban movement analysis (LMA), and hidden Markov modeling, to isolate salient kinematic and affective characteristics of movements based on which a desired affective movement is generated.

In order to enable the application of the Laban Effort and Shape components for computational movement analysis, measurable physical correlates of Effort and Shape components need to be identified and used for quantifying these components. In this chapter, an approach for quantifying Laban Effort (Weight, Time, Space, and Flow) and Shape Directional components of hand-arm movements is proposed and its efficacy is evaluated in comparison with annotations from a certified motion analyst (CMA). Subsequently, the verified quantification approach is adapted for annotating full-body movements, extended to include additional Laban Shape components (Shape Shaping and Shape Flow), and used for abstract movement representation in the proposed affective movement generation approach.

The proposed generation approach generates novel affective movements through a systematic blending of movement exemplars from available labeled affective movement datasets; hence, it is an example-based generation approach. Given a labeled movement dataset, to adapt a desired motion path to convey a target emotion, the proposed approach proceeds as follows: 1) using the Laban Effort and Shape abstraction, nearest neighbours to the desired motion path that belong to the target emotion class are identified, 2) an HMM for the identified nearest neighbours along with the desired motion path is learned, 3) the most likely state sequence of the learned HMM for the desired motion path is obtained using the Viterbi algorithm, 4) a deterministic generation approach [194] is used to generate a modulated version of the desired motion path that encodes the target emotion.

The proposed generation approach is tested with the full-body affective movement dataset described in Section 4.2.1.1. The expressivity of the generated movements is assessed objectively using the automatic recognition model presented in Chapter 4. Furthermore, the subjective nature of the affective movement perception necessitates subjective measures to confirm the expressivity of synthetically-generated movements. For this purpose, the generated movements were subjectively evaluated via a user study in which participants reported their perception of the generated movements.

5.1 Laban Effort and Shape Quantification

As discussed in Section 2.2, the Laban system is a movement notation and analysis system that provides a set of semantic components based on which both kinematic and expressive characteristics of movements can be described. Among Laban components, Effort and Shape are the most relevant for the study of affective movements. Effort describes the inner attitude toward the use of energy and Shape characterizes the bodily form, and its changes in space. Bartenieff presents Effort and Shape as a complete system for the objective study of movements, from behavioural and expressive perspectives [1]. Effort has four bipolar semantic components: Weight, Time, Space, Flow (see Table 2.1 in Chapter 2), and Shape has three components: Shape Flow, Directional, and Shaping/Carving (see Table 2.2 in Chapter 2).

In order to enable the application of LMA for computational movement analysis, the LMA components need to be quantified. In this section, the Laban Weight, Time, and Space Effort quantifications from [195], and Laban Weight, Time, Space, and Flow Effort quantifications from [196] are adapted for hand-arm movements. The approach from [195] applies the quantification to the aggregate set of body parts involved in a movement, whereas [196] quantifies Effort components for individual body parts. We propose a continuous measure of Laban Effort components, unlike [195] and [196], which quantify the components as boolean features. In addition, a quantification for Shape Directional of hand-arm movements is proposed. The quantified Laban components are evaluated using their statistical correlation with annotations provided by a CMA for a set of affective hand-arm movements. The adapted quantifications from [195] and [196] are referred to as Q1 and Q2, hereafter.

5.1.1 Weight Effort

Inspired by [195], the maximum of the sum of the kinetic energy of moving body parts is used to estimate Weight Effort. The higher the peak kinetic energy, the Stronger the Weight. The sum of the kinetic energy for the hand and arm at time t_i is:

$$\begin{aligned} E_{t_i} &= E_{t_i}^{\text{Hand}} + E_{t_i}^{\text{Arm}}, \text{ where} \\ E_{t_i}^{\text{Hand}} &= \sum_{k=\text{Fingers}} E_{t_i}^{k} \\ E_{t_i}^{\text{Arm}} &= E_{t_i}^{\text{Upper Arm}} + E_{t_i}^{\text{Forearm}}. \end{aligned} \tag{5.1}$$

For instance, the kinetic energy of the forearm at time t_i is computed as:

$$E_{t_i}^{\text{Forearm}} = \alpha_{\text{Forearm}} v_{t_i}^{\text{Forearm}^2}, \tag{5.2}$$

where α_{Forearm} is the mass coefficient for the forearm and $v_{t_i}^{\text{Forearm}^2}$ is the square of the speed of the forearm at time t_i. The Weight Effort for a sampled movement of length T (the movement is defined by T sequential observations, where each observation represents Cartesian positions of body joints at a particular $t_i, i \in [1, T]$) is then determined as

$$Weight_{Q1} = \max_{i \in [1,T]} E_{t_i}. \tag{5.3}$$

In [197], the mass coefficients for different body parts of a dancing pet robot were set based on their "visual mass and conspicuousness" in the range of 1 to 4 (e.g., $\alpha_{\text{Trunk}} = 4$ and $\alpha_{\text{Arm}} = 2$). In [195], the mass coefficients for different human body parts were set to 1. In the proposed quantifications in this chapter, mass coefficients are set to 1 for all the body parts.

In [196], Weight is quantified using deceleration of motion: Strong (large deceleration) or Light (little or no deceleration). In the proposed quantification for hand-arm movements, this is adapted such that for each hand-arm movement, the maximum of the deceleration time-series for the k^{th} body part is used as the Weight component for that body part ($Weight_{Q2}^k$).

5.1.2 Time Effort

Based on the quantification in [195], the weighted sum of the accelerations of the moving body parts is used to estimate the Time Effort for hand-arm movements. The weights are the mass coefficients similar to the $Weight_{Q1}$. The acceleration[1] for the k^{th} body part at time t_i is:

$$a_{t_i}^k = \frac{v_{t_i}^k - v_{t_{i-1}}^k}{t_i - t_{i-1}}. \tag{5.4}$$

Sudden movements are characterized by larger values in the weighted acceleration time-series as compared to Sustained movements. In particular, the maximum of the weighted acceleration time-series is used to describe Time Effort of hand-arm movements ($Time_{Q1}$).

[1] In the quantifications, the derivatives of motion trajectories, used to compute velocity, acceleration, and jerk, are low-pass filtered to remove the high-frequency noise.

Using [196], the Time Effort for a sampled movement of length T is determined as the net acceleration accumulated at the body parts over time (Equation 5.5). In Equation 5.5, $\mathbf{v}^k_{t_i}$ represents the Cartesian velocity of the k^{th} body part at time t_i.

$$Time^k_{Q2} = \sum_{i=2}^{T} \frac{\|\mathbf{v}^k_{t_i} - \mathbf{v}^k_{t_{i-1}}\|}{t_i - t_{i-1}}. \tag{5.5}$$

5.1.3 Space Effort

The Space quantification from [195] is adapted to measure the Directness of hand-arm movements using the inner product of tangents of torso and wrist trajectories at every time instance. If a movement is Direct, then the resulting inner product time-series would be strictly monotonic or constant, whereas an Indirect movement would demonstrate a fluctuating time-series with multiple peaks (or changes in direction). Therefore, to quantify Space Effort, the number of peaks in the resulting Space time-series is counted, and the Space Effort component is determined on a scale of 1 to 5 through a binning process.

$$Space_{Q1} = \#\text{ of peaks}(\mathbf{v}^{\text{Wrist}}_{t_i}.\mathbf{v}^{\text{Torso}}_{t_i},\ \forall\ i \in [1, T]), \tag{5.6}$$

where $\mathbf{v}^{\text{Wrist}}_{t_i}$ and $\mathbf{v}^{\text{Torso}}_{t_i}$ represent tangents to wrist and torso trajectories (Cartesian velocities) at time t_i, respectively.

Using the Space quantification from [196], the Space Effort of an upper body part is quantified as the ratio of the displacement to the net distance travelled by the body part. In our adaptation, $Space^k_{Q2}$ represents the Space component for the k^{th} body part and is computed for a sampled movement of length T as:

$$Space^k_{Q2} = \frac{\sum_{i=2}^{T} \|\mathbf{p}^k_{t_i} - \mathbf{p}^k_{t_{i-1}}\|}{\|\mathbf{p}^k_{t_T} - \mathbf{p}^k_{t_1}\|}, \tag{5.7}$$

where $\mathbf{p}^k_{t_i}$ is the Cartesian position of the k^{th} body part at time t_i.

5.1.4 Flow Effort

The Flow Effort for a body part k is computed as the aggregated jerk over time as proposed in [196] (Equation 5.8). Jerk is the third order derivative of the position.

$$Flow^k_{Q2} = \sum_{i=2}^{T} \frac{\|\mathbf{a}^k_{t_i} - \mathbf{a}^k_{t_{i-1}}\|}{t_i - t_{i-1}}, \tag{5.8}$$

where $\mathbf{a}_{t_i}^k$ is the Cartesian acceleration of the k^{th} body part at time t_i.

5.1.5 Shape Directional

The Shape Directional component is related to the pathway along which a movement is executed. The Shape Directional component can be 'Spoke-like' or 'Arc-like'. Spoke-like Directional movements travel on a straight line from the body center or return on a straight line to the body center [198]. The pathway of Arc-like movements is characterized by a flat arc starting at the side and moving forward and upward [198]. Arc-like Directional movements can have a flat arc in the horizontal plane while travelling forward and sideways (and likewise in the reverse direction).

Since the Shape Directional describes the transverse behaviour of the hand movements, we propose the average curvature of the hand movements in a 2 dimensional (2D) plane within which the largest displacement occurs as the measure for the Shape Directional. There is a separate Shape Directional component for each hand. To capture the 2D plane with maximum displacement, we apply multivariate principal component analysis (PCA) on the 3D Cartesian trajectories of the hands and extract the top two dimensions. The extracted dimensions span the 2D plane where the maximum displacement occurs. Next, the average 2D curvature within the PCA-extracted 2D plane (xy plane) for a sampled movement of length T is computed as follows

$$\kappa = \frac{1}{T}\sum_{i=1}^{T}\frac{\sqrt{(\ddot{y}_{t_i}\dot{x}_{t_i} - \ddot{x}_{t_i}\dot{y}_{t_i})^2}}{(\dot{x}_{t_i}^2 + \dot{y}_{t_i}^2)^{3/2}}, \tag{5.9}$$

where $\dot{x}_{t_i}$ and $\ddot{x}_{t_i}$ indicate the first and second derivatives of the x trajectory at time t_i, respectively.

5.1.6 Experiments

The quantifications adapted from Q1 and Q2 are evaluated in comparison with annotations from a CMA using the affective hand-arm movement dataset described in Section 4.2.1.2.

5.1.6.1 CMA Annotation

Laban movement annotations provided by the CMA are used to investigate the efficacy of the proposed Laban quantifications for the hand-arm movements by computing the statistical correlation between the CMA-annotated and quantified Laban components.

To this end, an annotation questionnaire was designed in consultation with the CMA in which Weight, Time, Space, and Flow Effort of a movement were rated by the CMA on a 5-point Likert scale, while watching the video of the movement. For each Effort component, two extra 5-point Likert scales were also provided to allow for annotating subintervals along the movement with distinct Effort qualities (if any). The 5-point Likert scale spans between poles of Effort components. Furthermore, an open-ended text box was provided for each Effort component to accommodate additional comments that the CMA might have on a particular movement and annotation. In the videos, the actor's face was masked and the videos were muted (Examples of hand-arm movements can be seen in [199]).

The questionnaire also included annotation of Shape Directional (options: Arc-like/Spoke-like), Shaping/Carving (options: yes/not significant). Shape Flow is self-referential and defines readjustments of the whole body for internal physical comfort [2]. Therefore, Shape Flow can not be observed in the isolated and prescribed hand-arm movements used in this study, and it was excluded from the annotation and quantification analysis. Furthermore, a visual observation of the hand-arm movements revealed to the CMA that Shaping/Carving occurs only slightly in the fingers for some of the movements, which is probably due to the prescribed motion paths and no-movement constraints placed on the wrist and fingers. Therefore, the Shaping/Carving is not significant for the hand-arm movements with the kinematic model used (Figure 4.4), and it was also excluded from quantification analysis.

A total of 44 movements were annotated, which include the following motion-emotion pairs[2]: 3 F-anger, 3 F-sadness, 3 F-disgust, 3 F-fear, 2 E-surprise, 2 E-fear, 1 E-disgust, 1 E-sadness, 1 E-happiness, 1 E-anger, 2 D-anger, 1 D-sadness, 2 D-fear, 2 C-anger, 2 C-happiness, 2 C-sadness, 2 C-disgust, 2 C-fear, 2 C-surprise, 1 B-anger, 1 B-happiness, 1 B-sadness, 1 B-disgust, 1 B-fear, 1 B-surprise, 1 A-happiness. We assume that each movement in the dataset has unique Time, Weight, Space, and Flow Effort qualities and therefore, we only consider the CMA-annotated movements with single Effort annotations (sample size column in Table 5.1). This assumption is important as we are after movement cues most salient to different Effort and Shape components that can uniquely represent varying levels of these components.

[2]For the description of the hand-arm motion paths, denoted by capital letters A to F, refer to Table 4.1

5.1.7 Results and Discussion

The correlations between the quantified and CMA-annotated Effort and Shape components are computed based on the Pearson linear correlation coefficient. Furthermore, the significance of the obtained correlations is evaluated at $p < 0.05$ with a null hypothesis H^0 stating: there is no correlation, $r = 0$, between the CMA-annotated and quantified Laban component L_i, where $\mathbf{L}$ ={Space, Time, Weight, Flow, Shape Directional}. p-values represent the probability of a correlation as large as the observed correlation by random chance, when the true correlation is zero. Table 5.1 shows the resulting correlations between the quantified and CMA-annotated Effort components and their significance. The Q1 Effort quantifications are obtained for aggregate set of body parts involved in the movements, whereas the Q2 Effort quantifications are obtained for each body part, separately.

There are high and significant correlations between the CMA-annotated and quantified Laban components for Weight and Time Effort, whereas there are lower correlations between the quantified and CMA-annotated Space and Flow Effort. Two approaches were used to measure Weight Effort: 1) the maximum of the kinetic energy of the upper body parts ($Weight_{Q1}$) , and 2) the maximum of the deceleration at different upper body parts ($Weight_{Q2}$). Considering a unit distance travelled by a body part, the deceleration of that body part is proportional to the rate of change in its kinetic energy. The maximum of the deceleration implies the maximum absorption of the kinetic energy [196] e.g., at the end of a punching movement. Therefore, the peak of the deceleration and the peak of the kinetic energy (the difference between the maximum and minimum of the kinetic energy as the minimum of the kinetic energy is zero) are both measuring the maximum absorption of the kinetic energy. However, the peak of the kinetic energy of the body parts ($Weight_{Q1}$) provides a more accurate measure for Weight Effort ($r = 81\%$) as it considers the maximum discharge of the kinetic energy of all the body parts.

Bartenieff described Sudden Time as *react quickly* or *immediately* and Sustained Time as *take time, moving slowly* and *lingering in time* [1]. Considering these descriptions, Sudden movements should demonstrate larger accelerations, whereas Sustained movements are characterized with lower or no acceleration. In the proposed quantification $Time_{Q1}$, the peak acceleration of the upper body parts results in a higher correlation with the CMA-annotated Time ($r = 77\%$) compared to the aggregated acceleration ($Time_{Q2}$: $r = 60\%$ at wrist). It is possible that the aggregated acceleration produces a large quantity for Sustained movements, which are generally characterized as lingering in time (resulting in a longer duration as compared to Sudden movements), and as a result, they are identified as Sudden movements. Normalizing the aggregated acceleration by the length of the movements results in an improved correlation ($r = 71\%$ at wrist) as compared to the

unnormalized aggregated acceleration used in [196]. However, this is still lower than the correlation for the maximum acceleration measure ($Time_{Q1}$).

Table 5.1: Correlations (%) between CMA-annotated and quantified Laban effort components along with their p-values. "*" indicates a significant correlation.

			Q2									Q1
	Sample size		Bk1*	Bk2	Clavicle	Chest	Shoulder	Elbow	Wrist	Center	Middle fingertip	
Weight	36	$r =$	68	75	63	58	72	77	81	78	75	81
		$p =$	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*
Time	34	$r =$	-14	-30	-29	-9	16	53	60	60	41	77
		$p =$	0.45	0.09	0.10	0.63	0.37	0.00*	0.00*	0.00*	0.02*	0.00*
Space	35	$r =$	8	3	9	-4	19	32	61	62	63	5
		$p =$	0.66	0.86	0.63	0.82	0.27	0.06	0.00*	0.00*	0.00*	0.75
Flow	21	$r =$	52	46	50	45	55	-2	67	57	-32	-
		$p =$	0.02*	0.04*	0.02*	0.04*	0.01*	0.94	0.00*	0.01*	0.15	

$\star$ Refer to Figure 4.4 for the the upper-body joints.

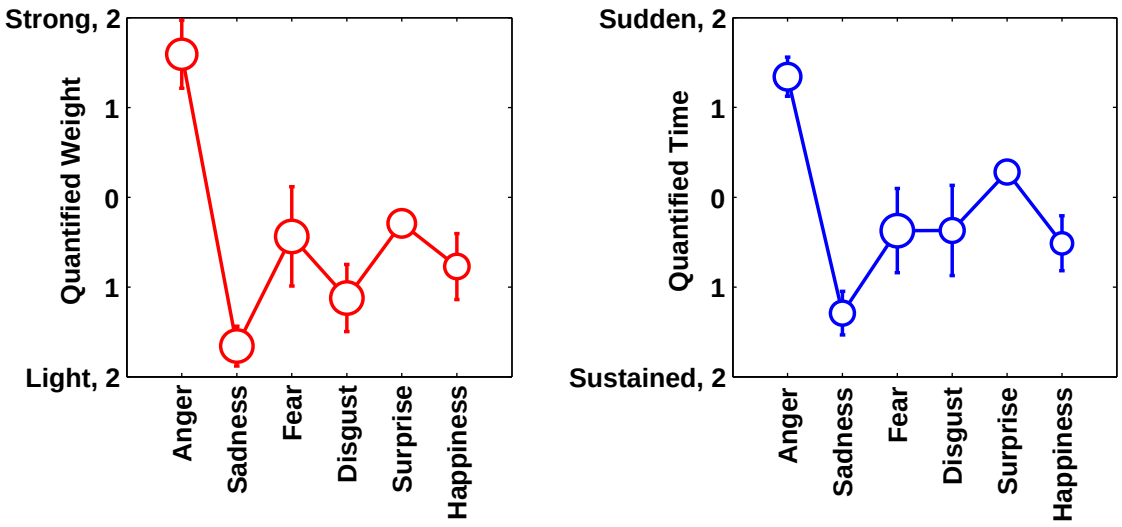

Figure 5.1: Average (and standard error) quantified Weight and Time for the affective hand-arm movements. y-axis labels are assigned to correspond to the 5-point Likert scale used in the annotation questionnaire. The circles and their sizes indicate the average value and sample size, respectively.

Furthermore, the normalized aggregated acceleration is not an appropriate measure for Time Effort as it represents the average acceleration and fails to capture abrupt changes in the velocity. Therefore, the maximum acceleration at the upper body parts ($Time_{Q1}$) is a more appropriate measure for Time Effort.

The kinematic cues (velocity and acceleration) used for the quantification of Weight and Time Effort are similar to those reported as relevant cues in the movement literature [64, 67, 200, 201, 202, 203]. The average quantified Weight ($Weight_{Q1}$) and Time ($Time_{Q1}$) Effort for the affective hand-arm movements are shown in Figure 5.1. The sad and angry movements are clearly distinguishable from the rest of the movements using the quantified Time and Weight. The angry movements are characterized by Strong Weight and Sudden Time, whereas the sad movements are characterized by Light Weight and Sustained Time. The average Weight for the remaining emotions is Light, however none of them exceeds the Lightness of the sad movements. The Time for the surprised movements is slightly oriented toward Sudden Time, whereas the Time for the fearful, disgusted, and happy movements is slightly Sustained. These results can be used to study the capability of the isolated hand-arm movements and their Laban encoding to convey distinct basic emotions.

The majority of the movements (31/35) are annotated by the CMA as Direct Space,

only 2 are annotated as Indirect and the remaining 2 are annotated as Spaceless (no attention to Space). Therefore, due to the imbalance in the sample size (lack of adequate samples at different levels of Space), it is difficult to validate the proposed Space quantifications for hand-arm movements. Furthermore, Space Effort describes the actor's focus (single-focused vs multi-focused) and other visual cues (facial movements) might be needed to better evaluate Space. For instance, an expansive arm movement can be used to greet several arriving parties (multi-focused, Indirect) or a single arriving person (single-focused, Direct), which would be difficult to annotate without additional contextual information.

The spatial stopping constraint in the motion paths (Table 4.1) contributes to having movements with multiple Flow qualities, as it turns a movement to Bound Flow toward the end even if it begins as a Free Flow movement. This resulted in Flow being the Effort component with the most multiple annotations for each movement (23 out of 44 movements). Using the magnitude of jerk for quantifying the Flow ($Flow_{Q2}$), a significant correlation as high as 67% is obtained at the wrist. Other studies also relate Flow to the jerkiness of the motion trajectory (e.g., [67, 201, 202]).

Using the average curvature of the movement projections in the PCA-extracted 2D planes, a correlation of 81% ($p < 0.0001$) between the quantified and CMA-annotated Shape Directional is obtained. Figure 5.2 shows the quantified values for Shape Directional, in which Arc-like Shape Directional (as annotated by the CMA) and a suitable threshold for labelling are shown as red circles and a dashed line, respectively. Applying the threshold, the movements are divided into two bins of Spoke-like and Arc-like, and a Phi correlation[3] of 93% is obtained between the binned-quantified and CMA-annotated Shape Directional. The high correlation results for Shape Directional verify the efficacy of the proposed measure for identifying the Arc-like or Spoke-like Shape of the movements.

Therefore, the experimental evaluation shows that the proposed quantifications for Weight ($Wight_{Q1}$), Time ($Time_{Q1}$), Flow ($Flow_{Q2}$), and Shape Directional are congruent with annotations provided by the CMA. These Laban quantifications are computationally efficient and allow for an efficient search for movement kinematically or affectively similar to a desired motion path. The proposed generation approach, described in Section 5.2, uses the quantified Laban Effort and Shape components for abstract representation of full-body movements.

[3]Pearson correlation computed between two binary variables returns Phi correlation.

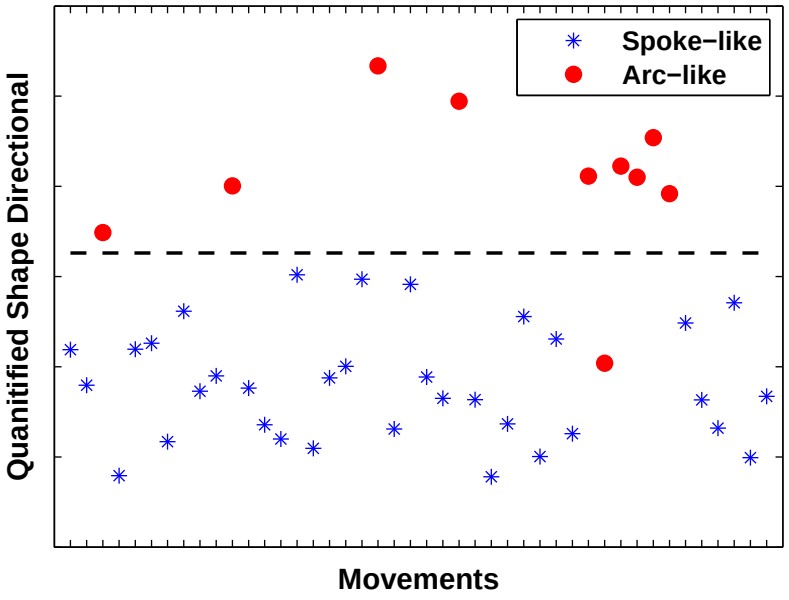

Figure 5.2: Quantified Shape Directional for different movements. The dashed line illustrates a suitable threshold value.

5.2 Full-Body Movement Generation Methodology

The proposed generation approach makes use of two movement abstractions: 1) Laban Effort and Shape components, and 2) hidden Markov modeling. Using these abstractions, a new movement is generated that is affectively constrained by a target emotion and kinematically constrained by a desired motion path. In this section, the Laban and HMM movement abstractions for full-body movements along with the proposed affective movement generation approach are described.

5.2.1 Laban Effort and Shape Components for Full-body Movements

The verified Laban Effort and Shape quantifications from Section 5.1 are adapted for full-body movements. In particular, the Weight and Time Effort quantifications ($Weight_{Q1}$ and $Time_{Q1}$) and the Flow Effort quantification ($Flow_{Q2}$) are adapted for the Laban full-body abstraction in the proposed generation approach. The results from Section 5.1 show that these Laban quantifications are highly correlated with the corresponding annotations from the CMA for hand-arm movements. Furthermore, the Laban quantification is extended to include Shape Shaping and Shape Flow components of full-body movements in the proposed generation approach.

Weight Effort: Based on $Weight_{Q1}$, the maximum of the sum of the kinetic energy of the torso and distal body limbs (end-effectors: head, hands, feet) is used to estimate the Weight Effort of the whole body. The sum of the kinetic energy at time t_i is

$$E_{t_i} = E_{t_i}^{\text{Torso}} + E_{t_i}^{\text{End-effectors}}. \tag{5.10}$$

Similar to the case of hand-arm movements, the mass coefficients α are set to 1 for all the body parts in the full-body analysis.

Time Effort: Based on $Time_{Q1}$, the weighted sum of the accelerations of the torso and end-effectors is used to estimate the Time Effort for full-body movements. The weights are the mass coefficients similar to the quantified Weight Effort.

Flow Effort: Based on $Flow_{Q2}$, the Flow Effort for a full-body movement is computed as the aggregated jerk over time for the torso and end-effectors.

Shape Shaping The Shaping in the vertical and sagittal planes (y and z coordinates) is represented as the maximum vertical and sagittal (back and forth) displacement of the

torso. The Shaping in the horizontal plane is expressed as the maximum of the area of the convex hull encompassing the body projected onto the horizontal plane.

Shape Flow The Shape Flow is measured as the maximum of the volume of the convex hull (bounding box) containing the body.

Shape Directional The Shape Directional quantification presented for hand-arm movements is used to quantify Shape Directional of full-body movements (Shape Directional is computed for the left and right hands).

5.2.2 Regularized Multinomial Logistic Regression

In the proposed approach, a regularized multinomial logistic regression (RMLR) is used to identify affectively non-discriminative Laban components. The multinomial logistic regression uses the inverse logit function to model the posterior of occurrence of different classes in a multi-class problem (K classes) given a set of explanatory variables. Given a dataset $\mathcal{D}_p = \{(\mathbf{x}_j, y_j)|\mathbf{x}_j \in \mathbb{R}^p, y_j \in \mathcal{Y}, j = 1, \ldots, n\}$, where $\mathbf{x}_j$ is the j^{th} multivariate datapoint of dimensionality p ($\mathbf{x}_j$ is represented by a vector of p features) and $\mathcal{Y}$ is the label set for the K classes, the posterior of the k^{th} class is approximated using the symmetric multi-class inverse logit function.

$$Pr(y_i = k|\mathbf{x}_i) = \frac{exp(\beta_{0k} + \beta_k^\top \mathbf{x}_i)}{1 + \sum_{l=1}^{K} exp(\beta_{0l} + \beta_l^\top \mathbf{x}_i)},\ k = 1, \ldots, K. \tag{5.11}$$

The posterior model parameters for all the classes are denoted by θ

$$\theta = \{[\beta_{01}; \beta_1], [\beta_{02}; \beta_2], \ldots, [\beta_{0K}; \beta_K]\}. \tag{5.12}$$

θ is the set of posterior coefficients for all the classes along with their corresponding intercepts (posterior coefficients of k^{th} class: $[\beta_{0k}; \beta_k] \in \mathbb{R}^{p+1}$, where β_{0k} is the k^{th} class intercept). The optimal parameters, θ, are found by solving the maximum log-likelihood problem

$$\underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{N} \log(P_{y_i}(\mathbf{x}_i; \theta)), \tag{5.13}$$

where $P_k(\mathbf{x}_i; \theta) = Pr(y = k|\mathbf{x}_i; \theta)$.

To identify dimensions of $\mathbf{x}_i$ most salient to discriminating between the K classes, the regularized multinomial logistic regression is proposed (RMLR) [204]. In this work, the multinomial logistic regression with elastic net regularization is used, which maximizes the following cost function to find the RMLR model parameters

$$\underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{N} \log(P_{y_i}(\mathbf{x}_i;\theta)) - \lambda \sum_{k=1}^{K} \sum_{j=1}^{p} (\alpha \|\beta_{kj}\|_1 + (1-\alpha)\|\beta_{kj}\|_2^2). \tag{5.14}$$

$\|.\|_n$ denotes the l_n norm. The tuning variable λ controls the strength of regularization; the larger λ, the larger the number of class-specific model parameters β_k that will be driven to zero (i.e., fewer Laban components are selected). The variable α is the mixing variable, which controls the contribution of l_1 and l_2 norms of the model parameters in the regularization ($\alpha = 1$ results in least absolute shrinkage and selection operator (LASSO) and $\alpha = 0$ results in ridge regression). The l_1 norm imposes sparsity over the model parameters, and the l_2 norm encourages the correlated features to be averaged. A detailed derivation and solution for regularized multinomial logistic regression can be found in [205].

5.2.3 HMM-based Movement Modeling

As described in Section 4.1.1, hidden Markov modeling is a generative technique that models a sequential observation as a stochastic process whose dynamics are described by a discrete hidden state variable. An HMM $\mathbf{\Lambda}$ has three parameters 1) the state transition matrix A of size of size $N \times N$, 2) the distribution of observations for the states of an HMM $\mathfrak{B}$, and 3) the initial state probability $\pi_i|_{i=1}^{N}$. The observation variables (outputs) are described by a vector of size m. The distribution of the observations for each hidden state (total of N hidden states) can be modeled as a mixture of M multivariate Gaussians. In the proposed approach, fully-connected hidden Markov models (HMM)s are used. The fully-connected models enable identifying an optimal set of states that best represent a desired motion path without the need to pass through all the states and being restricted with the left-to-right order of the states as is the case in the left-to-right models.

The Baum-Welch algorithm is used to train HMMs and the Viterbi algorithm [169] is used to generate the most likely state sequence for a desired motion path given an HMM of its Laban Effort and Shape nearest neighbours.

5.2.4 The Proposed Generation Approach

Figure 5.3 shows a schematic of the proposed generation approach. Consider a sample dataset of n time-series movements x, each labelled with one of K affective labels from a set $\mathcal{E}$, $\mathcal{D} = \{(x_j, e_j) | x_j \in \mathbb{R}^{m \times T_j}, e_j \in \mathcal{E}, j = 1, \ldots, n\}$, where x_j and e_j denote the j^{th} sampled movement of length T_j and its affective label, respectively. For a desired motion path $x_d \in \mathbb{R}^{m \times T_d}$ and a target emotion $e_t \in \mathcal{E}$, the proposed approach selects movements from $\mathcal{D}$ that are kinematically similar to the desired motion path and belong to the target emotion class e_t. The selected movements are referred to as *nearest neighbours*. x_d may be taken from the dataset $\mathcal{D}$ (e.g., if we wish to take a kinematic movement from the existing set and alter the emotion), or may be specified independently of $\mathcal{D}$.

The nearest neighbour search is performed in a space spanned by affectively non-discriminative Laban Effort and Shape components. Bartenieff emphasizes that in addition to kinematic characteristics, emotions and their intensity can be precisely studied via the Laban Effort and Shape components [1]. Therefore, different affective movements might occupy distinct regions of the space spanned by all the Laban Effort and Shape components and finding nearest neighbours of a desired motion path in such a space might be impossible.

To overcome this limitation, we restrict our nearest neighbour search to affectively non-discriminative Laban components for the target emotion class. We assume that affectively non-discriminative Laban Effort and Shape components are the components describing the kinematic characteristics of the movements. To identify affectively non-discriminative Laban components, the multinomial logistic regression with elastic net regularization is used [204], as summarized in Section 5.2.2. The regularized multinomial logistic regression (RMLR) uses the movements' Effort and Shape components and affective labels as the explanatory variables and class labels, respectively, and weights the Effort and Shape components according to their saliency for each affective class (RMLR assigns zero weights to affectively non-discriminative Laban components for each emotion class).

Given the identified affectively non-discriminative Laban components, the nearest neighbours (NN) of the desired motion path belonging to the target emotion class are next identified. The number of nearest neighbours is set to the number of movements from the target emotion class in the ϵ-neighbourhood of the desired motion path. The ϵ-neighbourhood is defined as a circle centered at the desired motion path with a radius equal to 10% of the distance from the desired motion path to the furthest movement from the target emotion class in a space spanned by the affectively non-discriminative Laban components.

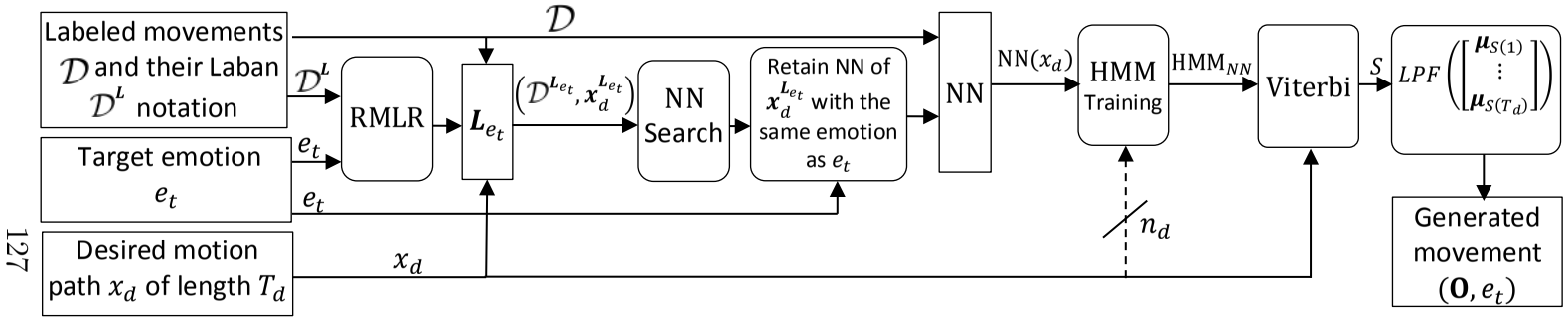

Figure 5.3: Schematic of the proposed approach. The rectangles represent input data or an outcome of a process and the processes are shown by rounded rectangles. $\mathcal{D}^{\mathbf{L}_{e_t}}$ and $x_d^{\mathbf{L}_{e_t}}$: the representation of $\mathcal{D}$ and x_d in terms of affectively non-discriminative Laban components $\mathbf{L}_{e_t}$ | NN: indices of nearest neighbours of x_d | NN(x_d): nearest neighbours of x_d from $\mathcal{D}$ | HMM$_{NN}$: HMM of NN(x_d) | S: the most likely HMM$_{NN}$'s state sequence for x_d | $\mu_{\mathbf{i}}$ is the mean of the Gaussian output of the i^{th} state of HMM$_{NN}$ | LPF: low-pass filter | $\mathbf{O}$: the modulated x_d conveying e_t | n_d number of copies of x_d encoded in HMM$_{NN}$.

In cases where the desired motion path (the motion path to be modulated in order to convey a target emotion) is affective, its encoded emotion (original emotion) can be recognized using a verified automatic affective movement recognition model (e.g., Chapter 4). Then, for a target emotion, Laban components discriminating between the pair of original and target emotions are identified using RMLR and excluded from the nearest neighbour search as these components contribute to maximum separability between the original and target emotion classes.

After identifying the NN of the desired motion path, they are encoded in an HMM (HMM_{NN}). To further preserve the kinematic specificities of a desired motion path in the generated movement, the HMM_{NN} can be augmented with n_d copies of the desired motion path (shown as dashed line in Figure 5.3). Encoding more copies of the desired motion path in the HMM_{NN} imposes extra constraints on the generated movement, favouring stronger adoption of the kinematics of the desired motion path at the expense of the affective modulation. Thus, n_d becomes a process parameter which controls the trade-off between strict adherence to either the affective constraint or the kinematic constraint. After training the HMM_{NN}, the Viterbi algorithm is used to estimate the most likely state sequence for the desired motion path given HMM_{NN}. A new motion path is generated by concatenating the means of the Gaussian outputs of the HMM_{NN} states, following the estimated Viterbi state sequence. The resulting sequence is low-pass filtered to remove discontinuities that might occur when transitioning to new states [194].

The output of the filtering stage is a novel movement kinematically close to the desired motion path and imbued with the target emotion ($\mathbf{O}$ in Figure 5.3). Kinematic characteristics of the desired motion path are preserved in two stages of the proposed approach: 1) by using an HMM abstraction of the nearest neighbours in a space spanned by affectively non-discriminative Laban components (HMM_{NN}), and 2) by using the most likely state sequence of HMM_{NN} for the desired motion path. Furthermore, in the case where $n_d > 0$, the resulting HMM_{NN} state sequence is further tailored toward kinematic specificities of the desired motion path; hence, the generated movement more closely mimics the kinematics of the desired motion path. The target emotion is overlaid as the movement is generated using an HMM of the nearest neighbours conveying the target emotion. Figure 5.4 shows the pseudo-code for the proposed approach.

5.3 Experimental Setup

The full-body movement dataset used in Chapter 4 is used here as well to demonstrate the performance of the proposed approach. A 10-fold division is used to divide the dataset into

Given: A dataset $\mathcal{D}$ with n movements x each labeled with one of K affective classes from a set $\mathcal{E}$.

Laban Effort and Shape representation

- **Define L**, a set of all Laban Effort and Shape components (a total of $n_{\mathbf{L}}$ components),
- **Compute** $\mathbf{x}_j^{\mathbf{L}} \in \mathbb{R}^{n_{\mathbf{L}}}$: Laban Effort and Shape representation, $\forall$ movements $x_j \in \mathcal{D}$, $\rightarrow \mathcal{D}^{\mathbf{L}} = \{\mathbf{x}_1^{\mathbf{L}}, \ldots, \mathbf{x}_n^{\mathbf{L}}\}$,

Regularized multinomial logistic regression:

- **Input:** $\{(\mathbf{x}_j^{\mathbf{L}}, e_j) | \mathbf{x}_j^{\mathbf{L}} \in \mathcal{D}^{\mathbf{L}}, e_j \in \mathcal{E}\}$,
- **Output:** the set of affectively discriminative Laban components for k^{th} class ($\mathbf{L}'_k$).
  - $\mathbf{L}_k = \mathbf{L} \setminus \mathbf{L}'_k$, affectively non-discriminative Laban components for k^{th} class.

Affective movement generation

- **Given** $\begin{cases} x_d, & \text{a desired motion path of length } T_d, \\ e_t, & \text{a target emotion,} \end{cases}$
- **Let** $\mathbf{L}_{e_t}$: the set of affectively non-discriminative Laban components for e_t,
- **Find** nearest neighbours of x_d in the space spanned by $\mathbf{L}_{e_t} \rightarrow \mathrm{NN}^{L_{e_t}}(x_d)$,
  - Retain those that belong to e_t: $\mathrm{NN}(x_d) = \{(x, e) \in \mathcal{D} | x \in \mathrm{NN}^{L_{e_t}}(x_d), e = e_t\}$,
- **HMM Training**
  - Train HMM_{NN} of N states to model $\mathrm{NN}(x_d)$ along with n_d copies of x_d; $n_d \geq 0$.
- **Viterbi algorithm using $\mathbf{HMM}_{NN}$ and x_d**
  S $\leftarrow$ most likely state sequence for x_d
  $S = [S_1, \ldots, S_i, \ldots, S_{T_d}], S_i \in \{1, \ldots, N\}$,
- **Motion generation**
  $\mathbf{O} = [\mu_{S_1}, \mu_{S_2}, \ldots, \mu_{S_{T_d}}]^*$
  $\mathbf{O} \leftarrow$ **Low-pass Filter**$(\mathbf{O})$

Return:

O: A modulated version of x_d conveying e_t.

$^*\mu_{\mathbf{i}}$ is the mean of the Gaussian output of the i^{th} state of HMM_{NN}.

Figure 5.4: Pseudo-code for the proposed generation approach.

testing and training movements. In each fold, a testing movement is considered the desired motion path and the proposed approach is applied to produce a new movement conveying the target emotion. Note that in these experiments, the testing movement already contains affective content. The target emotion is selected to be different than the original emotion encoded in the testing movement. The training movements in each fold establish the labeled dataset required for the proposed approach. The affective quality of the generated movements is objectively evaluated using the recognition model described in Chapter 4. Furthermore, a subjective evaluation is conducted via a user study in which participants rate the affective qualities of the generated movements. For a description of the full-body dataset, refer to Section 4.2.1.1.

5.3.1 RMLR Parameter Selection

In order to identify affectively non-discriminative Laban components for each target emotion, multinomial logistic regression with elastic net regularization (RMLR) is used. To set the tuning parameters of RMLR, α and λ in Equation 5.14, a two-dimensional cross-validation is used. α values in the range of $0.05 : 0.05 : 1$ and 100 values of λ are tested and the pair of α and λ with the minimum 10-fold cross-validation error (misclassification rate) is selected[4].

5.3.2 HMM Initialization and Model Selection

In order to reduce variability due to differences in demonstrators' physical builds (e.g., height), each movement in the dataset is normalized by the length of its demonstrator's hand and arm measured from his/her right shoulder to the right wrist.

Since all the movements in the full-body dataset start from a known T-pose, the hidden state sequence for the HMMs always starts at state 1; hence, the hidden state priors are: $\pi_i = 1$ for $i = 1$, and $\pi_i = 0$, otherwise. The full-body movements are generally very short and progress from the start toward the end with instances of cyclic movements. In this work, all transitions between states were allowed to capture the cyclic behaviours. Using fully-connected HMMs, the Viterbi algorithm identifies a sequence of HMM states that best represents a desired motion path without the need to pass through all the HMM states, sequentially.

[4]The generalized linear models toolbox from [206] was used to perform RMLR.

As described in Section 5.2, movements that are kinematically similar to a desired motion path along with n_d copies of the desired motion path are encoded in an HMM (HMM_{NN}). In the experimental evaluation of the proposed approach, no copies of the desired motion path are included in the HMM_{NN} abstractions ($n_d = 0$), favouring the affective constraint over the kinematic constraint. In the HMM_{NN} abstractions, the distribution of the movements for each hidden state is modeled as a single Gaussian output. Generally, for generation purposes, an HMM with a large number of hidden states is favoured to capture fine details of the motion path. We tested different numbers of states ranging between 10-14 states and chose 12 states by a visual inspection for the quality of the generated trajectories.

To initialize the HMM training process, the identified nearest neighbour movements for a desired motion path are divided into N (number of hidden states) equal segments and their means and covariances are used as the initial means and covariances of the output Gaussians associated with the hidden states. Full covariances are used to account for potential correlations between body joint trajectories.

5.3.3 Affective Movement Validation

To evaluate the performance of the proposed approach, the expressivity of the generated movements is assessed objectively using an automatic recognition model, and subjectively via a user study.

5.3.3.1 Objective Validation

The automatic recognition approach from Chapter 4 is used to objectively evaluate the synthetically-generated affective movements. Note that none of the generated movements in this study were used to train the recognition model.

5.3.3.2 Subjective Validation

Since affective movement perception is subjective, we have also evaluated the expressivity of the generated movements with a questionnaire-based user study. A questionnaire similar to the one presented in Section 3.3.1 is designed in which participants are asked to evaluate the expressivity of the generated affective movements in terms of the six basic Ekman emotions (anger, sadness, happiness, fear, disgust, and surprise) on a 5-point Likert scale ranging between "0: not conveyed at all" to "4: strongly conveyed".

Using 10-fold division of the full-body dataset into testing and training movements, a large number of movements (549 movements) are generated and evaluating the subjective perception of all these movements is not feasible. To reduce the number of movements for evaluation, we identify and remove movements which are similar to others, keeping only exemplars for evaluation. To identify similar movements in a conversion class, the generated movements from the conversion class are represented in terms of Laban Effort and Shape components and k-means clustered. Different numbers of clusters (2 to 5) are tested and the best number of clusters is selected to maximize the goodness of clustering metric [207] defined as

$$GOC = \frac{\sum_{i=1}^{k} n_i \sum_{j \neq i} \frac{n_i}{n - n_i} d_{ij}}{2 \sum_{i=1}^{k} n_i d_{ii}}, \tag{5.15}$$

where d_{ij} and d_{ii} are the average distances between clusters i and j and within cluster i, respectively.

Next, the popularity of the identified clusters for each conversion class is inspected and the closest movement to the center of the most populous cluster is selected as the exemplar of that conversion class for evaluation in the user study. Therefore, there are 12 generated movements used in the user study, each representing a conversion class. For the conversion class of 'sadness to happiness', there are two populous clusters with an equal number of members. In this case, the cluster whose representative displays a more kinematically distinct movement than the other generated movements was selected. This is done solely to include a wider range of kinematic movements in the user study. For each movement, participants were asked to rate the six basic Ekman emotions, each on a 5 point Likert scale. To test whether the target emotion is unambiguously recognized, the participants in the user study were asked to rate each basic emotion separately, and were not forced to chose between emotions; they could indicate that they observed more than one emotion, or that they observed no emotional content. The movements were presented to the participants in a randomized order.

A pilot user study was conducted with 6 participants to obtain an estimate of effects of the original and target emotions on the participants' perception. The estimated effects vary in size and range from very small effects (e.g., $\eta^2 = 0.001$) to large ones (e.g., $\eta^2 = 0.745$). Using the G*Power software [208], the sample size required to detect a potential significance of the estimated effects of medium size ($\eta^2 = 0.06$; as recommended by Cohen [164]) or larger at the statistical power of 90% is found to be 16 participants.

17 participants (11 male and 6 female) completed the questionnaire. Participants were healthy adults and recruited from among the students at the University of Waterloo via email. They were provided with the detailed information about the study and the procedure

to complete the questionnaire. The study received ethics approval from the Office of Research Ethics, University of Waterloo, and a consent form was signed electronically by each participant prior to the start of the questionnaire. Participants' responses are then analyzed to assess whether the generated movements communicate the target emotions and if there exists any confusion between the basic emotions.

5.4 Results

As described in Section 4.2.1.1, the dataset used for the experimental evaluation of the proposed approach contains affective full-body movements for 4 emotion classes: sadness, happiness, fear, and anger. Following the proposed approach, each testing movement is converted to convey three emotions other than its own, resulting in 12 conversion classes. The following naming convention is used hereafter to refer to the conversion classes: 'original to target' (e.g., 'sadness to happiness' indicate the conversion from the original emotion 'sadness' to the target emotion 'happiness').

As described in Section 5.3, a 10-fold division is used to divide the affective full-body movement dataset into testing and training movements. In the experimental evaluations, the testing movements are the desired motion paths, and the training movements form the labeled dataset $\mathcal{D}$ used in the proposed approach as described in Section 5.2.4.

Using the quantification described in Section 5.2.1, movements are annotated in terms of Effort components (Weight, Time, and Flow) and Shape components (Shaping, Directional, and Flow). For the Effort components, in addition to the whole-body annotation, these components are also computed for individual body parts: head, right hand, left hand, right foot, and left foot. Since all the movements (testing and training) used in the experimental evaluation are affective (intended to convey a specific emotion), the Laban components that discriminate between emotions encoded in the training movements and those encoded in the testing movement are first identified using RMLR, and excluded from the search for the nearest training movements. This will allow finding movements that are kinematically most similar to a desired motion path from within the target emotion class (training movements in the ϵ-neighbourhood of the desired motion path that belong to the target emotion class). Furthermore, this will result in a computationally efficient nearest neighbour search based on a few Laban components.

We denote the set of all Laban components as $\mathbf{L}$. The set of affectively non-discriminative Laban components for a pair of emotions (e_1, e_2), $\mathbf{L}_t$, is defined as

$$\mathbf{L}_t = \mathbf{L} \setminus (\mathbf{L}_1 \cap \mathbf{L}_2), \tag{5.16}$$

where $\mathbf{L}_1$, and $\mathbf{L}_2$ are the Laban components salient to e_1 and e_2, respectively. Table 5.2 shows the discriminative Laban components between each pair of emotions in the full-body dataset.

Table 5.2: Discriminative Laban components for each pair of emotions identified using regularized multinomial logistic regression.

	Laban Effort and Shape components
Sadness-Happiness	WeightRFoot, TimeAll, TimeTorso, TimeRHand, TimeLFoot, FlowRFoot, FlowLFoot, ShapeHor, ShapeFlow, ShapeDirRHand
Sadness-Fear	TimeTorso, TimeRHand, TimeLFoot, TimeHead, FlowRFoot, ShapeZ, ShapeHor, ShapeFlow
Sadness-Anger	WeightRFoot, TimeTorso, TimeRHand, TimeRFoot, TimeLFoot, ShapeHor, ShapeDirRHand
Happiness-Fear	WeightTorso, TimeRHand, TimeLFoot, FlowHead, ShapeZ, ShapeDirRHand
Happiness-Anger	TimeLHand, FlowLFoot, FlowHead, ShapeHor, ShapeFlow
Fear-Anger	WeightHead, TimeTorso, TimeLFoot, TimeHead, ShapeZ, ShapeHor

R:right, L:left, Hor: Horizontal, Dir: Directional.

Among the emotion classes in the full-body dataset, sadness and happiness have the largest number of Laban components discriminating between them (see Table 5.2), which could be due to the distinct natures of these two emotions. For instance, in the circumplex space [165], sadness and happiness are located at opposing extrema of the arousal and valence dimensions. Shape Horizontal and Time for the left foot are the most frequent discriminative Laban components presented in 5 out of 6 cases in Table 5.2. There are components that are found discriminative only for one pair of emotions: 1) TimeAll measures the Effort Time for the whole body and is found to be discriminative for the sadness-happiness pair. 2) Weight Effort for torso (WeightTorso) is found to be discriminative for the happiness-fear pair. 3) Time for the left hand (TimeLHand) is discriminative for the happiness-anger pair, and 4) Weight for the head (WeightHead) discriminates between fear and anger. It should be emphasized that the identified Laban components in Table 5.2 collectively contribute to maximum separability between the pairs of emotions and when considered individually may not result in maximum separation.

Next, the proposed approach is used to generate movements that convey target emotions. In the experimental evaluation presented in this section, no copies of the desired motion path are included in the HMM_{NN} abstractions ($n_d = 0$). Videos of movement exemplars generated by alternative implementations of the proposed approach ($n_d = 0$ and $n_d = 1$) can be found in [209].

5.4.1 Objective Validation

Table 5.3 shows the confusion matrix for the automatic recognition of generated full-body movements. The generated affective movements are recognized above chance with a 72% recognition rate, which is comparable with the 72% interpersonal recognition rate reported in [168] (Section 4). Similar confusions between emotion classes as those in Section 4 are observed here. Nevertheless, the automatic recognition rate of the generated movements is high and demonstrates the congruence of the generated emotions with the target ones.

Table 5.3: Confusion matrix (%) for the automatic recognition of the generated movements

			Recognized Emotion			
			Sadness	Happiness	Fear	Anger
Target	Emotions	Sadness	**83**	1	15	1
		Happiness	1	**61**	22	15
		Fear	3	8	**77**	13
		Anger	1	15	16	**67**

5.4.2 Subjective Validation

The performance of the proposed approach is also evaluated subjectively via a questionnaire-based user study as described in Section 5.3.3.2. Using a one-way ANOVA, the effect of the order in which the videos were presented to a participant was not found significant at $p < 0.05$, indicating no habituation was observed as a result of the presentation order.

To inspect the effects of the original and target emotions on the participants' perception of the generated movements, a mixed-model ANOVA with repeated measures was performed for each emotion rating. In the ANOVA tests, the original and target emotions were the independent variables and participant ratings of the basic emotions were the dependant variables. Initially, the participant's gender was also included as a between-subject variable, however, its effect was not found significant at $p < 0.05$, thus, the ANOVA tests

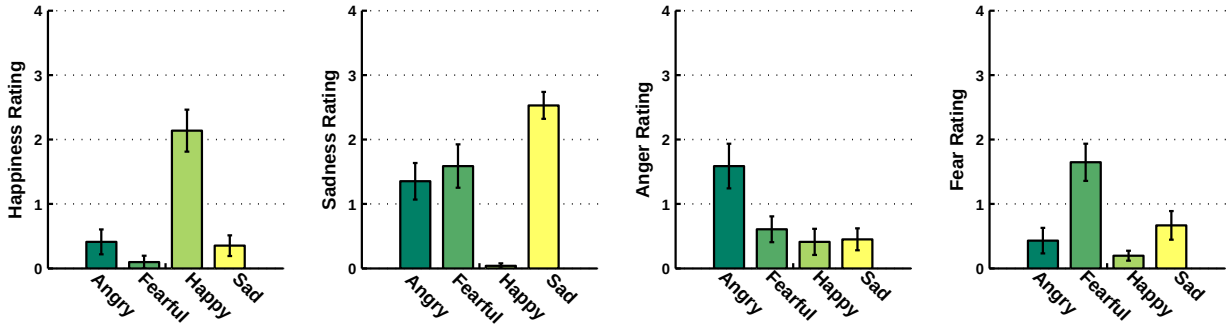

Figure 5.5: Average participants' ratings (mean±SE) for the target emotions. Each bar shows an average rating across all the movements with the same target emotion (x-axis is the target emotion).

were repeated without the gender variable. In the previous user study reported in Section 3.3, gender-specific differences were observed in the perception of affective movements displayed on different embodiments. In the present user study, a single embodiment (a stick figure) was used to display the generated affective movements and no significant effect of gender was observed. The SPSS statistical software package was used for the analysis.

It was found that the target emotion has a significant main effect in all cases ($p < 0.005$), and original emotion has a main effect on participants' rating of anger, sadness, fear, surprise ($p < 0.05$). There are also significant interaction effects between the original and target emotions in all cases except for the rating of disgust. Furthermore, participants' ratings of an emotion were found to be significantly higher (at a Bonferroni corrected significance level of $p < 0.0125$ with an effect size Cohen's $d > 0.67$) when the emotion is the target emotion (e.g., sadness rating for anger to sadness conversion) as compared with cases when the emotion is not the target emotion (e.g., sadness rating in anger to happiness conversion). This observation can also be seen in the average participants' ratings of target anger, sadness, fear, and happiness emotions shown in Figure 5.5.

The detected main effects of the target emotion indicate that the proposed approach successfully encoded the target emotion in the generated movements. The detected significant effects of the original emotion might indicate the presence of residues of the original emotion in the generated movements or may be an indicator of deficiencies in the perception of emotion from body movements.

A closer look at the participants' perception of individual movements reveals a few cases of confusion between the emotions, some of which are between the original and

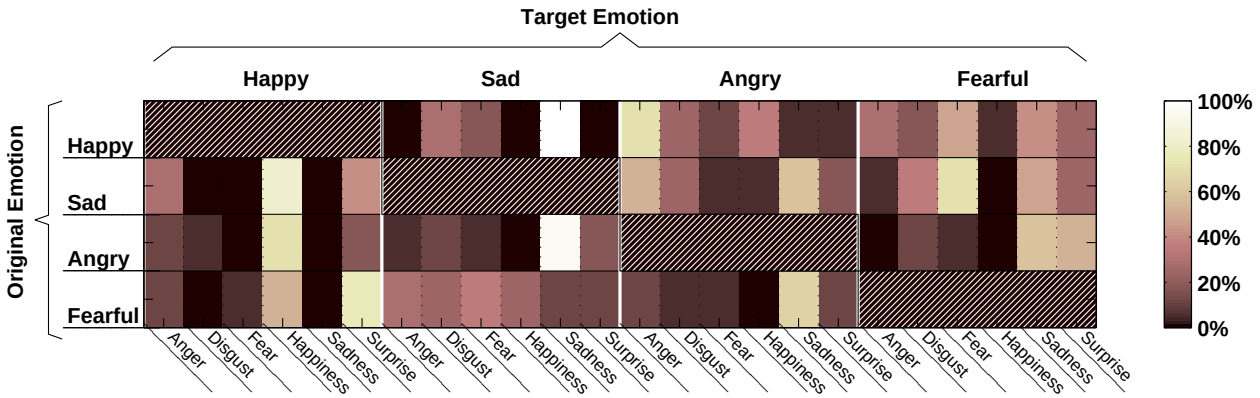

Figure 5.6: A heatmap showing the average recognition rates of different emotions for the conversion classes. An emotion is considered recognized if it is rated 2 or above on the Likert scale spanning between 0 to 4. Note that this recognition cut-off is applied for illustrative purposes in this figure only and all the subjective evaluation reported in this section is performed on the full scale of rating.

target emotions. Figure 5.6 is a heat map showing the average recognition of different basic emotions for each conversion class (confusion matrix). To explore significant pair-wise differences, for each movement, paired t-tests between participants' ratings of its target emotion and those of other emotions were performed. Figure 5.7 highlights the results of the paired t-tests. The green boxes are the target emotions, grey boxes are the emotions whose ratings are significantly different than the corresponding target emotion at $p < 0.05$ (i.e., there was no confusion), and red boxes are those emotions whose ratings are not significantly different than the target emotion at $p < 0.05$ (i.e., the emotions were confused). The Cohen's d effect size was also computed to evaluate the size of pair-wise differences between participants' ratings of target emotions and those of other emotions for each movement. Cohen's recommendation for effect size evaluation is used here (small: $d < 0.2$, medium: $d = 0.5$, and large: $d > 0.8$).

As can be seen in Figures 5.6 and 5.7, the target emotion is clearly perceived in the conversions from anger and sadness to happiness at a rate >71% (significant pair-wise differences are detected between ratings of happiness and other emotions at $p < 0.001$ with effect sizes $d > 1.22$), whereas in the fear to happiness conversion, there is no significant difference between ratings of happiness and surprise ($p = 0.13$ and small to medium effect size of $d = 0.41$) and both happiness and surprise are rated high.

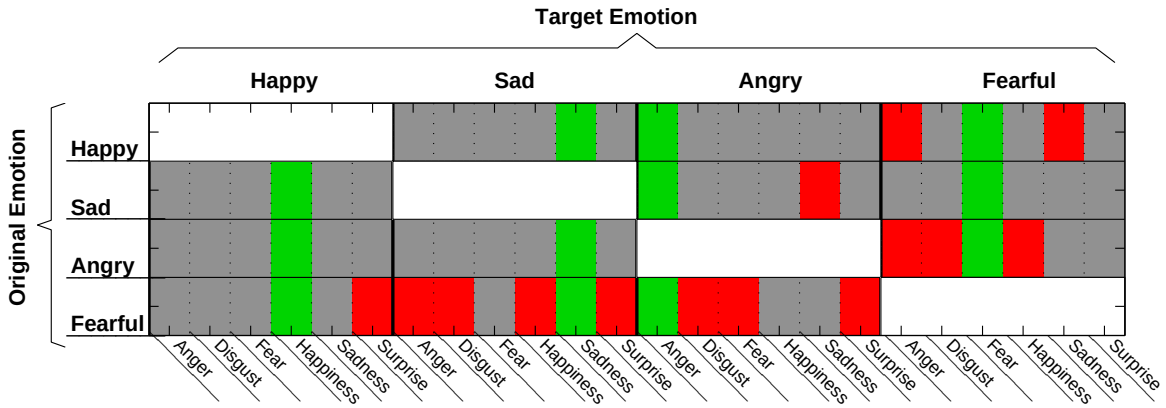

Figure 5.7: A heatmap showing significance of pair-wise differences between participants' ratings of target emotions and other emotions (paired t-tests). The green boxes highlight the target emotion, the grey boxes indicate significant differences to the ratings of the target emotion at $p < 0.05$, and a red box indicates that there is no significant difference to ratings of the target emotion at $p < 0.05$.

When the target emotion is sadness, it is correctly perceived for the conversions of happiness and anger to sadness at a rate >94%, $p < 0.001$, and $d > 2.76$ (2^{nd} column of Figures 5.6 and 5.7). For the fear to sadness conversion, the target emotion is only recognized at a 12% rate and no other emotion is rated high.

Anger is correctly perceived for the movement converted from happiness at a rate of 71% with significant pair-wise differences between ratings of the anger and other emotions at $p < 0.028$ with effect sizes $d > 0.93$ (3^{rd} column of Figures 5.6 and 5.7). For the sadness to anger conversion, anger and sadness are rated the highest with a small non-significant pairwise difference ($p = 0.47, d = 0.035$) between the ratings of anger (recognition rate of 53%) and those of sadness (recognition rate of 59%). This could have been caused by residues of sadness remaining in the converted movement (main effect of the original emotion on anger ratings observed in the ANOVA test). For the fear to anger conversion, the generated movement is confused as a sad movement with a recognition rate of 65% (Figure 5.6).

When the target emotion is fear, it is correctly recognized when converted from sadness at a rate of 71% with significant pair-wise differences between ratings of fear and those of other emotions at $p < 0.01, d > 1.07$. For happiness to fear conversion, despite the highest

rating for fear among other emotions (recognition rate of 47%), it is not significantly different than the ratings of sadness ($p = 0.46, d = 0.04$) and anger ($p = 0.17, d = 0.34$) (Figure 5.7). The fearful movement converted from anger is confused as sad (recognition rate of 59%) and surprise (recognition rate of 53%) (Figure 5.6).

5.5 Discussion

The performance of the proposed approach in generating movements with recognizable target emotions was evaluated objectively using an automatic recognition model from Chapter 4, and subjectively via a user study.

The encoded emotions are correctly recognized by the automatic recognition model from Chapter 4 for 72% of the generated movements, which is comparable with the recognition result achieved by the model for the affective movements from 13 demonstrators. The observed confusions reported in Table 5.3 are mainly between fear, happiness, and anger. Considering the circumplex model of emotion [165], the observed confusions seem to be related to the similarities between the affective expressions along the arousal and valence dimensions. For instance, anger and fear are both high arousal and negative valence expressions and are frequently confused.

In the subjective validation, the effect of the target emotion on the participants' perception was found significant. However, there are cases where the target emotion is confused with other emotions. In all of these cases, either the original or target emotion is fear with the exception of the sadness to anger conversion. One hypothesis is that the observed confusions could have resulted from inaccurate training movements. The movements in the full-body dataset are labeled by their demonstrators, and therefore, they might not fully communicate the demonstrator-intended emotions. Inclusion of these movements would degrade the performance of the proposed approach as the training movements used to modulate a desired motion path to convey a target emotion are not sufficiently (or accurately) affective. Therefore, the performance of the proposed approach depends on the availability of a dataset containing a diverse range of training movements whose expressivity is verified by different observers. In addition, the observed deficiencies in the perception of fear are congruent with reports from the literature [71, 73, 72], which indicate that other modalities in addition to body movements might be required for the perception of fear. After excluding the cases with original or target fear emotion, the participants recognized the target emotions in the generated movements at a rate of 71% in the user study.

A direct comparison between the recognition rates from the objective and subjective evaluation experiments is not possible as the objective evaluation is performed for all

the generated movements, whereas the subjective evaluation is done for twelve movement exemplars only (one for each conversion class). Nevertheless, the higher objective recognition rates (especially for generated fearful movements) are likely because the automatic recognition model was trained using the affective movements in the full-body dataset to discriminate between four emotions (anger, happiness, fear, and sadness), while in the user study, participants rated all the six basic Ekman emotions for each movement, which might have rendered the rating of target emotions more difficult.

We have excluded the cases in which the original emotion is fear and rerun the ANOVA tests. A significant main effect of the target emotion was found for all cases at $p < 0.01$. The original emotion only has a significant main effect on anger and fear ratings after excluding movements whose original emotion is fear. For fear ratings, there is a large variation between ratings of fear across different original emotions (e.g., last column of Figure 5.6). For the sadness to anger conversion, both anger and sadness are rated high ($p = 0.47, d = 0.035$; sadness to anger movement in Figures 5.6 and 5.7), which shows the main effect of the original emotion on the participants' rating of anger. Figure 5.8 shows the velocity changes for the sadness to anger movement used in the user study. It shows an overall drooped body posture, which is characteristic of sadness. Combined with the fast and sudden forward and inward movement of the hands and feet (see Figure 5.8), we hypothesize that the mixture of typically-angry extremity movements and the drooped posture of sadness may communicate frustration, which could explain the confusion between sadness and anger for the sadness to anger movement (see the supplementary video for sadness to anger conversion in [209]).

Nevertheless, after excluding the cases in which the original emotion is fear, in 6 out of 9 remaining cases, the target emotion is rated the highest by the participants with significant pairwise differences (at $p < 0.05$) between the ratings of the target emotion and those of other emotions. These significant pairwise differences are of medium to large size (Cohen's $d > 0.7$), which indicate that the target emotion was unambiguously recognized in these cases. Therefore, the proposed approach is capable of generating movements with recognizable target emotions.

In the proposed approach, the nearest neighbours to a desired motion path from a target emotion class in an available training dataset might not closely resemble the kinematic specificities of the desired motion path. In such cases, when $n_d = 0$, the kinematic trajectory of the generated movement might exhibit deviations from that of the desired motion path. To overcome this limitation, the n_d parameter of the proposed approach can be tuned to force the generated movement to adopt the kinematic specificities of its corresponding desired motion path. To illustrate the effect of encoding copies of a desired motion path in the HMM_{NN} ($n_d > 0$) on the kinematic trajectory of the generated

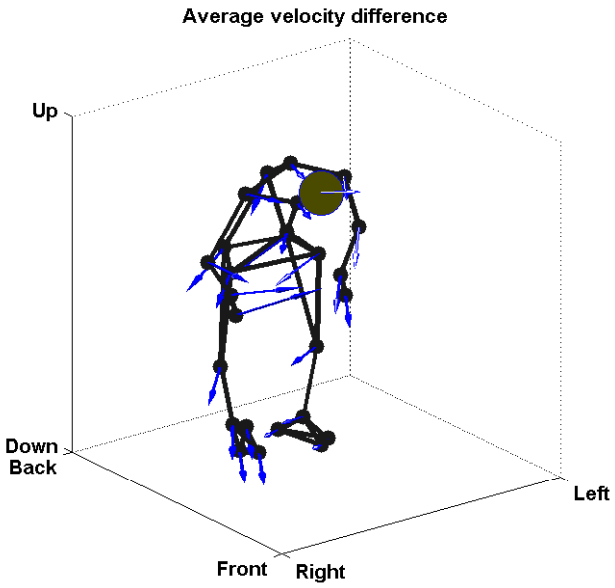

Figure 5.8: An illustration of the velocity changes (magnitude and direction) when converting from sadness to anger. The direction of the arrows shows the direction of the velocity and the length of the arrows is proportional to the increase in the velocity of their corresponding joints in the generated movement.

movement, we have generated exemplar movements by setting $n_d = 1$ and animated them side-by-side with their counterparts generated by setting $n_d = 0$ and corresponding desired motion paths (see videos [209]). We observed that the generated movements more closely mimic the kinematic trajectory of their corresponding desired motion paths when $n_d > 0$ as compared with the case where $n_d = 0$, as the HMM_{NN} explicitly encode the kinematics of the desired motion path when $n_d > 0$.

5.6 Summary

An example-based approach for automatic affective movement generation was proposed that makes use of two movement abstractions: the Laban Effort and Shape components and hidden Markov modeling. The proposed approach uses Laban nearest neighbours of a desired motion path in a labeled affective movement dataset to generate a new movement that conveys a target emotion and is kinematically close to the desired motion path. For

a desired motion path and a target emotion, a systematic approach is used (regularized multinomial logistic regression) to select a subset of affectively non-discriminative Laban Effort and Shape components for the target emotion based on which the nearest neighbours for the desired motion path are found. Therefore, the proposed approach does not require manual selection or definition of motion descriptors. Furthermore, the abstract semantic representation in terms of Laban components allows for an efficient search for nearest neighbours of a desired motion path as compared with the search in high-dimensional joint space. After identifying the nearest neighbours, the proposed approach encodes them in an HMM (HMM_{NN}) and uses the Viterbi algorithm to estimate the optimal state sequence for the desired motion path given the HMM of its nearest neighbours. The resulting state sequence is then used to generate a modulation of the motion path that conveys the target emotion.

Using only the nearest neighbours of a desired motion path reduces the dimensionality of the generation problem (HMM modeling), which in turn results in reduced computational complexity of the proposed approach. In comparison with other stochastic generative approaches that are non-parametric and assume continuous latent variables (e.g., Gaussian process dynamical model [144]), or multiple layers of interactive hidden states (e.g., conditional restricted Boltzmann machines [145]), HMMs are less expensive to learn and their complexity can be tuned (number of hidden states, transition between the states, training samples, and observation distributions) to accurately model movements.

The performance of the proposed approach is evaluated using a full-body affective movement dataset with movements conveying anger, sadness, happiness, and fear, and the expressivity of the generated movements are evaluated objectively using a state-of-the-art affective movement recognition model and subjectively using a perceptual user study. The generated movements were recognized at a rate of 72% by the automatic recognition model. There are cases in the user study where the target emotions were confused with other emotions. In all such cases, either the original or target emotion is fear. Excluding these confusions, participants in the user study were able to recognize target emotions at a rate of 71%. Therefore, the proposed approach is capable of generating movements with recognizable target emotions.

A favourable property of the proposed approach is that the HMM-based abstraction highlights specificities of the generated movements captured in hidden states (key postures) and the dynamics of transition between the states. This allows exploring how a movement is modified when converting from one emotion class to another.

A further advantage of the proposed approach is that generated movements can also serve as training movement exemplars to generate movements with an even larger vari-

ety of styles and encoded affective expressions. Furthermore, additional movements can be generated by a random walk in the proposed Laban space representation, and interpolating/extrapolating between the Laban representations of the movements in a labeled dataset.

Using the proposed approach, variances of body joints in an HMM_{NN} can be set to control the extent of spatial deviations in a generated movement from the nearest neighbours of the corresponding desired motion path. In cases where copies of a desired motion path are also encoded in the HMM_{NN}, tighter joint variances can be used to limit kinematic variabilities modeled by the HMM and as a result, the generated movement will more closely mimic the spatial trajectory of the desired motion path. In the present work the joint variances were not controlled.

Moreover, encoding additional copies of a desired motion path in HMM_{NN} ($n_d > 0$) enables a greater control on the spatial trajectory of the generated movement to mimic the desired motion path. This utility of the proposed approach is especially beneficial when there are no close kinematic neighbours from the target emotion class to the desired motion path in available training datasets. The possibility to augment the HMM_{NN} with the desired motion path, along with the possibility to control joint variances are further favourable properties of the proposed approach and demonstrate its flexibility.

Chapter 6

Conclusions

In this thesis, a computational analysis of time-series movements was carried out to design automatic models for recognition and generation of affective expressions in movements. To this end, feature extraction and stochastic modeling techniques were leveraged, adapted, and extended to derive abstract representations of affective movements that isolate features most salient to the discriminative or generative affective movement analysis. These movement abstractions were realized in terms of basis function expansion, stochastic transformation, and Laban Effort and Shape analysis of the body joint trajectories, and were exploited for salient movement feature identification and automatic affective movement recognition and generation. Furthermore, the thesis explores the human perception of affective movements displayed by different embodiments via a set of user studies. The contributions of the thesis are as follows:

A. *Affect recognition and generation based on functional movement representation* (**Chapter:** 3).

A systematic approach for transforming variable-length movements to fixed-length vectorial observations and recognizing affective expressions from movements was proposed based on functional movement representation. Various dimensionality reduction techniques were adapted to the class of functional observations and used to identify a subspace where class-specific affective movements are best separable. It was shown that functional supervised principal component analysis outperforms other functional reduction techniques in identifying a discriminative subspace where affective movement recognition can be performed. Furthermore, movement abstraction derived by functional PCA along with inverse functional PCA transformations were used to generate prototypical class-specific affective movements. The expressivity of the generated movements was confirmed via a user study.

This contribution is particularly important as it introduces a generic approach for fixed-length representation of sequential observations enabling the application of dimensionality reductions techniques and discriminative learning approaches to this class of data.

B. *Affect recognition based on hybrid generative-discriminative movement modeling* (**Chapter:** 4).

Despite the attractive features of the functional affective movement analysis, it is sensitive to temporal, stochastic, and interpersonal variabilities, and is limited to within-class kinematically similar movements and requires a tedious preprocessing stage of land-mark alignment and length normalization for generation. To address these shortcomings, a discriminative abstraction of affective movements based on a stochastic transformation of body joint trajectories was proposed. The proposed stochastic abstract modeling also enables the identification of features salient to affective expressions. In the proposed approach, movements belonging to an affective class are encoded in a single HMM with a mixture of Gaussian outputs and the resulting class-specific HMMs are used to obtain a Fisher score representation of the movements. The Fisher score representation is a fixed-length representation and allows the application of discriminative classifiers and dimensionality reduction techniques on the variable-length affective movements. An SVM classification in the resulting Fisher score space was shown to achieve higher recognition rates than other recognition models applied on the same affective movements. Using sPCA, Fisher scores were mapped to a low-dimensional subspace whose dimensions encompass the main modes of variation in the movements and along which the distribution of the movements is highly correlated with their affective expressions. The sPCA abstraction was used for automatic affective movement recognition and is shown to retain the high recognition rates achievable in the high-dimensional Fisher score space, but with only a few dimensions.

Furthermore, the resulting sPCA subspace dimensions form the salient movement features and were further explored to identify the corresponding movement kinematics and dynamics that communicate different affective expressions. For instance, for a full-body affective movement dataset, we found the following salient postures: 1) sad movements are generally characterized by a drooped head and shoulders, an overall slumped posture, and a protective crouch posture, 2) happy movements are characterized by open and expansive postures, upward hand and arm movements, and advancing movements, 3) angry movements are characterized by arms akimbo postures, angled forward postures, and clenching fists, and 4) fearful postures seem to show alert, retreating, and cowering movements. With regard to the speed of movements, we found that sad movements are slow while happy and angry movements are characterized by fast motions. Fearful movements are long in duration and are characterized by frequent jitteriness and freezing postures.

The proposed hybrid generative-discriminative modeling can be applied to any class of sequential observations to overcome the limitations associated with variable-length nature of these observations and irrelevant sources of variation and to optimize classification of the sequential observations into categories of interest.

C. *Affective movement generation based on Laban movement analysis and generative movement modeling* (**Chapter:** 5).

Another contribution of this thesis is an automatic example-based affective movement generation approach that makes use of two movement abstractions: 1) the Laban movement analysis, and 2) hidden Markov modeling. The Laban movement analysis provides a compact movement representation that captures both kinematic and expressive characteristics of movements and enables an efficient search for movements kinematically or affectively similar to a given movement. In collaboration with a CMA, this thesis proposes a quantification approach for the Laban Effort and Shape components and validates its efficacy against annotations from the CMA. For a desired motion path and a target emotion, the proposed generation approach uses the quantified Laban components to identify movements that are kinematically-similar to the desired motion path and belong to the target emotion class (nearest neighbours) in available affective movement datasets. An HMM for the identified movements is learned and used with the desired motion path to generate a novel movement that is a modulated version of the desired motion path, conveying the target emotion.

The expressivity of the synthetically-generated movements was verified objectively using the automatic Fisher-score recognition approach from Contribution B and subjectively with a user study in which participants rated observed affective expressions from the movements. The proposed approach is computationally efficient as it operates on a stochastic abstraction of a small number of nearest neighbours to a desired motion path and is capable of generating a wide range of affective movements provided that the required training data is available. A further advantage of the proposed approach is the ability to tune the relative importance of adopting the kinematic characteristics of a desired motion path or the expressive characteristics of a target emotion. The proposed approach can also be used to explore characteristic differences between affective classes in converting an affective movement from one class to another.

D. *Affective movement perception* (**Chapters:** 3, 5).

User studies were used in this thesis to evaluate the human perception of affective movements and study the impact of the display embodiment and the observer's gender on the perception of affective movements. It was found that the human perception of affective

movements is significantly influenced by the intended emotions and that the display embodiment and the observer's gender might effect the perception of affective movements. Furthermore, user studies were also used for subjective evaluation of the synthetically-generated affective movements.

6.1 Concluding Remarks and Future Work

The perceptual and computational affective movement recognition experiments in the thesis show that it is feasible to perceive and express affect via body movements. Furthermore, it appears that affective expressions can be perceived from movements of non-anthropomorphic structures and that the kinematic embodiment and physical appearance of a structure can influence the perception of its affective expressions. For example, if a structure doesn't have a human-like appearance, it may be more difficult for a human observer to ascribe the full range of human emotions to it (e.g., sad movements were frequently described as conveying happiness and positive valence when displayed on the frond-like structure).

The developed computational models for affective movement recognition and generation, the identified salient affective movement features, and insights about the human affective movement perception gained in this thesis can inform the development of affect-aware interactive machines. Such machines will be capable of recognizing affective expressions of their users and displaying appropriate actuated responses with recognizable affective content. However, there are challenges in developing such affect-aware machines. These challenges stem from appearance, kinematic, and dynamic differences between the machines and humans. Kinematic and dynamic differences constrain the range of motion executable by a machine, which along with the machine's physical appearance may limit the range of expressions conveyable or in some cases may alter the perception of intended affective movements displayed by the machine. Therefore, the physical structure and presentation of the display embodiment may influence the human interaction with them and merits further exploration. We have also observed gender-specific differences in the perception of affective movements displayed on different structures, which emphasize the importance of considering the user's gender in designing affect-aware machines and call for validating displayed affective movements by interactive machines with both male and female users. The datasets used in this thesis for evaluating the developed computational models contain movement exemplars from both male and female demonstrators and male and female participants were also used in the conducted user-studies to evaluate the performance of the proposed generation models. Additionally, larger studies should be conducted to iden-

tify if there exist physical structures and appearances that might limit (or modulate) the communication of affective expression with male or female observers.

The occupants' culture and the context within which the movements are presented might also influence the affective movement perception. In future work, it will be important to investigate and consider the impact of culture and context on the affective movement perception.

Combining other modalities such as facial expression, physiological activities, or verbal communication with body movements to improve affective communication is another future direction for this research. The efficacy of multi-modal affective communication should also be explored for non-anthropomorphic interactive machines.

In order to develop computational models that emulate human users in affective movement recognition, the perception of human users needs to be incorporated in establishing the ground truth. Kleinsmith and Bianchi-Berthouze [7] discuss challenges and propose solutions for establishing the ground truth for an affective expression [210]. In future studies, a ground truth for the affective expressions encoded in movements should be established following the recommendations in [7, 210] and used to further evaluate the correspondence between the performance of the proposed computational affective movement recognition and generation models and that of the human observers.

The proposed HMM-based Fisher score representation eliminates the need for temporal alignment of the movements, and can handle variable length kinematically similar/affectively different and kinematically different/affectively similar movements. The proposed approach is especially beneficial as it automatically identifies salient discriminative movement features; hence, no assumptions are needed about the importance of specific movement features for discriminating between affective movements. The sPCA subspace lends itself as a suitable basis for mapping between discriminative low-level movement features (the subspace dimensions) and high-level movement descriptors (e.g., obtained from movement notation systems such as Laban [58]) that would inform generic affective movement recognition and generation. In this thesis, the sPCA subspace was used to infer motion and postural cues salient to different affective expressions. A further exploration of correspondences between sPCA subspace and high-level movement features, and development of generic recognition models driven by high-level movement features are avenues for future research.

One of the main barriers to the development of reliable computational models for affective movement analysis is the lack of datasets that represent nuances in affective expression and are obtained under naturalistic settings [7]. As a result, movement features salient to these expressions are not well understood. The proposed Fisher-score-based

discriminative approach in this thesis is well suited to this problem as it automatically extracts salient movement features, which may not be obvious with more naturalistic and non-basic affective expressions. The application of the proposed approach to such affective movement datasets is a direction for future research.

Neutral expressions are common during human-machine interaction; hence, the importance of considering them in computational analysis of affect. The proposed discriminative modeling approaches can be applied to identify features specific to the neutral state, which along with the knowledge of salient features for other states could inform exploration of kinematic and dynamic transitions from the neutral state to other affective states. This is another potential future avenue for the research presented in the thesis.

The proposed approach for affective movement generation provides parameters (joint variances and number of copies of a desired motion path (n_d) encoded in HMM_{NN}), which can be set to control the trade-off between the adherence to specified affective constraints (a target emotion) or kinematic constraints (a desired motion path) in the generated movement. A systematic approach for setting the parameters of the proposed approach based on motion designer objectives is warranted and is a direction to be explored in the future. Furthermore, an approach to quantify the Laban Body and Space major components should be investigated and the quantified Body and Space components should be used along with the quantified Effort and Shape components in the search for movements kinematically and/or affectively similar to a desired motion path.

Aside from HCI/HRI applications, automatic recognition of affective expressions from body movements can also be used for patient monitoring and to tailor patient-specific treatments, especially in physical rehabilitation therapy, where patients commonly display protective body movements caused by pain and/or anxiety about exacerbating the pain [7]. Having a mechanism in place that can detect and recognize protective behaviours and adapt the rehabilitation exercises to the patient's comfort level could improve patient's adherence to the therapy, and in return expedite patient's recovery.

The research on computational affective movement analysis can also aid the research on autistic patients who suffer from impairment in understanding and conveying affect [211]. Social robotic companions with expressive body language can help encouraging basic communication and social skills in children with autism [212].

In high-demand working environments, work cycles can be adapted upon detection of anxiety, boredom, or lack of attention from worker's body movements to avoid mistakes or injury. For instance, by monitoring the worker's emotional state, robotic motions can be optimized to increase safety and decrease the worker's anxiety [213]. Finally, computational

models for automatic affective movement recognition can also aid surveillance systems in detecting suspicious individuals from distance in high-security locations (e.g., [214]).

APPENDICES

Appendix A

Summary Tables for the Literature Review in Chapter 2

A.1 Affective Movement Perception

Table A.1: A selective list of affective movement perception studies. A: Anger, C: Content, D: Disgust, F: Fear, H: Happiness, N: Neutral, P: Pride, S: Sadness, Ś: Surprise.

Study	Year	Movement	Emotions and dimensions	Visualization	Recog. rate
[69]	2004	Full-body	A, H, S, F, D	Static (dynamic) PL Static (dynamic) FL	39%(76%) 54%(85%)
[66]	1998	Dance	A, H, S, F	Video	96%
[81]	2007	Head	N, A, H, S	Animated agent	-
[85]	1937	Hand-arm	34 emotions*	Photographs and videos	

*Anger, Distrust, Scorn, Defiance, Worship, Grief, Surprise, Entreaty, Sympathy, Haughtiness, Terror, Thoughtfulness, Curiosity, Assurance, Wistful appeal, Resignation, Critical superiority, Quandary, Satisfaction, Pain, Scolding, Avoidance, Disgust, Bewilderment, Stubbornness, Admiration, Humility, Welcome, Supplication, Threatening, Strong anger, Despair, Determination, Contemplation.

Table A.1 continued . . .
Anger, C: Content, D: Disgust, F: Fear, H: Happiness, N: Neutral, P: Pride, S: Sadness, Ś: Surprise.

Study	Year	Movement	Emotions and dimensions	Visualization	Recog. rate
[72]	2004	Full-body	A, D, F, H, S, S	Static puppet	-
[74]	2007	Gait	N, A, C, H, S	Video	67%
[67]	2003	Dance	A, F, S, H	Video	-
[68]	1996	Dance	A, S, H, Ś, D, F	PL Full-light display	63% 88%
[87]	2004	Hand	N, A, H	Video	62 %
[75]	2010	Gait	N, A, S, H Arousal Valence Dominance	Animated puppet	63% 61% 55% 62%
[70]	2009	dance	A, H, S, F	Video	-
[73]	1987	Gait	A, H, S, P	Video	–
[29]	2000	Hand-arm	A, H, N, S, F	PL FL	59% 71%
[86]	1992	Hand	N, A, H, S, Ś	Video	77%
[215]	2008	Gait	N, A, H, S	Unspecified	88%
[76]	2009	Gait	A, H, S, F	Animated puppet	78%
[105]	2011	Hand	A, H, S	Animated human–like, stick and frond-like hand models	-
[88]	2003	Hand-arm	A, H, S	Video	-
[71]	2007	Full body	A, H, S, F	Gray-scale photographs	>80%

A.2 Affective Movement Recognition

Table A.2: Recognition of affective expressions from body movements. * indicates person-dependent rates. An earlier version of this table has been published in [8]. A: Anger, B: Boredom, C: Clam, Ć: Concentration, C̈: Confusion, D̈: Delight, Ď: Despair, E: Excited, F: Fear, F́: Frustration, I: Interest, H: Happy, Ḧ: High-intensity negative, Í: Irritation, Ĺ: Low-intensity negative, J: Joy, P: Pride, Ś: Surprise, S̈: Serene.

Study	**Year**	**Movement**	**Affective States**	**Feature Extraction**	**Classifier**	**Recog. Rate**
[216]	2012	Arm	Pleasure, arousal	Correlation Index	Linear regression	Pleasure: $R = .37$ Arousal: $R = .73$
[123]	2007	Knocking	N, H, A, S	5 selected features	SVM	50%, 81%*
[121]	2004	Dance	A, F, S, H	Amount and quality of movement	Decision tree	36%
[132]	2007	Arm	A, H, pleasant, S, Ď, I, Í, P	Correlation based FS, wrapper FS	DTW, NN, J48, HNB	4 emotions: 63%, 66%*, 8 emotions: 41%, 47%*
[125]	2008	Pianist	S, S̈, overexpressive	Quantity of upper-body movement and head velocity	Dicision trees, ...	58%
[124]	2009	Sitting	B, C̈, D̈, flow, F́	112 selected features	Bayesian, SVM, kNN, trees, ...	39%
[130]	2006	Gestures	S, F́, J, H	Key-Points	HMM	79%

Table A.2 continued . . .
Anger, B: Boredom, C: Clam, Ć: Concentration, ́D: Delight, Ď: Despair, E: Excited, F: Fear, Ŕ: Frustrated, I: Interest, H: Happy, Ĥ: High-intensity negative, Í: Irritation, Ĺ: Low-intensity negative, J: Joy, Ś: Surprise, ́S: Serene.
$\star$ indicates person-dependent rates
Hilarious, social, awkward, fake, non-laughter.

Study	**Year**	**Movement**	**Affective States**	**Feature Extraction**	**Classifier**	**Recog. Rate**
[119]	2013	Laughter	4 laughter types[a]	Kinematic and dynamic features	kNN, MLP, RF, Ridge regression, SVR	66%
[35]	2009	Upper-body	12 affective states	PCA	BayesNet, SVM, Random forest, Adaboost, HMM	77%
[126]	2010	Head	1)Pleasure, 2)Arousal, 3)Expectation, 4)Intensity, 5)Power	Head angle	HMM, SVR	MSE 1).15$\star$, .06$\star$, .10$\star$ 2).13$\star$, .07$\star$, .06$\star$ 3).13$\star$, .11$\star$, .09$\star$ 4).10$\star$, .05$\star$, .09$\star$ 5).14$\star$, .12$\star$, .12$\star$

Table A.2 continued . . .
Anger, B: Boredom, C: Clam, Ć: Concentration, ́D: Delight, Ď: Despair, E: Excited, F: Fear, Ŕ: Frustrated, I: Interest, H: Happy, Ĥ: High-intensity negative, Í: Irritation, Ĺ: Low-intensity negative, J: Joy, Ś: Surprise, ́S: Serene. ⋆ indicates person-dependent rates.

Study	**Year**	**Movement**	**Affective States**	**Feature Extraction**	**Classifier**	**Recog. Rate**
[114]	2008	Gait	N, H, S, A C, E, N	200 features 368 features	ANN ANN	84%⋆ 79%⋆
[75]	2010	Gait	N, H, S, A	PCA,LDA	SVM, NN, NB	69%, 95%⋆
[75]	2010	Gait	Pleasure, arousal, dominance	PCA, LDA	NN	Pleasure: 88%⋆, Arousal: 97%⋆, Dominance: 96%⋆
[117]	2005	Gestures	S, J, A, F	8 selected features	Logistic regression, SVM, J48, NB, ANN	85%, 93%⋆
[120]	2010	Gestures & Body Movement	A, Ď, I, pleasure, S, Í, J, P	Software EyesWeb[217]	Bayesian classifier	67%

Table A.2 continued . . .
Anger, B: Boredom, C: Clam, Ć: Concentration, ́D: Delight, Ď: Despair, E: Excited, F: Fear, Ḟ: Frustrated, I: Interest, H: Happy, Ḧ: High-intensity negative, Í: Irritation, Ĺ: Low-intensity negative, J: Joy, Ś: Surprise, ́S: Serene. ⋆ indicates person-dependent rates.

Study	**Year**	**Movement**	**Affective States**	**Feature Extraction**	**Classifier**	**Recog. Rate**
[115]	2006	Full-body postures	S, H, A, F	Pair-wise joint distances	Mixture discriminant analysis	85%
[116]	2007	Full-body postures	Valence, arousal, potency, avoidance	Pair-wise joint distances	NN	Valence: 79%⋆, Arousal: 79%⋆, Potency: 81%⋆, Avoidance: 81%⋆
[218]	2011	Shoulder	Valence, arousal (continuous)	5 points on shoulder and torso	SVR, BLSTM-NN	RMSE for valence: 0.21⋆ RMSE for arousal: 0.29⋆
[122]	2004	Dance	H, Ś, A, S	SVD	MLP	73%
[118]	2012	Full-body	H, Ć, Ḧ, Ĺ	Dynamic features	Recurrent Neural Network	55%, 61%⋆
[88]	2003	Arm	J, S, A	3 selected features	Discriminant analysis	78%
[131]	2007	Arm	7 emotions	PCA	SVM, NN	73%⋆

References

[1] I. Bartenieff, *Effort-Shape analysis of movement: The unity of expression and function.* Albert Einstein College of Medicine, Yeshiva University, 1965.

[2] P. Hackney, *Making connections: Total body integration through Bartenieff fundamentals.* Routledge, 2004.

[3] P. Beesley, "Hylozoic soil," *Leonardo*, vol. 42, no. 4, pp. 360–361, 2009.

[4] J. Ramsay, *Functional data analysis*, 2nd ed. New York: Springer Science+Media, 1997.

[5] A. Damasio, *Descartes' error: Emotion, reason and the human brain.* New York: Harcourt Brace, 1994.

[6] J. Burgoon, D. Buller, and W. Woodall, *Nonverbal communication.* Allyn & Bacon, 2010.

[7] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 15–33, 2013.

[8] M. Karg, A. Samadani, R. Gorbet, K. Kuhnlenz, J. Hoey, and D. Kulić, "Body Movements for Affective Expression: A Survey of Automatic Recognition and Generation," *IEEE Transactions on Affective Computing*, vol. 4, no. 4, pp. 341–359, 2013.

[9] C. Darwin, *The expression of emotions in man and animals.* Chicago university press, Chicago, 1872.

[10] W. James, "What is an Emotion?" *Mind*, vol. 9, no. 34, pp. 188–205, 1884.

[11] A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 18–37, 2010.

[12] R. Picard, *Affective Computing*. The MIT Press, 1997.

[13] C. Breazeal, *Designing sociable robots*. The MIT Press, 2004.

[14] T. Spyropoulos and S. Spyropoulos, "Minimaforms kinetic sculptures," available at www.minimaforms.com/petting-zoo-frac/.

[15] S. Park, G. Chae, C. MacDonald, R. Stein, S. Wiedenbeck, and J. Kim, "AMARA: The affective museum of art resource agent," in *CHI '12 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2012, pp. 1035–1038.

[16] A. Samadani, E. Kubica, R. Gorbet, and D. Kulić, "Perception and generation of affective hand movements," *International Journal of Social Robotics*, vol. 5, no. 1, pp. 35–51, 2013.

[17] A. Samadani, A. Ghodsi, and D. Kulić, "Discriminative functional analysis of human movements," *Pattern Recognition Letters*, vol. 34, no. 15, pp. 1829 – 1839, 2013.

[18] M. Pantic and L. Rothkrantz, "Automatic analysis of faial expressions: The sate of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000.

[19] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, 2009.

[20] C. Bethel and R. Murphy, "Survey of Non-facial/Non-verbal Affective Expressions for Appearance-Constrained Robots," *IEEE Transactions on Systems, Man, Cybernetics. C, Applications and Reviews*, vol. 38, no. 1, pp. 83–92, 2008.

[21] A. Ortony, *Affective Information Processing*. Springer, 2009, ch. 2: Affect and Emotions in Intelligent Agents: Why and How?, pp. 11–21.

[22] K. Scherer, *Blueprint for affective computing: A sourcebook*. Oxford University Press, 2010, ch. Emotion and emotional competence: Conceptual and theoretical issues for modelling agents, pp. 3–20.

[23] K. Scherer, "What are emotions? And how can they be measured?" *Social Science Information*, vol. 44, no. 4, pp. 695–729, 2005.

[24] C. Derbaix and M. Pham, "Affective reactions to consumption situations: A pilot investigation," *Journal of Economic Psychology*, vol. 12, no. 2, pp. 325–355, 1991.

[25] R. Cowie, N. Sussman, and A. Ben-Zéev, "Emotion: Concepts and definitions," in *Emotion-Oriented Systems*, ser. Cognitive Technologies, R. Cowie, C. Pelachaud, and P. Petta, Eds. Springer Berlin Heidelberg, 2011, pp. 9–30.

[26] N. Dael, M. Mortillaro, and K. Scherer, "The body action and posture coding system (BAP): Development and reliability," *Journal of Nonverbal Behavior*, vol. 36, no. 2, pp. 97–121, 2012.

[27] I. Mauss and M. Robinson, "Measures of emotion: A review," *Cognition and Emotion*, vol. 23, no. 2, pp. 209–237, 2009.

[28] J. Tracy and R. Robins, "Show your pride: Evidence for a discrete emotion expression," *Psychological Science*, vol. 15, no. 3, pp. 194–197, 2004.

[29] F. Pollick, H. Paterson, A. Bruderlin, and A. Sanford, "Perceiving affect from arm movement," *Cognition*, vol. 82, no. 2, pp. B51–B61, 2001.

[30] S. Marsella, J. Gratch, and P. Petta, *Blueprint for affective computing: A sourcebook*. Oxford University Press, 2010, ch. Computational models of emotion, pp. 21–41.

[31] S. Planalp, *Communicating emotion: Social, moral, and cultural processes*. Cambridge University Press, 1999.

[32] R. Buck, *The communication of emotion*. Guilford Press New York, 1984.

[33] R. Walk and K. Walters, "Perception of the smile and other emotions of the body and face at different distances," *Bulletin of the Psychonomic Society*, vol. 26, no. 6, pp. 510–510, 1988.

[34] B. de Gelder, J. Van den Stock, H. Meeren, C. Sinke, M. Kret, and M. Tamietto, "Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions," *Neuroscience and Biobehavioral Reviews*, vol. 34, no. 4, pp. 513–527, 2010.

[35] H. Gunes and M. Piccardi, "Automatic temporal segment detection and affect recognition from face and body display," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 39, no. 1, pp. 64–84, 2009.

[36] P. Ekman, "Darwin, Deception, and Facial Expression," *Annals of the New York Academy of Sciences*, vol. 1000, no. 1, pp. 205–221, 2003.

[37] P. Ekman and W. Friesen, "Detecting deception from the body or face," *Journal of Personality and Social Psychology*, vol. 29, no. 3, pp. 288–298, 1974.

[38] H. Aviezer, Y. Trope, and A. Todorov, "Body cues, not facial expressions, discriminate between intense positive and negative emotions," *Science*, vol. 338, no. 6111, pp. 1225–1229, 2012.

[39] J. Tracy and D. Randles, "Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt," *Emotion Review*, vol. 3, no. 4, pp. 397–405, 2011.

[40] P. Ekman and W. Friesen, "A new pan-cultural facial expression of emotion," *Motivation and Emotion*, vol. 10, no. 2, pp. 159–168, 1986.

[41] P. Ekman, "Are there basic emotions?" *Psychological Review*, vol. 99, no. 3, pp. 550–553, 1992.

[42] R. Plutchik and H. Conte, Eds., *Circumplex models of personality and emotions*. APA, 1997.

[43] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, no. 4, pp. 261–292, 1996.

[44] L. Barrett and J. Russell, "The structure of current affect: Controversies and emerging consensus," *Current Directions in Psychological Science*, vol. 8, no. 1, pp. 10–14, 1999.

[45] J. Mikels, B. Fredrickson, G. Larkin, C. Lindberg, S. Magold, and P. Reuter-Lorenz, "Emotional category data on images from the international affective picture system," *Behavior Research Methods*, vol. 37, no. 4, pp. 626–630, 2005.

[46] W. Havlena and M. Holbrook, "The varieties of consumption experience: Comparing two typologies of emotion in consumer behavior," *Journal of Consumer Research*, vol. 13, no. 3, pp. 394–404, 1986.

[47] J. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *Journal of Research in Personality*, vol. 11, no. 3, pp. 273–294, 1977.

[48] L. Barrett, "Solving the emotion paradox: Categorization and the experience of emotion," *Personality and Social Psychology Review*, vol. 10, no. 1, pp. 20–46, 2006.

[49] P. Ekman and W. Friesen, *The Facial action coding system.* Consulting Psychologists Press, Stanford Univ., Palo Alto, 1977.

[50] R. Birdwhistell, *Kinesics and context: Essays on body motion communication.* University of Pennsylvania Press, 1970.

[51] P. Ekman and W. Friesen, "The repertoire of nonverbal behavior: Categories, origins, usage, and coding," *SEMIOTICA*, vol. 1, no. 1, pp. 49–98, 1969.

[52] M. Ptaszynski, J. Maciejewski, P. Dybala, R. Rzepka, and K. Araki, "CAO: A fully automatic emoticon analysis system based on theory of kinesics," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 46–59, 2010.

[53] T. Shawn, *Every little movement: A book about François Delsarte.* Dance Horizons, 1968.

[54] M. Nixon, P. Pasquier, and M. El-Nasr, "DelsArtMap: Applying delsarte's aesthetic system to virtual agents," in *Intelligent Virtual Agents, LNCS 6356.* Springer Berlin Heidelberg, 2010, pp. 139–145.

[55] S. Marsella, S. Carnicke, J. Gratch, A. Okhmatovskaia, and A. Rizzo, "An exploration of delsarte's structural acting system," in *International conference on Intelligent Virtual Agents, LNCS 4133.* Springer-Verlag, 2006, pp. 80–92.

[56] N. Dael, M. Mortillaro, and K. Scherer, "Emotion expression in body action and posture." *EMOTION*, 2011.

[57] E. Velloso, A. Bulling, and H. Gellersen, "AutoBAP: Automatic coding of body action and posture units from wearable sensors," in *International Conference on Affective Computing and Intelligent Interaction*, 2013, pp. 135 – 140.

[58] R. Laban and F. Lawrence, *Effort.* Macdonald and Evans, 1947.

[59] R. Laban and L. Ullmann, *The mastery of movement.* ERIC, 1971.

[60] J. Rett, "Robot-human interface using Laban movement analysis inside a bayesian framework," Ph.D. dissertation, University of Coimbra, January 2009.

[61] D. Chi, M. Costa, L. Zhao, and N. Badler, "The EMOTE model for Effort and Shape," in *Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '00. ACM Press/Addison-Wesley Publishing Co., 2000, pp. 173–182.

[62] T. Nakata, T. Mori, and T. Sato, "Quantitative analysis of impression of robot bodily expression based on Laban movement theory," *Journal of Robotics Society of Japan*, vol. 19, no. 2, pp. 104–111, 2001.

[63] M. Kapadia, I. Chiang, T. Thomas, N. Badler, and J. Kider, "Efficient motion retrieval in large databases," in *ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, ser. I3D. ACM, 2013.

[64] T. Lourens, R. Van Berkel, and E. Barakova, "Communicating emotions and mental states to robots in a real time parallel framework using Laban movement analysis," *Robotics and Autonomous Systems*, vol. 58, no. 12, pp. 1256–1265, 2010.

[65] J. Harrigan, "Proxemics, kinesics, and gaze," *The new handbook of methods in nonverbal behavior research*, pp. 137–198, 2005.

[66] R. Boone and J. Cunningham, "Children's decoding of emotion in expressive body movement: The development of cue attunement." *Developmental Psychology; Developmental Psychology*, vol. 34, no. 5, pp. 1007–1016, 1998.

[67] A. Camurri, I. Lagerlöf, and G. Volpe, "Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques," *International Journal of Human-Computer Studies*, vol. 59, no. 1, pp. 213–225, 2003.

[68] W. Dittrich, T. Troscianko, and S. L. ans D. Morgan, "Perception of emotion from dynamic point-light displays represented in dance," *Perception*, vol. 25, no. 6, pp. 727–738, 1996.

[69] A. Atkinson, W. Dittrich, A. Gemmell, and A. Young, "Emotion perception from dynamic and static body expressions in point-light and full-light displays," *PERCEPTION*, vol. 33, no. 6, pp. 717–746, 2004.

[70] I. Lagerlöf and M. Djerf, "Children's understanding of emotion in dance," *European Journal of Developmental Psychology*, vol. 6, no. 4, pp. 409–431, 2009.

[71] J. Van den Stock, R. Righart, and B. de Gelder, "Body expressions influence recognition of emotions in the face and voice." *Emotion*, vol. 7, no. 3, pp. 487–494, 2007.

[72] M. Coulson, "Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence," *Nonverbal Behavior*, vol. 28, no. 2, pp. 117–139, 2004.

[73] J. Montepare, S. Goldstein, and A. Clausen, "The identification of emotions from gait information," *Journal of Nonverbal Behavior*, vol. 11, no. 1, pp. 33–42, 1987.

[74] E. Crane and M. Gross, "Motion capture and emotion: Affect detection in whole body movement," in *International Conference on Affective Computing and Intelligent Interaction, LNCS 4738*, 2007, pp. 95–101.

[75] M. Karg, K. Kühnlenz, and M. Buss, "Recognition of affect based on gait patterns," *IEEE Transactions on Systems, Man, Cybernetics: B, Cybernrtics*, vol. 40, no. 4, pp. 1050–1061, 2010.

[76] C. Roether, L. Omlor, A. Christensen, and M. Giese, "Critical features for the perception of emotion from gait," *Journal of Vision*, vol. 9, no. 6, pp. 1–32, 2009.

[77] J. Michalak, N. Troje, J. Fischer, P. Vollmar, T. Heidenreich, and D. Schulte, "Embodiment of sadness and depression - gait patterns associated with dysphoric mood," *Psychosomatic Medicine*, vol. 71, no. 5, pp. 580–587, 2009.

[78] P. Beesley, "Kinetic architectures and geotextile installations," *Riverside Architectural Press*, 2010.

[79] K. Takahashi, M. Hosokawa, and M. Hashimoto, "Remarks on designing of emotional movement for simple communication robot," in *IEEE International Conference on Industrial Technology (ICIT)*, 2010, pp. 585–590.

[80] P. Ekman and W. Friesen, "Head and body cues in the judgement of emotion: A reformulation," *Perceptual and Motor Skills*, vol. 24, no. 3, pp. 711–724, 1967.

[81] C. Busso, Z. Deng, M. Grimm, U. Neumann, and S. Narayanan, "Rigid head motion in expressive speech animation: Analysis and synthesis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1075–1086, 2007.

[82] H. Wallbott, "Bodily expressions of emotion," *European Journal of Social Psychology*, vol. 28, pp. 879–896, 1998.

[83] J. Fast, *Body language*. Pocket, 1988.

[84] M. Argyle, *Bodily communication* . Methuen, 1988.

[85] L. Carmichael, S. Roberts, and N. Wessell, "A study of the judgment of manual expression as presented in still and motion pictures," *The Journal of Social Psychology*, vol. 8, no. 1, pp. 115–142, 1937.

[86] J. Reilly, M. McIntire, and H. Seago, "Affective prosody in american sign language," *Sign Language Studies*, vol. 75, pp. 113–28, 1992.

[87] J. Hietanen, J. Leppänen, and U. Lehtonen, "Perception of emotions in the hand movement quality of finnish sign language," *Journal of nonverbal behavior*, vol. 28, no. 1, pp. 53–64, 2004.

[88] M. Sawada, K. Suda, and M. Ishii, "Expression of emotions in dance: Relation between arm movement characteristics and emotion," *Perceptual and Motor Skills*, vol. 97, no. 3, pp. 697–708, 2003.

[89] J. Hall, "Gender effects in decoding nonverbal cues." *Psychological Bulletin*, vol. 85, no. 4, pp. 845–857, 1978.

[90] J. Hall and D. Matsumoto, "Gender differences in of multiple emotions from facial expressions." *Emotion*, vol. 4, no. 2, pp. 201–206, 2004.

[91] G. Kirouac and F. Dore, "Accuracy of the judgment of facial expression of emotions as a function of sex and level of education," *Journal of Nonverbal Behavior*, vol. 9, no. 1, pp. 3–7, 1985.

[92] B. Montagne, R. Kessels, E. Frigerio, E. De Haan, and D. Perrett, "Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity?" *Cognitive Processing*, vol. 6, no. 2, pp. 136–141, 2005.

[93] H. Wagner, C. MacDonald, and A. Manstead, "Communication of individual emotions by spontaneous facial expressions." *Journal of Personality and Social Psychology*, vol. 50, no. 4, pp. 737–743, 1986.

[94] N. Rotter and G. Rotter, "Sex differences in the encoding and decoding of negative facial emotions," *Journal of Nonverbal Behavior*, vol. 12, no. 2, pp. 139–148, 1988.

[95] M. Mandal and S. Palchoudhury, "Perceptual skill in decoding facial affect," *Perceptual and Motor Skills*, vol. 60, no. 1, pp. 96–98, 1985.

[96] L. Brody, "Gender, emotional expression, and parent-child boundaries," *Emotion: Interdisciplinary perspectives*, pp. 139–170, 1996.

[97] U. Donges, A. Kersting, and T. Suslow, "Women's greater ability to perceive happy facial emotion automatically: Gender differences in affective priming," *PLoS ONE*, vol. 7, no. 7, p. e41745, 2012.

[98] A. Sokolov, S. Krüger, P. Enck, I. Krägeloh-Mann, and M. Pavlova, "Gender affects body language reading," *Frontiers in Psychology*, vol. 2, 2011.

[99] K. Johnson, L. McKay, and F. Pollick, "He throws like a girl (but only when he's sad): Emotion affects sex-decoding of biological motion displays," *Cognition*, vol. 119, no. 2, pp. 265–280, 2011.

[100] S. Halovic and C. Kroos, "Facilitating the perception of anger and fear in male and female walkers," in *AISB, Symposium on Mental States, Emotions and their Embodiment*, 2009, pp. 3–7.

[101] J. Spiegel and P. Machotka, *Messages of the body.* Free Press, 1974.

[102] Aldebaran-Robotics company., available at www.aldebaran-robotics.com.

[103] A. Beck, B. Stevens, K. Bard, and L. Cañamero, "Emotional body language displayed by artificial agents," *ACM Transactions on Interactive Intelligent Systems*, vol. 2, no. 1, pp. 1–29, 2012.

[104] M. Saerbeck and C. Bartneck, "Perception of affect elicited by robot motion," in *ACM/IEEE International Conference on Human-robot Interaction.* ACM, 2010, pp. 53–60.

[105] A. Samadani, B. DeHart, K. Robinson, D. Kulić, E. Kubica, and R. Gorbet, "A study of human performance in recognizing expressive hand movements," in *IEEE International Symposium on Robot and Human Interactive Communication*, 2011, pp. 93–100.

[106] F. Heider and M. Simmel, "An experimental study of apparent behavior," *American Journal of Psychology*, vol. 57, no. 2, pp. 243–259, 1944.

[107] J. Hodgins, J. O'Brien, and J. Tumblin, "Perception of human motion with different geometric models," *IEEE Transactions on Visualization and Computer Graphics*, vol. 4, no. 4, pp. 307–316, 1998.

[108] R. McDonnell, S. Jörg, J. McHugh, F. Newell, and C. O'Sullivan, "Evaluating the emotional content of human motions on real and virtual characters," in *ACM Symposium on Applied perception in graphics and visualization*. ACM, 2008, pp. 67–74.

[109] P. Blythe, P. Todd, and G. Miller, "How motion reveals intention: Categorizing social interactions," *Evolution and cognition*, pp. 257–285, 1999.

[110] J. Lee, J. Park, and T. Nam, "Emotional interaction through physical movement," in *International Conference on Human-Computer Interaction: Intelligent Multimodal Interaction Environments, LNCS 4552*. Springer, 2007, pp. 401–410.

[111] C. Bartneck, J. Reichenbach, and A. Van Breemen, "In Your Face, Robot! the influence of a character's embodiment on how users perceive its emotional expressions," in *Design and Emotion*, 2004, pp. 32–51.

[112] A. Powers, S. Kiesler, S. Fussell, and C. Torrey, "Comparing a computer agent with a humanoid robot," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2007, pp. 145–152.

[113] R. Blake and M. Shiffrar, "Perception of human motion," *Annual Review of Psychology*, vol. 58, pp. 47–73, 2007.

[114] D. Janssen, W. Schöllhorn, J. Lubienetzki, K. Fölling, H. Kokenge, and K. Davids, "Recognition of emotions in gait patterns by means of artificial neural nets," *Journal of Nonverbal Behavior*, vol. 32, no. 2, pp. 79–92, 2008.

[115] A. Kleinsmith, P. De Silva, and N. Bianchi-Berthouze, "Cross-cultural differences in recognizing affect from body posture," *Interacting with Computers*, vol. 18, no. 6, pp. 1371–1389, 2006.

[116] A. Kleinsmith and N. Bianchi-Berthouze, "Recognizing affective dimensions from body posture," in *International Conference on Affective Computing and Intelligent Interaction, LNCS 4738*. Springer Berlin Heidelberg, 2007, pp. 48–58.

[117] A. Kapur, N. Virji-Babul, G. Tzanetakis, and P. Driessen, "Gesture-based affective computing on motion capture data," in *International Conference on Affective Computing and Intelligent Interaction*, 2005, pp. 1–7.

[118] N. Savva, A. Scarinzi, and N. Bianchi-Berthouze, "Continuous recognition of player's affective body expression as dynamic quality of aesthetic experience," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 3, pp. 199–212, 2012.

[119] H. Griffin, M. Aung, B. Romera-Paredes, C. McLoughlin, G. McKeown, W. Curran, and N. Bianchi-Berthouze, "Laughter type recognition from whole body motion," in *International Conference on Affective Computing and Intelligent Interaction*, 2013, pp. 349–355.

[120] L. Kessous, G. Castellano, and G. Caridakis, "Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis," *Journal on Multimodal User Interfaces*, vol. 3, no. 1, pp. 33–48, 2010.

[121] A. Camurri, B. Mazzarino, and G. Volpe, "Expressive Interfaces," *Cognition, Technology and Work*, vol. 6, pp. 15–22, 2004.

[122] H. Park, J. Park, U. Kim, and W. Woo, "Emotion recognition from dance image sequences using contour approximation," in *Structural, Syntactic, and Statistical Pattern Recognition, LNCS 3138*. Springer, 2004, pp. 547–555.

[123] D. Bernhardt and P. Robinson, "Detecting affect from non-stylised body motions," in *Interntional Conference of Affective Computing and Intelligent Interaction, LNCS 4738*. Springer Berlin / Heidelberg, 2007, pp. 59–70.

[124] S. D'Mello and A. Graesser, "Automatic detection of learner's affect from gross body language," *Applied Artificial Intelligence*, vol. 23, no. 2, pp. 123–150, 2009.

[125] G. Castellano, M. Mortillaro, A. Camurri, G. Volpe, and K. Scherer, "Automated analysis of body movement in emotionally expressive piano performances," *Music Perception: An Interdisciplinary Journal*, vol. 26, no. 2, pp. 103–119, 2008.

[126] H. Gunes and M. Pantic, "Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners," in *Intenraional Conference on Intelligent Virtual Agents*. Springer, 2010, pp. 371–377.

[127] H. Monkaresi, R. Calvo, and M. Hussain, "Automatic natural expression recognition using head movement and skin color features," in *International Working Conference on Advanced Visual Interfaces*. ACM, 2012, pp. 657–660.

[128] M. Schroder, E. Bevacqua, R. Cowie, F. Eyben, H. Gunes, D. Heylen, M. ter Maat, G. McKeown, S. Pammi, M. Pantic, C. Pelachaud, B. Schuller, E. de Sevin, M. Valstar, and M. Wollmer, "Building autonomous sensitive artificial listeners," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 165 –183, 2012.

[129] S. Scherer, M. Glodek, G. Layher, M. Schels, M. Schmidt, T. Brosch, S. Tschechne, F. Schwenker, H. Neumann, and G. Palm, "A generic framework for the inference of user states in human computer interaction," *Journal of Multimodal User Interfaces*, vol. 6, no. 3-4, pp. 117–141, 2012.

[130] P. De Silva, M. Osano, A. Marasinghe, and A. Madurapperuma, "Towards recognizing emotion with affective dimensions through body gestures," in *International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 269 –274.

[131] C. Shan, S. Gong, and P. McOwan, "Beyond facial expressions: Learning human emotion from body gestures," in *British Machine Vision Conference*, 2007, available at www.dcs.warwick.ac.uk/bmvc2007/proceedings/CD-ROM/papers/paper-276.pdf.

[132] G. Castellano, S. Villabla, and A. Camurri, "Recognising human emotions from body movement and gesture dynamics," in *International Conference on Affective Computing and Intelligent Interaction*, 2007, pp. 71–82.

[133] M. Karg, R. Jenke, W. Seiberl, K. Kühnlenz, A. Schwirtz, and M. Buss, "A comparison of PCA, KPCA and LDA for feature extraction to recognize affect in gait patterns," in *International Conference of Affective Computing and Intelligent Interaction*, 2009, pp. 195–200.

[134] J. Lasseter, "Principles of traditional animation applied to 3D computer animation," *ACM SIGGRAPH on Computer Graphics*, vol. 21, no. 4, pp. 35–44, 1987.

[135] M. Masuda and S. Kato, "Motion rendering system for emotion expression of human form robots based on Laban movement analysis," in *IEEE International Symposium on Robot and Human Interactive Communication*, 2010, pp. 324–329.

[136] K. Nakagawa, K. Shinozawa, H. Ishiguro, T. Akimoto, and N. Hagita, "Motion modification method to control affective nuances for robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 5003–5008.

[137] B. Hartmann, M. Mancini, and C. Pelachaud, "Implementing expressive gesture synthesis for embodied conversational agents," in *Gesture in Human-Computer Interaction and Simulation, LNCS 3881.* Springer Berlin Heidelberg, 2006, pp. 188–199.

[138] O. Arikan and D. Forsyth, "Interactive motion generation from examples," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 483–490, 2002.

[139] G. Taylor, G. Hinton, and S. Roweis, "Modeling human motion using binary latent variables," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, Eds. The MIT Press, 2007, pp. 1345–1352.

[140] M. Unuma, K. Anjyo, and R. Takeuchi, "Fourier principles for emotion-based human figure animation," in *ACM International Conference of SIGGRAPH*. ACM, 1995, pp. 91–96.

[141] T. Mukai and S. Kuriyama, "Geostatistical motion interpolation," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 1062–1070, 2005.

[142] L. Torresani, P. Hackney, and C. Bregler, "Learning motion style synthesis from perceptual observations," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, Eds. The MIT Press, 2007, pp. 1393–1400.

[143] M. Brand and A. Hertzmann, "Style machines," in *Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '00. ACM Press/Addison-Wesley Publishing Co., 2000, pp. 183–192.

[144] J. Wang, D. Fleet, and A. Hertzmann, "Gaussian process dynamical models for human motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, vol. 30, no. 2, pp. 283–298, 2008.

[145] G. Taylor and G. Hinton, "Factored conditional restricted boltzmann machines for modeling motion style," in *Annual International Conference on Machine Learning*, ser. ICML '09. New York, NY, USA: ACM, 2009, pp. 1025–1032.

[146] G. Hinton, "A practical guide to training restricted boltzmann machines," in *Neural Networks: Tricks of the Trade, LNCS 7700*. Springer Berlin Heidelberg, 2012, pp. 599–619.

[147] A. Samadani, R. Gorbet, and D. Kulić, "Gender differences in the perception of affective movements," in *Human Behavior Understanding, LNCS 7559*. Springer Berlin Heidelberg, 2012, pp. 65–76.

[148] R. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Human Genetics*, vol. 7, no. 2, pp. 179–188, 1936.

[149] E. Barshan, A. Ghodsi, Z. Azimifar, and M. Zolghadri Jahromi, "Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds," *Pattern Recognition*, vol. 44, no. 7, pp. 1357–1371, 2011.

[150] J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[151] C. De Boor, *A practical guide to splines*. Springer Verlag, 2001, vol. 27.

[152] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002, vol. 2.

[153] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf, "Measuring statistical dependence with Hilbert-Schmidt norms," in *Algorithmic learning theory, LNCS 3734*. Springer, 2005, pp. 63–77.

[154] T. Cox and M. Cox, "Multidimensional scaling," in *Handbook of Data Visualization*, ser. Springer Handbooks Comp.Statistics. Springer Berlin Heidelberg, 2008, pp. 315–347.

[155] Y. Bengio, J. Paiement, P. Vincent, O. Delalleau, N. Le Roux, and M. Ouimet, "Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral clustering," *Advances in neural information processing systems*, pp. 177–184, 2004.

[156] Measurand, "Motion capture systems," 2009, available at http://www.measurand.com.

[157] J. Ramsay, "Functional data analysis software," 2008, available at http://www.psych.mcgill.ca/misc/fda/software.html.

[158] J. Tenenbaum, V. Silva, and J. Langford, "Isomap software," 2000, available at http://isomap.stanford.edu/.

[159] A. Martinez and A. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, vol. 23, no. 2, pp. 228 –233, 2001.

[160] G. Xin, Z. De-Chuan, and Z. Zhi-Hua, "Supervised nonlinear dimensionality reduction for visualization and classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,*, vol. 35, no. 6, pp. 1098 –1107, 2005.

[161] R. Krauss and U. Hadar, "The role of speech-related arm/hand gestures in word retrieval," *Gesture, speech, and sign*, pp. 93–116, 1999.

[162] A. Samadani, "Affective hand movements displayed on different embodiments," available at www.eng.uwaterloo.ca/~asamadan/HandSurvey.html.

[163] SPSS, "Spss for windows, rel. 19.0," 2010, chicago: SPSS inc.

[164] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Taylor & Francis, 2013. [Online]. Available: http://books.google.ca/books?id=cIJH0lR33bgC

[165] J. Russell, "A circumplex model of affect," *Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[166] W. Wood, N. Rhodes, and M. Whelan, "Sex differences in positive well-being: A consideration of emotional style and marital status." *Psychological Bulletin*, vol. 106, no. 2, pp. 249–264, 1989.

[167] P. Lang, M. Greenwald, M. Bradley, and A. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, no. 3, pp. 261–273, 1993.

[168] A. Samadani, R. Gorbet, and D. Kulić, "Affective movement recognition based on generative and discriminative stochastic dynamic models," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 4, pp. 454–467, 2014.

[169] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257 –286, 1989.

[170] D. Bernhardt and P. Robinson, "Detecting emotions from connected action sequences," in *Visual Informatics: Bridging Research and Practice, LNCS 5857*, H. Badioze Zaman, P. Robinson, M. Petrou, P. Olivier, H. Schrder, and T. Shih, Eds. Springer, 2009, pp. 1–11.

[171] T. Jaakkola and D. Haussler, "Exploiting generative models in discriminative classifiers," in *Advances in Neural Information Processing Systems 11*, M. Kearns, S. Solla, and D. Cohn, Eds. The MIT Press, 1999, pp. 487–493.

[172] R. Urtasun, D. Fleet, and P. Fua, "Temporal motion models for monocular and multiview 3d human body tracking," *Computer Vision and Image Understanding*, vol. 104, no. 23, pp. 157 – 177, 2006, special Issue on Modeling People: Vision-based understanding of a persons shape, appearance, movement and behaviour.

[173] S. Fine, J. Navratil, and R. Gopinath, "A hybrid GMM/SVM approach to speaker identification," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, 2001, pp. 417–420.

[174] A. Vinokourov and M. Girolami, "Document classification employing the fisher kernel derived from probabilistic hierarchic corpus representations," in *European Colloquium on Information Retrieval Research*, 2001, pp. 24–40.

[175] A. Holub, M. Welling, and P. Perona, "Combining generative models and fisher kernels for object recognition," in *IEEE International Conference on Computer Vision*, vol. 1, 2005, pp. 136 – 143.

[176] L. Chen, H. Man, and A. Nefian, "Face recognition based on multi-class mapping of fisher scores," *Patterb Recognition*, vol. 38, no. 6, pp. 799–811, 2005.

[177] O. Aran and L. Akarun, "A multi-class classification strategy for fisher scores: Application to signer independent sign language recognition," *Pattern Recognition*, vol. 43, no. 5, pp. 1776–1788, 2010.

[178] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Machine Learning: ECML-98, LNCS 1398*, C. Nédellec and C. Rouveirol, Eds. Springer, 1998, pp. 137–142.

[179] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: A local SVM approach," in *17th International Conference on Pattern Recognition*, vol. 3, 2004, pp. 32–36.

[180] C. Bishop and N. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006.

[181] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *14th International joint conference on Artificial Intelligence*, vol. 2, 1995, pp. 1137–1143.

[182] J. Rodriguez, A. Perez, and J. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 569–575, 2010.

[183] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 1–27, 2011.

[184] J. Fontaine, K. Scherer, E. Roesch, and P. Ellsworth, "The World of Emotions is not Two-Dimensional," *Psychological Science*, vol. 18, no. 12, pp. 1050–1057, 2007.

[185] J. Newhagen, "TV news images that induce anger, fear, and disgust: Effects on approach-avoidance and memory," *Journal of Broadcasting and Electronic Media*, vol. 42, no. 2, pp. 265–276, 1998.

[186] A. Kleinsmith, "Grounding affect recognition on a low-level description of body posture," Ph.D. dissertation, University College London, 2010.

[187] K. Brodersen, C. Soon, K. Stephan, and J. Buhmann, "The balanced accuracy and its posterior distribution," in *International Conference on Pattern Recognition*, 2010, pp. 3121–3124.

[188] P. De Silva, A. Kleinsmith, and N. Bianchi-Berthouze, "Towards unsupervised detection of affective body posture nuances," in *International Conference on Affective Computing and Intelligent Interaction, LNCS 4738*, J. Tao, T. Tan, and R. Picard, Eds. Springer, 2005, pp. 32–39.

[189] A. Kleinsmith, T. Fushimi, and N. Bianchi-Berthouze, "An incremental and interactive affective posture recognition system," in *International Workshop on Adapting the Interaction Style to Affective Factors*, 2005.

[190] A. Normoyle, F. Liu, M. Kapadia, N. Badler, and S. Jörg, "The effect of posture and dynamics on the perception of emotion," in *ACM Symposium on Applied Perception*, 2013, pp. 91–98.

[191] M. Meijer, "The contribution of general features of body movement to the attribution of emotions," *Journal of Nonverbal Behavior*, vol. 13, no. 4, pp. 247–268, 1989.

[192] A. Samadani, S. Burton, R. Gorbet, and D. Kulić, "Laban Effort and Shape analysis of affective hand and arm movements," in *International Conference on Affective Computing and Intelligent Interaction*, 2013, pp. 343 – 348.

[193] A. Samadani, R. Gorbet, and D. Kulić, "Affective movement generation using Laban Effort and Shape and hidden Markov models," *IEEE Transactions on Affective Computing*, 2014, under review.

[194] D. Kulić, W. Takano, and Y. Nakamura, "Incremental learning, clustering and hierarchy formation of whole body motion patterns using adaptive hidden Markov chains," *The International Journal of Robotics Research*, vol. 27, no. 7, pp. 761–784, 2008.

[195] K. Hachimura, K. Takashina, and M. Yoshimura, "Analysis and evaluation of dancing movement based on LMA," in *IEEE International Workshop on Robot and Human Interactive Communication, ROMAN.*, 2005, pp. 294–299.

[196] M. Kapadia, I. Chiang, T. Thomas, N. Badler, and J. Kider, "Efficient motion retrieval in large motion databases," in *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, 2013, pp. 19–28.

[197] T. Nakata, T. Mori, and T. Sato, "Analysis of impression of robot bodily expression," *Journal of Robotics and Mechatronics*, vol. 14, no. 1, pp. 27–36, 2002.

[198] T. Casciero, *Laban Movement Studies and Actor Training: An Experiential and Theoretical Course for Training Actors in Physical Awareness and Expressivity*. UMI Dissertation Services, 1998.

[199] A. Samadani, "Examples of movements used in the laban quantification study," available at www.eng.uwaterloo.ca/~asamadan/LMA_Annotation.html.

[200] M. Gross, E. Crane, and B. Fredrickson, "Methodology for assessing bodily expression of emotion," *Journal of Nonverbal Behavior*, vol. 34, no. 4, pp. 223–248, 2010.

[201] E. Barakova and T. Lourens, "Expressing and interpreting emotional movements in social games with robots," *Personal and Ubiquitous Computing*, vol. 14, no. 5, pp. 457–467, 2010.

[202] D. Petersen, "Space, Time, Weight, and Flow: Suggestions for enhancing assessment of creative movement," *Physical Education and Sport Pedagogy*, vol. 13, no. 2, pp. 191–198, 2008.

[203] D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Mortillaro, and K. Scherer, "Toward a minimal representation of affective gestures," *IEEE Transactions on Affective Computing*, vol. 2, no. 2, pp. 106–118, 2011.

[204] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2 2010.

[205] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, 2nd ed., ser. Springer series in statistics. Springer, 2009.

[206] J. Qian, T. Hastie, J. Friedman, R. Tibshirani, and N. Simon, "Glmnet for Matlab 2010," http://www.stanford.edu/~hastie/glmnet_matlab/.

[207] E. Pekalska and R. Duin, *The dissimilarity representation for pattern recognition: Foundations and applications*. World Scientific Publishing Company, 2005, vol. 64.

[208] F. Faul, E. Erdfelder, A. Lang, and A. Buchner, "G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behavior research methods*, vol. 39, no. 2, pp. 175–191, 2007.

[209] A. Samadani, "Videos of generated affective movement exemplars," available at www.eng.uwaterloo.ca/~asamadan/GenerationVideos.html.

[210] A. Kleinsmith, N. Bianchi-Berthouze, and A. Steed, "Automatic recognition of non-acted affective postures," *IEEE Transactions on Systems, Man, Cybernetics. C, Applications and Reviews*, vol. 41, no. 4, pp. 1027–1038, 2011.

[211] R. El Kaliouby, R. Picard, and S. Baron-Cohen, "Affective computing and autism," *Annals of the New York Academy of Sciences*, vol. 1093, no. 1, pp. 228–248, 2006.

[212] B. Robins, K. Dautenhahn, and P. Dickerson, "From isolation to communication: A case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot," in *International Conferences on Advances in Computer-Human Interactions*, 2009, pp. 205–211.

[213] D. Kulić and E. Croft, "Pre-collision safety strategies for human-robot interaction," *Autonomous Robots*, vol. 22, no. 2, pp. 149–164, 2007.

[214] W. Niu, J. Long, D. Han, and Y. Wang, "Human activity detection and recognition for video surveillance," in *IEEE International Conference on Multimedia and Expo*, vol. 1, 2004, pp. 719–722.

[215] C. Roether, L. Omlor, and M. Giese, "Lateral asymmetry of bodily emotion expression," *Current Biology*, vol. 18, no. 9, pp. 329–330, 2008.

[216] D. Amelynck, M. Grachten, L. V. Noorden, and M. Leman, "Toward e-motion-based music retrieval a study of affective gesture recognition," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 250–259, 2012.

[217] A. Camurri, B. Mazzarino, and G. Volpe, "Analysis of expressive gesture: The eyesweb expressive gesture processing library," in *Gesture-based Communication in Human-Computer Interaction, LNCS 2915*. Springer, 2004, pp. 460–467.

[218] M. Nicolaou, H. Gunes, and M. Pantic, "Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space," *IEEE Transactions on Affective Computing*, vol. 2, no. 2, pp. 92 –105, 2011.