We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

10

Laban Movement Analysis using a Bayesian model and perspective projections

Joerg Rett[1], Jorge Dias[1] and Juan-Manuel Ahuactzin[2]
[1]*Institute of Systems and Robotics - University of Coimbra,* [2]*Probayes SAS, Montbonnot*
[1]*Portugal,* [2]*France*

1. Introduction

Human movement is essentially the process of moving one or more body parts to a specific location along a certain trajectory. A person observing the movement might be able to recognize it through the spatial pathway alone. Kendon (Kendon, 2004) holds the view that willingly or not, humans, when in co-presence, continuously inform one another about their intentions, interests, feelings and ideas by means of visible bodily action. Analysis of face-to-face interaction has shown that bodily action can play a crucial role in the process of interaction and communication. Kendon states that expressive actions like greeting, threat and submission often play a central role in social interaction.
In order to access the expressive content of movements theoretically, a notational system is needed. Rudolf Laban, (1879-1958) was a notable central European dance artist and theorist, whose work laid the foundations for Laban Movement Analysis (LMA). Used as a tool by dancers, athletes, physical and occupational therapists, it is one of the most widely used systems of human movement analysis.
Robotics has already acknowledged the evidence that human movements could be an important cue for Human-Robot Interaction. Sato et al. (Sato et al., 1996), while defining the requirements for 'human symbiosis robotics' state that those robots should be able to use non-verbal media to communicate with humans and exchange information. As input modalities on a higher abstraction level they define channels on language, gesture and unconscious behavior. This skill could enable the robot to actively perceive human behavior, whether conscious and unconscious. Human intention could be understood, simply by observation, allowing the system to achieve a certain level of friendliness, hospitality and reliance. Fong, Nourbakhsh and Dautenhahn (Fong et al., 2003) state in their survey on 'socially interactive robots' that the design of sociable robots needs input from research concerning social learning and imitation, gesture and natural language communication, emotion and recognition of interaction patterns. Otero et al. suggest (Otero et al., 2006) that the interpretation of a person's motion within its environment can enhance Human-Robot Interaction in several ways. They point out, that a recognized action can help the robot to plan its future tasks and goals, that the information flow during interaction can be extended and additional cues, like speech recognition, can be supported. Otero et al. state that body motion and context provide in many situations enough information to derive the person's current activity.

1.1 Related works on computational Human Movement Analysis

There has been an interesting work which also used movement descriptors and a probabilistic framework. Bregler (Bregler, 1997) introduced mid-level descriptors embedded in a thorough probabilistic framework that produced a robust classification for human movements. The concept of multiple hypothesizes is kept from low-level motion clusters to high-level gait categories producing good classification results even for noisy and uncertain evidences in natural environments. Model parameters are learned from training data using the EM-algorithm. The work points towards the concept of atomic phonemes and words used in speech recognition. Bregler defines his 'movemes' as simple dynamical categories, i.e. a set of second order linear dynamical systems. A Hidden Markov Model (HMM) is used to classify three different gait categories: running, walking, and skipping. The critical point on this approach were the ‘movemes’ themselves. The 'movemes' appear limited in their expressiveness. This might have been caused by their simplicity and that no relations are drawn to models and data of physiological studies of human movements. To overcome this weakness we have tied our descriptors to a well established notational framework: Laban Movement Analysis.

That probabilistic methods can produce very good classification results was also demonstrated for the application of sign language recognition. Starner & Pentland (Starner & Pentland, 1995) based their system on real-time tracking of the hands using color gloves and a monocular camera with 5 frames per second. The learning and classification of the 40 words (signs) was embedded in 400 sentences (sequence of signs) for learning and 100 sentences for classification. The results compared the accuracy when using grammar rules (99.2%) or when not using them (91.3%). The results showed that when constraints can be applied like colored marker, a spatially well defined trajectory and rules that help to deal with a sequence of symbols, high accuracies can be reached. In this approach no mid-level descriptors were needed as the sign language has very well defined spatial pathways and grammars. The application of this approach as a general interface for Human-Robot Interaction is difficult, as it requires the person to learn the sign language.

1.2 Related works on computational Laban Movement Analysis

A long tradition in research on computational solutions for Laban Movement Analysis (LMA) has the group around Norman Badler, who already started in 1993 to re-formulate Labanotation in computational models (Badler et al., 1993). The work of Zhao & Badler (Zhao & Badler, 2005) is entirely embedded in the framework of Laban Movement Analysis. Their computational model of gesture acquisition and synthesis can be used to learn motion qualities from live performance. Many inspirations concerning the transformation of LMA components into physically measurable entities were taken from this work. As the final application was the back-projection of the LMA parameters to an animated character, Zhao (Zhao, 2002) made no attempt to address the problem of gesture recognition. For the same reason video capture was presented for a controlled environment (human wearing black cloth in front of black curtain). The application of LMA to the classification of movements, especially in unconstrained environments is the main goal of our contribution.

In (Nakata et al., 2002) Nakata et al. reproduced expressive movements in a robot that could be interpreted as emotions by a human observer. The first part described how some parameters of Laban Movement Analysis (LMA) can be calculated from a set of low-level features. They concluded further that the control of robot movements oriented on LMA

parameters allows the production of expressive movements and that those movements leave the impression of emotional content to a human observer. The critical points on the mapping of low-level features to LMA parameters was, that the computational model was closely tied to the embodiment of the robot which had only a low number of degrees of freedom. For our solution we have chosen low-level features that can be used for an arbitrary object (human full body, body parts, etc.).

1.3 The contribution of this work

This work poses the automatic movement classification task as a problem to recognize a sequence of symbols taken from an alphabet consisting of motion-entities. The alphabet and its underlying model is well defined though Laban Movement Analysis. The LMA parameters serve as mid-level descriptors that can be produced and understood by the system. Our *Tracking* process is two-fold, the technique use for learning based on the active markers of our positioning device produces a robust representation of each object as points. For the visual tracking we use the central point of a boundary box containing the pixels or regions found in the figure–ground segmentation process. The relationship between the two approaches is established through a geometric model (Rett & Dias, 2007-B). This work emphasizes probabilistic methods, i.e. Bayesian approaches, as a tool to model the concept of Laban Movement Analysis (LMA), learn its parameters and classify the movements. The process of segmentation and tracking of image data is also based on a probabilistic method, i.e. the CAMshift algorithm (Bradski, 1998). This work provides a new skill for machines that analyze human movements, i.e. computational Laban Movement Analysis. The system has been implemented in our social robot, 'Nicole' to test several human-robot interaction scenarios (Rett & Dias, 2007-A).

2. Laban Movement Analysis

Laban Movement Analysis (LMA) is a method for observing, describing, notating, and interpreting human movement. It was developed by a German named Rudolf Laban (1879 to 1958), who is widely regarded as a pioneer of European modern dance and theorist of movement education (Zhao, 2002). The general framework was described in 1980 by Irmgard Bartenieff a scholar of Rudolf Laban in (Bartenieff & Lewis, 1980). While being widely applied to studies of dance and application to physical and mental therapy (Bartenieff & Lewis, 1980), it has found little application in the engineering domain. Most notably the group of Norman Badler, who already started in 1993 to re-formulate Labannotation in computational models (Badler et al., 1993). More recently a computational model of gesture acquisition and synthesis to learn motion qualities from live performance has been proposed in (Zhao & Badler, 2005). Also recently but independently, researchers from neuroscience started to investigate the usefulness of LMA to describe certain effects on the movements of animals and humans. Foround and Whishaw adapted LMA to capture the kinematic and non-kinematic aspects of movement in a reach-for-food task by human patients whose movements had been affected by stroke (Foroud & Whishaw, 2006). It was stated that LMA places emphasis on underlying motor patterns by notating how the body segments are moving, how they are supported or affected by other body parts, as well as whole body movement.

The theory of LMA consists of several major components, though the available literature is not in unison about their total number. The works of Norman Badler's group (Chi et al., 2000); (Zhao, 2002) mention five major components shown in Figure 1.

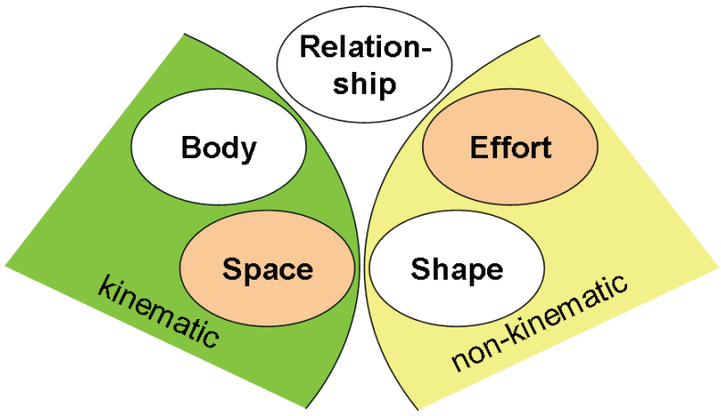

Figure 1. The major components of LMA are Body, Space, Effort, Shape and Relationship

Relationship describes modes of interaction with oneself, others, and the environment (e.g. facings, contact, and group forms). As *Relationship* appears to be one of the lesser explored components, some literature (Foroud & Whishaw, 2006) only considers the remaining four major components. *Body* specifies which body parts are moving, their relation to the body center, the kinematics involved and the emerging locomotion. *Space* treats the spatial extent of the mover's *Kinesphere* (often interpreted as reach-space) and what form is being revealed by the spatial pathways of the movement. *Effort* deals with the dynamic qualities of the movement and the inner attitude towards using energy. *Shape* is emerging from the *Body* and *Space* components and focused on the body itself or directed towards a goal in space. The interpretation of *Shape* as a property of *Body* and *Space* might have been the reason for Irmgard Bartenieff to mention only three major components of LMA. Like suggested in (Foroud & Whishaw, 2006) we have grouped *Body* and *Space* as kinematic features describing changes in the spatial-temporal body relations, while *Shape* and *Effort* are part of the non-kinematic features contributing to the qualitative aspects of the movement as shown in Figure 1. This article concentrates on the *Space* component in order to establish a basis for comparison with subsequent works that include also other components.

2.1 Space

The *Space* component presents the different concepts to describe the pathways of human movements inside a frame of reference, when "carving shapes in space" (Bartenieff & Lewis, 1980). *Space* specifies different entities to express movements in a frame of reference determined by the body of the actor. Thus, all of the presented measures are relative to the anthropometry of the actor. The concepts differ in the complexity of expressiveness and dimensionality but are all of them reproducible in the 3-D Cartesian system. The following definitions were taken from Choreutics (Laban, 1966) and differ in some aspects from those given in Labanotation (Hutchinson, 1970). The most important ones shown in Figure 2 are: I) The *Levels of Space* - referring to the height of a position, II) The *Basic Directions* - 26 target points where the movement is aiming at, III) The *Three Axes* - Vertical, horizontal and sagittal axis, IV) The *Three Planes* - *Door Plane* (vertical) π_v, *Table plane* (horizontal) π_h, and the

Wheel Plane (sagittal) π_s each one lying in two of the axes, and V) The *Icosahedron* - used as *Kinespheric Scaffolding*.

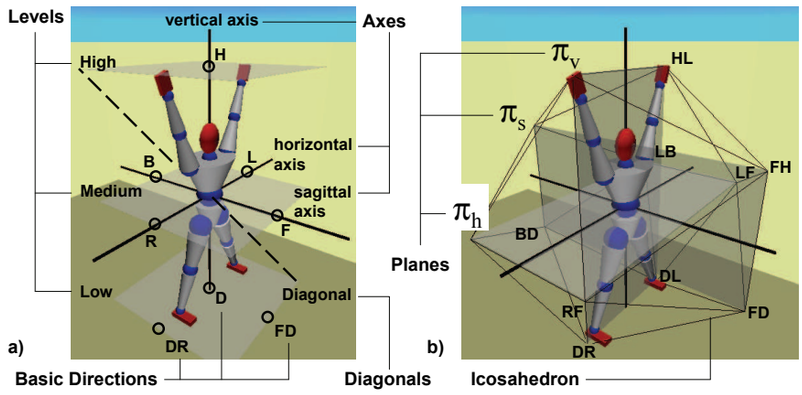

Figure 2. The *Space* component defines several concepts: a) Levels of Space, Basic Directions, Three Axes, and b) Three Planes and Icosahedron

The *Kinesphere* describes the space of farthest reaches in which the movements take place. Levels and Directions can also be found as symbols in modern-day Labanotation (Bartenieff & Lewis, 1980)

Labanotation direction symbols encode a position-based concept of space. Recently, Longstaff (Longstaff, 2001) has translated an earlier concept of Laban which is based on lines of motion rather than points in space into modern-day Labanotation. Longstaff coined the expression *Vector Symbols* to emphasize that they are not attached to a certain point in space. The different concepts are shown in Figure 3.

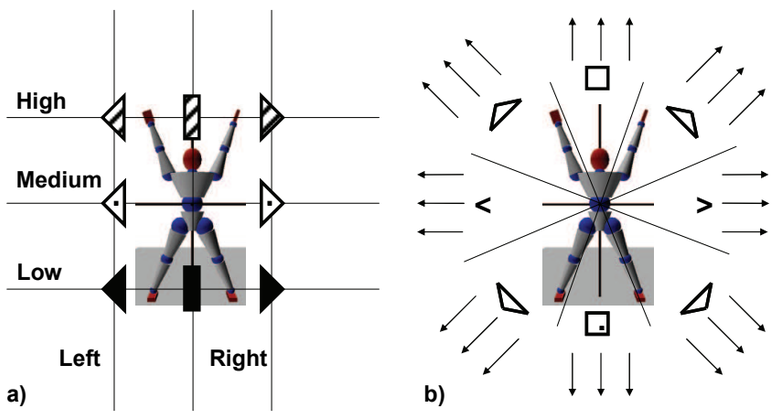

Figure 3. Two different sets of symbols to describe the *Space* component presented through the *Door Plane*. a) Position based symbols of Labanotation represent the 'height' through shading and the horizontal position through shape. b) Direction based vector symbols of Choreographie use different shapes for each direction

The symbols of Labanotation correspond to positions in space like *Left-High* while the *Vector Symbols* describe directions. Figure 3 represents a 2-D view of the vertical (door) plane π_v and thus shows only a fraction of the set of symbols (8) which describes movements in 3-D space. It was suggested that the collection of *Vector Symbols* provides a heuristic for the

perception and memory of spatial orientation of body movements. The thirty eight *Vector Symbols* are organized according to *Prototypes* and *Deflections*. The fourteen *Prototypes* divide the Cartesian coordinate system into movements along only one dimension (*Pure Dimensional Movements*) and movements along lines that are equally stressed in all three dimensions (*Pure Diagonal Movements*) as shown in Figure 2 a). Longstaff suggests that the *Prototypes* give idealized concepts for labeling and remembering spatial orientations. The twenty four *Deflections* are mentally conceived according to their relation to the prototype concepts. The infinite number of possible deflecting orientations is conceptualized in a system based on eight *Diagonal Directions*, each deflecting along three possible *Dimensions*.

2.2 Labanotation and Effort Notation

The need to develop some means of recording for the perceptions of movements led to a notation system known as *Labanotation*. It is built of symbols which describe the structure and progression of the movement (shown in Figure 4.). The spatial definitions (Hutchinson, 1970) vary from those stated in Choreutics (Laban, 1966). In Labanotation the three *Levels of Space* are circular causing the distances e.g. *centre-L* and *centre-LD* to be equal. Moreover, distinct frames of reference are defined for the different groups of body parts. e.g. placing the origin of the arm-hand group at the shoulder joint. The symbols reflect which body part does what in space and time and with what kind of dynamic stress. In particular it contains when the movement starts and its duration. The so called *Staff* organizes the body parts in columns where the time proceeds from the bottom up along the length. The placement of a symbol shows that the body part is active, its shape indicates the direction of the movement, its shading shows the level and its length, the duration of the movement. From a properly notated movement sequence, the skilled reader can see at one glance what is happening at any moment in every part of the body.

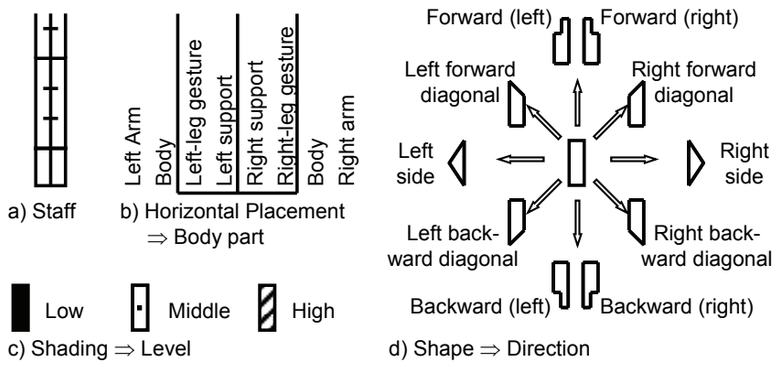

Figure 4. Labanotation: a) The staff is used to place the symbols. b) The horizontal placement of the symbol indicates the body part. c) Shading of the symbol is used to indicate the *Level* (height) of the 3-D position. d) Different shapes of the symbols are used to indicate the position in the *Table Plane*

The example in Figure 5 shows the ballet figure, *Port de Bras*. For the sake of readability we rotated the staff by 90 degrees. Reading from the right (usually bottom), one sees the basic position of neutral standing, arms hanging down. Then move your arms forward middle (shoulder level), followed by an open side movement (for two counts), followed by lowering the arms.

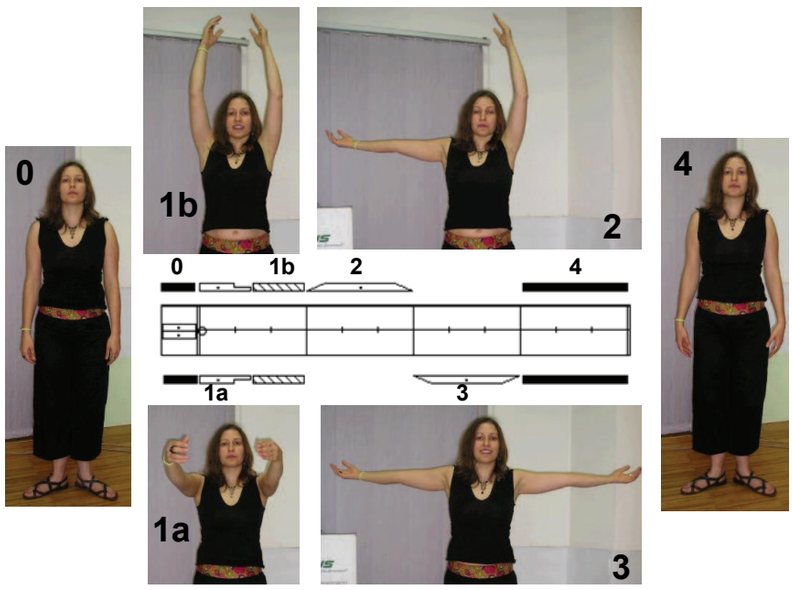

Figure 5. Example of a ballet "Port de Bras" figure. The staff in the center holds the symbols to represent the sequence of positions performed by the actor. Verify with the previous figure: Mainly the left and the right arm symbols are written, the sequence starts and concludes with *Level=low*

2.3 Database of Expressive Movements

We have created a database of 'expressive movements'. Some of the movements are based on suggestions mentioned in (Bartenieff & Lewis, 1980) and (Zhao, 2002) others are commonly used gestures with anticipated *Effort* qualities. In this work we will concentrate only on movements with distinct *Space* component. Table 1 shows such movements.

Movement	**Description**	π_{prin}
Lunging	Lunging for a ball	XY
Maestro	Conducting an orchestra	YZ
Stretch	Stretch to yawn	YZ
Ok	OK-sign gesture	YZ
Point:	Pointing gesture	XY
Byebye	Waving bye-bye	YZ
Shake	Reach for someone's hand	XZ
Nthrow	Waving sagittally (approach sign)	XZ

Table 1. Expressive movements from our database (HID) with principal plane π_{prin} i.e. the plane where the movement can be observed best

Their distinctive *Space* component can be verified by observing the trajectories (see Figure 6).

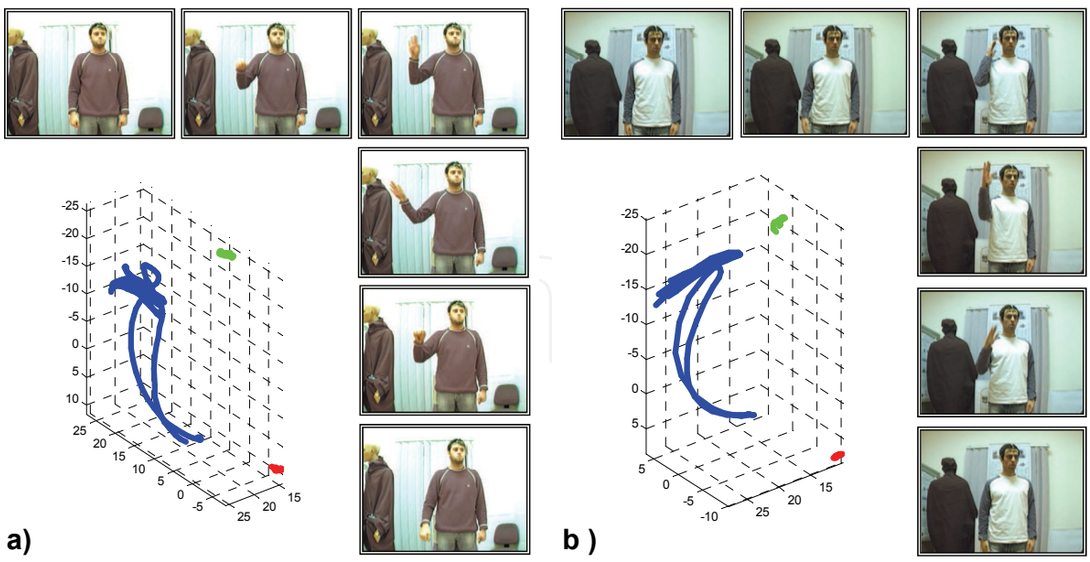

Figure 6. Two movements with distinct *Space* component. a) Horizontal waving (*byebye*) and b) Sagittal waving (*nthrow*)

The *byebye* gesture represents a horizontal waving, while *nthrow* represents a sagittal waving. Both movements are oscillatory and in the case of *byebye* the primary signal can be described by a sequence of left to right *R* and right to left *L Vector Symbols*. In the case of *nthrow* the primary signal would be described by a sequence of forward (*F*) and backward (*B*) *Vector Symbols*.

The case of non-oscillatory movements like the *ok* sign and reaching for someone's hand (*shake*) can be seen in Figure 7.

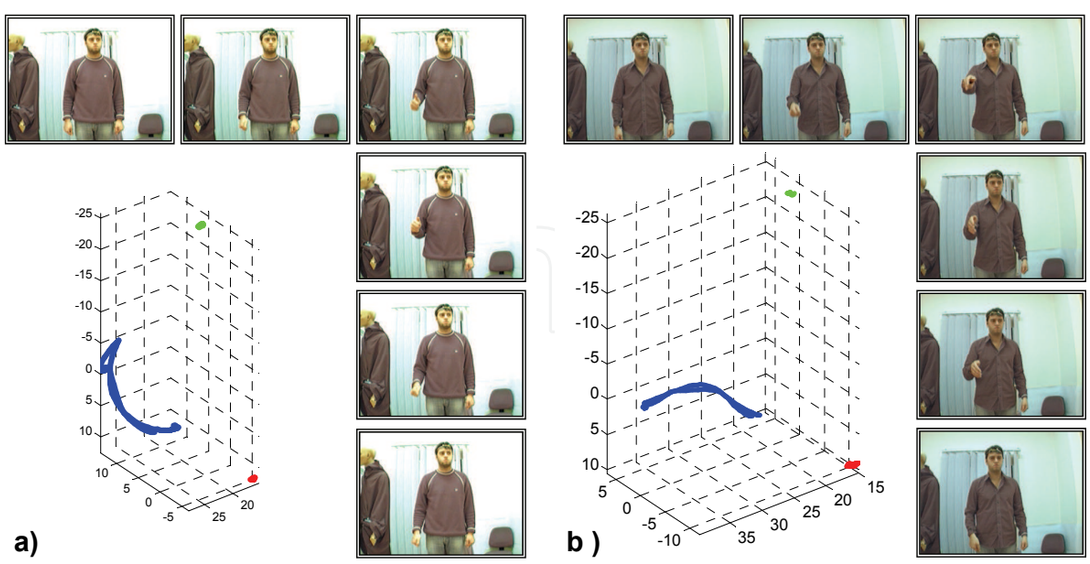

Figure 7. Two movements with distinct *Space* component. a) Showing the *ok* sign and b) Reaching for someone's hand (*shake*)

These two cases can be distinguished by a greater influence of forward (*F*) and backward (*B*) vector symbols in the case of *shake*. The shown trajectories present one trial of one person. The whole set of trials can be seen in (Rett, 2008).
In the case of lunging for a ball (*lunging*) the *Space* component consists mainly of forward (*F*) and backward (*B*) *Vector Symbols* for both hands as shown in Figure 8.

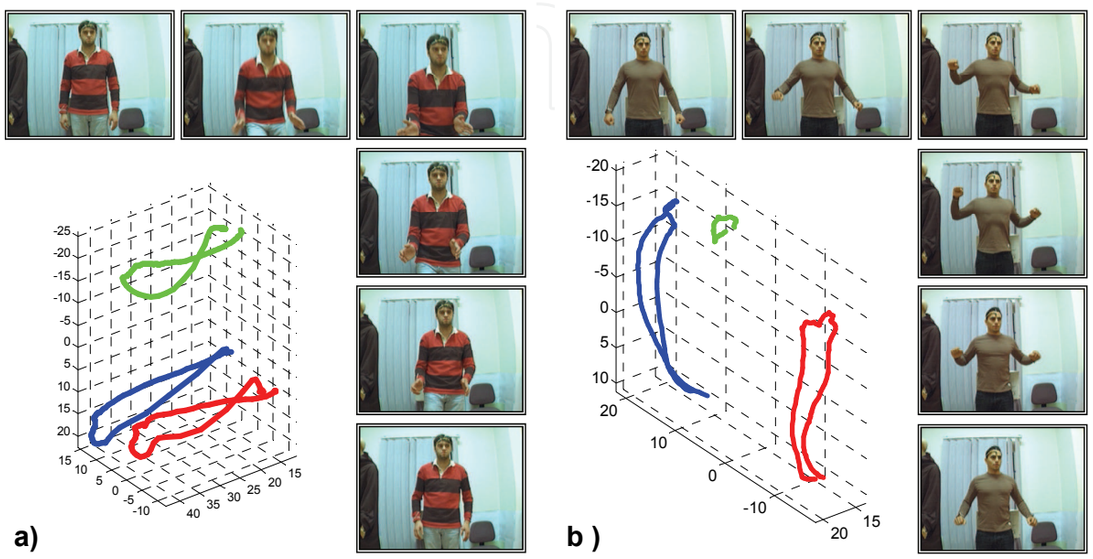

Figure 8. Two movements with distinct *Space* component. a) Forward dab (Lunging for a ball) and b) Upward wring (Stretching to yawn)

The mainly appearing *Vector Symbols* for the 'stretch to yawn' (stretch) movement are upward (*U*) and downward (*D*).

3. Human Movement Tracking

Laban Movement Analysis is essentially defined in a 3-D space related whether with a world frame of reference {W} or an egocentric frame of reference {H} of the human under observation. With a magnetic tracker precise movement data from body parts related to both {W} and {H} can be collected. This kind of sensor system is useful for a stationary Human-Computer-Interface, as it requires a certain preparation-effort from the user (e.g. attaching the sensors). For a mobile robot a visual-based system is more useful as it does not require any preparation, though on the cost of precision, as depicted in Figure 9. Our solution is based on a system which uses both, 3-D magnetic tracker data and 2-D visual data. The relationship between LMA parameters, Low-Level features and the types of movement will be learned by a synchronous acquisition of 3-D tracking and 2-D image data. Additionally, a probabilistic model for the geometry of the frames of reference will be established. The mobile robot will be equipped with a monocular camera only, but additionally with the knowledge from the previous learning.

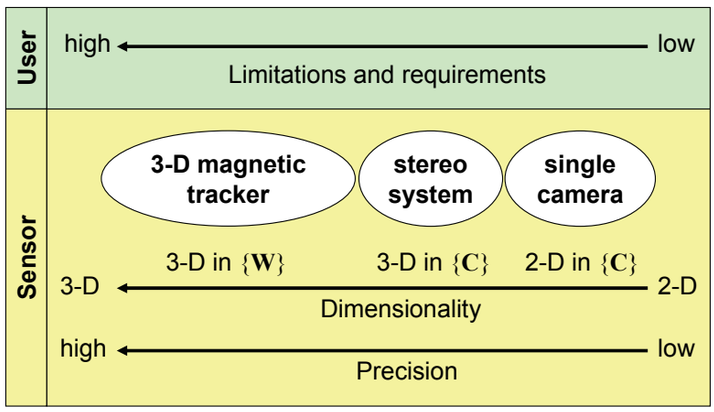

Figure 9. Different sensor modalities and their characteristics

Some sets of movement data have already been introduced in section 2.3. The corresponding database is called Human Interaction Database (HID) and is accessible through *WWW* (Rett, et al. 2007). The database consists of image sequences, high precision 3-D position data and results from our visual tracker and classifier.

Our geometric model needs to address the appearance of sensors and objects in the interaction scenery, i.e. define their frames of reference. Figure 9 shows the frames of reference: the camera referential {C} in which the image is defined and some world coordinate system {W}.

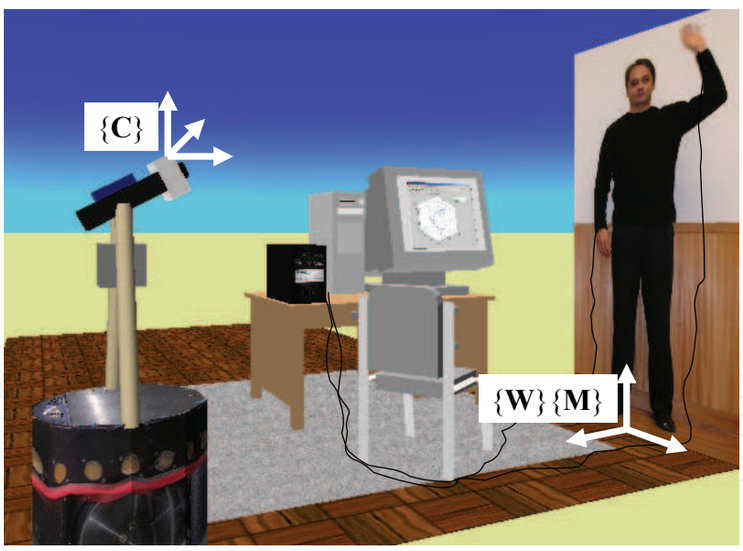

Figure 10. Frames of reference of the scene

In the experimental setup for collecting movement data we have the world frame of reference {W} coinciding with the one of the magnetic tracker {M}.

Using a 6-DoF magnetic tracker provides 3-D position data with a sufficiently high accuracy and speed (50Hz). We use a Polhemus Liberty™ system with sensors attached to several body parts and objects. From the tracker data a set of features is calculated and related to the Laban Movement Parameters (LMP). The 3-D position data is projected in the following step to 2-D planes from which the low-level features are computed.

3.1 Geometric model of the camera

As the learning of human movements is based on a synchronous acquisition of 3-D tracking and 2-D image data we need to establish the geometric relationship in a model. The presented model considers the frame of reference of the world {W} and of the camera referential {C}. As shown in Figure 11 we placed the origin of {W} on the ground level aligned with the gravitational vertical and the sagittal axis of the person.

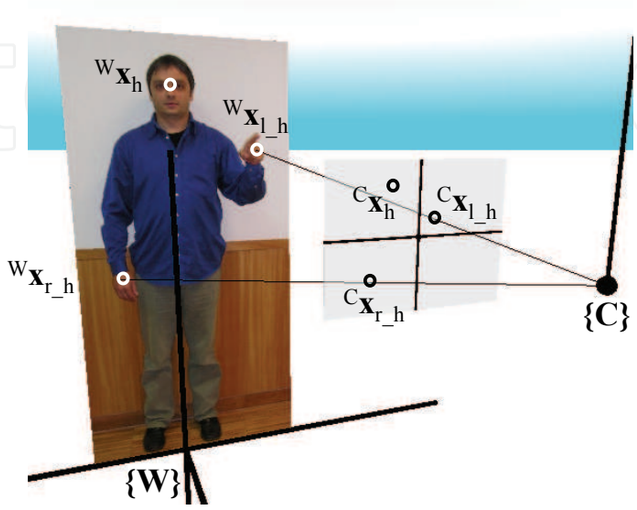

Figure 11. Projection of head and hands position in the camera plane

Any generic 3-D point $^{W}\mathbf{X} = [X \;\; Y \;\; Z]^{T}$ and its corresponding projection $^{img}\mathbf{X} = [u \;\; v]^{T}$ on an image-plane can be mathematically related using projective geometry and the concept of homogeneous coordinates through the following equation, the projective camera relation, where *s* represents an arbitrary scale factor (Hartley & Zisserman, 2000):

$$\begin{bmatrix} sv \\ su \\ s \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

Matrix **A** is called the projection matrix, and through its estimation it is possible to make the correspondence between any 3-D point and its projection in a camera's image-plane. We can likewise express the matrix **A** by using the parameters of the projective finite camera model, as stated in (Hartley & Zisserman, 2000).

$$\mathbf{A} = \mathbf{C}\left[{}^{\{C\}}\mathbf{R}_{\{W\}} \quad {}^{\{C\}}\vec{\mathbf{t}}_{\{W\}} \right] \tag{2}$$

Where {**C**} is the camera's calibration matrix, more frequently known as the intrinsic parameters matrix, while the camera's extrinsic parameters are represented by the rotation orthogonal matrix **R** and the translation vector **t** that relates the chosen {W} to the camera {C}.

The projective camera presents us, in fact, with the solution for the intersection of planes π_{cam1} and π_{cam2} which, assuming $\mathbf{X}^{*} = [X \; Y \; Z \; 1]^{T}$ (i.e. homogeneous coordinates), can be proven from its projection expression to be given by (3) (Dias, 1994).

$$\begin{cases} (\mathbf{a}_1 - u\mathbf{a}_3)^{\mathrm{T}} \ {}^{\mathrm{W}}\mathbf{X} + a_{1,4} - u = 0 \\ (\mathbf{a}_2 - v\mathbf{a}_3)^{\mathrm{T}} \ {}^{\mathrm{W}}\mathbf{X} + a_{2,4} - v = 0 \end{cases} \Leftrightarrow \begin{cases} \mathbf{\Pi}_{cam1} \ \mathbf{X}^* = 0 \\ \mathbf{\Pi}_{cam2} \ \mathbf{X}^* = 0 \end{cases} \tag{3}$$

This solution is called the projection or projecting line, which can be alternatively represented by equation (4) (Dias, 1994).

$$\vec{\mathbf{n}} = (\mathbf{a}_1 - u\mathbf{a}_3) \times (\mathbf{a}_2 - v\mathbf{a}_3) \tag{4}$$

These relations indicate that all 3-D points on the projecting line correspond to the same projection point on the image-plane. A unique correspondence between ${}^{\mathrm{W}}\mathbf{X}$ and ${}^{\mathrm{C}}\mathbf{X}$ could only be established through additional constraints, such as the intersection with the surface of a sphere, a plane, etc.

3.2 Low-level features

The selection of 'good' features is a general and long known problem in pattern recognition. For the description of the *Space* component we have chosen a feature based on the *displacement angle*. This physical measurable entity represents the *Space* component of LMA very well and the process of computation is simple. When using a low cardinality we can expect a good performance of the Bayesian method for learning and classification. Displacement angles, which also have been used by (Zhao, 2002) can be calculated easily from two subsequent positions. They describe the trace of a curve quite well and are independent from the absolute positions. As the position data is projected to planes, each plane produces a sequence of displacement angles with a certain sampling rate and discretization.

All computations are based on the raw tracking data inside our Human Interaction Database (HID). The tracking data consists of: I) the 2-D or 3-D position $\mathbf{X}_{\mathrm{bp}}$ of a point belonging to a body part *bp* and II) the timestamp t_{i} given by some timer function of the system. The position is defined in a frame of reference ϕ indicated by ${}^{\phi}\mathbf{X}$. This usually indicates the sensor used for input like the camera {C} or the commercial motion capture device {W}. With the sampling (frame) index *i* the sampling interval Δt_{i+1} can be calculated between two conescutive frames *i* and *i*+1. In order to treat 2-D and 3-D data equally the first step is to project the 3-D data to some suitable planes. Usually the three principal planes *Door Plane* (vertical) π_{v}, *Table plane* (horizontal) π_{h}, and the *Wheel Plane* (sagittal) π_{s} are used. To allow for a fast computation we are discretizing the low-level features to a low cardinality. The continuous *displacement angles* are discretized into *directions D* with a cardinality of eight.

$$\mathrm{D} \in \{180°,\ 135°,\ 90°,\ 45°,\ 0°,\ -45°,\ -90°,\ -135°\} < 8 > \tag{5}$$

With this we get one discrete variable *D* per body part and plane. Considering the two most important body parts 'left hand' *lh* and 'right hand' *rh* and the three principal planes we get six *directions*:

$$D_{xy}^{rh}, D_{yz,}^{rh} D_{xz}^{rh}, \ D_{xy}^{lh}, D_{yz,}^{lh} D_{xz}^{lh} \tag{6}$$

The angular values of *D* are then translated into the *Vector Symbols* A_{bp}, B_{bp} and C_{bp}. Figure 12 shows this transformation for the *Door Plane* (vertical) π_{v} and the right hand *rh* using a 'byebye' movement as an example.

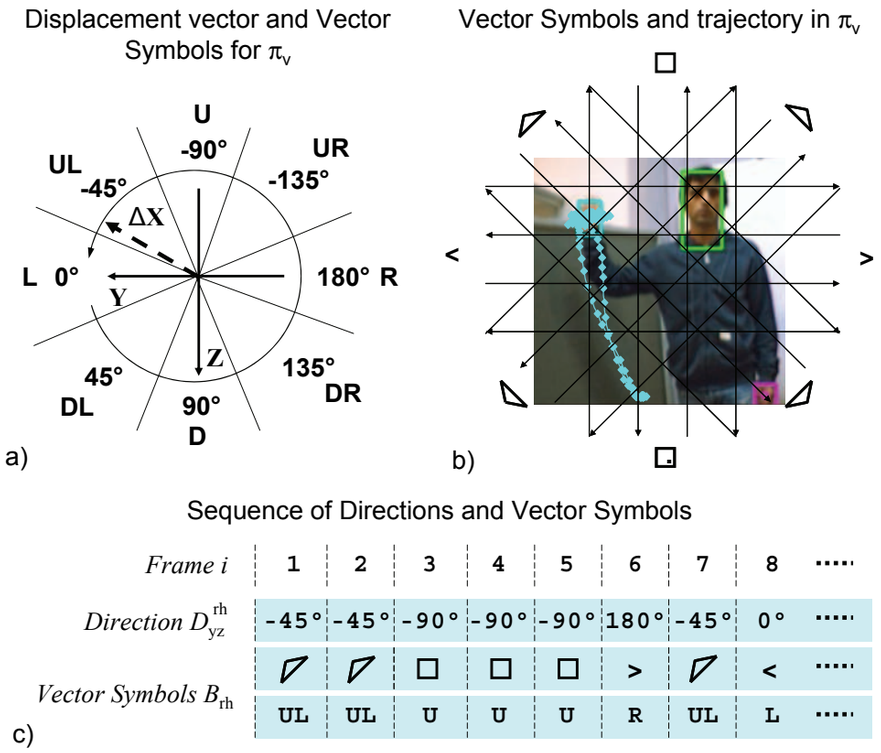

Figure 12. Vector Symbols for the *Door Plane* and the right hand by means of a 'byebye' movement. a) The *displacement* is converted into the *Vector Symbol* B_{rh}. b) Grid of *Vector Symbols* superimposed on the movement trajectory. c) The continuous computation results in a stream of *Vector Symbols*

In Figure 12 a) the scheme for the conversion from the *displacement* to the *displacement angles* to the *direction* D_{yz}^{rh} and finally to the *Vector Symbol* B_{rh} is shown. In Figure 12 b) the grid of *Vector Symbols* is superimposed on the movement trajectory. As a result of the continuous computation we get a stream of *Vector Symbols* as shown in Figure 12 c). Figure 12 shows both representtations for the Vector Symbols, the signs taken from (Longstaff, 2001) and the letters used by our algorithm.

4. Bayesian Models for Movement Perception

The concepts of Laban Movement Analysis (LMA) and the characteristics of our system to track human movements can be mathematically and computationally modeled using a common framework. The Bayesian theory gives us the possibility to deal with incompleteness and uncertainty, make predictions on future events and, most important, provides an embedded scheme for learning.

Included in the Bayesian framework are specialized models which have a long tradition in many areas and are known under the names, Hidden Markov Models (HMMs), Kalman Filters and Particle Filters. Bayesian models have already been used in a broad range of technical applications (e.g. navigation, speech recognition, etc.). Recent findings indicate that Bayesian models can also be useful in the modeling of cognitive processes. Research on the human brain (and in its computations for perception and action) report, that Bayesian

methods have proven successful in building computational theories for perception and sensorimotor control (Knill & Pouget, 2004).
In the course of our investigation and development we found, that the process of prediction and update during classification represents an intrinsic implementation of the mental concept of anticipation. Using the property of conditional independence the dimensionality of the parameter space that describes the human movements can be reduced. Bayesian nets offer the possibility to represent dependencies, parameters and their values intuitively understandable, which is a frequently expressed request from non-engineers (Loeb, 2001). Furthermore these methods have already proven their usability in the related field of gesture recognition (Starner, 1995); (Pavlovic, 1999).
Probabilistic reasoning needs only two basic rules. The first is the *conjunction rule*, which gives the probability of a conjunction of *propositions*.

$$\begin{aligned} P(a\ b) &= P(a)\times P(b \mid a) \\ &= P(b)\times P(a \mid b) \end{aligned} \tag{7}$$

The second one is the *normalization rule*, which states that the sum of the probabilities of a and $\neg a$ is one.

$$P(a) \wedge P(\neg a) = 1 \tag{8}$$

The two rules are sufficient for any computation in discrete probabilities. All the other necessary inference rules concerning variables can be derived such as the *conjunction rule*, the *normalization rule* and the *marginalization rule for variables* (Rett, 2008).

4.1 Global Model

The global model to describe the phenomenon of computational Laban Movement Analysis (LMA) is shown in Figure 13.

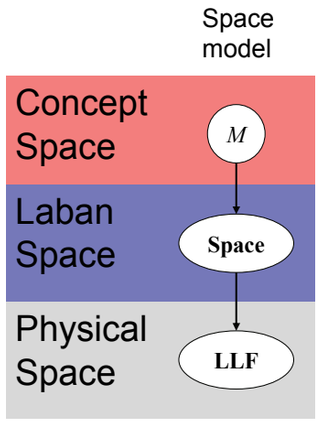

Figure 13. Bayes-Net of the global model with three levels of abstraction (i.e. Concept Level, Laban Space and Physical Space)

Having the concept of a movement represented by the variable M certain characteristics will be exhibited through the sets of variables of LMA (**Space**). The sets of LMA can be observed through the set of low-level features **LLF**. This concept is accompanied by different levels of abstraction by introducing the 'Concept space', the 'Laban space' and the 'Physical space'.

The nodes represent variables (e.g. movement *M*) and sets of variables (e.g. low-level features **LLF**). The arcs describe the dependencies between the nodes. The movement *M* represents the parent node which effects the child nodes in the 'Laban space'. The node on the 'Laban space' is a parent for the set of low-level features **LLF**. The dependencies can also be expressed as a joint distribution and its decomposition while omitting the conjunction symbol $\wedge$ as:

$$P(M\ \mathbf{Space}\ \mathbf{LLF}) \quad = \quad P(M)\ P(\mathbf{Space} \mid M)\ P(\mathbf{LLF} \mid \mathbf{Space}) \tag{9}$$

In the following section the *Space* model will be discussed in detail. Additionally a temporal model will be discussed which tackles issues concerning the duration of a movement and the frames of inflection (phase).

4.2 Space Model

The *Space* component of LMA is modeled using the concept of *Vector Symbols*. As defined in the 'global model' (see section 4.1) two sets of variables are used in the model:

$$\mathbf{LLF} \quad = \quad \mathbf{Space} \in \{A, B, C\} \tag{10}$$

It can be seen that **LLF** and **Space** are equal which is due to the fact that the variables {*A*, *B*, *C*} are both, LMA descriptors and low-level features.

The *Vector Symbols* receive one additional value from the velocity variable, i.e. the indication of no movement $v = 0$. As we describe the spatial pathway of a movement by 'atomic' displacements, we refer to the *Vectors Symbols* sometimes as *atoms*. Movements which are parallel to one of the axes are expressed as up, down, left, right, back and forward movement resulting in the values *U*, *D*, *L*, *R*, *B* and *F* respectively. This represents the concept of *Pure Dimensional Movements* within LMA, while the concepts of *Pure Diagonal Movements* and *Deflections* are described as combinations of *Pure Dimensional Movements*.

Of particular interest are the *atoms B*, occurring in the frontal *Door Plane* (*YZ*-plane) as they convey most of the information found in gestures. The variables and their sample space are shown in

$$\begin{aligned} M &\in \{byebye,\ldots,lunging\} \langle 8 \rangle \\ I &\in \{1,\ldots,I_{\max}\} \langle I_{\max} \rangle \\ A_{bp} &\in \{O,F,FR,R,BR,B,BL,L,LF\} \langle 9 \rangle \\ B_{bp} &\in \{O,U,UR,R,DR,D,DL,L,UL\} \langle 9 \rangle \\ C_{bp} &\in \{O,U,UF,F,DF,D,DB,B,UB\} \langle 9 \rangle \end{aligned} \tag{11}$$

The model of LMA-Space assumes that each movement $M = m$ produces certain *atoms* $A_{\text{bp}} = a$, $B_{\text{bp}} = b$ and $C_{\text{bp}} = c$ at a certain point in time, i.e. frame $I = i$ and for a certain *body part bp*. In this model a certain movement *m* is 'causing' the atoms *a*, *b* and *c* at the frame *i*. The *evidences* that can be measured are the atoms *a*, *b*, c and the frame *i*. The model might be applied to any number of body parts *bp* which are treated as independent evidences a thus expressed through a product as shown in the joint distribution as

$$P(M\ I\ A\ B\ C) \quad = \quad P(M) \quad P(I) \prod_{bp} \left\{ P(A_{bp} \mid M\ I) \quad P(B_{bp} \mid M\ I) \quad P(C_{bp} \mid M\ I) \right\} \tag{12}$$

Figure 14 shows the corresponding representation in a Bayes-net.

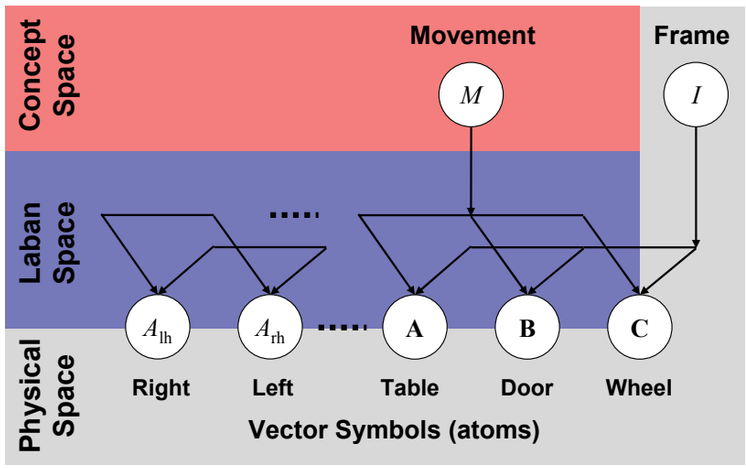

Figure 14. Bayes-Net for the *Space* component of LMA. The movement *M* belongs to the *concept space* while the *Vector Symbols* are part of both, the *Laban space* and the *physical space*. Their instances are in the principal planes *Table*, *Door* and *Wheel* and the left and right hand. The frame *I* is associated with the *physical space* only

Table 2 summarizes the variables used in this model.

Variable	Symbol	Description
Movement	M	Set of movements
Frame	I	Frame index
Body part	bp	e.g. *rh* (right hand)
Vector symbol	A_{bp}	*Vector Symbols* (*Atoms*) in π_h
	B_{bp}	*Vector Symbols* (*Atoms*) in π_v
	C_{bp}	*Vector Symbols* (*Atoms*) in π_s

Table 2. Space variables

4.3 Temporal Model

The *Space* model is based on the temporal sequence of *atoms*. Different paces and number of repetitions while performing the movement influence the classification result. One solution to deal with this problem is to introduce an additional uncertainty model. For each trial of movements the total length in frames i_{max} can be determined. For all trials the mean and variance can be calculated. The uncertainty about the length i_{max} of a performance can be expressed as a an uncertainty concerning the frame i itself. One may think of this as stretching and shrinking the length of the frames i so they may fit in a static length i_{max}.
Technically one can map an observed frame i_obs to a normal frame i, probabilistically we define a conditional probability

$$P(i\_obs \mid i) \quad = \quad N(i\_obs; \sigma_i) \tag{13}$$

where for a certain frame i get probability values for all possible values of i_obs. It makes sense to assign the highest probability to the case $i\_obs = i$ and model the relationship as a

Gaussian distribution. The mean of the Gaussian will be the observed frame *i_obs* = *i* itself and the standard deviation may have a value $0 < \sigma_i \leq \sigma_{i\_max}$.
For each newly observed frame *i_obs* the mean of the distribution slides one step further as shown in Figure 15.

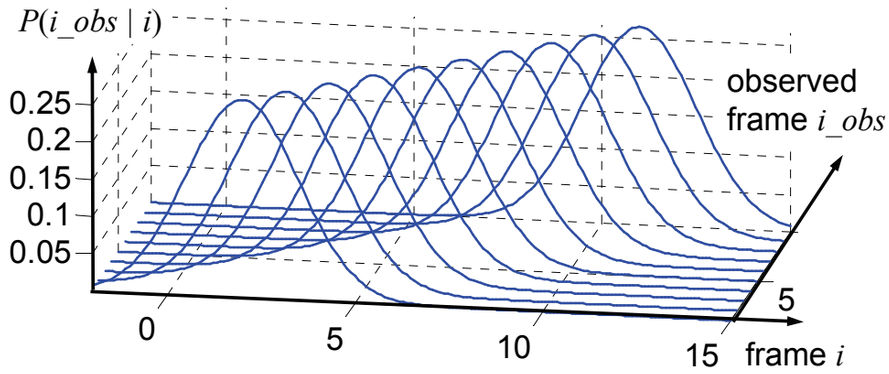

Figure 15. P(*i_obs* | *i*=*i_obs*) as a Gaussian distribution with 'sliding' mean

One might notice that the standard deviation does not change, producing the relation of probabilities e.g. between P(*i_obs* | *i*=*i_obs*) and P(*i_obs* | *i*=*i_obs*+1) for any observed frame in the interval.
The variables and their sample space are shown in (14).

$$\begin{aligned} I &\in \{1,\ldots,I_{\max}\}\langle I_{\max}\rangle \\ I\_obs &\in \{1,\ldots,I\_obs_{\max}\}\langle I\_obs_{\max}\rangle \end{aligned} \tag{14}$$

For the *temporal* model we assume that each frame *I* can show up as an observed frame *I_obs* with a certain probability. Thus, we have a conditional dependency of *I* and *I_obs* as can be seen in Bayes-net of Figure 16.

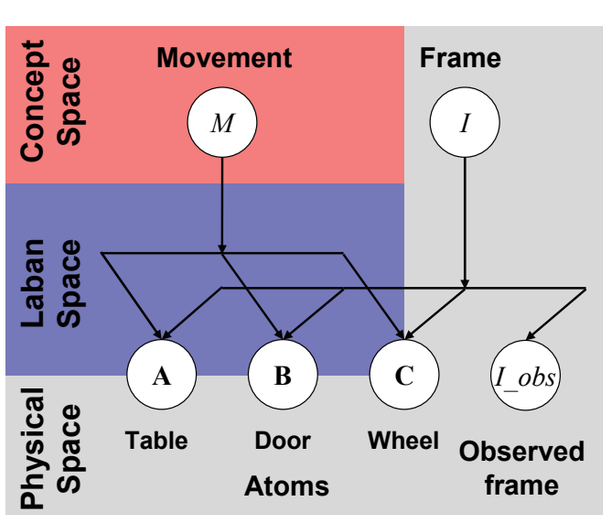

Figure 16. Bayes-Net of the *temporal* model which connects to the *Space* model

The *I_obs* variable is measured directly as hard evidence, the frame *I* can be interpreted as a soft evidence for the *Space* model. The joint distribution embedded *Space* model while omitting the body part index can be expressed as

$$\begin{aligned} &P(M\ I\_obs\ I\ A\ B\ C) \\ &= P(M)\quad P(I)\quad P(I\_obs \mid I)\quad P(A \mid M\ I)\quad P(B \mid M\ I)\quad P(C \mid M\ I) \end{aligned} \tag{15}$$

Table 2 summarizes the variables used in this model.

Variable	Symbol	Description
Frame	*I*	Computed (soft) evidence
Observed Frame	*I_obs*	Measured (hard) evidence

Table 3. Variables used in the *temporal* model.

4.4 Learning of probability tables

The previous sections presented models through Bayesian nets and joint distributions. The latter appeared as a product of several conditional distributions which links the hypothesizes to the data. Thus, the distributions need to represent the data given a certain condition. The question is 'How can we find this distribution?' and the answer is 'Learning'. Many different techniques and forms of representations exist.

The probability distribution can be learned by counting the observations a variable has a certain value (Histogram Learning). For a finite number of discrete values the process can be described as building a histogram. By dividing the counts for each value *i* of the variable *V* (*V*=*i*) by the total number of samples *n* a probability distribution can be computed as

$$P^*([V = i]) = \frac{n_i}{n} \tag{16}$$

The assumptions that apply are: i) All samples *n* come from the same phenomenon. ii) All samples are from a single variable *V*. iii) The order of the samples is not important.

When learning a probability distribution through the histogram some values of *V* might have zero probability, simply because they have never been observed. Whenever these values occur in the later classification stage the corresponding hypothesis(es) will receive also a zero probability. In continuous classifiers, that are based on multiplicative update of beliefs, this leads to an immediate and definite out-rule of the hypothesis(es). Most of the time this is not desirable and appears 'unnatural'.

One way of solving this is to use an equation which produces a minimum probability for non-observed evidences. Equation (17) is based on the Laplace Succession Law and it can be seen that it will produce a minimum probability of $1/(n+\lfloor V \rfloor)$.

$$P^*([V = i]) = \frac{n_i + 1}{n + \lfloor V \rfloor} \tag{17}$$

The atom variable A_{rh} has nine values $\lfloor V \rfloor = 9$ and by learning from six samples n=6 each non-observed value will receive a probability of $P^*(V) = 0.0667$ for all values i where $n_i = 0$.

The learned table P (*Atom* | *M I*) holds the probability distribution of the Variables *Atom* e.g. *Table Plane* right hand atom A_{rh}. The variable has two conditions, the movement *M* and the frame *I*. Figure 17 represents this multidimensional table.

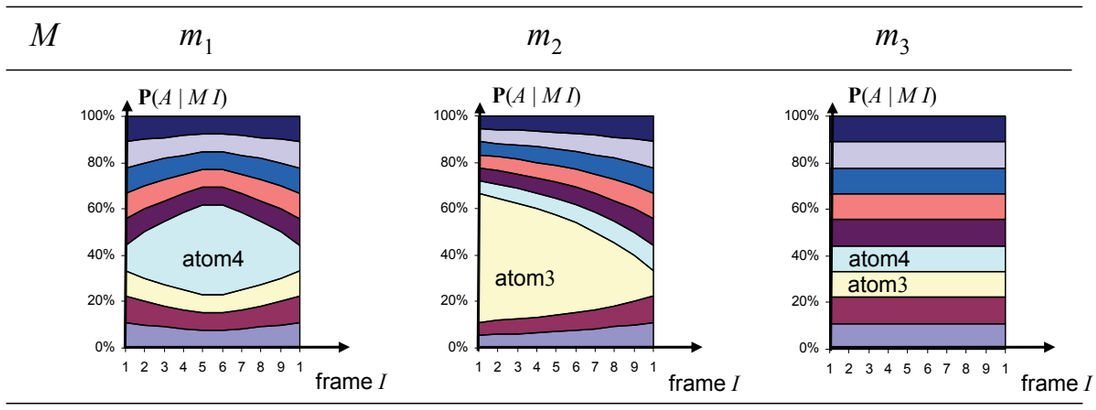

Figure 17. Learned table for generic movements of the type *P(Atom | M I)*. The movements m_1 and m_2 have a dominating *atom* (4 and 3) during certain phases (middle and beginning) while movement m_3 shows no spatial pattern at all

When stacking the probabilities for each value one over each other, patterns can be observed along the time given by the frame *I* and between the movements *M*. From the hypothetical example one can conclude, that in movement $M = m_1$ at frame $I = 5$ most probably *atom* 4 will show up. This shows that after learning the data can be presented in a way that allows an evaluation of both, the hypothesis and the data. The generic movement m_2 has its dominating *atom* 3 at the beginning. Movement m_3 can be seen as a 'white noise' movement where no spatial pattern can be observed along the time. The size of the table is given by the cardinality of *Atom* i.e. nine, the maximum number of frames, e.g. forty and the number of movements, e.g. four. In this example the table will have 1440 entries.

4.5 Continuous classification of movements

Classification is the final step after the model has been established and the tables have been learned. Given our joint distribution (17) we need to formulate a question, i.e. what we want to classify and what we can observe. In our case we are interested to classify an unknown movement from the evidences observed in the *Physical Space*. In the following we continue with a simplified question, i.e. classifying a movement *M* taking into account only the *Vector Symbols* (atoms) *A* and the frame *I*.

The previous step of learning provided us with the possibility to determine the probability that the *atom A* has value *a* given a frame *i* from all possible frames *I* and a given a movement *m* from all possible movements *M*, i.e. *P(a | m, i)*. The table **P** (*A* | *M*, *I*) holds the probability distribution for all possible values of atom *A* given all possible movements *M* and frames *I*.

Knowing the conditional probability **P** (*A* | *M*, *I*) together with the prior probabilities for the movements **P** (*M*) we are able to apply Bayes' rule and compute the probability distribution for the movements *M* given the frame *I* and the *atom A* with

$$\mathbf{P}(M \mid I\ A) \propto \mathbf{P}(M)\quad \mathbf{P}(A \mid M\ I) \tag{18}$$

It is possible to compute how likely it is that an observed sequence of *n atoms* was caused by a certain movement *m*. To compute the *likelihood* we assume that the observed *atoms* are independently and identically distributed (i.i.d.). In (19) the sequence of *n* observed values

for *atom a* is represented by $a_{1:n}$. For each movement m the joint probability will be the product of the probabilities from frame $i = 1$ to $i = n$, where the j-th frame of the sequence is indicated by i_j.

$$P(a_{1:n} \mid m\, i_{1:n}) = \prod_{j=1}^{n} P(a_j \mid m\, i_j) \tag{19}$$

We can formulate (19) in a recursive way and for all movements M and get

$$P(a_{n+1} \mid M\ i_{1:n+1}) = P(a_n \mid M\ i_{1:n})\quad P(a_{n+1} \mid M\ i_{n+1}) \tag{20}$$

The *likelihood computation* (20) can be plugged in our question (18). Assuming that each frame i a new observed direction symbol arrives we can continuously (online) update our classification result.

$$P(M_{n+1} \mid i_{1:n+1}\ a_{1:n+1}) \propto P(M_n)\quad P(a_{n+1} \mid M\ i_{n+1}) \tag{21}$$

We can see that the prior of step n+1 is the result of the classification of step n. Given a sufficient number of evidences (*atoms*) and assuming that the learned tables represent the phenomenon sufficiently good, the classification will converge to the correct hypothesis. This will happen, regardless of the probability distribution of the 'true' prior for n=0, if there are no zero probabilities assigned to any of the hypothesizes.
The final classification result is given by the maximum a posteriori (MAP) method. Several questions can be formed and compared against each other. The following Table 4 presents some questions and their decompositions.

Movement using 2-D (horizontal) Space model	
Question	$P(M \mid i\ a)$
Decomposition	$P(M)\ \ P(a \mid M\ i)$
Movement using 2-D (vertical) Space with temporal model	
Question	$P(M \mid i\_obs\ b)$
Decomposition	$P(M)\ \ P(i\_obs \mid i)\ \ P(b \mid M\ i)$
Movement using 3-D Space model	
Question	$P(M \mid i\ a\ b\ c)$
Decomposition	$P(M)\ \ P(a \mid M\ i)\ \ P(b \mid M\ i)\ \ P(c \mid M\ i)$

Table 4. Questions for classification and their decompositions

In this example the query variable (usually M) is held in a capital letter, while the observed evidences have small letters.

5. Online Movement Recognition System

The previous sections of this article reflect the steps of designing a probabilistic model. The implementation of the processes and its results can also be organized in steps. The first step is the extraction and computation of the low-level features. In the second step probabilistic variables and conditional kernel tables need to be defined. The third step of learning fills the tables with data from a number of trials. In the fourth step several joint distributions and questions can be defined to investigate different types of models. The fifth and final step is to run the continuous classification and discuss the evolution of the probabilities and the

confusion table for several trials. To emphasize the important characteristic of the system it is called *Online Movement Anticipation and Recognition (OMAR)* system. As this section emphasizes the technological aspects of the solution some technical terms will be used such as conditional kernel maps and -tables to represent a probability distribution.

5.1 Learning conditional kernel maps

The second step of implementation is the definition of the probabilistic variables and conditional kernel tables. As this has been done already in section 4 we can proceed to the third step of learning those tables. Figure 18 shows the flow chart of the learning process.

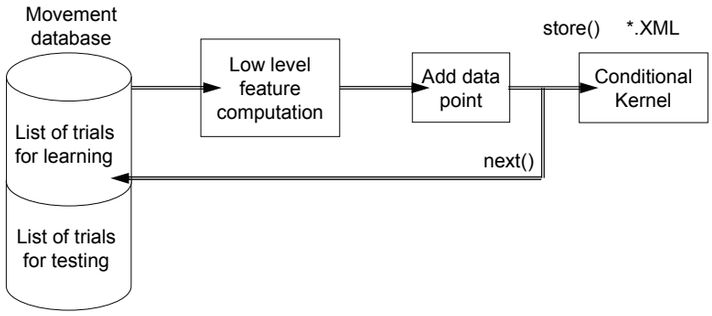

Figure 18. Learning process: Low level features are extracted each frame and 'adds' points to the histogram. After all trials are processed the conditional kernel maps are stored, e.g. in XML-format

From the movement database (HID) a set of trials for learning is chosen and fed into the system for low-level feature extraction. The database consists of five trials per person and movement. Three trials are usually chosen for learning. Each trial produces one data point per feature and frame. Learning based on an histogram approach creates probabilistic tables simply by adding those points until all trials are processed.

5.2 Results: Probability tables for Space

Some of the movements from our database can be described as 'gestures'. Figure 19 shows tables for two 'gestures', i.e. *byebye* and *pointing* with the nine *atoms* of the right hand.

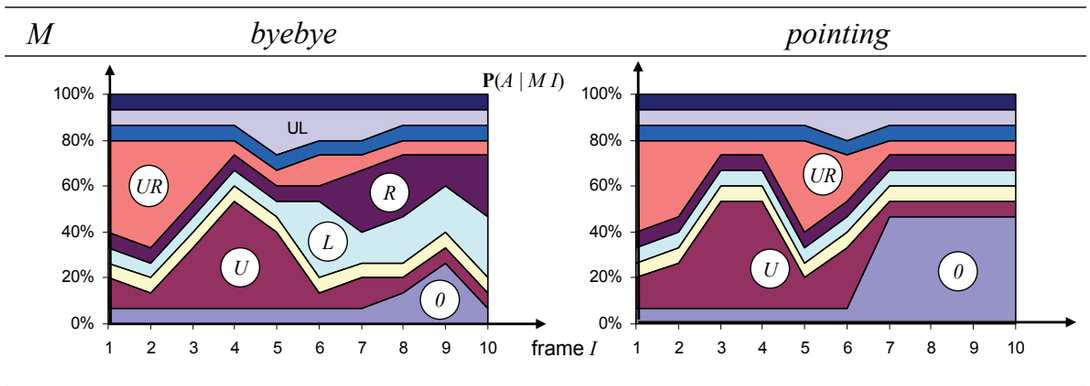

Figure 19. Learned Table $\mathbf{P}(B \mid M\,I)$ for gesture *byebye* and *pointing*

It represents the 'fingerprint' of the gesture prototype for waving *byebye*. The table preserves the possibility to evaluate what has been learned. Figure 19 uses a stacked representation of the probabilities to show which atoms are dominant during certain phases. Two gestures are to be compared: *byebye* on the left and *pointing* on the right. During the first frames the most likely *atoms* to be expected are the ones that go upward and to the right, i.e *UR* and *U*. This is coincides with our intuition, that while we are starting to perform a gesture with the left hand we tend to move up and to the left to gain space to perform the gesture. This is similar for both gestures. From the fifth or sixth frame on, the gestures become distinct. The gesture *byebey* has mainly movements to the left and right (*L* and *R*) with some zero *atoms* 0 at the points of inflection. The gesture *pointing* has mainly non-movement *atoms* (0) leaving the other probabilities at their minimum given by the Laplace assumption. It can be concluded that the movement set 'gestures' has a high spatial distinctiveness and can be used for simple but robust command interaction with a robot.

5.3 The process of recognition

The fourth step of implementation has been presented in section 4. The joint distributions of interest are:

- Movement classification using 2-D Space.
- Movement classification using 3-D Space.

The fifth and final step is the investigation of the evolution of probabilities and the confusion table that can be obtained for all trials of the test set.
Figure 20 shows the flow chart of the classification process.

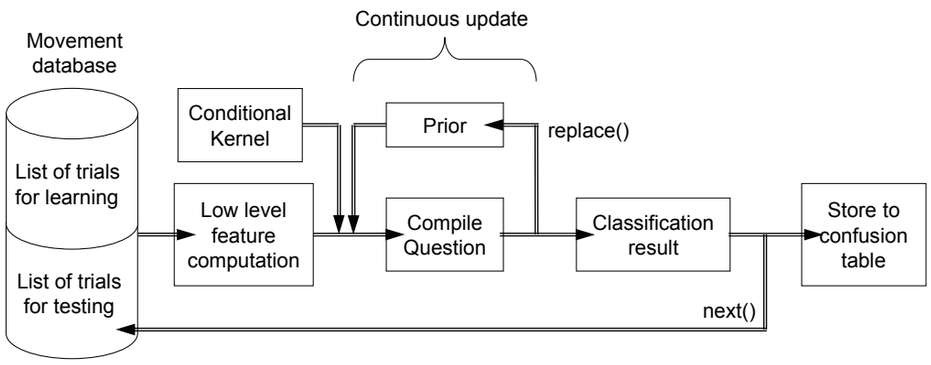

Figure 20. Classification process: The inner loop of continuous update produces the evolution of probabilities, the outer loop of next trial produces the confusion table

The inner loop of continuous update produces the evolution of probabilities, the outer loop of 'next trial' produces the confusion table. Classification uses the same process for the computation of low level features as learning before. With the low level features and the previously stored conditional kernel maps it possible to compute the desired probability distribution. This goes according to the defined joint distribution and the desired question. Inside the probabilistic library the step is known as 'compiling the question'. Through feeding in (replacing) the result of the compiled question as the new prior a continuous update of the classification results for all frames can be obtained. The result of the 'last' frame gives the final result and while looping through all trials for testing a confusion table can be built.

We can conclude that the two processes of learning and classification are based on the same type of observations as shown in Figure 21.

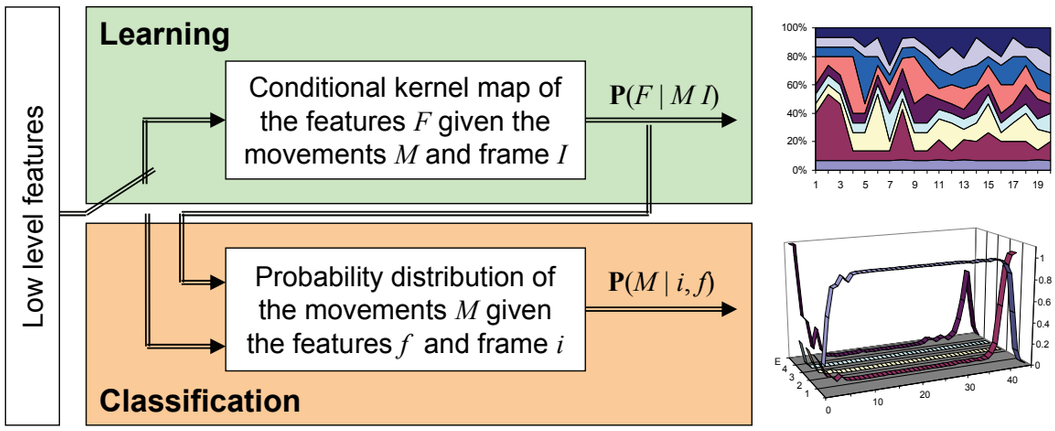

Figure 21. Switching between Bayesian learning and classification

The previously presented scheme starts by learning and, after the conditional kernel maps have been build, continues with classification. An important feature of Bayesian histogram learning is that we can 'switch back' at any time to learn new and more data. This opens the possibility to create artificial agents that are able to continuously learn new data during their daily operation.

5.4 Results: 3-D versus 2-D Recognition

By representing movements using only one plane, some of the spatial information gets lost. This section investigates the loss by comparing the results of classification when using the *Vector Symbols* (*atoms*) of all the tree planes *A*, *B*, *C* to the results gained when only the *atom B* of the *Door Plane* (vertical) plane π_v is used. The results are compared through a confusion table for eight movements, as already presented in Table 1. Table 5 shows the results for using the *B* atoms of the vertical plane π_v.

Movement	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	Σ_e
1 lunging	7			5				1	6
2 maestro		5				8			8
3 stretch			12				1		1
4 ok				8	1		4		5
5 pointing				1	10		1		2
6 byebye						13			0
7 shake				4			9		4
8 nthrow				4			1		5
									31

Table 5. Confusion table using only the 2-D (vertical) *atoms*

The sum of all numbers in each row usually adds up to thirteen, though some movements have fewer trials. In the 2-D case 31 of 95 trials are classified wrongly leaving a recognition rate of 67%. The highest false-rate has the *maestro* sequence which is confused with the

byebye sequence shown in the second row. By comparing the traces of the two movements (see Figure 22) we can see that they are quite similar, though the vertical plane π_v appears as the most distinctive one.

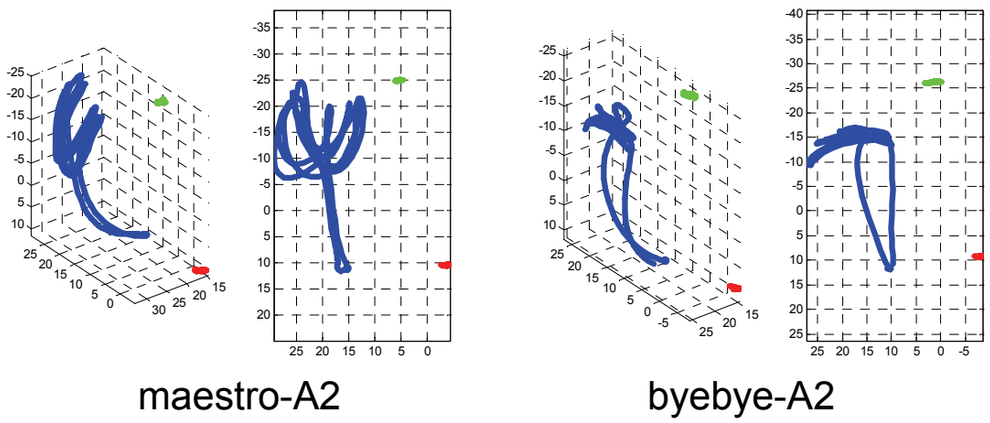

Figure 22. Comparing the traces of the movements *byebye* and *maestro* in 2-D and 3-D

The confusion between the two-hand movement *lunging* and the one-hand gesture *ok* indicated in the first row of the table is partly due to the traces but also due to the model. From Figure 23 it can be seen that for 2-D the right hand traces (blue) are similar leading partly to the confusion.

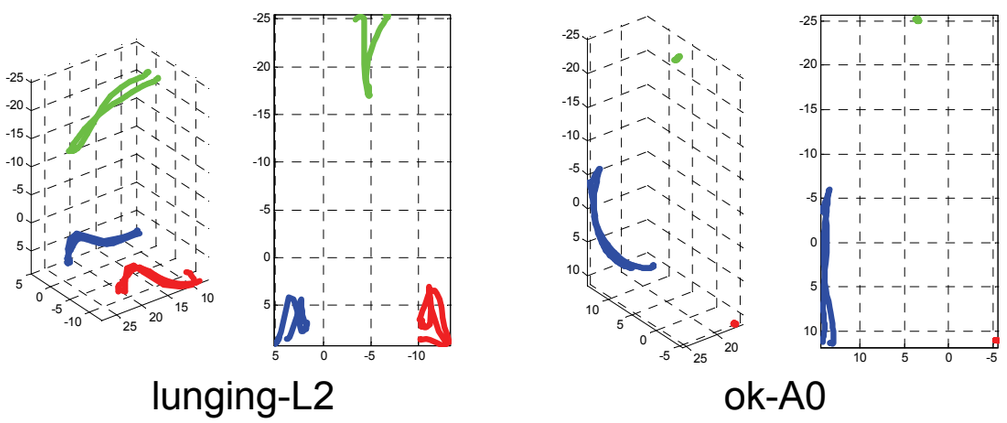

Figure 23. Comparing the traces of the movements *lunging* and *ok* in 2-D and 3-D

The model for the left hand is based on the assumption that we get mostly non-movement *atoms* which is true for both cases. The model can be easily improved by adding an evidence for not having moved at all.

The confusion between the gesture *ok* and the movement *shake* indicated in the fourth row of the table is due to the traces for some trials. From Figure 24 it can be seen that trials where the hand does not reach towards the middle (sagittal plane), but goes straight forward, the 2-D projection can be confused easily.

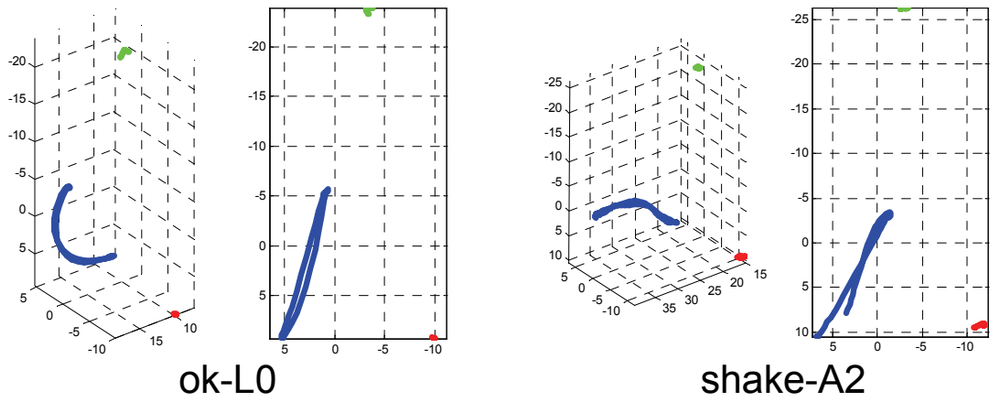

Figure 24. Comparing the traces of the movements *lunging* and *ok* in 2-D and 3-D

In this case the confusion goes in both directions as can be seen in the seventh row.
The final confusion occurs in the eight row. From Figure 25 it can be seen that the 2-D projecttion does not convey the information on the sagittal oszillation the right hand is performing.

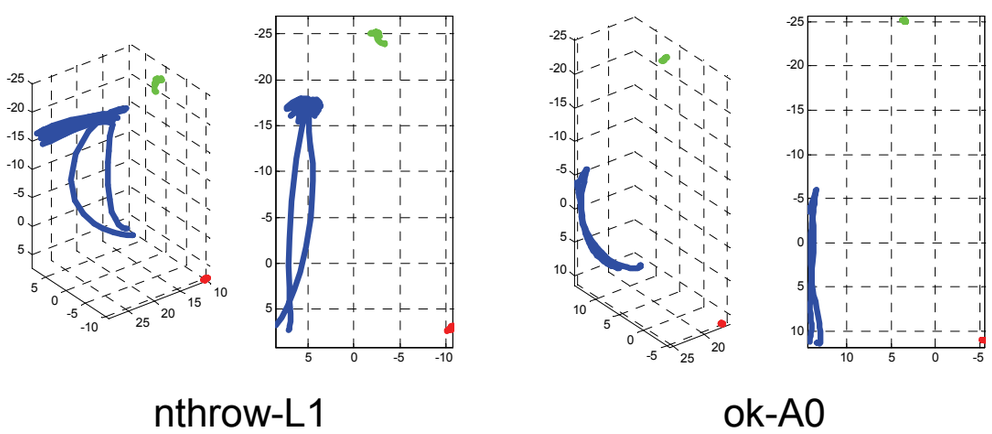

Figure 25. Comparing the traces of the movements *nthrow* and *ok* in 2-D and 3-D

Table 6 shows the results for using the *A*, *B* and *C atoms* of the all planes.

Movement	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	Σ_e
1 lunging	11			1			1		2
2 maestro		2				11			11
3 stretch			12	1					1
4 ok				9			4		4
5 pointing					10		2		2
6 byebye						13			0
7 shake				1			12		1
8 nthrow								5	0
									21

Table 6. Confusion table using 3D (all planes) *atoms*

In the first row it can be seen, that the recognition rate has improved due to the additional evidences indicating the movement in the *x*-dimension. Similar is true for the seventh row where the hands are usually reaching further in the *x*-dimension when performing the *shake* movement as compared to the *ok* gesture. The *nthrow* movement is now recognized in all trials as the evidences of the sagittal waving are now processed. In the 3-D case 21 of 95 trials are classified wrongly leaving a recognition rate of 78%. The *maestro* movement of the second row is significantly worse in 3-D which may be due to the fact that the *x*-dimension does not add additional information for distinction.
We can conclude that the recognition rate improves in general when using evidences from all three planes (from 67% to 78%). Some movements can not be seen in certain planes, e.g. *nthrow* in the vertical plane π_v. It appears that apart from the 'pure' spatial pattern also evidences from the temporal model effect the classification result. A further tuning of the temporal model (sliding mean was used) should improve the results. Further improvements are expected from a variable that indicates if a hand has not moved at all.

6. Conclusions and future works

The work presented in this article started with the premise that the field of computational Human Movement Analysis is in need of an annotated database for human movements. The second section of this article showed that Laban Movement Analysis (LMA) is a good choice for this descriptor. After a brief overview the *Space* component was discussed in detail and the descriptive language Labanotation was presented. This section concluded with examples from our Human Interaction Database (HID) to outline the applicability of LMA for an annotated database.
To allow applications which involve autonomous mobile robots (e.g. 'social robots') a technical solution needed to be found to bring together monocular cameras and high precision data from a 3-D tracking device. The third section showed that we are able to base the computation of our low level features on two very different sensor types, i.e. monocular camera and commercial motion tracking device. This allows to work with a database of rich 3-D position data and sensory input from (a) 2-D projection(s). The section presented the extraction and computation of low level features based on displacement angles.
A suitable framework needed to be chosen which provides a scheme for learning as well as for classification and takes into account that LMA is based on human observations where incompleteness and uncertainty are issues. By suggesting a Bayesian approach in the fourth section the former issues have been taken into account. A probabilistic scheme for learning and classification using models that can be represented as Bayesian nets was shown.
The fifth section presented the implementation as flow charts with some links to functions of the probabilistic library used. It was shown that the probabilistic approach provides the learned data in a way that allows its visual inspection and evaluation. The chosen histogram-based approach for learning provides a simple way of adding data points. The Human Interaction Database (HID) provides several sets of 'expressive movements'. It was shown that 'gestures' has a high spatial distinctiveness and can be used for simple but robust command interaction with a robot. As a benefit of the modularity of the system results for movement classification could be presented and compared using 3-D *Space* and 2-D *Space.* The recognition rate improved in general when using evidences from all three planes (from 67% to 78%). Some movements could not be seen in certain planes, e.g. *nthrow* in the *Door Plane* π_v.

For the future the database of movements with annotated Laban Movement Analysis (LMA) descriptors will be extended by certain classes. The Bayesian models will be extended by additional components taken from LMA. A socially assistive robot will be designed to be used in rehabilitation that records human movements annotated with LMA descriptors. An interface for the smart infrastructure and the socially assistive robot will be designed to show the results of the recorded movement and the evolution of the rehabilitation process.
The main goals of the future research will be to establish Laban Movement Analysis (LMA) as a general tool for the evaluation of human movements and provide those communities that collect large amounts of experimental data with technical solutions for labeled data sets. The research will be justified by showing that rehabilitation processes do benefit from evaluations based on LMA. That comparison of experimental data with very distinct experimental set-ups is possible by using the descriptors of LMA. Data from computational LMA opens the possibility to cluster motor deficits and neurological disorders that are similar with regards to LMA.

7. Acknowledgements

The authors would like to thank Luis Santos from the Institute of Systems and Robotics, Coimbra for his work on the implementation. This work is partially supported by FCT-Fundação para a Ciência e a Tecnologia Grant #12956/2003 to J. Rett and by the BACS-project-6th Framework Programme of the European Commission contract number: FP6-IST-027140, Action line: Cognitive Systems to J. Rett and L. Santos.

8. References

Badler, N.I.; Phillips, C.B. & Webber, B.L. (1993). *Simulating Humans: Computer Graphics, Animation, and Control*, Oxford Univ. Press

Bartenieff, I. & Lewis, D. (1980). *Body Movement: Coping with the Environment*, Gordon and Breach Science, New York

Bradski, G.R. (1998). Computer Vision Face Tracking For Use in a Perceptual User Interface, *Intel Technology Journal*, Q2, 15

Bregler, C. (1997). Learning and recognizing human dynamics in video sequences, *Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico,

Chi, D.; Costa, M.; Zhao, L. & Badler, N. (2000). The EMOTE model for Effort and Shape, SIGGRAPH 00, *Computer Graphics Proceedings, Annual Conference Series*, 173-182 ACM Press

Dias, J. (1994). *Reconstrução Tridimensional Utilizando Visão Dinâmica*, University of Coimbra, Portugal,

Fong, T.; Nourbakhsh, I. & Dautenhahn, K. (2003). A survey of socially interactive robots, *Robotics and Autonomous Systems*, 42, 143–166

Foroud, A. & Whishaw, I.Q. (2006). Changes in the kinematic structure and non-kinematic features of movements during skilled reaching after stroke: A Laban Movement Analysis in two case studies, *Journal of Neuroscience Methods* 158, 137-149

Hartley, R. & Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*, Cambridge University Press,

Hutchinson, A. (1970). *Labanotation or Kinetography Laban*, Theatre Arts, New York

Kendon, A. (2004). *Gesture: Visible Action as Utterance*, Cambridge University Press,

Knill, D.C. & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation, *TRENDS in Neurosciences*, 27, 712-719

Laban, R. (1966). *Choreutics*, MacDonald & Evans., London

Loeb, G.E. (2001). Learning from the spinal cord, *Journal of Physiology*, 533.1, 111-117

Longstaff, J.S. (2001). Translating vector symbols from Laban's (1926) Choreographie, 26. *Biennial Conference of the International Council of Kinetography Laban*, ICKL, Ohio, USA, 70-86

Nakata, T.; Mori, T. & Sato, T. (2002). Analysis of Impression of Robot Bodily Expression, *Journal of Robotics and Mechatronics*, 14, 27-36

Nakata, T. (2007). Temporal segmentation and recognition of body motion data based on inter-limb correlation analysis, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, IROS,

Otero, N.; Knoop, S.; Nehaniv, C.; Syrda, D.; Dautenhahn, K. & Dillmann, R. (2006). Distribution and Recognition of Gestures in Human-Robot Interaction, *The 15th IEEE International Symposium on Robot and Human Interactive Communication*, 2006. ROMAN 2006., 103-110

Pavlovic, V.I. (1999). *Dynamic Bayesian Networks for Information Fusion with Applications to Human-Computer Interfaces*, Graduate College of the University of Illinois,

Rett, J. & Dias, J. (2007-A). Human-robot interface with anticipatory characteristics based on Laban Movement Analysis and Bayesian models, *Proceedings of the 2007 IEEE 10th International Conference on Rehabilitation Robotics*,

Rett, J. & Dias, J. (2007-B). Human Robot Interaction Based on Bayesian Analysis of Human Movements, *EPIA 07*, Neves, J.; Santos, M. & Machado, J. (ed.) 4874, 530-541 Springer, Berlin,

Rett, J. (2008). *Robot-Human Interface using Laban Movement Analysis inside a Bayesian framework*, University of Coimbra,

Rett, J.; Neves, A. & Dias, J. (2007). *Hid-human interaction database*: http://paloma.isr.uc.pt/hid/,

Sato, T.; Nishida, Y. & Mizoguchi, H. (1996). Robotic room: Symbiosis with human through behavior media, *Robotics and Autonomous Systems*, 18, 185-194

Starner, T. (1995). *Visual recognition of american sign language using Hidden Markov Models*, MIT,

Starner, T. & Pentland, A. (1995). Visual recognition of american sign language using hidden markov models, In *International Workshop on Automatic Face and Gesture Recognition*, Zurich, Switzerland, 189-194

Zhao, L. (2002). *Synthesis and Acquisition of Laban Movement Analysis Qualitative Parameters for Communicative Gestures*, University of Pennsylvania,

Zhao, L. & Badler, N.I. (2005). Acquiring and validating motion qualities from live limb gestures, *Graphical Models* 67, 1, 1-16

Brain, Vision and AI

Edited by Cesare Rossi

ISBN 978-953-7619-04-6

Hard cover, 284 pages

Publisher InTech

Published online 01, August, 2008

Published in print edition August, 2008

The aim of this book is to provide new ideas, original results and practical experiences regarding service robotics. This book provides only a small example of this research activity, but it covers a great deal of what has been done in the field recently. Furthermore, it works as a valuable resource for researchers interested in this field.

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen