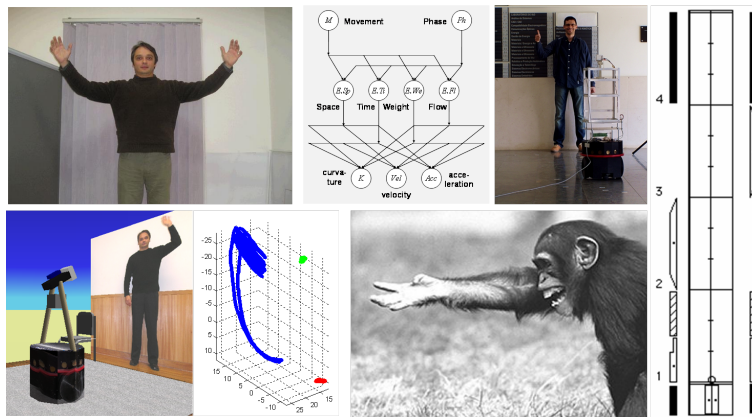# ROBOT-HUMAN INTERFACE USING LABAN MOVEMENT ANALYSIS INSIDE A BAYESIAN FRAMEWORK

PhD THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND
COMPUTER ENGINEERING
UNIVERSITY OF COIMBRA

Jörg Rett

June 2008

To Dorota and Alan Peter and Peter and Elisabeth.

iv

Cutting of the possibilities to move is like cutting off live itself.

For some presumptuous reason, man feels the need to create something of his own that appears to be living, that has an inner strength, a vitality, a separate identity - something that speaks out with authority - a creating that gives the illusion of live.

Frank Thomas and Ollie Johnston of Disney Studios

The thesis was first revised in February 2009.

Acknowledgments

First of all, I would like to thank Professor Jorge Dias, my supervisor, who has accepted me as his student and thus gave me the chance to become a researcher and pursue the goal to receive the degree of a PhD. I thank him for giving me inspiration and balancing out dream and reality. His sparkling energy was never fading.

I would also thank Dr. Juan-Manuel Ahuactzin from Probayes, Grenoble for working, advising and discussing all issues concerning the Bayesian approach. He has been a shiny beacon in the ocean of uncertainty.

There are several people that have worked with me on the technical realization of this work. The first was Enguerran Boissier from from Universite Paul Sabatier in Toulouse, France who has worked on the Human Tracking Module. He was followed by Bernardo Sousa from ISR, Coimbra who realized the gesture recognition system and the first version of Nicole. During that time Pedro Nuñez from the University of Málaga was visiting our lab and created the *Action Planner* and the *Navigation* system for Nicole. Their cooperation resulted in a public demonstration at the entrance hall of the University of Coimbra, Department of Electrical Engineering in Summer 2006. Their work was taken up by Alberto Neves from the University of Coimbra who created the Human Interaction Database (HID) and integrated the commercial motion tracking device and the process of low feature extraction. His work was finally taken up by Luis Santos from the University of Coimbra who extended the HID, finalized the Effort module and automated and improved many processes. Without them it would have been impossible to give this

dream an embodiment.

Most of all I would like to thank my wife Dorota who gave me counseling and love all along the way. My son Alan-Peter who was born during my studies and reminded me of the joy away from all work. Finally my parents Elisabeth and Peter whose love was always apparent even over the long distance.

Coimbra, Portugal   Jörg Rett
May 9th, 2008

Abstract

This work presents the implementation of a computational method for Human Movement Analysis based on a notational system by Laban using a Bayesian framework.

The research field of computational Human Movement Analysis is lacking a general underlying modeling language. A semantic descriptor allows to pose the classification task as a problem to recognize a sequence of symbols taken from an alphabet consisting of motion-entities. Systems which are based on such a modeling language can use it as a ground truth for recoding and labeling training data automatically. The inherent constraints of a modeling language can be used to make the task of movement recognition more tractable. The knowledge gained in Human Science can contribute tremendously to an automatic human movement analysis. In order to apply this knowledge effectively it is crucial that a theoretical 'fit' in description and modeling exists. The main contribution of this thesis is to provide this theoretical 'fit' through: i) Laban Movement Analysis as a semantic description and ii) Bayesian approaches as the tool for modeling.

Laban Movement Analysis (LMA) has been proven successful in areas where humans are observing other humans' movements like studies of dance and its application to physical and mental therapy. LMA provides models for the observation and description of human body movements. What makes the framework of LMA so special is its ability to describe an additional 'expression' that accompanies the spatial trajectory known as *Effort*. This might be the key to retrieve some evidences about the emotional state or the

intention of the performer.

By using a Bayesian approach for modeling a coherent framework for designing, learning and 'online' classification can be used. The probabilistic approach allows to introduce 'certainty' to evaluate how strong the system believes in a hypothesis. The 'online' characteristic allows the system to continuously update and eventually change its belief. Emerging from the probabilistic approach this thesis introduces also the characteristic of 'anticipation'.

For the technical task of human movement tracking a 'hybrid' approach is followed. Through this, the system benefits from the simpler processing and higher precision of an 'active sensor' during the recoding step, while using the attractive touch-free alternative of computer vision for the classification step. The correspondence is established through mapping the 3-D data collected from the 'active sensor' to 2-D planes aligned with the camera plane. Along the work of the thesis a 'Human Interaction Database' (HID) was established which contains the labeled movement data.

Possible applications that could benefit from the computational solution for automatic LMA are 'Smart Houses', 'Surveillance' and 'Rehabilitation'. A 'social robot' can be seen as an instance of the former applications, where the interaction with a person is key-skill. Thus, the social robot Nicole was chosen to demonstrate the feasibility of a Bayesian movement classification based on LMA descriptors. The robustness of the probabilistic algorithm allowed several trials in a natural environment with untrained actors.

Keywords: Human Movement Analysis, Laban Movement Analysis, Bayesian models, Visual Tracking, Social Robots, Human-Robot Interaction

Resumo

Neste trabalho é apresentada a Análise Computacional de Movimentos de Laban usando uma rede Bayesiana.

A pesquisa no campo de Análise Computacional de Movimentos Humanos não tem ainda uma linguagem modelo associada, i.e., uma maneira de mapear características em símbolos. Existindo tal descritor de semântica, o problema do reconhecimento pode ser colocodo como o reconhecimento de uma sequência de símbolos retirados de um alfabeto que consiste em entidades de movimento. Este alfabeto intruduz limitações que tornam o problema de reconhecimento de movimentos realizável. O método abre a possibilidade de se dedicar mais esforço em gravar e etiquetar dados, como tem acontecido na área de 'reconhecimento de voz' no passado.

A Análise de Movimentos com notação de Laban (AML) tem provas dadas e áreas onde humanos observam movimentos de outros humanos tais como o estudo da dança, e a sua aplicação a terapias de foro físico de mental. AML fornece modelos para observação de descrição de movimentos do corpo humano. O sistema notacional (notação Laban) consegue descrever os conteúdos estáticos bem como dinânimos de um movimento. Para implementar a semântica da AML num computador foi escolhida uma aproximação Bayesiana e o formalismo da 'Programação Bayesiana'. A estrutura permite modelar o processo, aprender e guardar as dependências entre características e símbolos e classificar, enquanto corre, um movimento usando as etiquetas da AML. O modelo completo é desenhado sob uma rede Bayesiana composta por pequenos sub-modelos, onde cada um representa um certo aspecto

do problema. A classificação pode ser resolvida juntando alguns (ou todos) sub-modelos num modelo completo e colocando a questão. A ligações entre sub-modelos são construídas através de evidências incertas. A rede Bayesiana pode ser interpretadas como a estrutura de comunicação entre as evidências e as suas mensagens.

Possíveis aplicaçoes que pode benificiar da Análise Computacional de Movimentos de Laban (ACML) são 'Casas Inteligentes', 'Vigilância' e 'Rehabilitação'. Neste trabalho robôs sociais foram escolhidos para demonstrar a nossa solução. O factor chave que os robôs sociais partilham com outras aplicações mencionadas é a interacção homem-máquina. Com ACML foi criada uma nova capacidade para o sistema interactivo de maneira a melhor entender o movimento humano observado.

O sistema é capaz de reconhecer um movimento (acção) capturado através de uma sequência de imagens capturas através de uma câmara de uma maneira muito eficaz. A aprendizagem é feita 'offline' baseada na nossa *Human Interaction Database* (HID). Essa base de dados guarda as sequencias de imagens bem como dados 3D de alta precisão. A técnica de constantemente actualizar a estimativa anterior através de observação é reminiscente da antecipação humana, enquanto que os descritores AML permitem a extracção de conteúdos expressivos de um movimento.

Palavras chave: Human Movement Analysis, Laban Movement Analysis, Bayesian models, Visual Tracking, Social Robots, Human-Robot Interaction

Notation

Probabilistic Laban Variables

Symbol	Description
A_{bp}	Vector Symbol *atoms* in π_h
Acc	Speed gain e.g. $Acc = high$
B_{bp}	Vector Symbol *atoms* in π_v
Body	Body set - set of constraints on human kinematics, e.g. $dist(head, lefthand)$
bp	Body part e.g. rh (right hand)
C_{bp}	Vector Symbol *atoms* in π_s
$E.Sp$	Effort Space e.g. $E.Sp = direct$
$E.Ti$	Effort Time e.g. $E.Ti = sudden$
$E.We$	Effort Weight e.g. $E.We = strong$
$E.Fl$	Effort Flow e.g. $E.Fl = free$
Effort	Effort set - set of *Effort* variables including *Time*, *Space*, *Weight* and *Flow*
I	Frame index
I_obs	Observed Frame - Measured (hard) evidence
K	Curvature e.g. $K = small$
l_{Tr}	Triangle change e.g. *pos*
LLF	Low-level feature set, may be split into sub-sets
LLF$_{Sp}$	Low-level feature sub-set *Space*
LLF$_{Ef}$	Low-level feature sub-set *Effort*

Symbol	Description
M	Movement variable - type of movement, e.g. $M = pointing$
Ph	Phase e.g. $pre - stroke$
Relshp	Relationship set - set of constraints on the human-robot space, e.g. $dist(hum, rob)$
$Sh.H$	Spatial Shaping e.g. $spreading$
$Sh.V$	Spatial Shaping e.g. $rising$
$Sh.S$	Spatial Shaping e.g. $advancing$
$Sh.Flo$	Shape Flow e.g. $growing$
Shape	Shape set - set of *Shape* variables including *Spatial Shaping* and *Shape Flow*
Space	Space set - set of direction symbol variables (*atoms*), e.g. $A = U$
Sx_{bp}	Direction sign e.g. pos
Vel	Speed e.g. $Vel = fast$
ΔI_z	Triangle change e.g. pos

Contents

List of Tables

List of Figures

xxix

Chapter 1

Introduction

Contents

A brief and compact glance over the main issues of this thesis and their relationships is shown in Fig. 1.1. The core contribution is a model for automatic Laban Movement Analysis (LMA). Automatic LMA supports the analysis of human movements by providing an (automatic) annotation for movement data. By introducing the LMA-model into algorithms for movement classification, the feasibility of using LMA descriptors can be shown. Finally, the algorithms are used for Human-Robot Interaction in order to show its applicability.
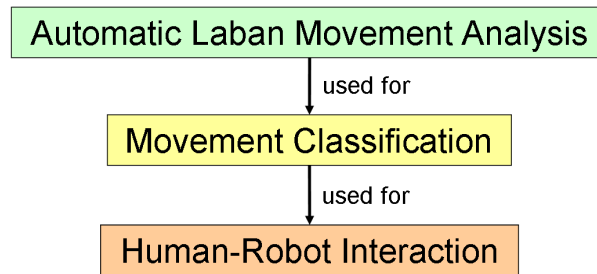
Figure 1.1: Main issues of this thesis. i) A model for automatic Laban Movement Analysis (LMA) is provided to support analysis of human movements. ii) The model is used for algorithms that classify movements. iii) The algorithms are used for the interaction between a robot and a person.

The endeavor to provide a model for automatic LMA can be seen as the initial task of this work. The second task, movement classification, can be see as the subsequent step. The task of movement classification is not decoupled from automatic LMA but uses its descriptors (features). Nevertheless, it also adds additional issues like a temporal segmentation (Phase) and the technical approach to human movement tracking.

By implementing movement classification through a Bayesian approach additional characteristics emerge like the measure of certainty and anticipation as shown in Fig. 1.2. The characteristics of Bayesian movement classification will finally pass into the qualities (skills) of a robot interacting with a human.

This chapter starts by presenting some human body movements which exhibit *expressiveness*. By suggesting to use a notational framework to capture this *expressiveness* several systems are presented afterward. Then the importance of movement analysis for the field of robotics is regarded. One step further, the possible impact of LMA for the *believability* of social robots is discussed. After this different aspects of the research area of 'Computational Human Movement Analysis' (C-HMA) are introduced that allow the interrelation with the work presented in this thesis. The first section deals
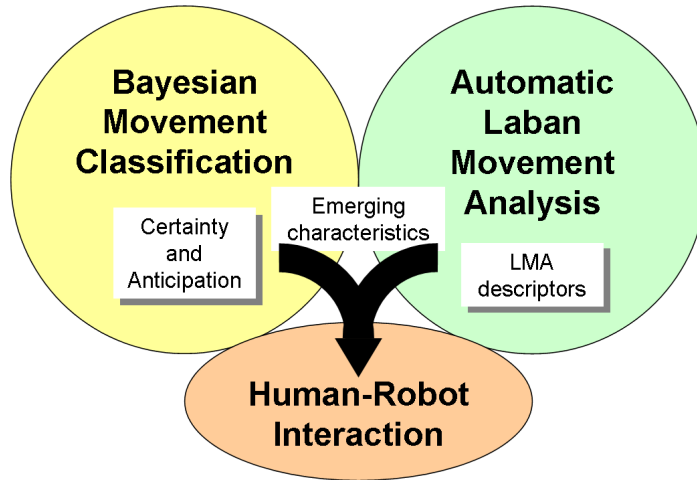
Figure 1.2: Main issues of this thesis. From automatic Laban Movement Analysis (LMA) emerge the descriptors. From the Bayesian movement classification emerge Anticipation and Certainty. These characteristics are passed to Human-Robot Interaction.

with applications, skills (functionalities) and methods for C-HMA. The following two sections will organize C-HMA through processes and interesting issues. The related works that are presented afterward are organized by i) contributions that explicitly use the concepts of LMA, ii) others that fall within the bigger scope of C-HMA and iii) implementations of social robots. The following sections present the contributions of this work followed by publications that were published along this thesis. The chapter concludes with a roadmap of the remaining chapters. Figure 1.3 summarizes the goal of the work described in this thesis. Laban Movement Analysis is implemented as a skill for Human-Robot Interaction.

1.1 Expressive human body movements

Human body movement is essentially the process of moving one or more body parts to a specific location along a certain trajectory. In some cases, a person observing the movement might be able to recognize it through the spatial
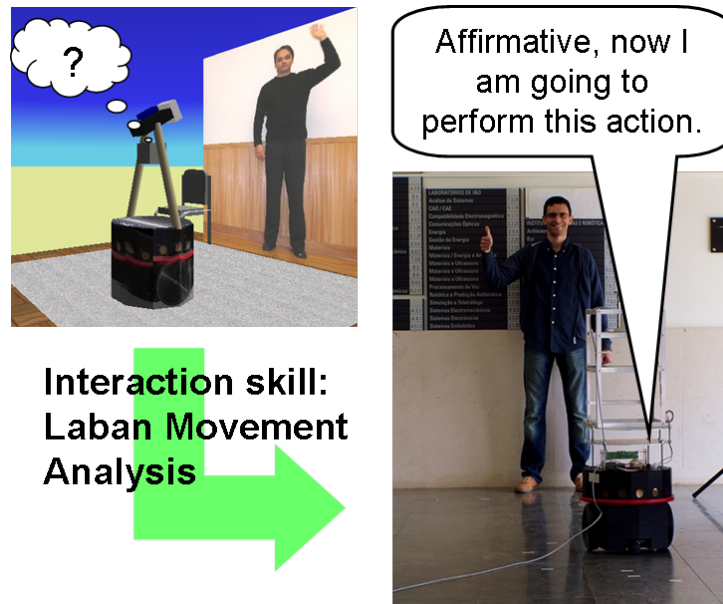
Figure 1.3: The final goal: Creating a social robot through implementing the interaction-skill of automatic Laban Movement Analysis.

pathway alone. Usually, the task of classification is supported by additional evidences which can not be retrieved from the kinematic information alone. These evidences can be interpreted as the *expressiveness* of movements. A typical case in which humans try to master the expressiveness of 'daily-life' movements is shown in Fig. 1.4 a). Mimes like Marcel Marceau try to carve out the expressive content of human actions and reproduce it on the stage [1].

Expressive movements might be produced unconsciously during sports as shown in Fig. 1.4 b) of Emil Zatopek who won several medals in the Olympics [2]. They might be the crucial part as in dance shown in Fig. 1.4 c) by Jane Franklin [3], Virginia. The study of expressiveness on artificial characters is interesting, as the crucial features of expressiveness need to be found and

---

[1]Photograph: David Hecker/Getty Images

[2](courtesy of http://www.omundodacorrida.com)

[3]Jane Franklin Dance studios in Arlington

Figure 1.4: Expressiveness a) carved out in mime, b) exhibiting physical effort, c) while performing a dance, d) in animated characters, e) in a monkey stretching his hand and f) for social interaction like political speeches.

transformed to another object as shown in Fig. 1.4 d)[4]. Thomas and Johnson of Disney Studios present the techniques of traditional animation and how to evoke emotional state through motion [JT81].

Kendon singles out some aspects of bodily action and treats them as 'gesture'. He states that 'gestures' have a certain manner of action and an expressive significance. The actor is deemed to exercise at least some degree of voluntary control over the movement and what it expresses. These definitions make 'gestures' distinct from movements due to being nervous, habitual or involuntary movements and necessary actions during interaction. Gestures are not only found with humans, but also appear frequently in apes. The data of Pollick and de Waal [PdW07] suggest that facial/vocal signals are more tied to a specific context than gestures. Furthermore that gestures seem to be less closely tied to particular emotions. The communicated need of a chimpanzee stretching out his open hand as shown in Fig. 1.4 e), depends

---

[4]Painting by Trevor Carlton

on the social context.

The social interaction of humans is full of expressive body movements, where politics is just one example for the need to convey expressive content. Figure 1.4 f) shows the speech of Nikita Khrushchev, premier of the Soviet Union (1960). Kendon [Ken04] 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.

1.2   Notational Systems for movements

In order to access the expressive content of movements theoretically, a notational system is needed. Early notational systems are known from the 17th century. Pierre Beauchamp and Raoul Auger Feuillet began in 1700 a program of publishing notated dances [LM92]. The basic elements of the *Beauchamp-Feuillet Notation* are the tract, symbols for the foot position, symbols for the step, signs for actions, turns and rhythm. Figure 1.5 b) shows the bar of a French courante.

The *Benesh Movement Notation* was invented in the late 1940's by Joan and Rudolf Benesh to document any form of dance or human movement [BB83]. It uses a five line stave that reads from left to right with bar lines to mark the passage of time, resembling a music score. All information about body and limb positions is shown within the five-line stave. *Movement Lines* trace the paths made by the extremities. *Locomotion Lines* link the positions of the feet, showing whether the performer steps, jumps or slides from one position to the next. Rhythm, phrasing and movement quality are shown above the stave. Figure 1.5 a) shows an example of Giselle Act 1, Peasant Pas de Deux, Male Variation.

a) *Ballet: Pas de Deux (Male)*

b) *Dance: French courante*   **c)** *Karate: Front Snap Kick*   **e)** *Ballet: Port de Bras*

Lower Arm	$\binom{0}{0}$			$\xleftarrow{4}$	$\left(\frac{2}{6}\right)$
Right Arm	$\binom{0}{0}$	$\uparrow 2$	$\left(\frac{2}{2}\right)$		

d)   *Movement: Arm pointing down - raising arm - bringing forearm to shoulder*
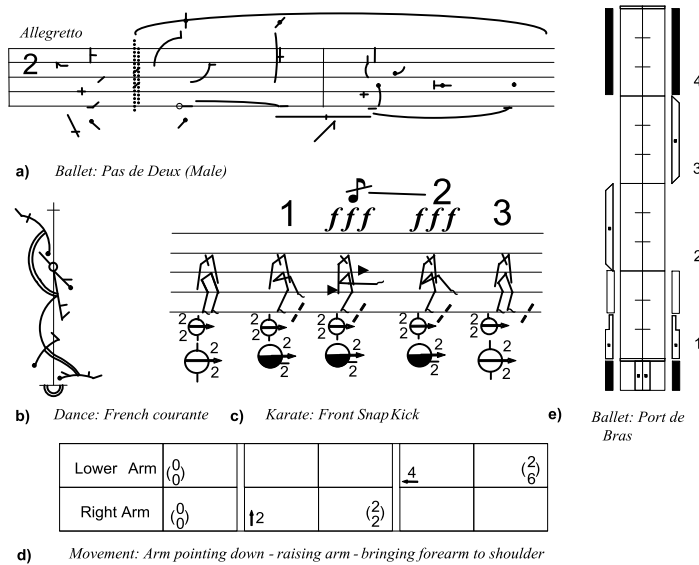
Figure 1.5: Different notational systems for movements a) The Benesh Movement Notation, b) the Beauchamp-Feuillet Notation, c) MovementWriting, d) The Eshkol-Wachman Movement Notation and e) Labanotation.

The *DanceWriting* was first developed in 1966 by Valerie Sutton and extended to a greater body of work called *MovementWriting*[Sut82]. In this system stick figures are placed on a five-line staff where the movement proceeds from left to right. The lines of the staff represents levels of height (feet, knee, hip and shoulder). Counts coordinating with the written music are placed above the staff and the dance is divided into dance measures. Figure 1.5 c) shows the notation of a Karate kick.

The *Eshkol-Wachman Movement Notation* was developed by the choreographer Noa Eshkol and architect Abraham Wachman [EW58]. It has been used to analyze animal behaviour [Gol76] as well as dance. The system uses a stick figure representation to describe the orientation of the different body limbs. The positions are expressed in a spherical coordinate system using a horizontal and a vertical component. It is often discretized in units of 45° segments that are numbered from 0 to 7. The horizontal and vertical coordinates given by the sphere are written one above the other. Units of time are

represented in grids from left to right, and limbs are written on a different line from top to bottom. The *Eshkol-Wachman Movement Notation* represents a good descriptor for spatial positions and the kinematic chains are not limited to the human body alone. The expressive content of movements are not addressed. Figure 1.5 e) shows the movement of raising an arm that was pointing down, then bringing the forearm close to the shoulder.

Rudolf Laban, (1879-1958) was a notable central European dance artist and theorist, whose work laid the foundations for Laban Movement Analysis. One of his great contributions to dance was his 1928 publication of Kinetographie Laban, which is still used as one of the primary movement notation systems in dance. Laban Movement Analysis (LMA) is a system and language for understanding, observing, describing and notating all forms of movement. Devised by Rudolf Laban, LMA draws on his theories of effort and shape to describe, interpret and document human movement. Used as a tool by dancers, athletes, physical and occupational therapists, it is one of the most widely used systems of human movement analysis. Figure 1.5 d) shows the notation of a Port de Bras. The major components of LMA, *Body*, *Space*, *Effort* and *Shape* will be described in detail in the following chapter. The *Space* component is probably the most important component to distinguish one movement from another. The use of spatial descriptors is a common characteristic for all above mentioned notational systems. What makes the framework of LMA so special is its ability to describe an additional 'expression' that accompanies the spatial trajectory. This might be the key to retrieve some evidences about the emotional state or the intention of the performer. Thus, the *Effort* component can be seen as the key descriptor to solve the task of analyzing 'expressive movements' like the ones shown in the previous section.

Anne Hutchinson-Guest compares thirteen historical and present-day dance notation systems (with visual examples) and through 'one to one' comparisons illustrates the advantages, and disadvantages of each system [Gue89].

1.3 Human-Robot Interaction through movements

Robotics has already acknowledged the evidence that human movements could be an important cue for Human-Robot Interaction. Sato et al. [SNM96], 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 [FND03] 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. Research on social robots also indicates an interest to reproduce expressive movements by a robot [MST$^+$97] and within this scope to utilize also Laban Movement Analysis [NMS02]. It seems to be appealing to move one step further from expressive movements to emotional movements. Attempts have been made to relate body movements to a number of basic emotions [Fri69].

Otero et al. [OKN$^+$06] suggest that the interpretation of a person's motion within its environment can enhance Human-Robot Interaction by i) helping the robot to plan its future tasks and goals, ii) extending the information flow during interaction and iii) supporting additional cues, like speech recognition. They state that body motion and context provide in many situations enough information to derive the person's current activity. Furthermore, that, in order to determine an optimal set of gestures for a given Human-Robot Interaction context, not only the technical limitations but also the user's preference should be taken into account. The set should

be efficient, serve a general purpose and be robust.

Regards for expressive movements can also be found in the adjacent area Human-Computer Interaction where research on expressive agents (avatars) inhabiting virtual environments is conducted. Fabri et al. [FMH02] state that in face-to-face interactions, facial expression, posture and gesture play an important and significant role. They give various references from psychology underlining that: i) More than 65% of the information exchanged during a person-to-person conversation takes place in the non-verbal band. ii) Hundreds of expressive movements are employed every day as part of the social interaction events of a typical day. iii) These non-verbal signals might be even more important than verbal information.

1.4   Interaction with social robots

Traditionally, the term *social robots* was applied to multi-robot systems where the dominant inspiration came from the collective behavior of insects, birds, fish such as flocking, foraging etc. Apparently, the term *social* has changed over the years to become more strongly associated with anthropomorphic (i.e., attribute human or animal-like qualities) social behavior [Bre03].

Autonomous mobile robots will soon assist in a range of tasks that are unpleasant, unsafe, taxing, confusing, low paid, or boring to people. For example, nurses making rounds in assisted living facilities spend much of their time sorting and administering medications. A robotic assistant could do some of this work. More than that, a robot whose task is to sort and dispense medications could interact with an elderly client. It could to sense its clients' state, using indicators such as unusual posture, gestures, or eye movement indicating illness. Autonomous robots like these will need to carry out social and intellectual as well as physical tasks. Ideally, these robots will create a comfortable experience for people, gain their cooperation, encourage healthy rather than overly dependent behavior in clients, customers, and co-workers, and provide appropriate feedback to remote operators and others

involved in the robotic system [KH04].

As robots start acting in human social environments issues of agency, believability and sociality become very important [Dau98]. Humans will expect that robots inhabiting their social spaces will conform as much as possible to their expectations. Sociality can improve the acceptance of tools (like virtual environments) in the first place. It seems to be desirable that artificial agents primarily adapt to human needs and human ways of interaction and living, and not vice versa. For instance, a personal assistant (a robot or a software agent) which works in close contact to a single human user, 'cares' about the user and which shows a humanstyle form of social intelligence is a desirable socially intelligent agent.

1.5 Laban Movement Analysis and believable social robots

In this section we want to present arguments that lead to the conclusion that LMA will be a very useful 'skill' for social robots. For this we will present a number of statements taken from a framework for 'socially intelligent agents' by [Dau98] and relate them to LMA. The statements can be interpreted as the basic elements of an architecture which carries the idea of 'believable social robots' built upon the framework of LMA (see Fig 1.6).

Dautenhahn argues that the interactions need to be 'acceptable' and 'comfortable' to humans, that humans are active agents who want to use their body and explore the environment and that the interfaces should serve the natural human needs of activity. We conclude that: i) *A social robot that is able to perceive human body movements provides a more comfortable interface to the person.*

Dautenhahn states, that *Cognitive Technology* has to understand human perception in order to optimize cognitive fit of technologically constructed tools, that the study of biological life and living can further research on artificial agents and that research on social robots could learn from human
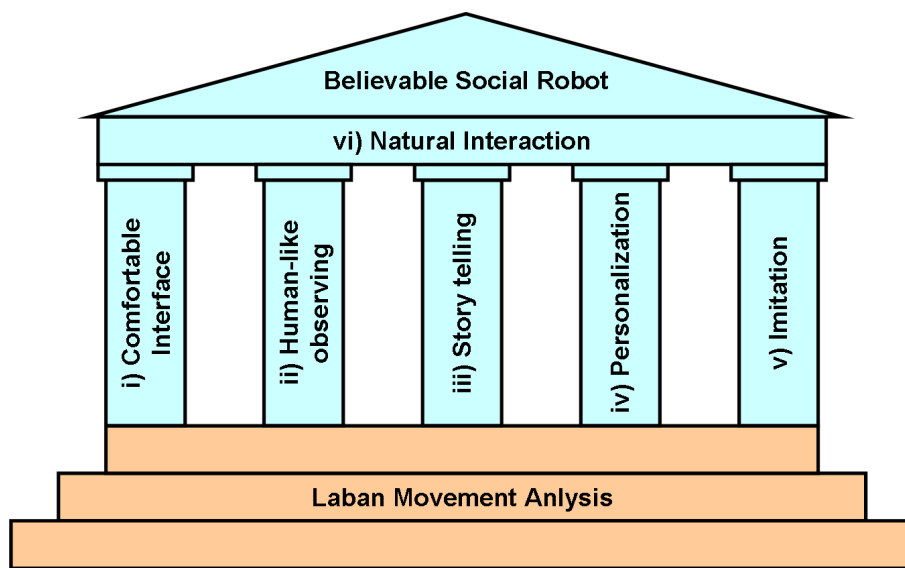
Figure 1.6: Using a foundation that is oriented on human behavior the pillars of having a comfortable interface, human-like observation, story telling and personalization leads to a natural interaction on which a believable social robot is based.

factors (ergonomics) about the study of how humans and machines interact in order to design technology that works well in 'human terms'. We conclude that: ii) *Laban Movement Analysis is natural because it is based on humans observing other humans' movements.*

Dautenhahn suggests that a central set of mechanisms which constitutes 'social intelligence' in humans are 'stories'. She concludes that, if stories are fundamental to human (social) intelligence, then social robots have to be good at telling and listening to stories. In dance movements, the art of choreography can also be interpreted as a an attempt to tell a story and whole choreographies are written by using Labanotation. We conclude that: iii) *Laban Movement Analysis is natural because its symbols can be used to tell a story.*

Dautenhahn argues that social robots should support individualized interactions, when personality, character, individual relationships are desirable (e.g. a personal assistant). A social robot which can adapt to our habits can be more robust through a smaller number of probable hypothesizes. The style of movement is a personal attribute, thus a descriptor of expressive movement is also a descriptor of the person itself. We conclude that: iv) *Laban Movement Analysis is natural because it allows a personalization of the system to the user.*

Dautenhahn states that imitation and social learning mechanisms might make socially intelligent agents more 'like us' and make them individuals. A descriptor for human movement opens the possibility to use it also to reproduce the movement while imitating. In this sense the skill of Laban Movement Analysis will archive personalization through expressiveness. We conclude that: v) *Laban Movement Analysis can support social robots in the task of imitation.*

Dautenhahn states that believable technology is 'familiar' to humans, it meets their cognitive and social, typically human needs. The question of believing in the interaction with a social robot might turn out to be a matter of believing that we can be understood. We believe, for example, in

the keyboard of an Automated Teller Machine (ATM) and that our access code can be understood by the system. We believe in the ATM because it seems 'natural' to us that we can be understood. A system that uses expressive movements as an input modality needs to provide a 'natural' way of understanding them. We finally conclude that: vi) *The skill of Laban Movement Analysis appears to be natural, thus a social robot having this skill will also be believable.*

1.6   Computational Human Movements Analysis: Applications, skills and methods

This thesis is in the research area of 'Computational Human Movement Analysis' (C-HMA). In literature this area has been named 'Visual Analysis of Human Movements' [Gav99], 'Looking at people' [Pen00], 'Human Motion Analysis [AC99]', 'Computer Vision-Based Human Motion Capture' [MG01] or 'Vision-based Human Motion Capture and Analysis' [MHK06]. The cited references refer to surveys only, indicating that this area has already received a lot of attention. Gavrila [Gav99] points out that, the ability to recognize humans and their activities by vision is key for a machine to interact intelligently and effortlessly with a human-inhabited environment.

This and the following two sections introduce different aspects of C-HMA that allow the interrelation with the work presented in this thesis. The organization of this section will start with i) the *application* as the final goal of any research and development in the engineering domain. Then follows ii) the main functionality (*skill*) that is contributed to the *application*. Finally, iii) the *method* that is chosen to implement the *skill*.

1.6.1 Applications

The area of C-HMA has a wide range of applications, thus the literature [MG01], [Gav99] clusters them in more or less the following groups as summarized in Fig. 1.7. *Analysis* are tracking tasks of human body parts with a deep analysis of the spatio-temporal signals for diagnostics. Such tasks might be performed by means of a robot, where the merging of *Assistive Robotics* and *Socially Interactive Robotics* recently led to the definition of *Socially Assistive Robotics* [FSM05]. Assistive robotics solves tasks ranging from physical therapy and daily life assistance to stimulation of emotional expression. The user-group goes from individuals in convalescent care and elderly people to individuals with cognitive disorders. Socially assistive robotics serves the very same purpose but acts through social interaction. One example is stroke rehabilitation with the assistive task of achieving measurable progress in convalescence. A robot capable of social interaction can then be used to repeatedly remind and coach the patient to use the affected limb(s)[FSM05]. In this example, the observation and analysis of the user's movements become a key mode of interaction. Figure 1.7 a) shows a case for post-stroke rehabilitation [MEFSW07].

Surveillance usually deals with long term monitoring and tracking tasks of a certain space or facility. Its main task is to classify actions of one person or groups of people. Figure 1.7 b) shows an image sequence from the European project CAVIAR of the entrance hall at INRIA, Grenoble. Related with this application, but once again aiming toward assistance are *smart rooms*. One of the main arguments for the need of *smart rooms* are the problems related with an increasingly aging society. Robotics offers a possible solution through robots which share a private environment with a person. In [SNM96] Sato et al. propose a 'Robotic Room' as a room which supports a person by means of robotic functions. The robotic elements consist of multiple surrounding sensors and actuators to monitor the behavior of the person and give support. Sato et al. state that the requirements on the functional level are: i) A system being able to use verbal as well as non-verbal communication. ii) A system
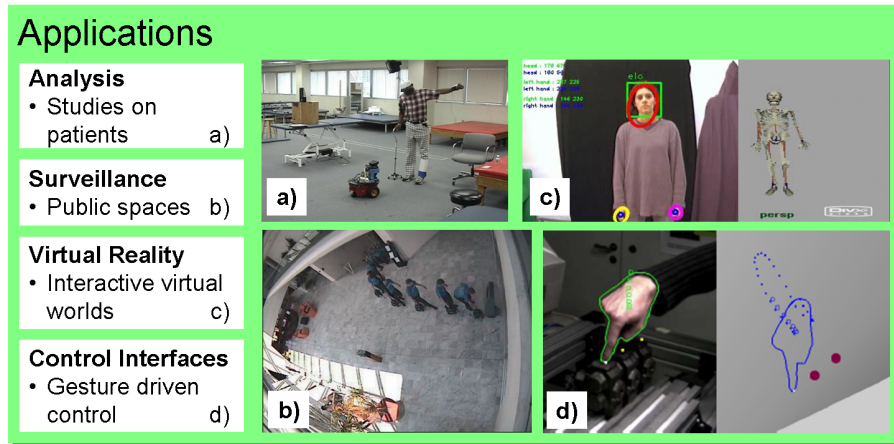
Figure 1.7: Some applications for C-HMA. a) Analysis as in post-stroke rehabilitation (image by Maja Mataric [MEFSW07]). b) Surveillance of public spaces (image sequence from the CAVIAR project). c) Virtual Reality as in animation of avatars. d) Control Iterfaces as for industrial robots (image by C. Eberst [ENUT06]).

that can adapt to the personality of the user. iii) Recognition of behavior in Real-Time. As input modalities on a higher abstraction level they define channels on language, gesture and unconscious behavior. They suggest inputs that are not invasive and unrestrained to not impose any physical burden on the human like CCD cameras, infrared camera and microphones. Also pressure sensors and acceleration sensors are considered. They also point out the importance to preserve historical events (e.g. states of the objects, the humans, and the room) to generate behavior based on past experiences.

Virtual Reality is often based on the animation of avatars. Figure 1.7 c) shows works by two visiting students from Universite Paul Sabatier in Toulouse, France. *Control* are tracking tasks of hands, arms and facial movements, where movements are classified to control an artificial process. Figure 1.7 d) shows an interface for an industrial inspection task [ENUT06].

Figure 1.8: Some skills through C-HMA. a) Human Motion Capture using 3-D tracking (image by R. Urtasun [UF04]). b) Face Recognition presented by two visiting students from Universite Paul Sabatier. c) Gesture Recognition presented by a former member of our team who contributed to the system. d) Laban Movement Analysis presented by a member of our team who contributed to the system.

1.6.2 Skills

For a machine that deals with human movements it is necessary to have certain 'skills'. Which skill is required might depend on the application. Figure 1.8 shows four possible skills. *Human Motion Capture* is the task of tracking several body parts. Figure 1.8 a) shows 3-D tracking for the application of gait analysis [UF04]. *Face Recognition* is a skill that is closely connected to human movements as they are personal characteristics. Figure 1.8 b) shows works by two visiting students from Universite Paul Sabatier. *Gesture Recognition* is the skill to classify a meaningful movement. Figure 1.8 c) shows a former member of our team who contributed to the gesture recognition system. *Laban Movement Analysis* is the novel skill discussed in this thesis. Figure 1.8 d) shows a member of our team who contributed to the system for automatic Laban Movement Analysis.
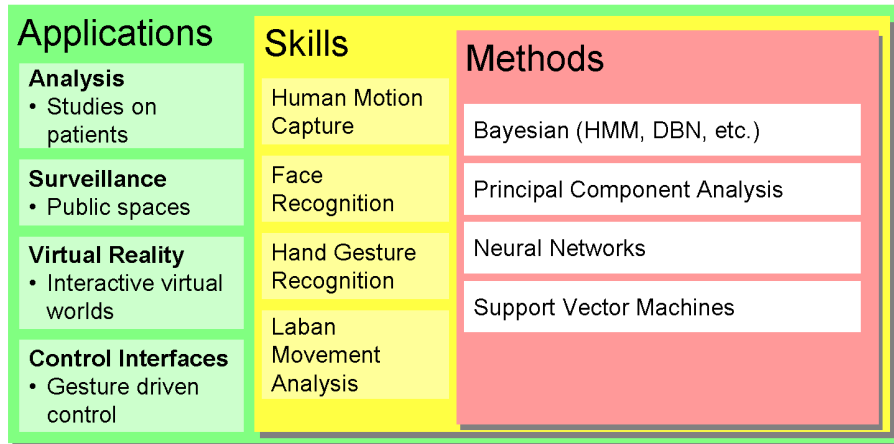
Figure 1.9: Some frequently used methods in C-HMA are Bayesian approaches, Neural Networks (NNs), Principal component analysis (PCA) and Support Vector Machines (SVMs).

1.6.3 Methods

There are certain methods that have been used frequently in C-HMA. Some of them are presented in Fig. 1.9.

Neural networks (NNs) are mathematical models that are based on biological neural networks. It consists of an interconnected group of *neurons* and processes information. In most cases a NN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. What has attracted the most interest in neural networks is exactly this possibility of learning.

Principal component analysis (PCA) is a vector space transform often used to reduce multidimensional data sets to lower dimensions for analysis. PCA was invented in 1901 by Karl Pearson [Pea01]. Now it is mostly used as a tool in exploratory data analysis and for making predictive models. PCA involves the calculation of the eigenvalue decomposition of a data covariance matrix or singular value decomposition of a data matrix, usually after mean centering the data for each attribute.

Support vector machines (SVMs) are a set of related supervised learning

methods used for classification and regression. They belong to a family of generalized linear classifiers. A special property of SVMs is that they simultaneously minimize the empirical classification error and maximize the geometric margin; hence they are also known as maximum margin classifiers.

The purpose of Bayesian inference and learning is to tackle the problem of reasoning in the presence of incompleteness and uncertainty. Moeslund and Granum [MG01] expect in the future more probabilistic models due to the methods' ability to handle uncertainties and to suppress noise. The Bayesian framework offers a great variety of probabilistic tools and has proven successful in building computational theories for perception and sensorimotor control of the human brain [KP04]. Diard et al. [DBM03] predict that in the future probabilistic reasoning will provide a new paradigm for understanding neural mechanisms and the strategies of animal behavior. Further that it will raise the performance of engineering artefacts to a point where they are no longer easily outperformed by the biological examples they are imitating.

In the context of this work Bayesian methods will be used to design the model for automatic LMA, for learning and for classification of movements. Through the Bayesian approach characteristics like anticipation and certainty will emerge and become qualities of the Human-Robot Interaction.

1.7 General scheme for the processes of C-HMA

In the following, the area of 'Computational Human Movement Analysis' will be presented through a scheme of taxonomy suggested by Moeslund and Granum [MG01]. It divides the problem into the four processes *initialization*, *tracking*, *pose estimation* and *recognition* as shown in Fig. 1.10 a). Most of the contributions in this area can be organized by the approaches used for each of the processes [MG01], [MHK06], though the processes might not always show up as distinct objects due to fusion and simplification. This taxonomy can be used to organize the approaches used in this work along those processes as shown in Fig. 1.10 b).
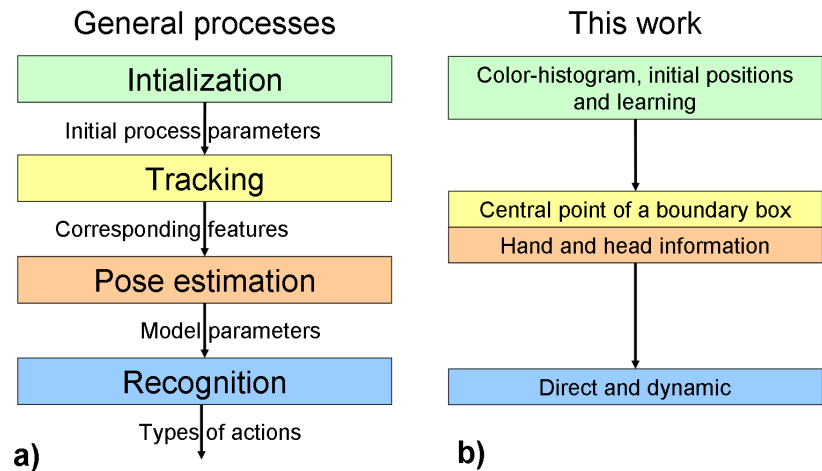
Figure 1.10: a) General processes. During the *initialization* process the initial model parameters are set. The *tracking* process extracts corresponding features. The *pose estimation* process computes the model parameters. Finally an action is classified in the *recognition* process. b) The approaches found in this work using the taxonomy of the general processes.

1.7.1   Initialization

The initialization process establishes the model that is used for the observed object. During the *initialization* process parameters of the prior (offline) calibration process are uploaded to the system or may be updated by actual data. The *Initialization* process represents the knowledge of the system, its representation, its models. It is the process to remember the prior knowledge and compare it to the actual situation.

In the context of this work initialization happens when i) the color-histogram for skin color is generated, ii) as an initial position for the hands and the face and iii) the process of learning the features that represent the movements.

1.7.2   Tracking

The *tracking* process holds the methods for segmentation of the subject in the image from the background, extracting 'good' features that can be used to represent the subject and/or its motion and finally correspondence of the segments of features between consecutive images. The segmentation can be based on a statistical approach, using characteristics such as colors of groups of pixels to extract the figure from the background. The features used for the *tracking* process may already be related to an object, like a point or box, or taken from the pixels of the image directly, like spatial features. To track the features over time and tackle the problem of occlusions predictive methods can be used. A commonly used method for prediction is the Kalman filter, which is also capable of estimating the uncertainties of the prediction. These uncertainties may be used to determine the regions-of-interest. Methods that overcome the problem of unimodal distributions have been proposed to support the tracking multiple hypotheses. One example is the Particle filter method used for tracking of 3-D gestures presented in [MLD06].

The visual tracking used in this work yields the central point of a boundary boxes containing the pixels or regions found in the figure–ground segmentation process.

1.7.3   Pose estimation

The *pose estimation* process defines to which extent a model of the human body is used. Through the model, prior knowledge like anthropomorphic data can be incorporated into the process. Some surveys on C-HMA have based their taxonomy on the human models used [AC99], [Gav99]. Moeslund and Granum [MG01] define three levels of complexity of the model used in the *Pose estimation* process. The *model-free* level still uses a sort of model to represent the pose. The model usually yields only coarse estimations about the subject's hands and head, such as points. In this case, the *Pose estimation* process melts into one with the object oriented *Tracking* process. The

indirect model level uses the model as a reference to constrain and guide the interpretation of measured data. Anthropomorphic data like the hight or dynamic information are used and the model yields positions of the head, hands, and feet or a rough description of the entire human body. In the *direct model* level the precise pose in terms of positions, orientation, width, etc. of each limb is estimated and updated by the data. A common scheme for this process is known as analysis-by-synthesis and is used in a predict–match–update fashion. The state space representing all possible model poses is pruned by some constraints to make the problem tractable.

This work uses a *Pose estimation* process which simply yields information about the subject's hands and head. Thus this approach can be called *model-free* and merges with the *tracking* process.

1.7.4   Recognition

The *recognition* process analyzes some features or variables to classify the movement. The recognition is usually carried out by classifying the captured motion as one of several types of actions. The actions are normally simple, such as walking and running, but more advanced actions such as different ballet dance steps have also been studied. Two paradigms can be found, that relate to the previous step of *Pose estimation*. In the first one, called *Indirect* recognition, the scene is reconstructed building some kind of geometric model, before it is classified. The second one, *Direct* recognition, may be based upon a motion model and classifies directly using low-level data. The approaches used for the *Recognition* process can also grouped in *static* and *dynamic* recognition. *Static* recognition is concerned with spatial data, one frame at a time. The goal of *static* recognition is mainly to recognize various postures, e.g., pointing [KSPF96]. *Dynamic* recognition use temporal characteristics in the recognition task. From the different types of actions, gestures have always received special attention. The results of the gesture recognition are used to control some sort of devices or interfaces based on sign language. The latter is appealing, as many constraints can be applied, which help

to make a vision based solution feasible. The early works already yielded good classification results for sets of 40 signs (words). Surveys specialized on gesture interfaces along the last ten years reflect the development and achievements [PSH97, MN03].

This work contributes mainly to the *recognition* process. Though, the approach uses features derived from LMA it is rather a *direct* recognition as no geometric model is used. The spatial pattern is classified through a *dynamic* recognition.

1.8 Important issues for C-HMA

Some important issues that occur in C-HMA do not explicitly fit into the scheme of general processes. In the following those issues that are of significance for this work will be introduced. First an issue concerning the technical approach for tracking, i.e. the *sensor modalities*, is presented. Then the necessity to pose some *constraints* to the system to make a solution feasible. The special case of using *mobile robots* is introduced afterward. Then the relationship of *moving light displays* to this work is discussed. Finally one main contribution of this work, the automatic generation of labeled data sets is presented.

1.8.1 Sensor modalities

Two main streams of sensor modalities for gathering human movement data can be found. One relies on computer vision and sensor signals generated by some kind of camera, the other is based on devices that are placed on the subject which transmit or receive generated signals. In [MG01] the former is called passive, the latter active sensing. Active sensing allows for simpler processing and is widely used when the applications are situated in well-controlled environments [MG01]. Passive sensing is mainly used in situations where mounting devices on the subject is not an option. Sometimes markers

are attached to the subject to ease the motion capture process based on computer vision. This can be seen as a compromise between active sensing and 'pure' passive sensing [MG01]. Moeslund and Granum [MG01] foresee that advances in active sensors are making them cheaper, smaller, more precise, and generally easier to use. Nevertheless, they will still be cumbersome and limited in their use, therefore, computer vision could provide an attractive touch-free alternative.

Given the benefits and drawbacks of the main streams this work will follow a hybrid approach. The system benefits from the simpler processing and higher precision of the commercial motion capture device ('active sensor' [MG01]) during the recoding step, while using the attractive touch-free alternative of computer vision for the classification step.

1.8.2   Constraints

Some problems are tractable only by establishing constraints concerning 'where' or 'how' the movement is performed. Common constraints are that the subject remains inside the workspace while the movement is performed, that the lighting of the environment is constant or that the start pose is known [MG01]. The complete table of constraints taken from [MG01] can be seen in the Table A.1 of the annex.

Constraints that apply to this work are i) the color of the cloth of the performer (non-skin color), ii) an initial at-ease position, iii) the person stays within the workspace and iv) the camera is static

1.8.3   Mobile robots

Moeslund and Granum [MHK06] point out, that it is difficult to discuss the classifier isolated from the lower level feature extraction. Again, the extracted features depend on the devices that are used which are determined by the application. When applied to medium sized mobile platforms (e.g. social robots) it is preferable to use small and light devices for the sensory input as

well as for the means of computation. For mobile robots it is also desired to have a fast classification result to be able to react accordingly.

The social robot Nicole which was developed along this thesis uses a small (web-cam sized) monocular camera and carries a notebook which performs all necessary computations (e.g. tracking, navigation and classification) online.

1.8.4 Moving Lights Displays

Johansson [Joh73] showed in his moving lights displays (MLD) experiments that the actions of a human may be recognized solely on motion (of the lights). The experiment suggests that it should be sufficient to extract motion features from a 2-D plane to classify movements. Later experiments showed that some transformations affect the classification result significantly. The inverted observation corresponding to a viewing angle of 180° was first found to yield negative responses [Sum84]. Laxmi et. al. [LDC06] showed later that positive results can be obtained, possibly depending on the number and placement of the markers. They also report neutral responses (i.e., deciding by chance) with a view obtained from the top (horizontal plane). Though, the findings have shown that humans can recognize movements based on motion it does not explain the fact that humans may also recognize different postures from a single frame, i.e., without any motion cues [MG01].

In Fig. 1.11 four frames from the performance of 'drinking from a mug' are presented. The left side shows the captured images, the right side the *Moving Light Displays* version. An additional marker (red) was used for the object. Though, representing a MLD-sequence through a still image as in this figure lessens somewhat the impression, one can imagine that the action is recognizable. The findings derived from the MLD experiments supported the concept of 2-D projections used in this work. Furthermore, that our approach is based on the motion of some points (hands and face) rather than limb orientation can be related to MLD. Finally, the experiments presented in thesis regarding different perspectives (Section 5.4)
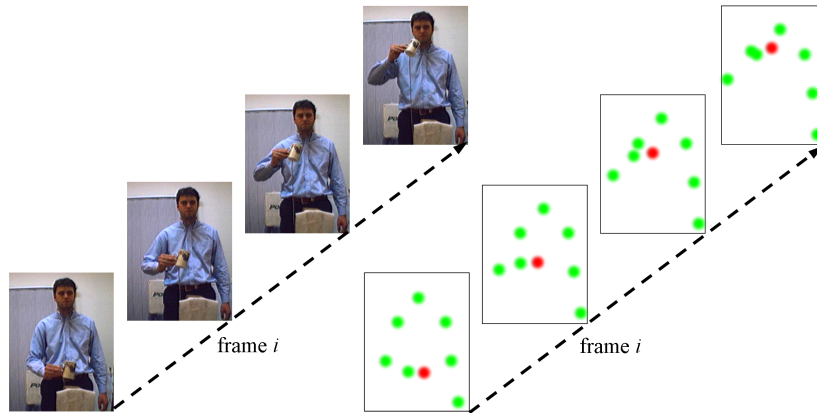
Figure 1.11: Four frames taken from the performance of 'drinking from a mug'. The left side shows the captured images, the right side the *Moving Light Displays* version. An additional marker (red) was used for the object.

1.8.5   Labeled data sets

The research field of computational Human Movement Analysis is lacking a general underlying modeling language [MG01]. A semantic descriptor allows to pose the classification task as a problem to recognize a sequence of symbols taken from an alphabet consisting of motion-entities. Systems which are based on such a modeling language can use it as a ground truth for recoding and labeling training data. The inherent constraints of a modeling language can be used to make the task of movement recognition more tractable.

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.

1.9 Related Works

1.9.1 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 [BPW93]. Zhao and Badler presented in [ZB05] a computational model of gesture acquisition and synthesis which can be used to learn motion qualities from live performance. A more detailed version of their work can be found in the thesis of Zhao [Zha02]. Zhao based his work on the earlier implementation of the 3-D animation control module EMOTE [CCZB00]. EMOTE allowed the manipulation of end-effector positions (*keypoints*) through *Shape* and the movement along the points through *Effort* parameters. As the *Shape* component is defined along the Cartesian axes, it can be assumed that the system has a latent *Space* component. Zhao's main contribution was the automatic classification of *Effort* qualities from movement data obtained from an active sensor as well as from a (stereoscopic) visual tracker. For classification of *Effort* qualities from the low-level features a three-layered feedforward neural network (NN) was used. For each of the four *Effort* qualities one network was trained with movement data from an active sensor. In the first experiment movement data from the active sensor was used. The *Effort* qualities were tested separately, yielding recognition rates of 95.9%, 88.2%, 93.2% and 92.3% for the *Time*, *Weight*, *Flow* and *Space* network respectively. Feeding in data from the visual tracker yielded good recognition rates of 82.4%, 84.6%, 81.8% and 88.5% for the *Time*, *Weight*, *Flow* and *Space* network respectively.

The work of Zhao [Zha02] represents the basic reference for this thesis. Particularly the relationship of LMA components with physically measurable entities was inspired by his thesis. The final application only required a back-projection of the LMA parameters to an animated character. Thus, gesture recognition stayed out of the scope of his thesis. In contrast, movement classification plays a major role in the work presented in this thesis. His thesis

presents the classification results independently for each *Effort* quality. The movement classification presented in this thesis goes one step further and evaluates also the correct combination of the *Effort* qualities. The fact that the classifiers are trained by data from the active sensor, while the classification also accepts data from a visual tracker is similar to the approach presented in this thesis. This thesis goes one step further by learning the 2-D projections, which allows to feed in multiple single-camera data.Zhao concluded that determining *Effort* qualities from a single-camera reliably and consistently needs to be explored further. The concept presented in this thesis closes this gap by following a 2-D approach, though with the option of multiple perspectives. Though, the neural networks yielded good results for classification two problems got apparent: i) For the design of the network, specially the number of layers, one relies on empirical knowledge. It is difficult to discuss the model apart from a black-box behavior which has its focus primarily on the effects (results). In contrast, Bayesian models and their representation as Bayesian nets offer the possibility to discuss the phenomenon in terms of dependencies, observations and probabilities. ii) When feeding in data from the visual tracker the classification results decreased. This was partially due to the noise generated by occlusions and the tracker getting lost. Probabilistic approaches, like the one used in this work, usually perform better with under such circumstances. For the scope of his thesis it was sufficient to use the visual tracker in a controlled environment with a human wearing black cloth in front of black curtain. In this thesis a more natural environment,like the entrance hall of our university, is tackled.

In [NMS02] 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. From the major components of LMA *Effort* and *Shape* were used. The second part of the work was the evaluation of the expressive content of the robot's movements. The robots dance movements (six types) where evaluated by a questionnaire to 21 persons in

terms of type of emotion (e.g. sad). They concluded with a significant correlation between Laban features and the observed impression of basic emotions.

Though, the reproduction of expressive movements based on LMA is out of the scope of this thesis it represents an interesting topic for future research on social robots. Similar is true for the relationship between 'emotions' and LMA. The critical point in [NMS02] is the mapping of low-level features to LMA parameters. The computational model is closely tied to the embodiment of the robot which has only a low number of degrees of freedom. The major physical entities were chosen subjectively by the designer without experimental data for evaluation. Finally, each low-level entity carried a weight factor for adjustment, whose setting was not evaluated. This thesis investigates the framework of LMA as deeply as possible to chose 'good' candidates for low-level features. Furthermore, the low-level features are chosen according to the concept of 'generality', which means that they are not tied to any particular object. This thesis claims to be useful for the classification of a moving human hand as well as for the control of a moving robot.

In [HNT⁺01] a method to describe the movement of humanoids based on Labanotation was proposed. The main contribution is a system which transforms movement data to Labanotation by using the example of a marionette system. Five body parts are used while performing one action in three different ways. A certain level of expressiveness was reported by using different timings for the duration of the symbols.

The work of Hattori et al. [HNT⁺01] is another example that the descriptive language of LMA can also be used for movement analysis of non-humans. Their computational descriptor is designed in the fashion of Labanotation with a 45 degree discretization like the *Space* component of the work presented in this thesis. As the other components of LMA (specially *Effort*) were not regarded it is questionable if their system is able to capture expressive movements.

In [CSJ07] an *Effort* related measure is estimated for 55 dance poses. Constraints for the human kinematics are introduced for the joint angles of

the hip and the foot distance and gravitational forces acting on the arms, legs and torso. The resulting eight-dimensional feature vector is computed from pose data extracted by a commercial motion capture device using 35 markers. The classification algorithm is based on Support Vector Machine (SVM) regression.

The allocation of *Effort* to human poses is difficult as LMA is essentially about the 'movement' itself. Thus, Chen et al. needed to interpret *Effort* in a broader sense and the common descriptors *Space*, *Weight*, *Time* and *Flow* were not used. One important aim of this thesis is to establish a common set of descriptors which allows labeling movement data and comparison of results.

1.9.2   Human Movement Analysis

There has been an interesting work which also used movement descriptors and a probabilistic framework. Bregler [Bre97] 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.

Bregler used a probabilistic method for the same reasons it is used in this thesis: i) noisy input images, ii) spatial and temporal ambiguities, iii) occlusion, iv) cluttered environments and v) large variability. The critical point in his 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 the work presented in this thesis ties the descriptors to a well established notational framework: Laban Movement Analysis.

In [RS00] 3-D data from an active sensor was used to obtain a set of movement sequences. Then 2-D projections from several orientations are generated. For the same orientations, projections of a 3-D model to images are created. Treating the two sets as input-output data a neural network was trained. The problem of data association is not trivial due to ambiguity and was solved by clustering statistically homogeneous data points in the 2-D projected marker space. They achieved good results for training five sequences sampled at 32 orientations.

As their system only provides the pose of a human body, the classification of movements still remained an open issue and consequently no descriptor was introduced. Also, online or real-time behavior was not addressed in their work.

That probabilistic methods can produce very good classification results was also demonstrated for the application of sign language recognition. Starner and Pentland [SP95] 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 probabilistic approach used a Hidden Markov Model (HMM). 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. Starner et al. [SWP98] later removed the constraint of using color gloves and added relative displacement as a feature. The results compared again the accuracy when using grammar rules (91.9%) or when not using them (74.5%). Changing to an egocentric camera position improved the recognition results (97.8% with grammar, 96.8% without grammar) which was due to less occlusions

and effects from body rotations.

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. The problem of personalization was not addressed and it appears that the dataset may be recorded from a single person. Starner first used a five state HMM with three skip transitions [Sta95], later a four state HMM with two skip transitions [SP95] and finally reduced to one skip transition [SWP98]. This reflects a main problem when using HMM: The design of the 'best' topology. The topologies presented in this thesis are designed without explicit temporal transitions. Through this Bayesian version of 'Occams razor' i) the focus is kept on modeling the phenomenon itself, ii) no empirical knowledge is needed like in the case on neural networks and iii) the multidisciplinary discussion of the models is eased as a certain level of generality is kept.

1.9.3   Social robots

Nourbakhsh et al. [NKW03], [NBG$^+$99] developed a family of robot guides serving at the Carnegie Museum of Natural History as docents for five years. All were using the holonomic Nomadic Technologies XR4000 base. Starting with the robot *Chip* (aka Sage) in 1998 which was equipped with a navigation system based on polygons that served as highways and a case-based chain for obstacle avoidance. Burgard et al. presented the autonomous tourguide/tutor robot *RHINO* which was deployed in the 'Deutsches Museum Bonn' in 1997 [BCF$^+$98], [BCF$^+$99]. The robot's control software integrates low-level probabilistic reasoning with high-level problem solving embedded in first order logic. Siegwart et al. [Sea03] presented the mobile robot *Robox* which operated at the Swiss National Exhibition Expo.02. The robot is a fully autonomous mobile platform with multi-modal interaction capabilities, a novel approach to global localization using multiple Gaussian hypotheses, and obstacle avoidance.

The guide robots of [NKW03], [BCF$^+$98] and [Sea03] have already developed quite advanced methods for autonomous navigation and provide various output modalities for the interaction with the human. The fact that their input modalities mainly relies on pressing some buttons shows the need for a more intuitive way of interaction. Burgard et. al [BCF$^+$99] contributes this to a lack of a convincing methodology for 'intuitive' human robot interaction where it is not desirable to expect the person to learn a large variety of control gestures. This thesis suggests LMA as a convincing methodology and expressive gestures as the intuitive modality for interaction.

1.10 The contribution of this work

The contribution of this work for the field of computational Human Movement Analysis encompasses the following points:

i) *This work provides semantic descriptors for the automatic analysis of human movements based on the framework of Laban Movement Analysis (LMA).* A semantic descriptor allows to pose the 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 through LMA. Systems which are based on this modeling language can use it as a ground truth for recoding and automatic labeling of training data. The inherent constraints of this modeling language can be used to make the task of movement recognition more tractable.

ii) *This work shows the design of probabilistic models which relate physically measurable entities obtained from movement tracking to the descriptors of LMA.* Both points i) and ii) use a common language which allows to discuss and incorporate knowledge gained in Human Science. The knowledge gained in Human Science can contribute tremendously to an automatic human movement analysis. In order to apply this knowledge effectively it is crucial that a theoretical 'fit' in description and modeling exists. The main contribution of this thesis is to provide this theoretical 'fit'.

iii) *This work implements classifiers for human movements based on the descriptors of LMA.* By using the descriptors of LMA, specially *Effort*, the process of movement recognition can reach a higher level of sophistication. Instead of only recognizing the sign of a gesture, now the expressiveness of a movement can be understood as well. This might be the key to retrieve some evidences about the emotional state or the intention of the performer.

iv) *This work uses a Bayesian approach for the process of learning and classification.* The Bayesian approach provides the classifier with the ability to better deal with the always apparent incompleteness of the real-world data. Additionally it creates the characteristic of anticipation and a measure of certainty which will pass as emerging qualities to Human-Robot Interaction.

v) *This work applies the movement classifier based on LMA to Human-Robot Interaction.* To show that the new skill of Laban Movement Analysis can survive in the real-world of Human-Robot Interaction, the social robot 'Nicole' was developed. Nicole has the ability to approach a person autonomously and engage him or her in an interaction. During the interaction the movements of the person are analyzed and an action of Nicole is triggered. The interaction scenario has been tested successfully in several public spaces. By giving the 'natural' skill of LMA to the social robot Nicole, she will be 'believable'.

vi) *This work follows hybrid approach for human movement tracking by recording the data with an active sensor and then mapping it to a camera plane.* With this the classification step can be based on features from a single 2-D projection (e.g. monocular camera) as well as on features from multiple 2-D projections (e.g. multi-ocular system). Our system benefits from the simpler processing and higher precision of the 'active sensor' during the recoding step, while using the attractive touch-free alternative of computer vision for the classification step.

1.11 Publications

In order to develop an autonomous platform as the base for the social robot 'Nicole' contributions to the navigational system were presented in [RD04]. A module for obstacle avoidance using a monocular camera was presented. The geometric models were based on projective geometry, log-polar transformation and optical flow.

In the next step the architecture of 'Nicole' was extended with an interaction system [RD05]. Geometric models using projective geometry, the *Space* component of LMA and a feature model inspired by 'Marionettes' were presented. The implementation shows the tracking of human movements from a gesture library using a camera and an active sensor.

The next step mapped actions to these gestures, presented a Dynamic Bayesian Network (DBN) for recognition, refined the architecture of the perception system, proposed the use of the inertial sensor for normalization of the gesture trajectories [RD06].

In the next step an overview of Laban Movement Analysis was given and the *Space*, *Effort* and *Shape* component explained in detail [RD07b]. The concept of *Phases* known from gestures was presented. Low-level features based on local displacement vectors called *Gesture Atoms* were introduced. A Bayesian model to learn and classify gestures using the *Gesture Atoms* was shown. The skill of *Anticipation* was introduced as emerging from online classification. The architecture of the 'Nicole' system was refined. An Interaction scenario with the social robot 'Nicole' was tested and shown. The geometric relation between camera and commercial human motion capture device was shown [RD07a].

A deeper investigation of the two LMA components *Space* and *Effort*, the design of their Bayesian models and the results for single and joint model classification was presented in [RDA08a]. In [RDA08b] the problem of classification using single and multiple 2-D projections for the *Space* component was presented. Following this approach a classifier for multi-ocular systems was presented in [RSD08].
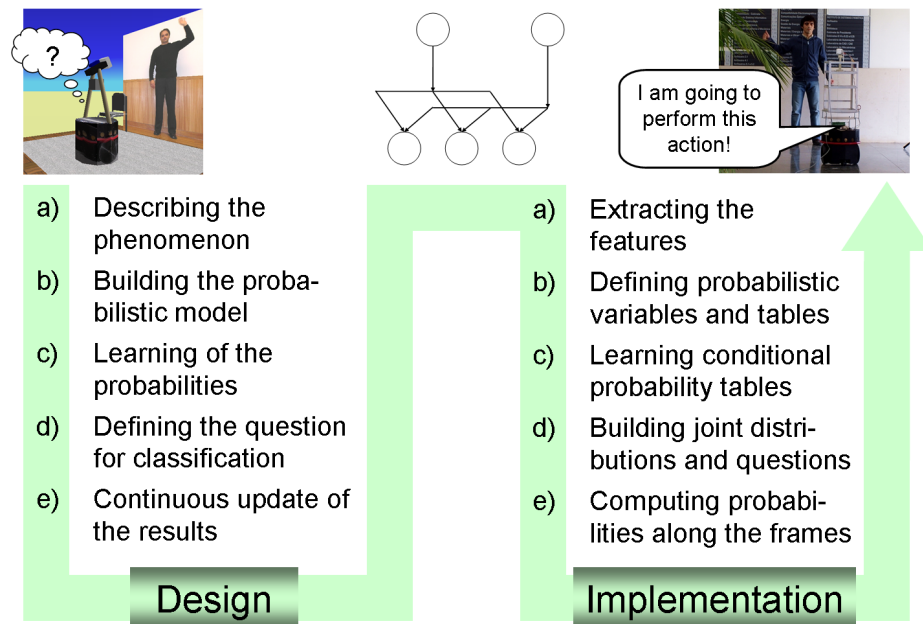
Figure 1.12: The organization of this thesis follows the process of designing and implementing the desired skill of interaction.

1.12   The organization of this thesis

The organization of this thesis follows the process of designing and implementing the desired skill of interaction as shown in Fig. 1.12. Several steps need to be taken from a robot that is not able to interact because he does not understand human movements, over a Bayesian model derived from the design-process to the final prototype: A robot who interacts with a person through classification of movements.

The following Chapter 2 will unfold the framework of Laban Movement Analysis (LMA) as one part of the description of the phenomenon. In this first step of the design-process the five components of LMA will be discussed in detail and prototypical movements will be presented. The concept of Labanotation will be explained.

Human Movement Tracking will be explained in the subsequent Chapter 3 as the second part of the description of the phenomenon. The chapter

presents the technique and models related with computational retrieval of human movement data. It shows the database of human movements, i.e. Human Interaction Database (HID), the process of tracking from camera input, the process of tracking using a commercial human motion capture system and the geometric model that relates the two tracking methods.

Chapter 4 will contain all remaining steps of the design-process. First the reader will be prepared with the Bayesian approach and the *Bayesian program* formalism. Then follows the second step of the design-process while building the models for computational Laban Movement Analysis (C-LMA). Each section will present one type of model. The subsequent section presents the third step of the design-process, i.e. the learning of probabilities. The following section defines the *questions* for classification which is the fourth step. The final step is shown in the subsequent section when the equations for continuous classification of movements are presented. The last section of Chapter 4 will discuss an emerging characteristic of the presented approach, i.e. anticipation and certainty. With this, the theoretical framework is settled and the process of designing is concluded.

Next, the process of implementation will be discussed in Chapter 5 where each step is represented by one section. The last section presents the desired prototype: The social robot, Nicole. Chapter 6 ends this thesis with conclusions of the presented work and developments that are planned for the future.

Chapter 2

Human Movement Analysis

Contents

This chapter starts with the *description of the phenomenon* being the first step (a) in the design process as shown in the global roadmap (see Fig. 2.1).

After a general introduction into Laban Movement Analysis (LMA) this chapter will address four LMA components as shown in Fig. 2.2. Then a special relationship (*Affinities*) between the *Effort* and *Shape* component will be discussed. The descriptive language which is included in the framework of LMA, i.e. *Labanotation* and *Effort Notation* will follow next. Then the database of movements is presented from where the LMA descriptors will be extracted. Two sets hold movements which are specially chosen for the *Effort* and *Space* evaluation. Figure 2.2 also shows where this chapter connects to
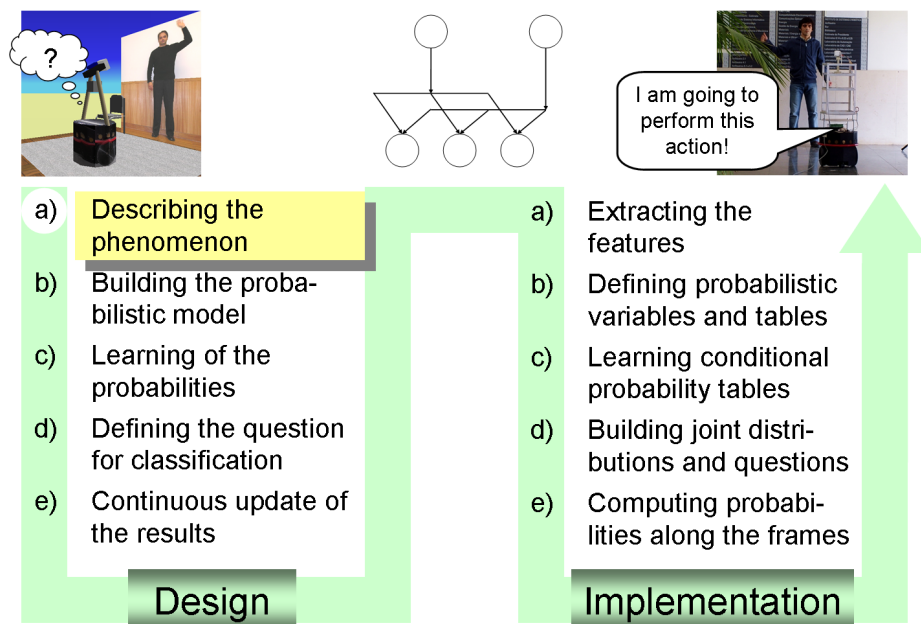
Figure 2.1: This chapter presents step (a) in the design process: The description of the phenomenon from the view-point of Laban Movement Analysis (LMA).
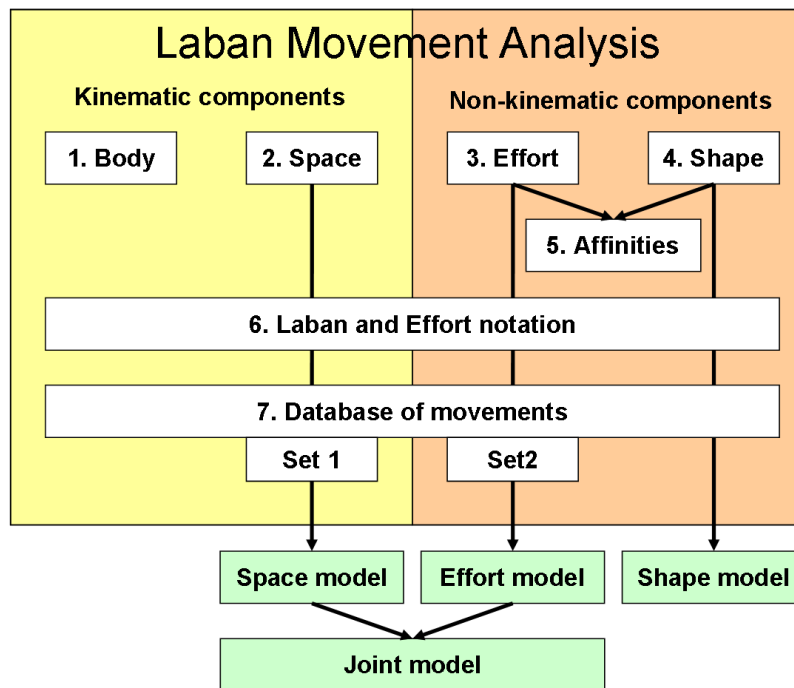
Figure 2.2: Laban Movement Analysis (LMA) addresses four LMA components. *Affinities* are defined between *Effort* and *Shape*. *Labanotation* and *Effort Notation* represent the descriptive language. The database contains the movements for testing.

the following chapters. Three of the components will be used to build models, while two will finally be integrated into a joint model. Thus the results can be compared when using *Space*, *Effort* and both together. The special attention Space and Effort receives is due to the fact that each one represents one group of components: kinematic and non-kinemtic.

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 [Zha02]. The general framework was described in 1980 by Irmgard Bartenieff a scholar of Rudolf Laban in [BL80]. While being widely applied to studies of dance and application to physical and mental therapy [BL80], it has found little application in the engineering domain.

Most notably the group of Norman Badler [CCZB00, Zha02, ZB05], who already started in 1993 to re-formulate *Labanotation* in computational models [BPW93]. Recently 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 [FW06]. 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 [CCZB00, Zha02] mention five major components shown in Fig. 2.3. *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 [FW06] only considers the remaining four major components. *Body* specifies which body parts are moving, their relation to the body center, the
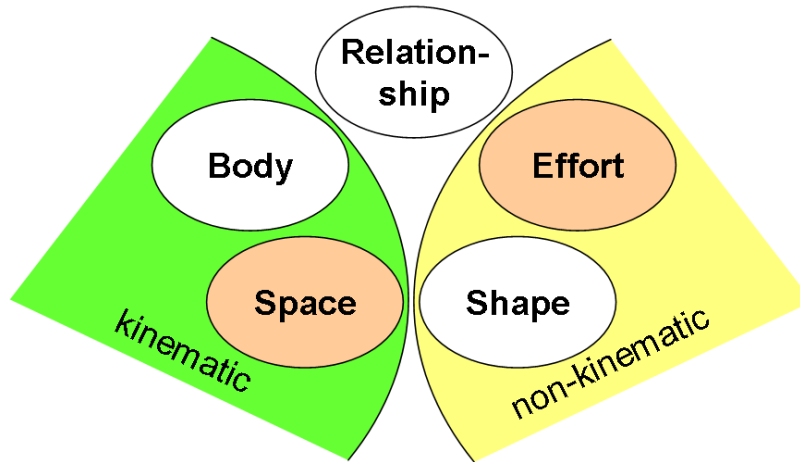
Figure 2.3: The major components of LMA are *Body*, *Space*, *Effort*, *Shape* and *Relationship*. As *Space* represents the kinematic and *Effort* the non-kinematic group, these components will receive special attention.

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 [FW06] 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 Fig. 2.3. The two components *Space* and *Effort* will receive special attention as they represent the two groups (i.e. kinematic and non-kinematic). The *Relationship* component will not be discussed here as not established definitions were found. It can be imagined the parameters like an *interaction distance* and the *focus of attention* will play a role.
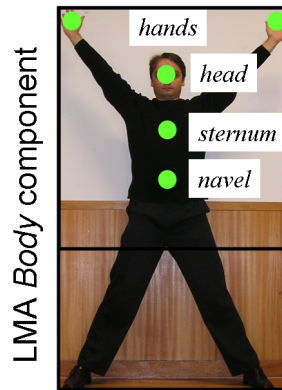
Figure 2.4: The *Body* component defines which of the body parts are moving and how their movement is related to the body center, e.g. navel or *sternum.*

2.1 Body

The *Body* component of LMA deals with the question which of the body parts are moving and how their movement is related to the body center. It also addresses issues concerning locomotion and kinematics. In LMA the kinematic chains are observed with relation to spatial *Shaping* possibilities and the dynamic qualities (*Effort*) accompanying them [BL80]. Laban placed the centre of the egocentric reference system usually at the body center which is approximately at the navel (see Fig. 2.4). Other known locations for the center of the frame of reference are at the sternum, near the belly-button, and in the pelvis [Lon96]. The lower unit (e.g. hip, leg and foot) essentially serves locomotor activity and postural changes. The upper unit (e.g. head, arm and hand) serves essentially exploring, manipulating and gesturing activities [BL80]. Our research is focused on the movement of the hands and the head.

2.2 Space

The *Space* component presents the different concepts to describe the pathways of human movements inside a frame of reference, when 'carving shapes
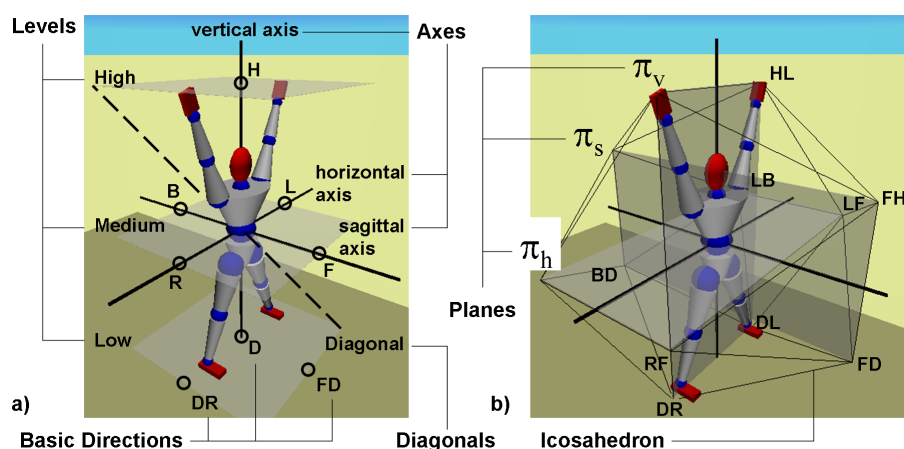
Figure 2.5: The *Space* component defines several concepts: a) Levels of Space, Basic Directions, Three Axes, and b) Three Planes and Icosahedron.

in space' [BL80]. *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 [Lab66] and differ in some aspects from those given in *Labanotation* [Hut70]. The most important ones shown in Fig. 2.5 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*. 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* [BL80].

Labanotation direction symbols encode a position-based concept of space. Recently, Longstaff [Lon01] has translated an earlier concept of Laban which is based on lines of motion rather than points in space into modern-day
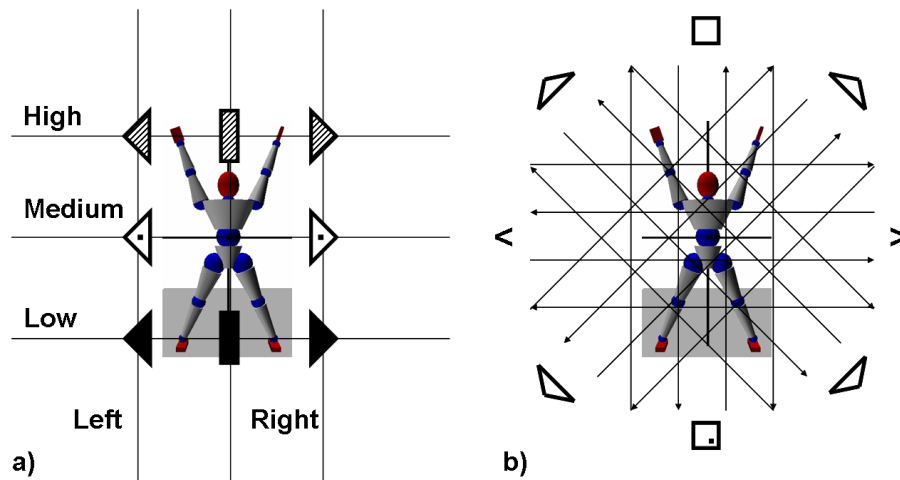
Figure 2.6: Comparison of *Labanotation* and *Vector Symbols* for the *Door Plane* π_v. a) *Labanotation*: The symbols are position-based and refer to the level (height) and left-right. b) *Vector Symbols*: A grid of lines in space describe the principal direction of the movement.

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 Fig. 2.6. The symbols of *Labanotation* correspond to positions in space like *Left-High* while the *Vector Symbols* describe directions. Fig. 2.6 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 Fig. 2.5 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

Table 2.1: *Effort* qualities and their subjects

Effort	Cognitive process	Subject	Extremes
Space	Attention	The spatial orientation	focused or non-focused
Weight	Intention	The impact	strong or light
Time	Decision	The urgency	urgent or non-urgent
Flow	Progression	How to keep going	free or careful

deflecting orientations is conceptualized in a system based on eight *Diagonal Directions*, each deflecting along three possible *Dimensions*.

2.3 Effort

The *Space* component discussed in the previous section is probably the most important component to distinguish one movement from another. The use of spatial descriptors is a common characteristic for all notational systems introduced in Chapter 1. What makes the framework of LMA so special is its ability to describe an additional 'expression' that accompanies the spatial trajectory. This might be the key to retrieve some evidences about the emotional state or the intention of the performer. Thus, the *Effort* component can be seen as the key descriptor to solve the task of analyzing 'expressive movements'.

Effort describes the dynamic qualities of the movement and the inner attitude towards using energy. It consists of four *Effort* qualities: *Space*, *Weight*, *Time*, and *Flow*. Table 2.1 shows the *Effort* qualities, the underlying cognitive process, the subject and the two extremes the qualities can have [BL80]. It can be seen that the *Space* quality requires an awareness, a consciousness to deal with the space either direct or indirect.

Each quality is bipolar and can have values between two extremes. The

Table 2.2: *Basic Effort Action Drives*

Action	Example	Space	Weight	Time
Punch	Forward punch	Direct	Strong	Sudden
Slash	Slash a sword	Indirect	Strong	Sudden
Dab	Lunging for a ball	Direct	Light	Sudden
Flick	Clean with a brush	Indirect	Light	Sudden

values for the *Effort* qualities are shown in (2.1)

$$
\begin{aligned}
Space &\in \{direct, neutral, indirect\} \\
Time &\in \{sudden, neutral, sustained\} \\
Weight &\in \{strong, neutral, light\} \\
Flow &\in \{bound, neutral, free\}
\end{aligned}
\tag{2.1}
$$

Movements are described and distinguished by those qualities that are close to an extreme, e.g. a *Punch* has *Strong Weight*, *Sudden Time* and *Direct Space*. For this movement the *Flow* quality is considered to be *neutral* which means that it lies in between *bound* and *free*. It is often synonymous to describe that no *Flow* quality can be observed or - to simply omit it.

Combinations of all four qualities close to an extreme rarely occur as they produce extreme movements (e.g. tearing something apart) [BL80]. Also single-quality movements are rare [BL80] and even for a trained Laban performer (i.e. Laban notator) difficult to perform [Zha02]. Combinations of three qualities, with the fourth considered to be neutral, appear to be the most natural way to perform an action. With *Flow*, *Space*, *Weight* or *Time* being neutral, the combinations are called *Action Drive*, *Spaceless*, *Timeless* or *Weightless*, respectively. Laban associated some *Basic Effort Actions* to the *Action Drive* which are shown in Table 2.2. The literature on Laban Movement Analysis like [Zha02] often gives some exemplary movements like those shown in the second column . More examples are given in the tables B.1,B.2, B.3, B.5 and B.4 of the annex.

For the remaining three combinations no explicit actions were defined but
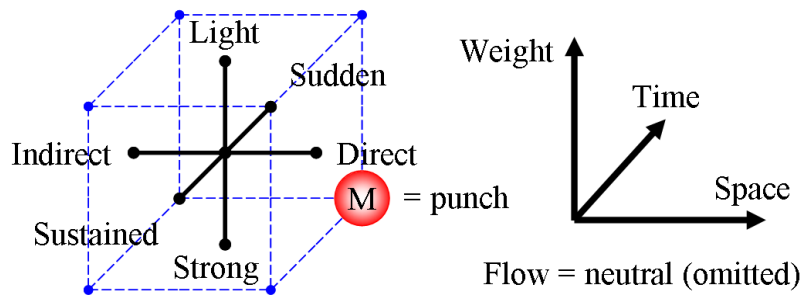
Figure 2.7: The bipolar *Effort* qualities of the *Action Drive*, i.e. *Flow = neutral* (omitted) represented as a cube. The position of the movement *M* (*punch*) indicates its qualities, i.e. *direct*, *sudden* and *strong*.

some examples were given in [BL80]. Using the concept of *Effort* combinations a movement can be defined by its position in one of the four 3-D spaces. The space of the *Action Drive* is often modeled as a cube where each vertex represents an action. The edge length represent the distance between two extremes (e.g. *sudden* and *sustained*). Figure 2.7 shows the space of *Action Drive* with some movement *M*, in this case a *Punch*.

Movements with only two *Effort* qualities are called *Incomplete* or *Inner States* as they occur often during transitions between two three-quality combinations. They can also reflect a failure to produce a certain three-quality action (e.g. an attempt to perform a *Punch* fails due to weakness).

2.4 Shape

In [BL80] *Shape* was not defined as a component of its own but rather a set of qualities that are emerging from the *Body* and *Space* components. Two *Shape* qualities were mentioned particularly: *Shape Flow* describes movements that are focused on the body itself, going towards or away from the body center and using descriptors like shrinking and growing, bulging and hollowing (also including breathing). The term *Spatial Shaping* is used for
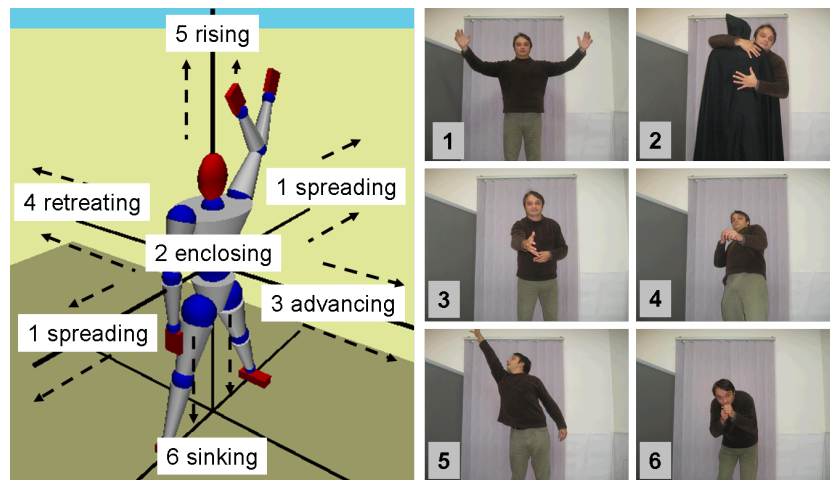
Figure 2.8: The *Shape* component with its spatial qualities and some exemplary movements: 1. *embracing*, 2. *hugging*, 3. *shake*, 4. *retreating* 5. *reaching* and 6. *ducking*.

movements that are going towards a goal in space (e.g. reaching). It is usually described in a Euclidean frame of reference that is aligned with an initial position of the egocentric frame of reference. Due to this, movements can be described by using the vertical, horizontal and sagittal axes and relating them to bipolar descriptors like sinking and rising, enclosing and spreading, and retreating and advancing, respectively as shown in Fig. 2.8. As for the *Effort* component, some exemplary movement can be found in literature [Zha02], also for the *Shape* component. The right side of Fig. 2.8 shows such movements: 1. Opening arms to embrace, 2. Clasping someone in a hug, 3. Reaching forward to shake hands, 4. A boxer retreating to avoid a punch, 5. Reaching for something on a high shelf and 6. A boxer ducking to avoid a punch. The numbers link the movement to the corresponding *Spatial Shaping* quality. Sometimes a third quality is mentioned in literature [Zha02] that is described as carving or moulding when the body with the environment (e.g. moving through a crowd).
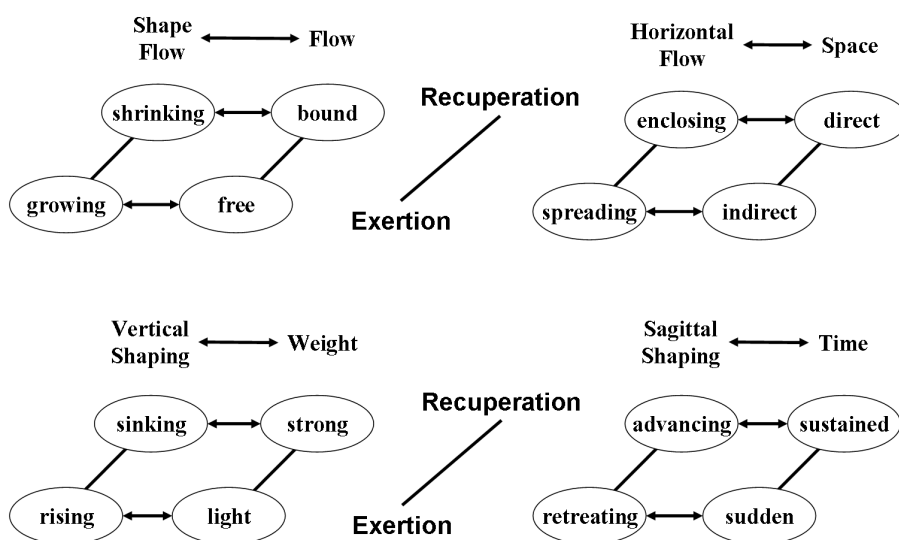
Figure 2.9: Each *Shape* quality has an *Affinity* to an *Effort* quality, e.g. *sinking* with *Weight = strong*. The qualities in the 'back' belong to *Recuperation*, the ones in the front to *Exertion*.

2.5 Affinities

Some *Effort* qualities and *Spatial Shapes* appear so frequently together that they have been defined as *Affinities* [BL80] presented in Fig. 2.9. Affinities describe naturally dominant *Effort* factors given a certain direction in space (e.g. smashing an object with the fist in downward direction naturally exhibits *Weight = strong*, for the performance of a punch upward additional energy must be exerted to overcome the affinity to *Weight = light*. Together with *Affinities* also goes the concept of 'cycles' where phases of *Exertion* are followed by phases of *Recuperation*. Figure 2.9 also presents which qualities belong to which phase.

2.6 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
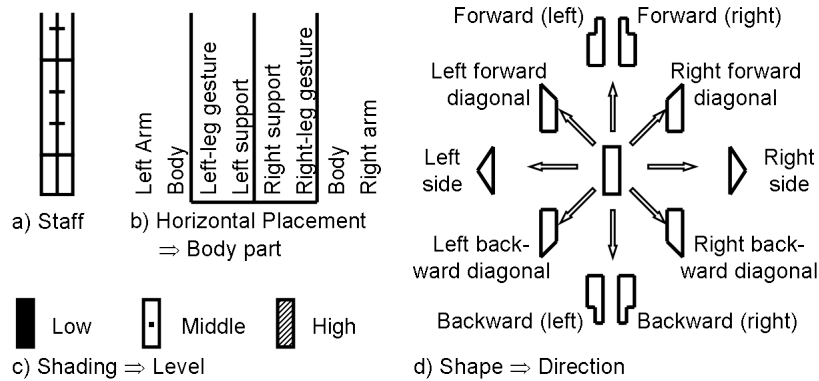
Figure 2.10: *Labanotation*: a) The staff is used to place the symbols. b) The horizontal placement of the symbol indicates the body part. c) Shading of the symbol is used to indicate the *Level* (height) of the 3-D position. d) Different shapes of the symbols indicate the position in the *Table Plane* π_v.

which describe the structure and progression of the movement (shown in Fig. 2.10). The spatial definitions (see [Hut70]) vary from those stated in Choreutics (see [Lab66]). 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. The example in Fig. 2.11 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),
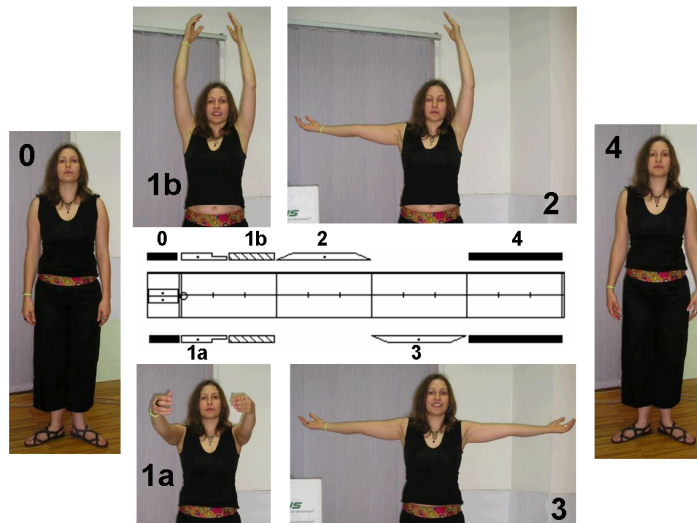
Figure 2.11: Example of a ballet 'Port de Bras' figure. The staff in the center holds the symbols to represent the sequence of positions performed by the actor. Verify with the previous figure: Mainly the left and the right arm symbols are written, the sequence starts and concludes with *Level = low*.

followed by an open side movement (for two counts), followed by lowering the arms.

To complement *Labanotation* with a description of '*how the movement is performed*', the Effort Notation was created. The symbols are based on the signature shown in Fig. 2.12 the small diagonal line is the common element in all symbol instances (see examples surrounding the circle).

2.7 Database of Expressive Movements

From the 32 possible movements with distinct *Effort* qualities that can be derived from the LMA definitions of *Action Drive*, *Spaceless*, *Weightless* and *Timeless* we have created a database of 'expressive movements'. Some of the movements are based on suggestions mentioned in [BL80] and [Zha02] others are commonly used gestures with anticipated *Effort* qualities.

From the database sets of movements were put together. The first set is
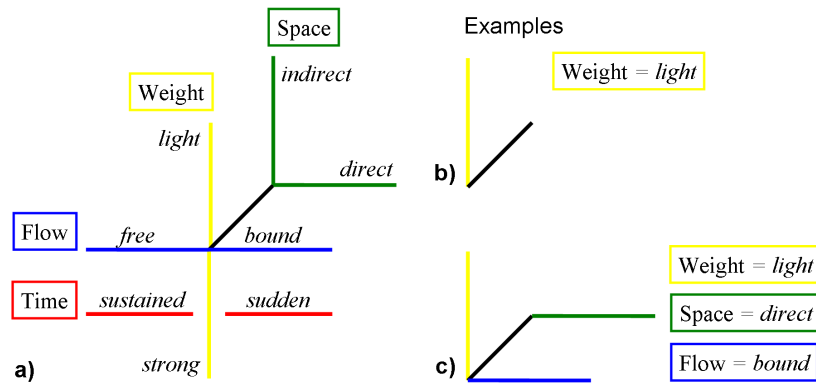
Figure 2.12: Effort Notation: a) Colored version of the *Basic Signature*. Each color represents an *Effort* quality, each branch an extreme. By drawing the observed qualities around the trunk (black line) a movement can be described.

called 'expressive movements' and holds movements which show some interesting spatial patterns. For some the *Effort* qualities are known as can be seen in Table 2.3. The main purpose of this set is to test the relationship between spatial distinctiveness and the chosen 2-D projection. The second set is called 'Bye-Byes' and has four movements where each one has a distinct combination of the *Space* and *Time* qualities. The get a first impression of the characteristics in the following a two by two comparison is presented.

Figure 2.13 shows the two movements *byebye* and *nthrow* from the *set 1* (expressive movements). The *byebye* gesture represents a horizontal waving, while *nthrow* represents a sagittal waving. Their spatial pattern is distinct for a given projection. Both movements have a oscillatory character. 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*. Both movements can loose their oscillatory character when viewed from a certain perspective. This happens with *nthrow* when observed from the Door Plane π_v and with *byebye* when observed from the *Wheel Plane* π_s.

The case of non-oscillatory movements like the *ok* sign and reaching for

Table 2.3: Movements from the database (HID) with *Effort* qualities

Movement	Flow	Space	Weight	Time	Set
ok					
byebye					
shake					1 Ex-
nthrow					pressive
pointing	Neutral	Direct	Strong	Sudden	Move-
lunging	Neutral	Direct	Light	Sudden	ments
stretch	Neutral	Indirect	Strong	Sustained	
maestro	Free	Indirect	Light	Neutral	
byebye_dab	Neutral	Direct	Light	Sudden	
byebye_Flick	Neutral	Indirect	Light	Sudden	2 Bye-
byebye_Glide	Neutral	Direct	Light	Sustained	Byes
byebye_Float	Neutral	Indirect	Light	Sustained	
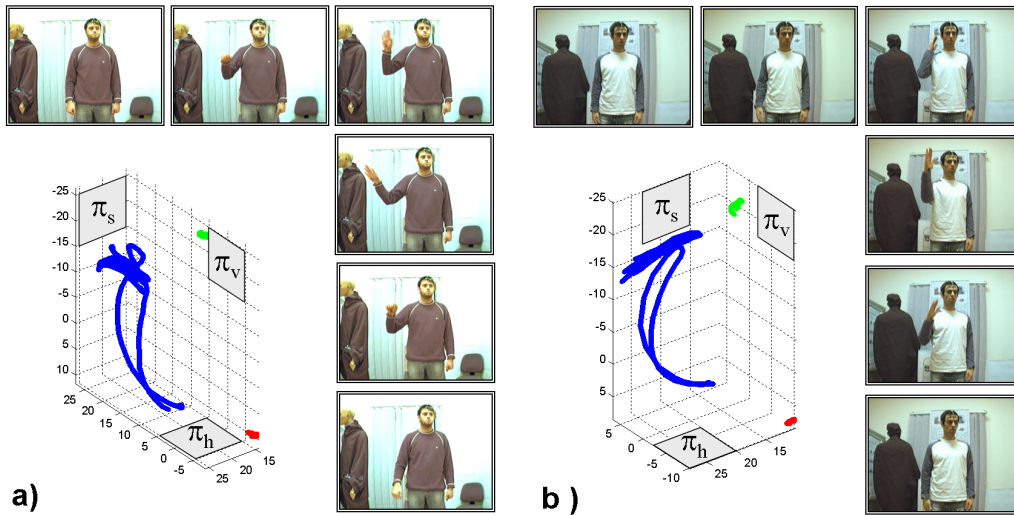
Figure 2.13: Two movements from *set 1* 'expressive movements'. a) Horizontal waving (*byebye*) and b) Sagittal waving (*nthrow*). Both can loose their oscillatory character when viewed from a certain perspective.
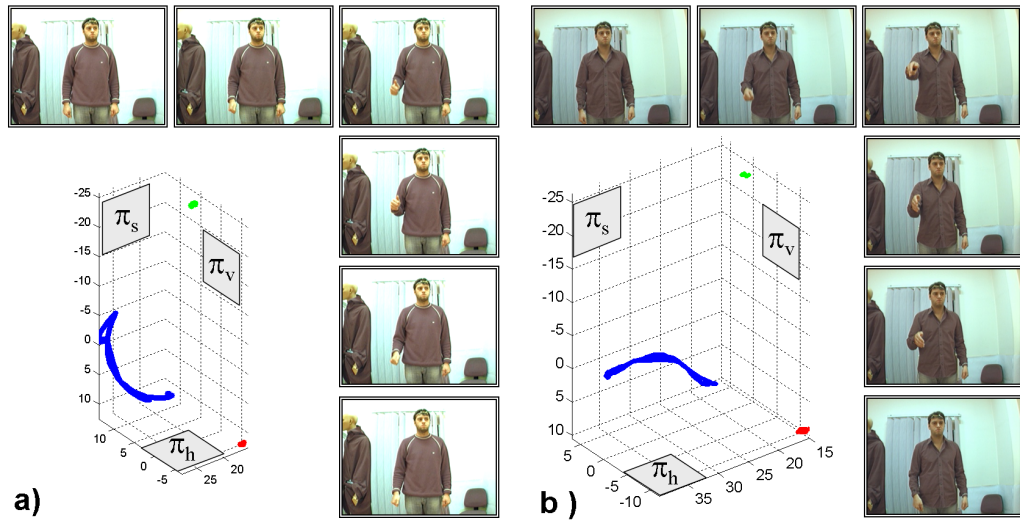
Figure 2.14: Two movements from *set 1* 'expressive movements'. a) Showing the *ok* sign and b) Reaching for someone's hand (*shake*). Both movements appear similar when observed from the *Door Plane* π_v only.

someone's hand (*shake*) can be seen in Fig. 2.14. The two movements are quite similar when observed from the *Door Plane* π_v. A distinction can best be done when observing both movements in the *Wheel Plane* π_s. In this case a larger number of forward (F) and backward (B) Vector Symbols can be observed for *shake*. The shown trajectories present one trial of one person. The whole set of trials can be seen in Fig. C.1, C.2 and C.3 of the annex.

For the two movements *lunging* (lunging for a ball) and *stretch* (stretching to yawn) shown in Fig. 2.15 also the *Effort* qualities are known. Using the terms of *Basic Effort Action Drives lunging* would be described as a *dab* in forward direction. The *Effort* qualities of *stretch* are completely inverse and would be described as a *wring* in the upward direction. Still, the two movements have been mainly involved in spatial experiments as their membership to *set 1* suggests. One thing in common is that these movements are performed with both hands. The distinctiveness is that *lunging* produces mainly *Vector Symbols* like forward (F) and backward (B) while *stretch* produces
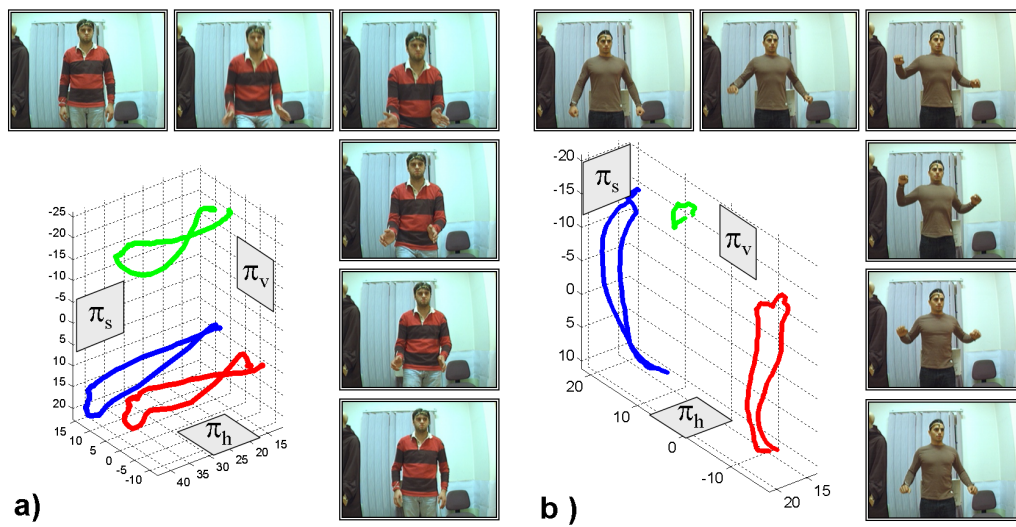
Figure 2.15: Two movements from *set 1* 'expressive movements'. a)Lunging for a ball (*lunging*) and b)Stretching to yawn (*stretch*). The former can be expressed as a forward *dab*, the latter as a upward *wring* in terms of *Basic Effort Action Drives*
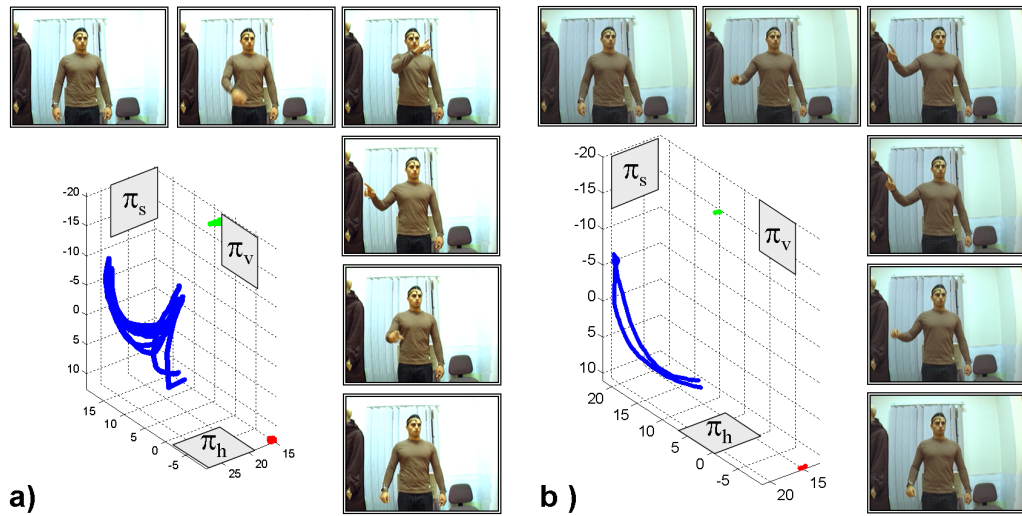
Figure 2.16: Two movements from set 1 'expressive movements'. a)pointing forward with strong impact (*pointing*) and b)Conducting an orchestra (*maestro*). The former can be expressed as a forward *punch* in terms of *Basic Effort Action Drives*.

mainly upward (U) and downward (D).

For the two movements *pointing* (pointing forward with strong impact) and *maestro* (conducting an orchestra) shown in Fig. 2.16 the *Effort* qualities are known. Using the terms of *Basic Effort Action Drives pointing* would be described as a *punch* in forward direction. This is one 'flavour' of performing a pointing - with a strong impact - as if someone sending away. The second movement *maestro* does not belong to the *Basic Effort Action Drives* as it also has a *Flow* quality. The *Free Flow* is represented by the idea of a 'layman' *conducting an orchestra*. Again, the two movements have been mainly involved in spatial experiments. Spatially, they can be confused easily with *byebye* in the case of *maestro* and with *stretch* in the case of *pointing*. This is specially true when observed only from the *Door Plane* π_v.

Tough some of the previously presented movements already have some distinguishable *Effort* qualities a second set of movements ('bye-byes') was
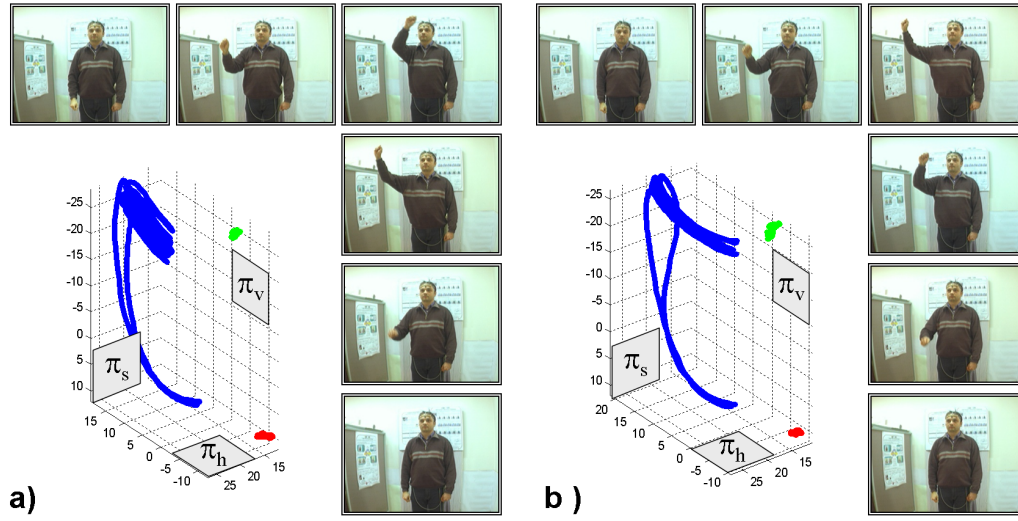
Figure 2.17: Two movements from set 2 ('bye-byes'). a)Bye-bye performed in a 'flick' way and b)Bye-bye performed in a 'glide' way.

put together as shown in Table 2.3. The reason for this was to have spatially one single movement pattern (*byebye*) but four different 'flavours' of performing it. As prototypes four out of the eight *Basic Effort Action Drives* as they are well described in literature [BL80, Zha02]. Consequently the *Flow* quality will be considered as *neutral*. The *Weight* quality was not considered for the experiments as it is difficult to gain evidences from the curve alone [Zha02]. Figure 2.17 show the first two 'bye-byes' one performed as a 'flick', the other as a 'glide'. It can be seen that the two trajectories are quite similar, thus a classification should rely mainly on the *Effort* qualities. Given our set of four movements 'flick' exhibits the inverse qualities (*Space = indirect* and *Time = sudden*) compared to 'glide' (*Space = direct* and *Time = sustained*).

Also the trajectories for 'dab' and 'float' appear similar as can be seen in Fig. 2.18. Also these two movements are inverse to each other as 'dab' has the qualities of *Space = direct* and *Time = sudden* and 'float' the qualities of *Space = indirect* and *Time = sustained*.
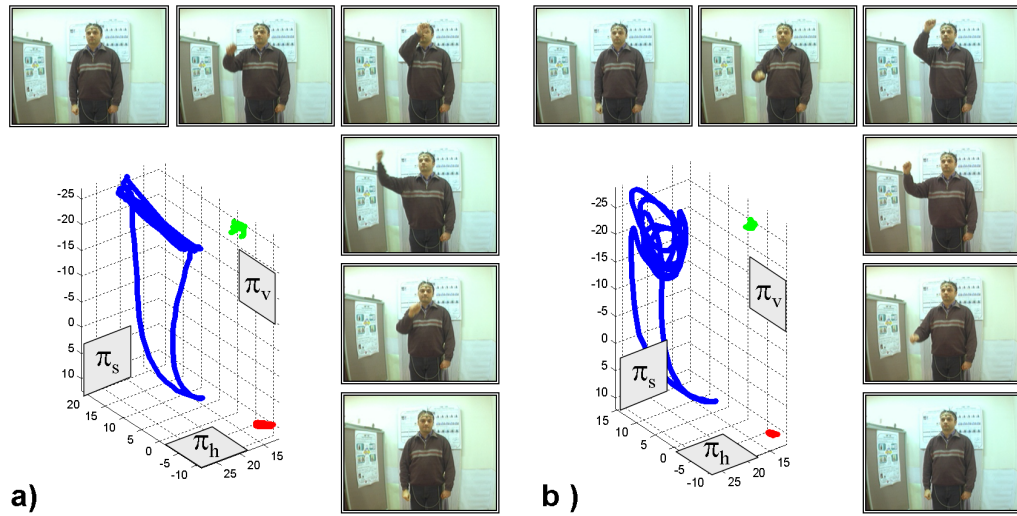
Figure 2.18: Two movements from set 2 ('bye-byes'). a)Bye-bye performed in a 'dab' way and b)Bye-bye performed in a 'float' way.

2.8   Conclusions

This chapter presented a *description of the phenomenon* from the view-point of Laban Movement Analysis (LMA). Four major components have been presented in detail with a special emphasis on *Space* and *Effort*. These two stand for the two groups of kinematic and non-kinematic components. It can be expected that the goal of classifying movements with an expressive content may already be achieved by considering these two components.

Additionally the two components *Body* and *Shape* and the special relationship between *Space* and *Shape*, i.e. *Affinities* were presented. With this, a description of the phenomenon can already be provided for a larger body, though the final classification presented in this thesis will concentrate on *Space* and *Effort*.

The emerging notational system, i.e. *Labanotation* and *Effort Notation* was introduced with an example. This proves that LMA provides an already established descriptive language. The semantics used in this work, though

takes also advantage of less known descriptors like *Vector Symbols*. By using the concept of *Vector Symbols* the descriptor is not bound to a position in space.

The presented database holds movement data, labeled with the descriptors of LMA. Movements can be put together to sets to perform experiments inside the framework of LMA. In this thesis the sets were put together aiming at the *Space* and *Effort* component.

The *Effort* component is the focal point of this work, it is the key to a system which classifies movements with expressive content. The *Effort* qualities are creating a rich descriptor with a theoretical number of $3^4 = 81$ distinctions. In practice the number is lower as some combinations do not, or only rarely occur. In any one action the *Effort* qualities appear in sequence and combination [BL80]. The 'Action Drive' combinations defines a set of 8 prototypical movements which have well defined *Effort* qualities, thus being a candidates for a database of movements. For the other 'Drives' sparse tables can be created using some movements where the *Effort* qualities are known.

It can be expected that, when LMA descriptors are used for movement classification, the phase of the movement needs to be taken into consideration. We can expect distinct LMA qualities for, e.g. when reaching for a cup and when (subsequently) carrying the cup.

As critics the question if the movements are performed 'correctly' in terms of LMA qualities can be posed. Bartenieff [BL80] states that the limitation of observing the *Effort* qualities is the fact, that their distinction requires practice. Furthermore, that sometimes signs are barely visible. The occurrence of a certain quality like *Sudden Time* may obscure the observation of another quality like *Strong Weight*. The occurrence of *Direct Space* and *Sudden Time* might be interpreted as *Strong Weight*. Bartenieff [BL80] suggests to make *Effort* observations always in relation to *Space* and *Shape* while being aware of *Body*. For this thesis, which is seen as a seminal work, this issue was given a lower priority. With some empathy the descriptors of LMA can be understood and with some mastership over the body the desired qualities

can be exhibited. Future contributions to the database may also be given by masters of bodily expressions like dancers or professional Labanotators.

Chapter 3

Human Movement Tracking

Contents

This chapter continues with the *description of the phenomenon* being the first step (a) in the design process as shown in the global roadmap (see Fig. 3.1). Chapter 2 has presented the phenomenon from the descriptive viewpoint of Laban Movement Analysis (LMA). This Chapter 3 Human Movement Tracking will present a more technical point of view. Both chapters will add up to a sufficient description of the phenomenon in order to proceed with the following design-steps. As a further result this chapter will already provide the first step (a) in the implementation process, i.e. extraction of features (see Fig. 3.1.

The main idea transported in this chapter is a hybrid approach concerning the sensor technology to obtain human movement data. By using a commercial motion capture device ('active sensor' [MG01]) during the recoding step the system can benefit from simpler processing and higher precision. For the
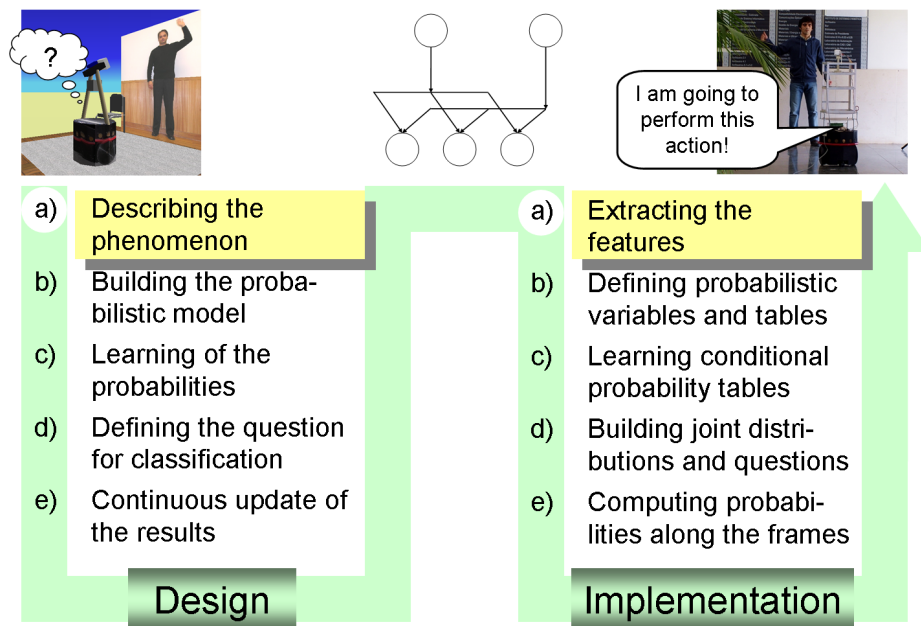
Figure 3.1: Global roadmap: step a) in the design process: Description of the phenomenon and step a) in the implementation process: Extracting the features.
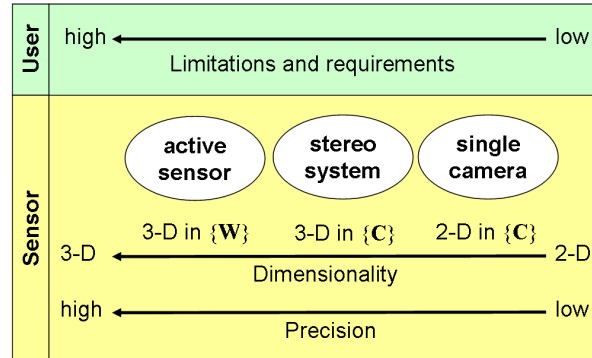
Figure 3.2: Different sensor modalities and their characteristics. Visual tracker: low limitation but also low precision, Active sensor: High precision but also high limitation.

classification step the attractive touch-free alternative of computer vision can be used. The characteristic of the two modalities are also shown in Fig. 3.2.

Once the correspondence between the different sensor modalities is established the recorded and labeled 3-D movement data obtained from the active sensor can be mapped to 2-D plane(s) aligned with the camera plane(s). After this the same algorithm can be used to calculate low-level features and perform a temporal segmentation. As shown in Fig. 3.3 features originated from the active sensor will be used in the following step of learning, while those generated from the visual tracker will be used for classification.

Four important issues can be singled out for this approach and related to the sections of this chapter as shown in Fig. 3.3.

- The sensor technology and modality that is used to obtain data from human movements represented by Section 3.1 and 3.2.

- The geometrical relationship between the sensory data presented in Section 3.3.

- Computable low-level features and their relationship to LMA descriptors shown in Section 3.4.
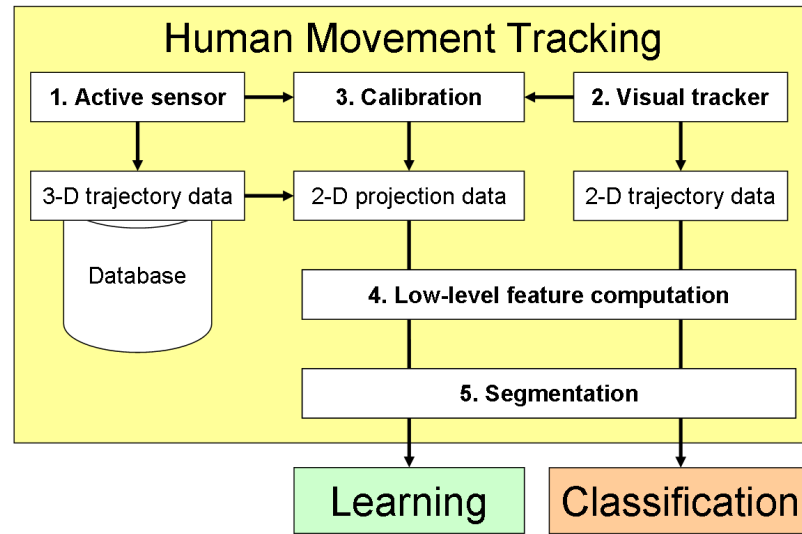
Figure 3.3: Important issues concerning the tracking of human movements: i) Sensor characteristics (1 and 2), ii) relationship of the sensory data (3), iii) 'good' features (4) and temporal segmentation (5).

- A concept for segmentation of movement data taking into account temporal phases introduced in 3.5.

For the first issue the frames of reference, active sensor and visual tracker will be presented. For the second issue the geometrical relationship between the two sensor modalities is shown. The different frames of reference are related through projective geometry and the technical process of calibration is shown. With this any 3-D position can be related to a point in the 2-D projection. The spatial concept of 2-D planes has also been shown earlier in Chapter 2 through the principal planes (emphDoor Plane π_v, *Table plane* π_h, *Wheel Plane* π_s). For the third issue low-level features that can be calculated from the raw position data are introduced. The selection of 'good' features is oriented on both, the concept of 'generality', i.e. features that can describe in general objects in motion and on the concept of 'specificity', i.e. features that are well suited to describe the parameters of LMA. For the fourth issue a general problem of movement classification is presented, i.e. the temporal
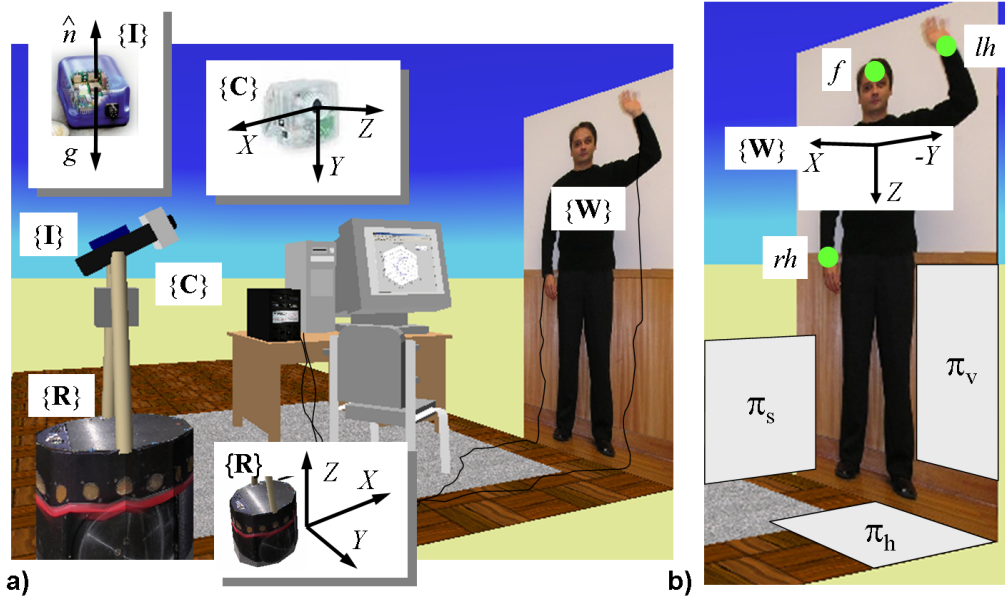
Figure 3.4: Example of a scene for Human-Robot Interaction. a) Frames of reference can be defined for a camera {**C**}, an inertial sensor {**I**}, a robot platform{**R**} and the world {**W**}. b) World frame of reference {**W**} with orientation of axes, principal planes and sensor positions.

segmentation of the sequence into *phases.*

3.1   Active sensing in 3-D

The scenery for an interaction of a human with a robot or, more generally, a machine can be described by frames of reference. Figure 3.4 a) shows an example with a mobile robot equipped with a camera and an inertial sensor to perform online classification of movements. Furthermore, the active sensor for the preceding step of 3-D movement recording can be seen. The world frame of reference {W} might be placed at any position in the 3-D scenery. For the experimental set-up used in this work the world frame of reference {W} coincides with the frame of reference of the active sensor. This is why {W} is close to the center of the *Kinesphere* defined by the hands and the

face. By using the inertial referential $\{I\}$ which allows to register the image data in the vertical and the robot referential $\{R\}$ more complex interaction scenarios can be addressed. The kind of experiments that were conducted did not require the introduction of $\{I\}$ and $\{R\}$. Neither, a human centered frame of reference attached to the human body, like the body center suggested in LMA, was introduced.

Figure 3.4 b) presents a closer look at $\{W\}$ by showing the orientation of the axes and the three principal planes (emphDoor Plane π_v, *Table plane* π_h and *Wheel Plane* π_s) as mentioned in LMA. The figure also indicates the position where the three active sensors are attached, i.e. right hand rh, left hand lh and face f. Though, for the sensors the smallest possible design (teardrop) was chosen the actors can sense their presence, not least due to the attached cables. Attaching the sensors also showed that a certain effort of preparation is required from the user. The six degrees of freedom magnetic tracker (Polhemus Liberty$^{\text{TM}}$) provides 3-D position data with a sufficiently high accuracy and speed (50Hz). The pose sensors work as a receiver of magnetic fields with an own referential attached to it. Taking measurements is the process of reading the sensor's position and orientation relative to the reference coordinate system of the transmitter $\{W\}$. More details can be found in Fig. D.1 of the annex.

Some recordings of the movement trajectories have already been presented in chapter 2. The corresponding database is called Human Interaction Database (HID) and is accessible through WWW [1]. The database consists of image sequences, high precision 3-D position data and results from our visual tracker and classifier. Besides the already presented category of *expressive movements* an earlier set of movements exists named *gestures*. Four examples of the six gestures are shown in Fig. 3.5. These gestures were chosen for a control oriented interaction with the mobile robot platform 'Nicole'. They are robust in terms of distinctiveness when observed from the emph-Door Plane π_v perspective. They are also intuitive when mapped to a certain

---

[1]HID - Human Interaction Database, `http://paloma.isr.uc.pt/hid/`, Institute of Systems and Robotics, Coimbra, PT, 2006 - 2008.
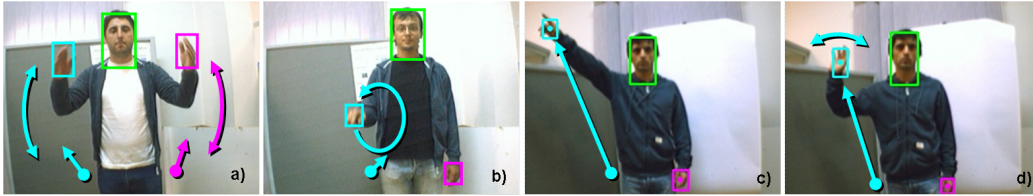
Figure 3.5: Four movement examples from the *gesture* category: a)*come closer* b)*circle* c)*pointing* and d)byebye. The gestures have a high distinctiveness when observed from π_v and allow intuitive mapping to actions.

action of the robot, e.g. performing a *circle* will make the robot rotate 360 degrees. All six movements of the gesture set with some characteristics can be seen in Table C.1 in the annex.

3.2 Visual Tracking in 2-D

For the interaction with the human a visual-based system is very attractive, as it does not require any preparation and the movements can be performed 'freely'. Cameras can be placed on the wall of a room, in front of a display or installed on a mobile robot. This freedom (lunch) does not come for free: Machine vision has a much lower precision, monocular systems produce only 2-D data and tracking is prone to occlusions and lighting conditions. The hybrid approach overcomes these drawbacks for the learning step, though they will remain an issue for the classification process.

To collect the data with 15Hz we use the gesture perception system (GP-System) [RD05] of our social robot Nicole. The system performs skin-color detection and object tracking based on the continuously adaptive mean shift (CAMshift) algorithm presented in [Bra98]. CAMshift extends the mean shift functionality by being adaptive to the position and size of a color object. Thus, providing a solution for continuous tracking of objects.

The images are converted from a RGB to a Hue Saturation Value (HSV) representation to decouple the color itself (hue) from its intensity (saturation) and the brightness. Using the facial area extracted by our Haar-like feature
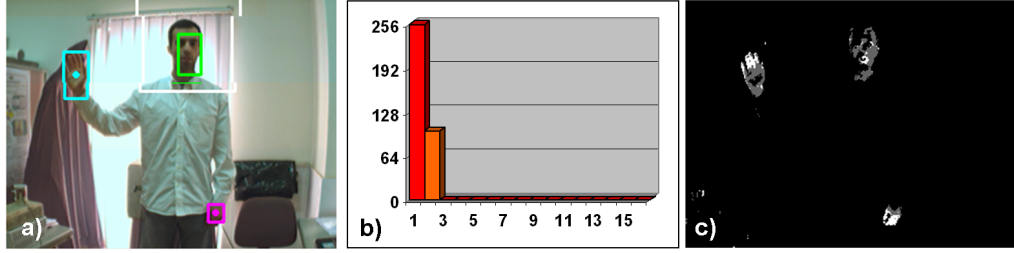
Figure 3.6: Tracking with the CAMshift algorithm. a) The captured image with three tracked color objects marked by rectangles. b) Hue histogram representing the color of the object (skin). c) Probability of the pixels to represent skin color.

detector [VJ01, LM02] we can calculate the actual hue histogram. In the following process of detection and tracking this histogram is used to convert the captured image to a probability image. Each pixel has a probability between 0 and 100% to represent skin color, while pixels with low or high brightness or low saturation are ignored. The mean shift algorithm itself, is based on finding the mean location x_c, y_c of a search window by computing the zeroth M_{00} and first moments M_{10}, M_{01} of the probability distribution within the window.

$$M_{00} = \sum_x \sum_y I(x,y); \qquad M_{10} = \sum_x \sum_y x I(x,y)$$

$$M_{01} = \sum_x \sum_y y I(x,y); \qquad x_c = \frac{M_{10}}{M_{00}}; \qquad y_c = \frac{M_{01}}{M_{00}} \tag{3.1}$$

The search window is centered at x_c, y_c and the computation repeated. Convergence is assumed if the shifting of the search window is below a certain threshold. The CAMshift algorithm will adapt in addition, the search window size as a function of M_{00}. Figure 3.7a) shows the trajectories for both hands taken from the performance of a *lunging for a ball* movement.

The chosen method has the advantage that it comes 'out of the box'. A
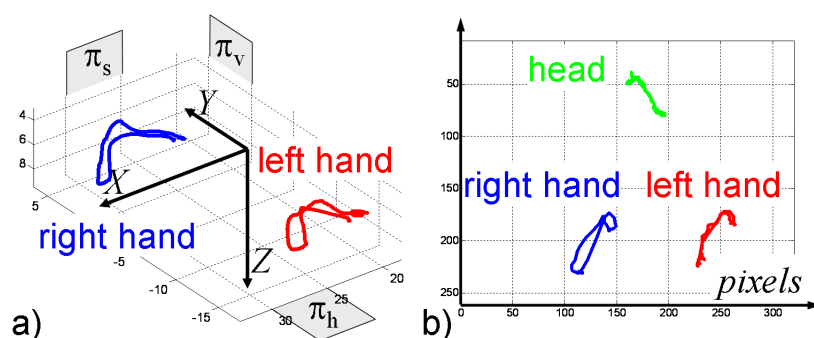
Figure 3.7: Tracking of hands movement for *lunging for a ball*. a) Data from the magnetic tracker b) Data from the vision tracker.
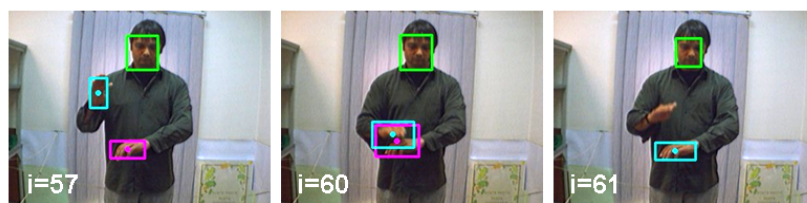
Figure 3.8: Effect of 'gluing' while tracking three skin-colored objects. Frame 57) Both hands are tracked as distinct objects (cyan = right hand; pink = left hand). Frame 60) Hands are close and both bounding boxes grow. Frame 61) Both tracker stay with the hand that was kept still.

system which provides the functionality of tracking skin-color can be implemented rapidly. This comes at the expense of robustness. One particular issue of this method is the effect of 'gluing'. Figure 3.8 shows that in frame 57 both hands are tracked as distinct object. The tracked object 'left hand' is represented by a pink bounding box, while the 'right hand' is represented by a cyan bounding box. In frame 60 the right hand is approaching the left hand. Due to their vicinity both bounding boxes grow in size as the tracked objects now embraces both hands. In frame 61 the right hand is lifted again but the tracker 'sticks' with the left hand. To which of the two objects the tracker will 'stick' depends on circumstances (velocity, size and shape of skin-color distribution, etc.).

Figure 3.9: Effect of fast motion while tracking three skin-colored objects. Frame 48) Both hands are in rest position and tracked (cyan = right hand; pink = left hand). Frame 49-51) Hands are lifted with a fast motion and the tracker gets 'stuck'.

A general problem for trackers is that of fast motion of the tracked objects. It arises when the tracker establishes a correspondence between the position of an object for each frame. The CAMshift algorithm calculates the search window size for a frame based on the position and size of the object in the previous frame. If the object moves to fast it will be outside the search window and thus 'lost'. Figure 3.9 shows the situation where both hands are lifted suddenly from the rest position. The tracker gets 'stuck' in frame 49 at a certain position while the hands are still moving. It can be seen that the left tracker (pink rectangle) is still changing size and position. In this case the tracker was caught by a slight skin-color probability around the silhouette of the person's body.

Another common problem is that of occluded objects, again posing a problem to establish correspondence. When a continuous motion can be assumed predictors like the Kalman filter [Kal60] can be a solution. In our case only some segments of the movement trajectories fall under this assumption. The performance of the tracker was mainly improved based on heuristics for the position, size and the aspect ration of the bounding boxes.
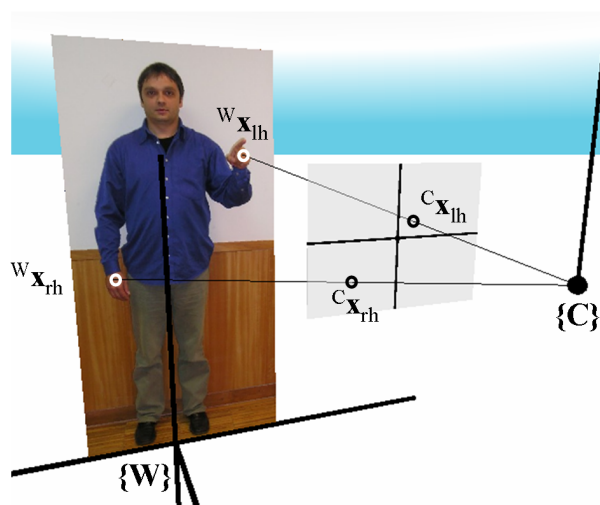
Figure 3.10: Projection of head and hands position in the camera plane.

3.3 Geometric model and calibration

This section will describe the model and process involved in the transformation of 3-D data to 2-D data. 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. Though, the presented model only considers the frame of reference of the world $\{W\}$ and of the camera referential $\{C\}$ it can be extended to include also $\{I\}$, $\{R\}$ and $\{H\}$. As shown in Fig. 3.10 we placed the origin of $\{W\}$ on the ground level aligned with the gravitational vertical and the sagittal axis of the person.

Any generic 3-D point $^{W}\mathbf{X} = [X\ Y\ Z]^{\top}$ and its corresponding projection $^{I}\mathbf{X} = [u\ v]^{\top}$ 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 [HZ00]:

$$\begin{bmatrix} sv \\ su \\ s \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{3.2}$$

Matrix $\mathbf{A}$ 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 $\mathbf{A}$ by using the parameters of the projective finite camera model, as stated in [HZ00].

$$\mathbf{A} = \mathbf{C} \left[{}^{\{C\}}\mathbf{R}_{\{W\}} \ \ {}^{\{C\}}\mathbf{t}_{\{W\}} \right] \tag{3.3}$$

Where $\mathbf{C}$ 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 $\mathbf{R}$ and the translation vector $\mathbf{t}$ that relates the chosen $\{W\}$ to the camera frame $\{C\}$.

The projective camera presents us, in fact, with the solution for the intersection of planes Π_{cam1} and Π_{cam2} which, assuming $\tilde{\mathbf{X}} = [X \ Y \ Z \ 1]^\top$ (i.e. homogeneous coordinates), can be proven from its projection expression to be given by 3.4) (see [Dia94]).

$$\begin{cases} (\mathbf{a}_1 - u\mathbf{a}_3)^\top \, {}^W\mathbf{X} + a_{1,4} - u = 0 \\ (\mathbf{a}_2 - v\mathbf{a}_3)^\top \, {}^W\mathbf{X} + a_{2,4} - v = 0 \end{cases} \iff \begin{cases} \Pi_{cam1}\tilde{\mathbf{X}} = 0 \\ \Pi_{cam2}\tilde{\mathbf{X}} = 0 \end{cases} \tag{3.4}$$

This solution is called the projection or projecting line, which can be alternatively represented by equation (3.5) [Dia94].

$$\mathbf{n} = (\mathbf{a}_1 - u\mathbf{a}_3) \times (\mathbf{a}_2 - v\mathbf{a}_3) \tag{3.5}$$

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 ${}^W\mathbf{X}$ and ${}^C\mathbf{X}$ could only be established through additional
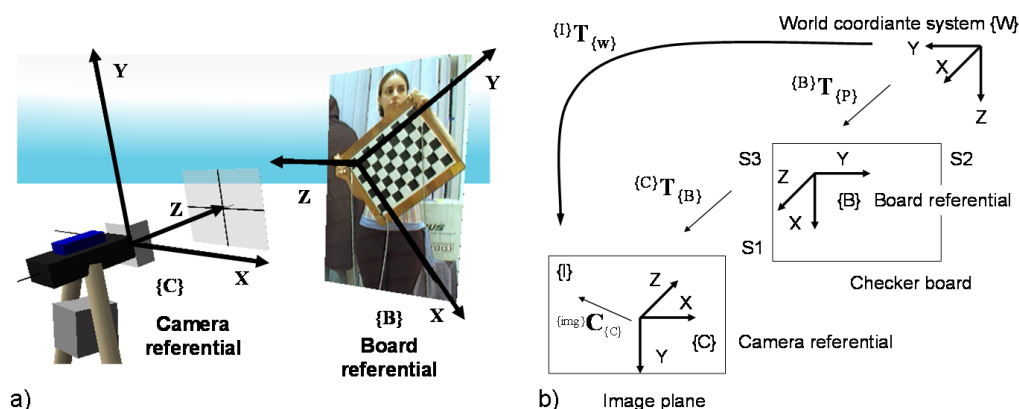
Figure 3.11: a) The calibration process is based on presenting a checker-board target with a frame of reference $\{B\}$ to the camera. b) The process yields the transformation matrix $^{\{I\}}\mathbf{T}_{\{W\}}$ which relates the referential of the world $\{W\}$ and the image $\{I\}$. The process uses the referential of the board $\{B\}$ and the camera $\{C\}$ to obtain intermediate transformations.

constraints, such as the intersection with the surface of a sphere, a plane, etc.

3.3.1   Camera Calibration

Camera calibration consists of the estimation of its intrinsic parameters; using these and knowing the camera's pose, it becomes possible to determine the projection matrix that relates points in a three-dimensional scene and their projections in the image-plane, as was defined in the previous section. With this purpose in mind, a robust camera calibration software package based on Intel's Open Source Computer Vision Library [Int] was used. The OpenCV's calibration method, based on [Zha99] and [HS97], is an iterative algorithm applied to the camera model described earlier. This algorithm uses a sequence of images of a checker-board pattern that supplies 3D points with well-known coordinates as can be seen in Fig. 3.11. During the process of camera calibration a frame of reference of the checker-board $\{B\}$ is established. The resulting extrinsic parameters are the parameters of the

transformation between the two referentials. By attaching sensors of the commercial motion capture device it is also possible to recover the transformation between $\{B\}$ and the world coordinate system $\{W\}$.

For each i-th presentation of a sequence of the checker-board pattern we can establish the relation between a point $^B\tilde{\mathbf{X}} = [X\ Y\ Z\ 1]^\top$ (in homogeneous coordinates) in the board frame of reference $\{B\}$, the corresponding point $^C\tilde{\mathbf{X}}$ in the camera frame of reference $\{C\}$ and the corresponding point $^W\tilde{\mathbf{X}}$ in the world coordinate system $\{W\}$. Where $\mathbf{T}$ represents a homogeneous matrix of an arbitrary transformation. With this we can establish the relation between a point $^W\tilde{\mathbf{X}}$ and the corresponding point $^C\tilde{\mathbf{X}}$ as shown in the following equations:

$$^C\tilde{\mathbf{X}} = {}^{\{C\}}\mathbf{T}_{i\{B\}}\ {}^B\tilde{\mathbf{X}} \tag{3.6}$$

$$^B\tilde{\mathbf{X}} = {}^{\{B\}}\mathbf{T}_{i\{W\}}\ {}^W\tilde{\mathbf{X}} \tag{3.7}$$

$$^C\tilde{\mathbf{X}} = {}^{\{C\}}\mathbf{T}_{i\{B\}}\ {}^{\{B\}}\mathbf{T}_{i\{W\}}\ {}^W\tilde{\mathbf{X}} \tag{3.8}$$

The relationship is also depicted in Fig. 3.11 b).

It is now possible to map the position of each sensor PS1, PS2 and PS3 into the image plane and calculated the error as the difference between the position in the image $^I\mathbf{X}$ and its estimate $^I\hat{\mathbf{X}}$. Figure 3.12 shows the results of the back-projection. The position of the three sensors PS1, PS2 and PS3 can be seen in Fig. 3.12 a) as being on three of the corners of the checker-board pattern. In the magnified Fig. 3.12 b) the vicinity of sensor PS1 can be inspected. The red circle marks a back-projection using the specific transformation matrix of trial i which coincides with position of the central point of the sensor. The blue cross marks a back-projection using the mean transformation matrix.

3.3.2   Raw tracking data

The previously discussed process of calibration allows us to treat movement data from the active sensor and from the camera similarly. Figure 3.13 shows
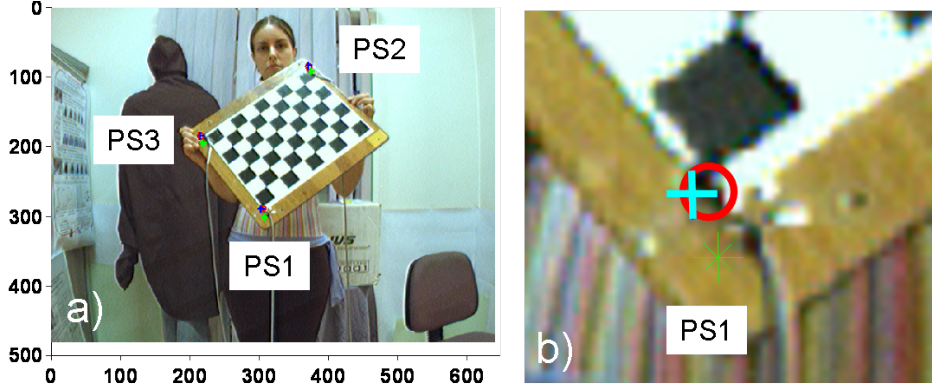
Figure 3.12: a) Back-projecting the position of the three sensors to the image. b) Vicinity of sensor PS1: Red circle: back-projection using the transformation matrix of the specific trial i; Blue cross: back-projection using the mean transformation matrix.

images of the two main sensor units: the active sensor (bottom left) and the monocular camera (top right) with the optionally usable stereo head. Below the images, the trajectories of the sensor data can be found. The primary processes in form of a flow chart are shown in the center of Fig. 3.13. The active sensor collects data from the human body parts while synchronously images are captured from the camera. The position of a part bp with respect to frame ϕ is indicated by $^{\phi}\mathbf{x}_{bp}$.

Frames of reference are the world $\mathbf{W}$ referential and the camera $\mathbf{C}$ referential. Body parts are the face (f), the left (lh) and the right hand (rh). Thus, the active sensor captures position data every i_W-th frame as $^{W}\mathbf{x}_f$, $^{W}\mathbf{x}_{lh}$ and $^{W}\mathbf{x}_{rh}$. The images are captured and processed every i_C-th frame. Our skin color based tracker produces 2-D position data in the camera referential as $^{C}\mathbf{x}_f$, $^{C}\mathbf{x}_{lh}$ and $^{C}\mathbf{x}_{rh}$.

For each frame i_C there is an associated frame i_W which will merge after the low-level feature computation into the common frame i. Through the planar projection we can also align the 3-D position data from the active sensor $^{W}\mathbf{x}_{bp}$ with the 2-D data from the camera $^{C}\mathbf{x}_{bp}$. Thus, the process of ow-level feature computation will use the same module for the data from both sensors.
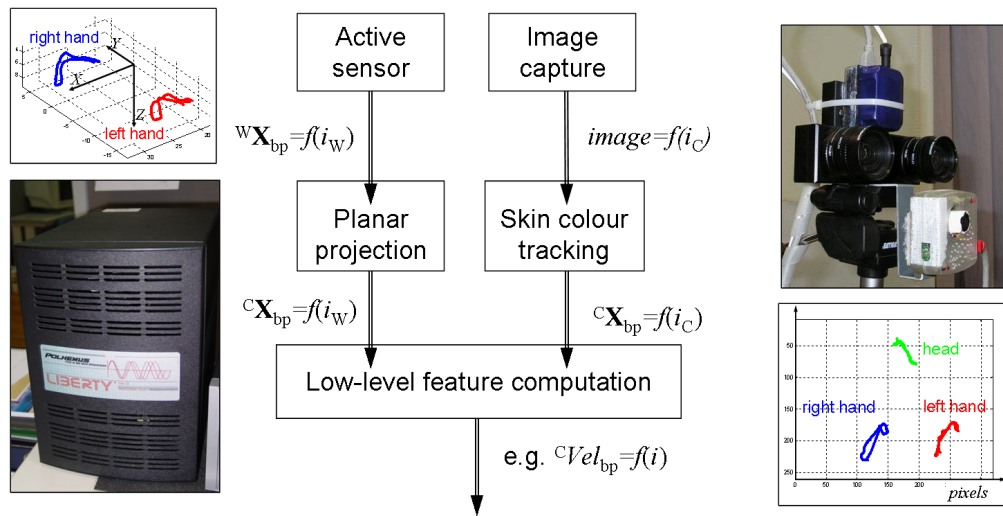
Figure 3.13: Image of the commercial human motion capture device (bottom left) and its sensor data (top left). Image of the monocular camera with stereo head (top right) and its sensor data (bottom right). Flow diagram of the primary processes (center).

Figure 3.14 shows that this concept can be extended to a multi-ocular system. Two separately mounted cameras (e.g. on two robots) capture the 3-D movement from two different camera perspectives $\mathbf{C}_1$ and $\mathbf{C}_2$. Through the transformation matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ the corresponding projections can be generated.

3.4   Low-level features

The selection of 'good' features is one of the great mysteries in pattern recognition. Three paradigms guided the selection of features in this work. Features were chosen

   i) by interpreting the parameters of Laban Movement Analysis (LMA) through physical measurable entities that could describe them best.

   ii) by predicting an optimal performance when using a Bayesian method for learning and classification.
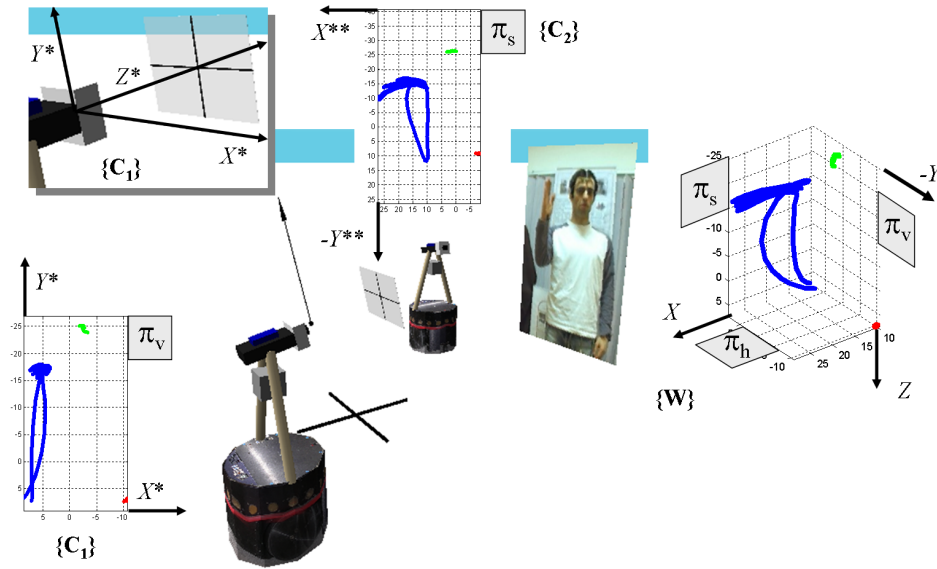
Figure 3.14: Two different camera perspectives $\mathbf{C}_1$ and $\mathbf{C}_2$ can be related to the correct projection of the 3-D data through the transformation matrices $\mathbf{A}_1$ and $\mathbf{A}_2$.

iii) according my interpretation of 'Ockham's Razor', that is simple features, low cardinality and a small number of them.

Correspondences between LMA and low-level features. The initial hypotheses of correspondences between LMA parameters and physical entities are expressed as shown in Table 3.1. The hypotheses were established with having our primary paradigm in mind.

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 [Zha02] 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

Table 3.1: Initial hypotheses of correspondences between LMA parameters and physical entities

LMA parameter	Physical entities
Space	Displacement angle
Time.sudden	High acceleration, High velocity
Time.sustained	Low acceleration, Low velocity
Space.direct	Small curvature, Small angular velocity
Space.indirect	High curvature, High angular velocity
Weight.strong	Muscle tension, Medium acceleration
Weight.light	Muscle relaxed
Flow.free	High curvature, High angular velocity
Flow.bound	Low acceleration, Low velocity
SpatialShaping	Displacement angle
ShapeFlow	Position and shape of hand-head-hand triangle

sampling rate and discretization.

For the *Effort* component of LMA, the assumption of a high acceleration when *Time.sudden* occurs seems to be a logical choice. The high velocity might follow as a consequence of the high acceleration. The inverse situation is assumed during *Time.sustained* when low acceleration and velocity is assumed. Interpreting *Space.direct* as reaching towards a target we can assume a straight trajectory of the hand. This suggests to take the curvature into consideration as a measure of 'directness'. The mathematical definition of curvature though, requires a parametrized curve which is independent of time t. We decided to approximate the curvature k by calculating the change of displacement angles (angular change; angular velocity).

The *Weight* quality is related with muscle tension but the attempt to give a measure for this can not be achieved within the scope of paradigm three. The only relation we draw is that highly tensed muscles can not exert high acceleration. The *Flow* quality is interpreted as the attempt of consciously following a planned trajectory or not. We assume that a *Flow.bound* movement will have a low acceleration and velocity, while in *Flow.free* a high curvature is expected. As it can be seen the mapping is not perfectly one-to-one and for *Weight.light* no (feasible) evidence is given at all.
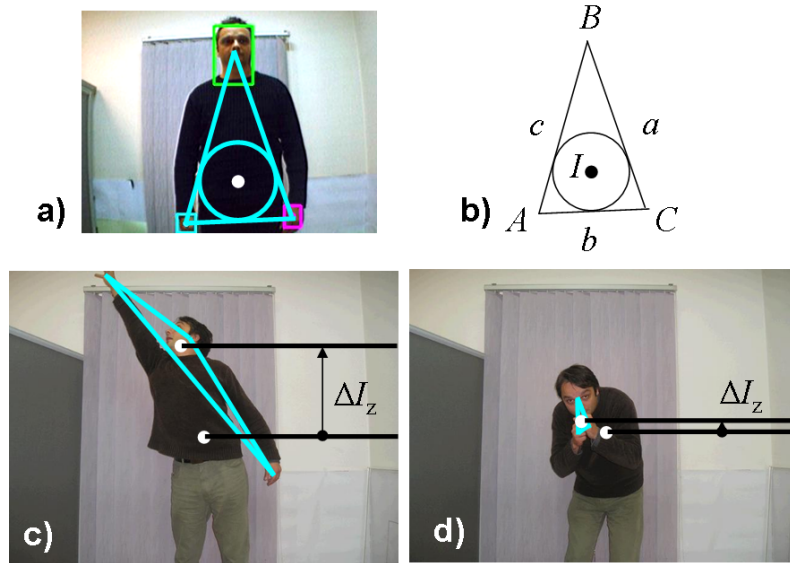
Figure 3.15: The triangle formed by the hand-head-hand positions is used to express *Shape*. a) Triangle of initial position. b) Measures on the initial triangle. c) *Growing*, while reaching for a shelf. d) *Shrinking* while ducking to avoid a punch.

Finally the *Shape* component might also be based on *displacement angles*. Additionally it requires the definition of the signs relative to a defined 'body center' which allows descriptions like *spreading* or *sinking*. As *Spatial Shaping* can be expressed for the three principal planes π_v, π_h and π_s separately, the corresponding *displacement angles* can be used. *Shape Flow* is interpreted as being more focused on the body itself, while going towards or away from the body center (see section 2.4). This behavior is expressed through the geometrical form of a triangle as shown in Fig. 3.15 b). Two measures are calculated one is the vertical position of the *Incenter* I_z of the triangle formed by the hands and the head. Upward displacements relative to the initial *Incenter* are indicating *growing*, while downward displacements indicate *shrinking*. Additionally, the total length of the triangle $l = a + b + c$ is used as an evidence for *shrinking* and *growing*.

Zhao [Zha02] used in his computational LMA the displacement D, the estimated velocity $\hat{v}_i$, the estimated acceleration $\hat{a}_i$ at t_i and the average

velocity and acceleration over a segment. The so called percentage of accelerations and decelerations (PAD) relates to the Number of zero-crossings (ZCR). The ZCR is a feature commonly used in sound analysis where it has shown to be one of the most relevant features when used e.g. in music genre classification [AML04].

Computation of the low-level features. 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 $\mathbf{X}_{bp}$ 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 consecutive 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. As body part bp the right hand rh, the left hand lh and the head (face) f is regarded.

The process of computation from a sequence of data points is show in Fig. 3.16 including some intermediate variables. The variables of the uppers level (light green) are the positions of the tracking data. The index bp indicating the body part is omitted here. The mid-level (light yellow) represents intermediate variables that are calculated for each plane. The final step (light red) is the discretization which yields the low-level feature variables that will be used as probabilistic variables of the Bayesian net. It can be seen that this approach uses only four discrete variables per body part and plane. The four discrete variables, their value space and cardinality are also shown in
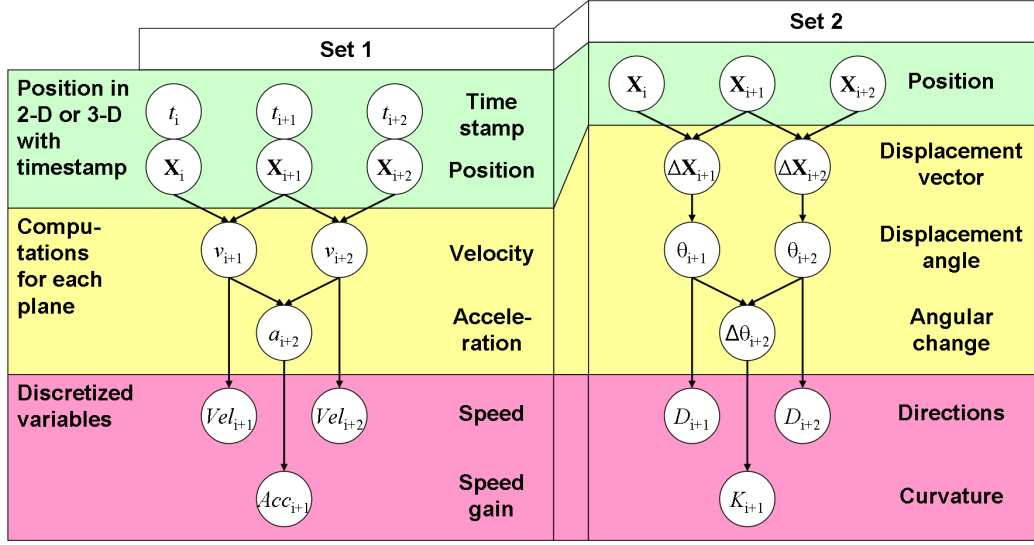
Figure 3.16: Scheme on computation of the low-level features. Set 1 holds variables derived from the magnitude of the displacement. Set 2 holds variables derived from the angle of the displacement.

(3.9).

$$
\begin{aligned}
D &\in \{180, 135, 90, 45, 0, -45, -90, -135\} \ \langle 8 \rangle \\
Vel &\in \{slow, medium, fast\} \ \langle 3 \rangle \\
Acc &\in \{no, low, medium, high\} \ \langle 4 \rangle \\
K &\in \{zero, small, medium, big\} \ \langle 4 \rangle
\end{aligned}
\tag{3.9}
$$

The thresholds that have been used during computation can be found in Table E.1 of the annex.

The computation yields one instance of the variables per body part and plane. The variable direction D undergoes a further conversion into the variable *Vector Symbol*. Table 3.2 shows the conversion for the three principal planes and the two most important body parts (right hand rh and left hand lh). It can be seen that the three directions D_{xy}, D_{yz} and D_{yz} are calculated for the *Table* π_h, *Door* π_v and *Wheel Plane* π_s and converted to the *Vector Symbols* A, B and C, respectively.

Table 3.2: Conversion between *directions* and *Vectors Symbols* for two body parts (right hand *rh* and left hand *lh*) and the tree principal planes.

Variable	Possible instances					
Plane	Table π_h		Door π_v		Wheel π_s	
Directions	D_{xy}^{rh}	D_{xy}^{lh}	D_{yz}^{rh}	D_{yz}^{lh}	D_{xz}^{rh}	D_{xz}^{lh}
Vectors Symbols	A^{rh}	A^{lh}	B^{rh}	B^{lh}	C^{rh}	C^{lh}

The procedure of the conversion is shown in Fig. 3.17 exemplary for the *Door Plane* π_v and the right hand *rh* by means of a *byebye* movement. From the *displacement vector* $\Delta\mathbf{X}$ first the continuous variable *displacement angle* θ and then the discrete variable *direction D* is calculated. Figure 3.17 a) shows the scheme of conversion from the *displacement vector* $\Delta\mathbf{X}$, over the *displacement angle* θ, to the *direction D* and finally to the *Vector Symbol* B_{rh}. In the annex E the figures corresponding to the vectors symbols *A* and *C* can be seen. In Fig. 3.17 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 Fig. 3.17 c). Figure 3.17 shows both representations for the *Vector Symbols*, the signs taken from [Lon01] and the letters used by our algorithm. A summary of all variables involved in the computation of the low-level features are also shown in Table 3.3

Results: Evaluation of low-level features. To test the usefulness of our features as evidences for the *Effort* qualities we conducted the following experiment. Through the tables of exemplary movements for the Effort qualities we can collect sets of movements with a certain *Effort* quality present. As an example all movements that are known to be *Time.sudden* were collected to one set. Then the low-level feature values were computed for each frame (i) and each trial. A suitable representation was to create a histogram of the value space of each low-level variable. Figure 3.18 a) and b) shows the evaluation of the low-level feature curvature when used to describe the *Effort* quality *Space*. It can be seen that the majority of the curvature has value *no* when performing the *Space.direct byebye_ dab*. In the case of a *Space.indirect byebye_flick* performance the majority of the curvature has
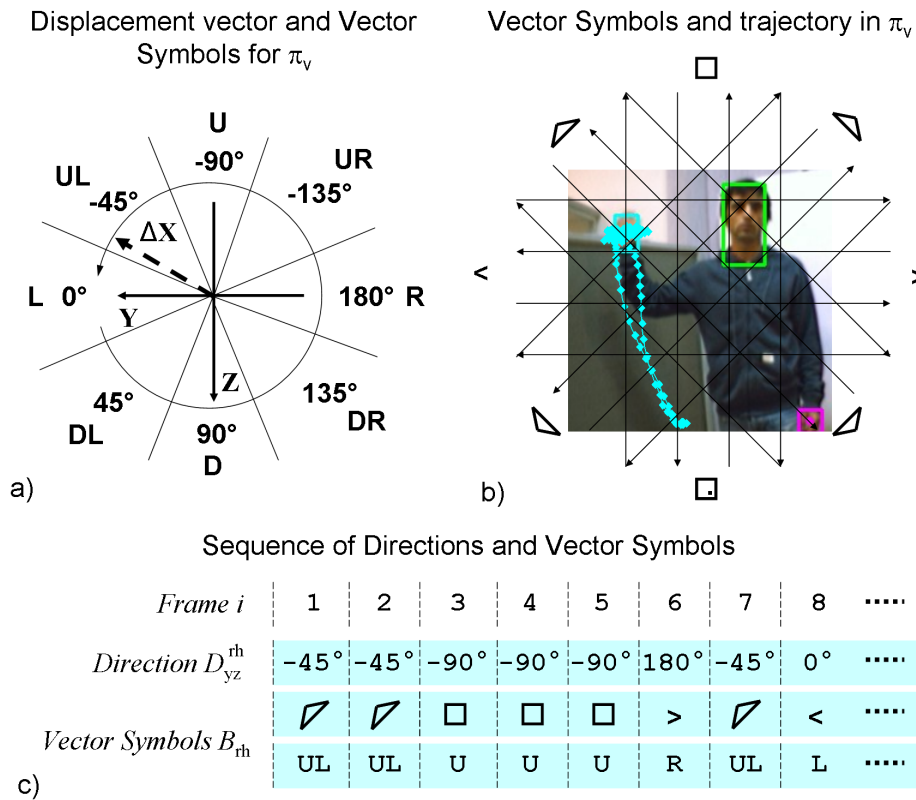
Figure 3.17: *Vector Symbols* for the *Door Plane B* and the right hand *rh* by means of a *byebye* movement. a) The *displacement vector* $\Delta \mathbf{X}$ is converted into the *Vector Symbol* B_{rh}. b) Grid of *Vector Symbols* superimposed on the movement trajectory. c) The continuous computation results in a stream of *Vector Symbols*.

Table 3.3: Description of the variables used for low-level feature computation.

Variable	Symbol	Description
Frame	i	index of the data frame
Position	$\mathbf{X}$	Position in 2 or 3-D
Time	t	time stamp of the frame
Displacement vector	$\Delta\mathbf{X}$	Displacement vector in 2 or 3-D
Direction angle	θ	Planar angle of the displacement
Angular change	$\Delta\theta$	Change of the angle
Norm velocity	v	Velocity of the norm position
Norm acceleration	a	Acceleration of the norm
Direction	D	Direction in 45 deg. interval
Curvature	K	Change of displacement angles
Speed	Vel	Velocity level e.g. *low*
Speed Gain	Acc	Acceleration Level e.g. *high*
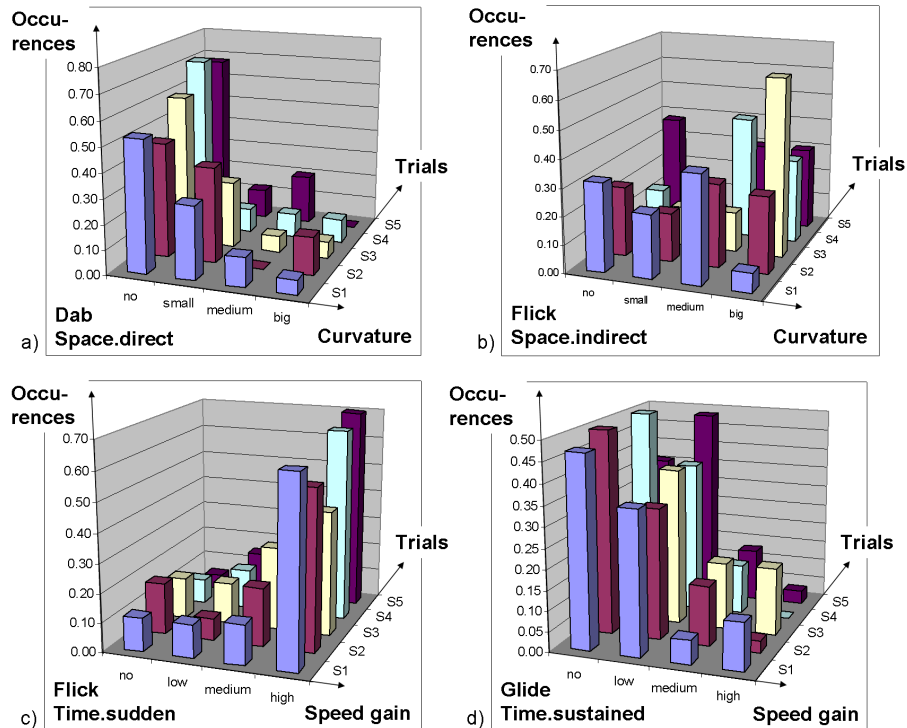
Figure 3.18: Evaluating the usefulness of the low-level features *Curvature* and *Speed gain* for the *Effort* qualities *Space* and *Time* on five trials.
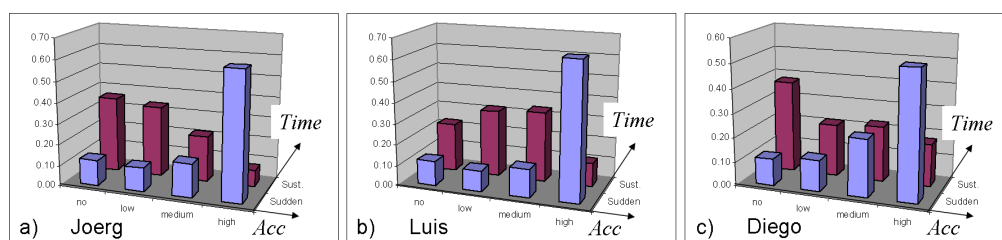
Figure 3.19: Feature histogram for *Speed gain Acc* given by all trials with *Time.sudden* against all trials with *Time.sustained.* Each diagram a), b) and c) represents trials of a single person.

value *big* or *medium.* Another example is shown in Fig. 3.18 c) and d) for the evaluation of *Speed gain* for *Time.* It can be seen that nearly all *Speed gain* values are showing *high* when performing the *Time.sudden byebye_ flick.* In the case of *Time.sustained* represented by *byebye_ glide* the majority is with the *no* or *low* value. The last diagrams gave a good impression about the major tendencies and patterns of each movement. They also showed the similarity and variance between trials of the same movement. Both examples have shown that the set of low-level features that was chosen is useful to describe the *Effort* qualities.

The diagrams of Fig. 3.18 were performed by the same person ('Joerg'). The next experiments investigates the variance between persons. For this, the trial histograms are fused (summed) to a single movement histogram and afterward movement histograms with the same *Effort* quality are created. Figure 3.19 shows the feature histogram for *Speed gain Acc* as a fusion of all trials with *Time.sudden* (blue bars in the front) against all trials with *Time.sustained* (red bars in the back). Each Fig. 3.19 a), b) and c) represents trials of a single person. It can be seen that the patterns for *Time.sudden* are similar for all three persons, while 'Diego' (Fig. 3.19 c)) shows a higher number of *medium* values. The patterns for *Time.sustained* appear different as 'Joerg' trials are monotonically decreasing from *no* values, 'Luis' trials are more 'Gaussian' with a mean between *low* and *medium* values and 'Diego'
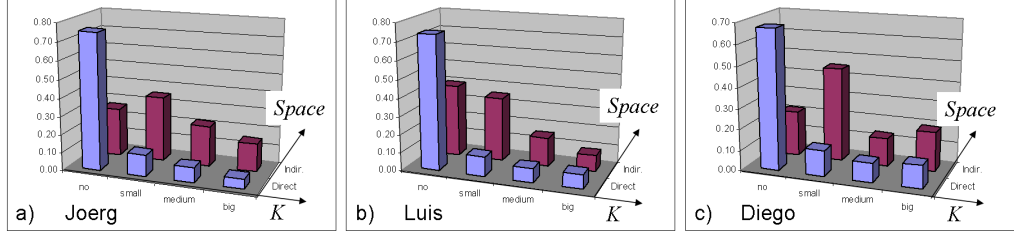
Figure 3.20: Feature histogram for *Curvature K* given by all trials with *Space.direct* against all trials with *Space.indirect*. Each diagram a), b) and c) represents trials of a single person.

trials have a peak at *no* values.

Figure 3.20 shows the feature histogram for curvature K as a fusion of all trials with *Space.direct* (blue bars in the front) against all trials with *Space.indirect* (red bars in the back). It can be seen that the patterns for *Space.direct* are similar for all three persons, without any significant distinction. The patterns for *Space.indirect* appear different as 'Joerg' trials are 'Gaussian' with a mean at *small* values, 'Luis' trials are monotonically decreasing from *no* values and 'Diego' trials have a peak at *small* values. Our conjecture is that some groups of patterns are more 'personal' than others. *Time.sudden* and *Space.direct* seem to have a pattern which is unaffected by the actor. For *Time.sustained* and *Space.indirect* qualitative changes can be observed depending on the performing person. It appears that histograms with 'Dirac'-like distributions are less effected by the variance of the performing person. Histograms with 'Gaussian' or monotonic distributions are in general effected by the actor. Though, we can see that an additional variance will be added when using a multiple person approach, the important distinction between opposite *Effort* qualities like *indirect* and *direct* and *sudden* and *sustained* still holds.

For the final evaluation 'person' histograms are also fused. Figure 3.21 shows the results for all trials, movements and persons for the *Effort* quality *Time*. The bars in the front (blue) show *Time.sudden* while the bars in the
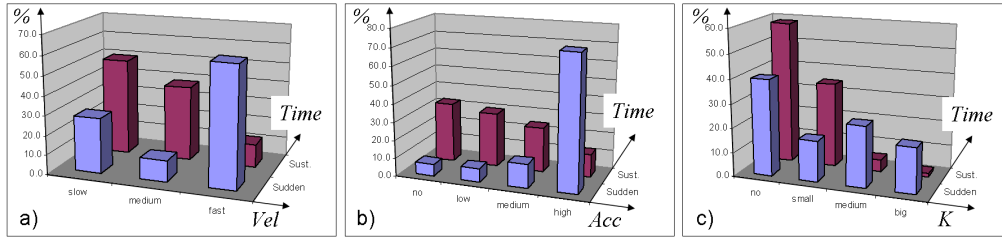
Figure 3.21: Behavior of the low-level features for the *Effort* quality *Time*.
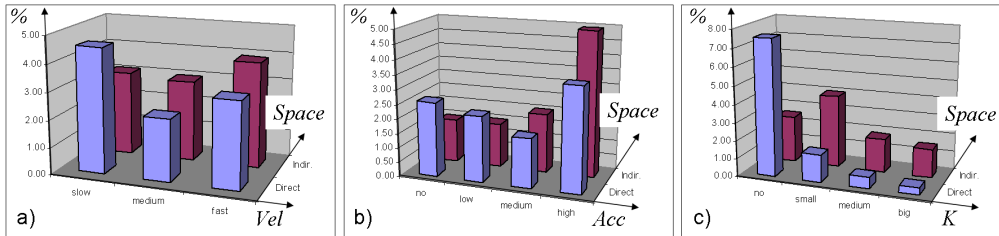
Figure 3.22: Behavior of the low-level features for the *Effort* quality *Space*.

back (red) show *Time.sustained*. For the *Speed* (*Vel*) (see Fig. 3.21 a)) we can see a distinct pattern *Time.sustained* is monotonically decreasing from the *low* values. For *Time.sudden* we can see that the majority of *Speed* is at *fast* values. Our conjecture is that low-level feature *Speed* can be used as an evidence for *Time*. The low level feature *Speed gain* (*Acc*) shown in Fig. 3.21 b) we also see a monotonically decreasing pattern for *Time.sustained* and a monotonically increasing pattern for *Time.sudden*. These patterns are even more distinct for low-level feature *Speed gain* and can also be used as an evidence for *Time*. The patters for *Curvature K* are not distinct as both are monotonically decreasing (see Fig. 3.21 c)), so this evidence can be assumed independent from *Time*.

Figure 3.22 shows the results for all movements and trials for the *Effort* quality *Space*. The bars in the front (blue) show *Space.direct* while the bars in the back (red) show *Space.indirect*. For *Speed* (*Vel*) (see Fig. 3.22 a)) we can see no strongly distinct pattern as the values are more or less equally

distributed, so this evidence can be assumed independent from *Space*. The equally distributed values can also be observed for the *Speed gain* (*Acc*) and *Space.direct* as shown in Fig. 3.22 b). When having *Space.indirect* though, the distribution shows a monotonically increasing behavior towards *high* values. As *indirect* movements often manifest as curved trajectories the linear acceleration changes constantly. Thus, *Speed gain* might be used as an evidence for *Space*. Finally *Curvature K* shows a strong pattern when it comes to *Space.indirect* as nearly all values are *no*. This distinguishes clearly from the 'Gaussian' shaped distribution around the *low* values when it comes to *Space.indirect*. Our conjecture is that the low-level feature *Curvature* can be used as an evidence for *Space*.

In order summarize the important findings of this sub-section it can be said that

- i) the chosen set of low-level features is useful to describe the Effort qualities.

- ii) some groups of patterns are more 'personal' than others.

- iii)low-level feature *Speed* can be used as an evidence for *Time*.

- iv) the low-level feature *Curvature* can be used as an evidence for *Space*.

3.5   Movement segmentation

In order to learn and classify expressive movements correctly, low-level features need to be related to a certain 'phase' of the movement. A person might raise his hand in a normal pace, then wave wildly and finally lower it slowly. A person might be lunging for a ball in a *Time.Sudden* leap forward but then when caught bring it back *Time.Sustained* to the start position. In the domain of gesture analysis an established model is that of dividing the movement into three phases [Ros04]:

- Pre-stroke (preparation)
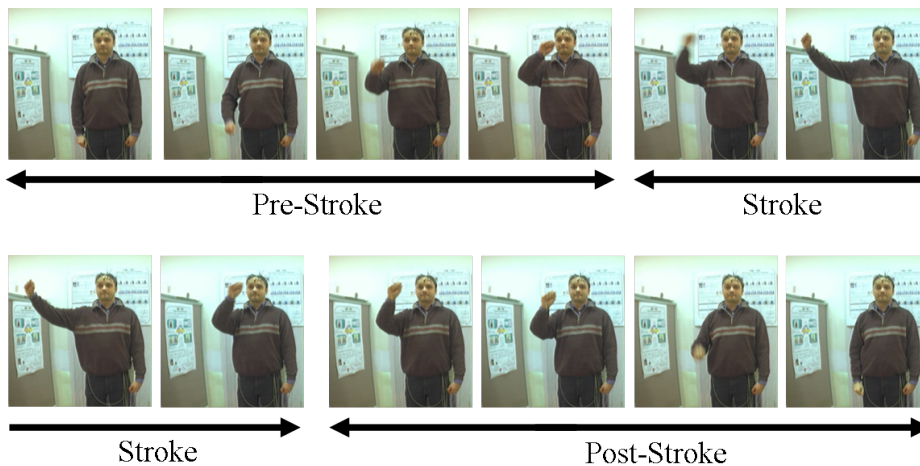
Figure 3.23: Gesture phases: a) Pre-Stroke b) Stroke c) Post-Stroke.

- Stroke

- Post-stroke (retraction)

The first phase is triggered when the hand starts leaving its rest position. The person starts to acquire the hand shape for gesture performance. After the hand reaches the area of meaningful gesture the second phase is triggered. During this phase which may be oscillatory the full hand shape and maximum gestures size can be perceived. In the retraction phase the hand will move again to a rest position.

We adopted this three-phase model for the segmentation of our expressive movements. Figure 3.23 shows three phases of the performance of the *byebye_dab* movement. The first four images belong to the *pre-stroke* phase where the hand moves from the rest position upward and slightly left to position the hand and prepare for the waving. The following four images show the performance of the waving during the *stroke* phase. The final four images belong to the *post-stroke* phase where the hand moves back to the rest position. Gesture recognition systems have often adopted this temporal composition [Sta95, Pav99]. In [KYS02] the phases are called 'phonemes' following the terms used in phonology to describe the principal sounds in
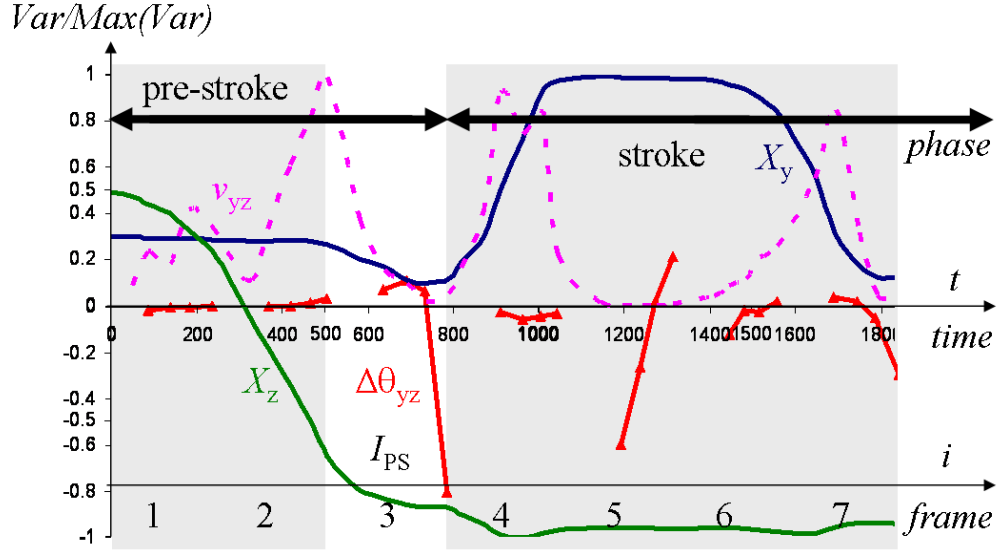
Figure 3.24: signals of the position X_y, the velocity v_{yz} and the angular change $\Delta\theta_{yz}$.

human languages.

From a computational viewpoint the problems is the detection of the frame i in which the movement passes from one phase to the next. In the case of the above mentioned three-phase model we would need to detect two frames of inflection I_{ps} in which the movement passes from (p)re-stroke to (s)troke and I_{sp} where the movement passes from (s)troke to (p)ost-stroke. Here we neglect the problem of detecting the start I_{wp} and the end I_{pr} of a gesture. We constrain the problem in a way that each movement starts from an at-ease position and that the movement concludes with the return of both hands to that position. In order to find a good 'signal' that could trigger the detection of I_{ps} and I_{sp} the variables that have been exposed in the last section are revisited. Figure 3.24 shows the signals of the position X_y and X_z, the velocity v_{yz} and the angular change $\Delta\theta_{yz}$ which related also the first four images of Fig. 3.23. All variables have continues values and are taken from the vertical plane π_v, i.e. YZ. What we are searching for are basically

moments of rest where the transition from one to the next phase happens. Moments of rest are directly represented by zero-velocity values. The dotted pink line shows the velocity signal v_{yz} and we can see its values close to zero is at $t = 736ms$, which is indeed the point of inflection, the frame I_{ps} where the phase changes from *pre-stroke* to *stroke*. The problem with the velocity signal is that one needs to find a appropriate threshold that defines what is zero, that the signal might linger for some time at this value and that we try to measure a signal at its lowest energy. Much better for detection is a high peak that gives a sharp trigger signals. Indeed, through inspection of the angular change $\Delta\theta_{yz}$ represented by a red line in Fig. 3.24 we can see that it produces high values during times of rest. During a movement, even a *Indirect* one, the angular change $\Delta\theta_y z$ hardly gets values greater than $\|75\|$ degrees per sample time. Only during the points of rest values can reach $\|180\|$ degrees. The problem of finding a consistent threshold for all types of movements has also been a matter in [Zha02] and was solved deterministically by using the second derivative and the curvature. Thus, $\Delta\theta_{yz}$ serves in our case as a trigger for I_{ps}. The detection of the second frame of inflection I_{sp} where the movement passes from the *stroke* to the *post-stroke* is difficult to chose deterministically without using hindsight or foresight. This means we whether decide after reaching the rest position what was the last 'peak' of $\Delta\theta_{yz}$ or we use a probabilistic way of prediction. We decided to start with the former approach and work on a probabilistic model for segmentation later.

3.6   Conclusions

The first part of this chapter presented the technical solution for human movement tracking using a monocular camera and an active sensor. The frames of reference for the geometric model that describes the relationship between a 3-D position and its 2-D projections were introduced. The calibration of the two sensors and the process of visual tracking was presented. Given the benefits and drawbacks of the sensor modalities this work follows

a hybrid approach by i) recording labeled 3-D movement data using a commercial motion capture device ('active sensor'), ii) mapping the data to 2-D planes aligned with the camera planes and iii) learning the LMA descriptors of these planes. With this the classification step can be based on features from a single 2-D projection (e.g. monocular camera) as well as on features from multiple 2-D projections (e.g. multi-ocular system). The system benefits from the simpler processing and higher precision of the active sensor during the recoding step, while using the attractive touch-free alternative of computer vision for the classification step.

The presented system does not require an accurate camera calibration, specially the effect of lens distortions has been neglected. This robustness is partially due to the type of descriptors (*Vector Symbols*) and partially due to the probabilistic approach. The probabilistic approach adds to this robustness by learning features from several trials and persons. This produces probabilities of a certain amount for neighboring values. Additionally, a less accurate calibration allows us to pose less demands on the accuracy of the visual tracking.

The problems that arise when movements are observed by vision, e.g. occlusions, are solved through the Bayesian sensor fusion. A camera which detects the disappearance of a tracked object (e.g. through occlusion) will continue its update with uniformly distributed descriptors resulting in an increasing uncertainty. The algorithm for sensor fusion will assign a lower confidence value having the effect that observers without occlusions are weighted stronger.

In the second part of this chapter the calculated features were chosen by interpreting the parameters of LMA, regarding optimal performance of the Bayesian classifier and applying 'Ockham's Razor'. For the *Space* component of LMA the sequence of planar displacement angles was used. For the *Effort* component of LMA, acceleration, velocity and curvature were chosen. The *Shape* component was based on the position and size of a triangle formed

by head and hands. To allow for a fast computation all features were dis-
cretized to a low cardinality (4 to 8). The presented approach used only four
discrete variables per body part and plane. The usefulness of the features as
evidences for the *Effort* qualities was evaluated by histograms derived from
experiments. Four spatially similar but qualitatively (in terms of effort) dif-
ferent movements were performed by three persons. The histograms showed
distinct patterns of the features for the different *Effort* qualities. Finally
the concept for segmentation into the three phases pre-stroke, stroke and
post-stroke was presented. With this concept distinct *Effort* qualities can be
related to the different *phases*. As a consequence LMA descriptors can be
used more effectively for classification of movements.

Chapter 4

Bayesian Models for Movement Perception

Contents

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
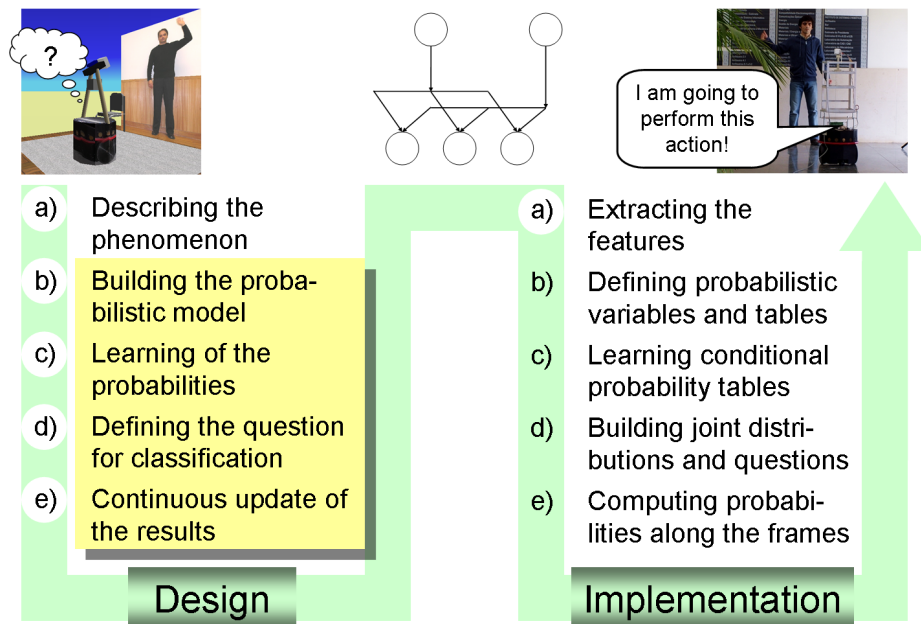
Figure 4.1: The organization of this thesis follows the process of designing and implementing the desired skill of interaction. This chapter presents the steps b) to e) of the design-process and concludes with a probabilistic framework for Laban Movement Analysis (LMA).

gives us the possibility to deal with incompleteness and uncertainty, make predictions on future events and, most important, provides an embedded scheme for learning.

The preceding chapters have concluded the first step (a) of the design-process (see Fig. 4.1). Both, the Laban Movement Analysis (LMA) shown in Chapter 2 and the Human Movement Tracking shown in Chapter 3 are important parts to describe the phenomenon. With this knowledge we can now proceed in this Chapter 4 with the remaining steps of the design-process.

First the reader will be prepared with the Bayesian approach, the *Bayesian program* formalism and the measure of entropy through the Section 4.1 to 4.3.

After this preparation the chapter continues with second step (b) of the design-process while building the models for computational Laban Movement

Analysis (C-LMA). Each section will present one type of model (Section 4.4 to 4.9). Through these models the relationship (dependency) between the different variables is defined.

The next step is to define the distributions that are associated to the relationships. Thus, the subsequent Section 4.10 presents the third step (c) of the design-process, i.e. the learning of probabilities.

In order to access the knowledge introduced by the previous steps a *questions* needs to be defined. The following Section 4.11 defines the *questions* for classification which is the fourth step (d).

The final step is shown in the subsequent Section 4.12 when the equations for continuous classification of movements are presented.

Section 4.13 will discuss an emerging characteristic of the presented approach, i.e. anticipation and certainty. With this, the theoretical framework is settled and the process of designing is concluded.

4.1 Basic Bayesian Concepts

Included in the Bayesian framework are specialized models which have a long tradition in many areas. Some examples of these models are 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.). Specially in the closely related field of *gesture recognition* these models have proven their usability [Sta95, Pav99]. 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 [KP04].

The first concept that will be presented is the usual notion of *logical proposition* ([BG03]). A logical *proposition* can be either true or false. *Propositions* are denoted by lowercase names. *Propositions* may be composed to obtain

new propositions using the usual logical operators: $a \wedge b$ denoting the conjunction of *propositions* a and b, $a \vee b$ their disjunction and $\neg a$ the negation of *proposition* a.

Variables are denoted by names starting with one uppercase letter. A discrete variable X is a set of logical *propositions* x_i which means that the variable X takes its ith value. These logical *propositions* are mutually exclusive and exhaustive. $\langle X \rangle$ denotes the cardinal of the set X. A variable $\mathbf{X}$ that is a vector or a set of variables is indicated by bold letters.

To be able to deal with uncertainty, we attach probabilities to *propositions*. To each *proposition* a a unique real value P(a) in the interval $[0, 1]$ is assigned. As this probability is conditioned by some preliminary knowledge, this fact is sometimes expressed by a *proposition* π which sums up this knowledge.

One interesting case is the probability of conjunctions $P(a \wedge b)$, disjunctions $P(a \vee b)$ and negations $P(\neg a)$of *propositions*. Another is the probability of *proposition* a conditioned by some other *proposition* b i.e. $P(a|b)$.

Probabilistic reasoning needs only two basic rules. The first is the *conjunction rule of propositions*, which gives the probability of a conjunction of *propositions*. While replacing the conjunction symbol by a comma we can state:

$$
\begin{aligned}
P(a,\ b) &= P(a) \times P(b \mid a) \\
&= P(b) \times P(a \mid b)
\end{aligned}
\tag{4.1}
$$

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

$$
P(a) + P(\neg a) = 1 \tag{4.2}
$$

The two rules (4.1, 4.2) are sufficient for any computation in discrete probabilities. All the other necessary inference rules concerning *variables* can be derived.

Conjunction (Bayes) rule for variables:

$$P(X, Y) = P(X) \times P(Y \mid X)$$
$$= P(Y) \times P(X \mid Y)$$

(4.3)

Normalization rule for variables:

$$\sum_X P(X) = 1$$

(4.4)

Marginalization rule for variables:

$$\sum_X P(X, Y) = P(Y)$$

(4.5)

4.2 Bayesian Programming

A *Bayesian program* is a generic formalism to build probabilistic models and to solve decision and inference problems on these models. Diard et al. presented the *Bayesian Programming* methodology as a unifying framework for most common probabilistic models [DBM03]. The generic formalism was applied to Kalman Filter (KF), Hidden Markov Model (HMM), Bayesian Network (BN), Dynamic Bayesian Network (DBN), Bayesian Filter (BF) and Particle Filters (PF). Lebeltel et al. [LBDM04] showed the application of *Bayesian Programming* to selected problems of autonomous robots, like reactive behaviors and sensor fusion. We used a *Bayesian program* to implement our human-robot interface and to build the underlying models. The single formalism allows to compare different solutions with one another and is used to automate our probabilistic computations.

A *Bayesian program* consists of two main parts (see Fig. 4.2). One is the *description* which holds the probabilistic model of the process (e.g. Laban Movement Analysis) and the other is the *question* that specifies an inference problem to be solved using this model. The *description* is, again divided into two parts. In the *specification* part the known entities and their relationships

Relevant variables:

$X_l, X_2, ... X_N$

Decomposition:

$P(X_1, X_2, \ ... \ , X_N) =$

$P(L_0) \times P(L_1|R_1) ... \ \times P(L_K|R_K)$

Parametric forms:

$P(L_0)$: type of distribution;

$P(L_1|R_1)$: type of distribution;

$P(L_2|R_2)$: type of distribution;

...

$P(L_K|R_K)$: type of distribution;

Identification:

Learning the free parameters of the parametric forms.

Questions:

$P(Search|known)$
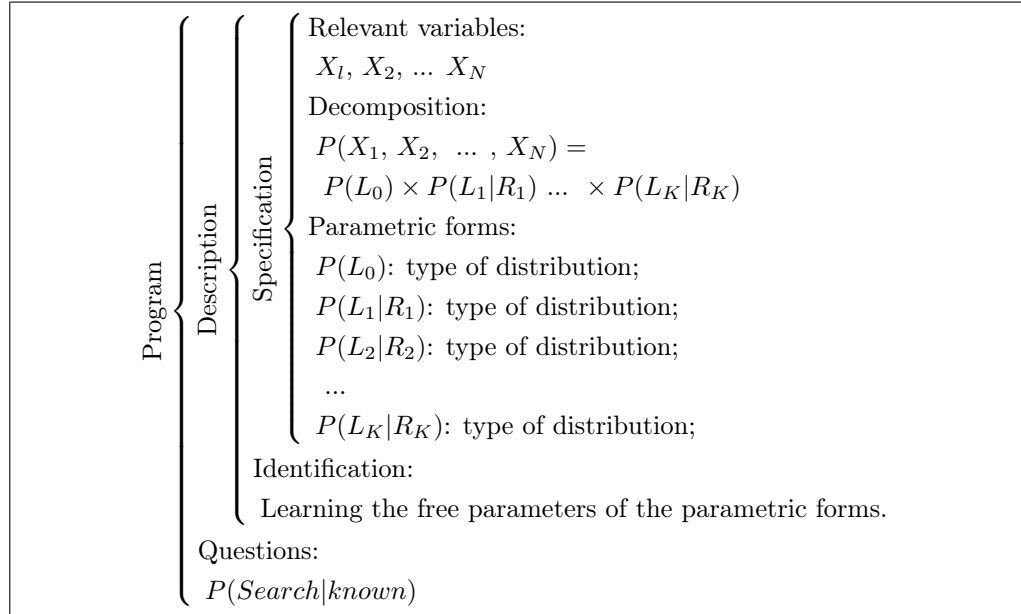
Figure 4.2: Generic Bayesian Program. The two main parts are the *description* and the *question*. The *description* is divided into the *specification* part and the *identification* part. The *specification* part holds the *relevant variables*, the *decomposition* part and the *parametric forms*.

are specified while the *identification* part holds the free parameters which are learned from experimental data. Finally, the *specification* part has three parts. The first part declares the *relevant variables*, the *decomposition* part states the joint distribution simplified by the dependency model (Bayes net) and in the *parametric forms* distributions are specified by e.g. functions or tables.

It can be seen that the *Bayesian programming* syntax allows a much richer description of a probabilistic problem than a graphical representation like a Bayesian network. Distribution with two or more variables on the left part of the conditioning mark or recurrences of sub-models are difficult to represent in Bayesian networks. Bayesian programming can be seen as a generalization of Bayesian networks [BG03] and has been successfully used as a compact tool for communication in European projects like BACS[1].

However, having said this, the following sections will use Bayesian nets

---

[1]BACS-project-6th Framework Programme of the European Commission contract number: FP6-IST-027140, Action line: Cognitive Systems

and joint distributions to represent the models for LMA. One reason is that structure of the models is kept simple without recurrences. Another is the possibility to show additional organizational levels within the graphical representation. Finally this thesis aims more at a deep understanding than a compact representation of the problem.

4.3 Entropy - A measure of uncertainty

Shannon [Sha49] extended the information theory by proposing a measure - *entropy* — wherein symbols have unequal probability of occurring. This measure associates information with uncertainty using the concept of probability. Being X a discrete random variable over a sample space the entropy was defined as

$$H(X) = -\sum_X P(X) \, \log P(X) = E \left[\log \frac{1}{P(X)} \right], \quad (4.6)$$

wherein the second form means that entropy can be defined as the expected value of $\log \frac{1}{P(X)}$. In equation (4.6), the logarithm's base determines in what unit entropy is measured. If it has base 2, entropy is expressed in *bits*. Hereafter, the logarithm's base will be omitted and it will be assumed that entropy is always computed by taking all logarithms to base 2. It will be also assumed the convention $0 \log 0 = 0$, since $x \log x \to 0$ as $x \to 0$, which means that adding terms with zero probability does not change the entropy.

The Shannon's definition of entropy is related with the definition of entropy in thermodynamics and it has the same mathematical expression. Shannon derived it axiomatically by defining certain properties that an information measure should have [CT91].

Entropy is a continuous function. Being $p_i = P(X = x_i)$ the probability of each possible value of the discrete random variable X, the entropy function H should be continuous in the p_i. Figure 4.3 shows the graph of the entropy of a binary random variable, having outcomes x_1 and x_2 with probabilities
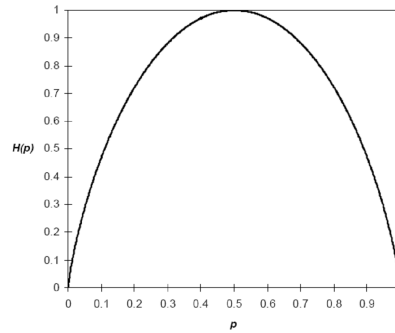
Figure 4.3: Entropy $H(p) = H(p, 1 - p)$ of a binary random variable having two outcomes with probabilities $p_1 = p, p_2 = 1 - p, 0 \le p \le .1$, as a function of p.

$p_1 = p$ and $p_2 = 1 - p$, respectively, with $0 \le p \le 1$. The graph shows that the probability distribution has *maximum entropy* if it is uniform, i.e. $p = \frac{1}{2}$ and $p_1 = p_2 = p$. Moreover, the entropy function is a concave function of the probability distribution, being null when $p = 0$ or $p = 1$. This means that:

> Entropy can be viewed as a measure of how much random a variable is, i.e. a measure of its uncertainty.

Entropy is a monotonic function and a formal measure of uncertainty. If all samples of a random experiment have the same probability, i.e. if $p_i = \frac{1}{n}$, wherein n is the number of possibilities (the cardinality of X), H is a monotonic increasing function of n. It can be easily proven from equation (4.6) that the entropy of a random variable X_u uniformly distributed is given by

$$H(X_u) = \log n \tag{4.7}$$

It also can be proven that the entropy of a random variable with n possible outcomes verifies the condition

$$0 \le H(X_u) \le \log n \tag{4.8}$$
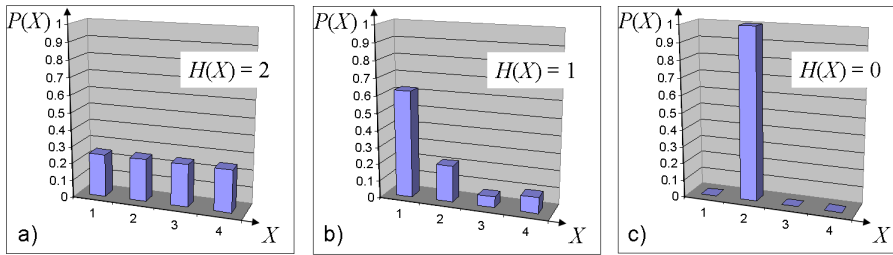
Figure 4.4: Entropy is an absolute measure of uncertainty. Three different probability distributions of the random variables X with cardinality $n = 4$ and their entropy values $H(X)$ are shown. a) Uniform distribution: maximum entropy b) Lower dispersion results in lower entropy and thus higher certainty c) Deterministic case: minimum entropy.

It can be easily proven that entropy is always non-negative, because in each of the terms $-p(x) \log p(x)$ in (4.6) the condition $0 \leq H(X_u) \leq \log n$ is always satisfied. This means that entropy can be viewed as the average length of the shortest description of a random variable. For example, in a fair coin toss experiment it is equal to 1 *bit*. As Fig. 4.3 shows, if the binary random variable is not evenly distributed, that shortest description is less than 1 *bit*, being the uniform distribution the *maximum entropy* case. The *minimum entropy* $H(X) = 0$ is achieved if $\exists_{i \in \{1,\ldots,n\}}, P(X = x_i) = 1$, i.e. if the variable X is deterministic.

Figure 4.4 depicts three different probability distributions and their entropy values. The entropy of the variables X falls somewhere between 0 and 2, since all random variables have $n = 4$ possible outcomes. The *maximum entropy* value $\log 4 = 2$ is achieved for an uniform distribution with the highest dispersion as shown in 4.4 a). If the dispersion is smaller as shown in Fig. 4.4 b), the entropy will also be smaller. This example makes also clear that the entropy of a discrete random variable is a bounded, formal measure of the uncertainty contained on it: it may evaluate to a value between 0 — null uncertainty, deterministic case as shown in Fig. 4.4 c) — and $\log n$ — maximum uncertainty, uniform distribution.
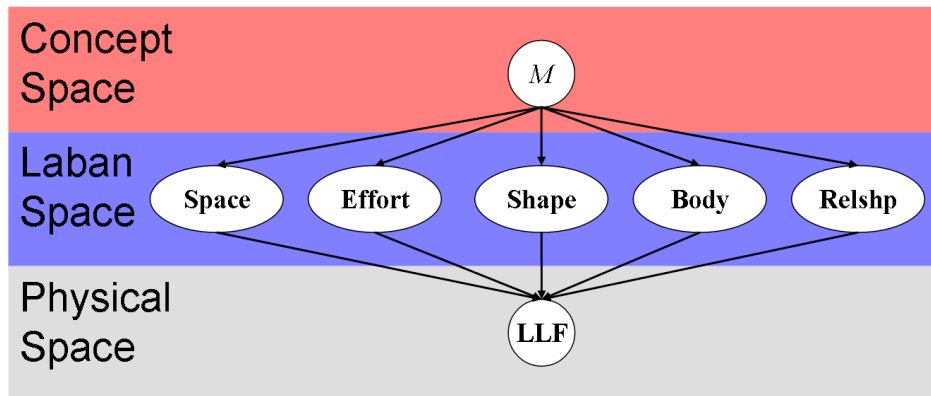
Figure 4.5: Bayes-Net of the *Global model*. Each component of LMA is represented as a node in the *Laban space* and contains a set of variables. The *physical space* holds the set of all low-level feature variables *LLF*. The variables of the *LLF* set depend on the variables of the *Laban space*. The high level concept of movement is represented by the variable *M*.

4.4   Global Model

The *Global model* which describes the phenomenon of computational Laban Movement Analysis (LMA) is shown in Fig. 4.5. Having the concept of a movement represented by the variable M certain characteristics will be exhibited through the sets of variables of LMA (**Space**, **Effort**, **Shape**, **Body** and **Relshp**). 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*. Each of the nodes on the *Laban space* is a parent for the set of low-level features **LLF**. The dependencies can
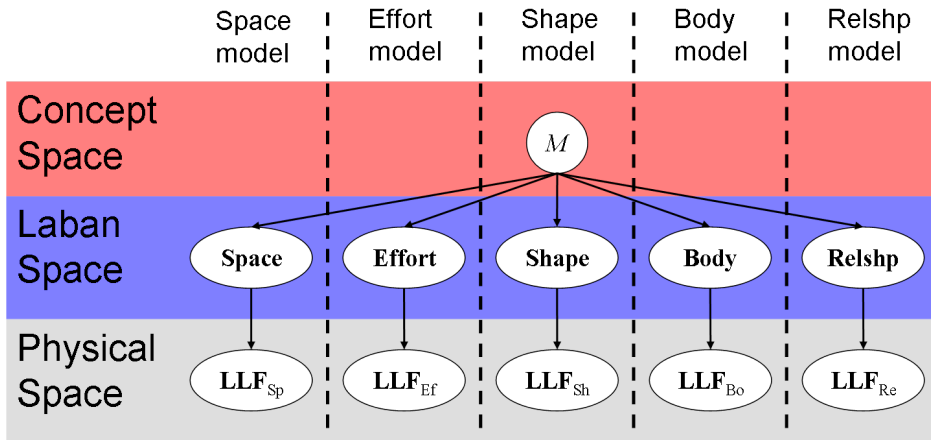
Figure 4.6: Bayes-Net of the *Global model*. By assuming n independent subsets for the low-level feature set $\mathbf{LLF}_n$ joint models can be composed from any combination of the n sub-models.

also be expressed as a joint distribution and its decomposition as:

$$P(M, \mathbf{Space}, \mathbf{Effort}, \mathbf{Shape}, \mathbf{Body}, \mathbf{Relshp}, \mathbf{LLF})$$
$$= P(M) \; \mathbf{P}(\mathbf{Space} \mid M) \; \mathbf{P}(\mathbf{Effort} \mid M)$$
$$\mathbf{P}(\mathbf{Shape} \mid M) \; \mathbf{P}(\mathbf{Body} \mid M) \; \mathbf{P}(\mathbf{Relshp} \mid M) \tag{4.9}$$
$$\mathbf{P}(\mathbf{LLF} \mid \mathbf{Effort}, \mathbf{Space}, \mathbf{Shape}, \mathbf{Body}, \mathbf{Relshp})$$

It appears that given the movement M the sets inside the *Laban space* are independent from each other. In fact, the *Affinities* between *Effort* and *Shape* are not modeled here. If we assume that for each LMA set n there exists an independent subset of the low-level feature set $\mathbf{LLF}_n$ we can decompose the *Global model* into an number of sub-models as shown in Fig. 4.6. The joint

Table 4.1: Global variables

Variable	Symbol	Description
Movement	M	Type of movement, e.g. $M = pointing$
Space set	**Space**	Set of direction symbol variables (*atoms*), e.g. $A = U$
Effort set	**Effort**	Set of *Effort* variables including *Time*, *Space*, *Weight* and *Flow*
Shape set	**Shape**	Set of *Shape* variables including *Spatial Shaping* and *Shape Flow*
Body set	**Body**	Set of constraints on human kinematics, e.g. $dist(head, lefthand)$
Relationship set	**Relshp**	Set of constraints on the human-robot space, e.g. $dist(hum, rob)$
Low-level feature set	**LLF**	Set of low-level features, may be split into sub-sets.

distribution and its decomposition would yield in this case:

$$
\begin{aligned}
&P(M, \textbf{Space}, \textbf{Effort}, \textbf{Shape}, \textbf{Body}, \textbf{Relshp}, \textbf{LLF}) \\
&= P(M) \ \textbf{P}(\textbf{Space} \mid M) \ \textbf{P}(\textbf{Effort} \mid M) \\
&\quad \textbf{P}(\textbf{Shape} \mid M) \ \textbf{P}(\textbf{Body} \mid M) \ \textbf{P}(\textbf{Relshp} \mid M) \\
&\quad \textbf{P}(\textbf{LLF}_{Sp} \mid \textbf{Space}) \ \textbf{P}(\textbf{LLF}_{Ef} \mid \textbf{Effort}) \\
&\quad \textbf{P}(\textbf{LLF}_{Sh} \mid \textbf{Shape}) \ \textbf{P}(\textbf{LLF}_{Bo} \mid \textbf{Body}) \\
&\quad \textbf{P}(\textbf{LLF}_{Re} \mid \textbf{Relshp})
\end{aligned}
\tag{4.10}
$$

Table 4.1 summarizes the sets, and variables used in the *Global model*.

As already argued in Chapter 2, we can expect that the strength of computational LMA can already be proven by regarding one component from each group (i.e. kinematic and non-kinematic). Thus, in the following section the *Space* and *Effort models* will be discussed in detail as they represent one component from each group. Due to its close relationship with *Space* and *Effort* the *Shape model* was also developed. Additionally a *Temporal model* will be discussed which tackles issues concerning the duration of a movement

and the frames of inflection (*phase*).

4.5 Space Model

The *Space* component of LMA is modeled using the concept of *Vector Symbols*. From the sets of variables presented in the *Global model* (see section 4.4) two sets of variables are used in this model:

$$\mathbf{LLF}_{Sp} \in \{A, B, C\}$$
$$\mathbf{Space} = \mathbf{LLF}_{Sp}$$

(4.11)

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

It was shown in Chapter 3 that the angular values of the *directions D* are translated into the *Vector Symbols* A_{bp} (*Table Plane* π_h), B_{bp} (*Door Plane* π_v) and C_{bp} (*Wheel Plane* π_s). The index *bp* corresponds to the *bodypart* like the right hand *rh*, the left hand *lh* and the head (face) *f*. The *Vector Symbols* receive one additional value from the velocity variable, i.e. the indication of no movement $v = 0$. The actual threshold for $v = 0$ can be seen in Table E.1 of the annex. 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* π_v (YZ) as they convey most of the information found in gestures. The temporal dependency of the *Vectors Symbols* is indicated
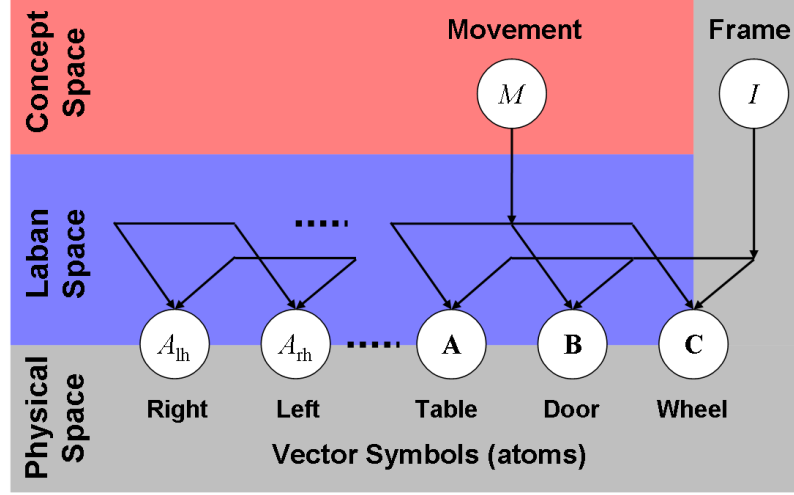
Figure 4.7: Bayes-Net for the *Space* component of LMA. The movement M belongs to the *concept space* while the *Vector Symbols* are part of both, the *Laban space* and the *physical space*. Their instances are in the principal planes *Table*, *Door* and *Wheel* and the left and right hand. The frame I is associated with the *physical space* only.

by the frame I. The variables and their sample space are shown in (4.12).

$$
\begin{aligned}
M &\in \{byebye\_dab, byebye\_glide, byebye\_flick, byebye\_float\} \ \langle 4 \rangle \\
I &\in \{1, \ldots, I_{max}\} \ \langle I_{max} \rangle \\
A_{bp} &\in \{O, F, FR, R, BR, B, BL, L, LF\} \ \langle 9 \rangle \\
B_{bp} &\in \{O, U, UR, R, DR, D, DL, L, UL\} \ \langle 9 \rangle \\
C_{bp} &\in \{O, U, UF, F, DF, D, DB, B, UB\} \ \langle 9 \rangle
\end{aligned}
\tag{4.12}
$$

The *Space model* assumes that each movement $M = m$ produces certain atoms $A_{bp} = a$, $B_{bp} = b$ and $C_{bp} = c$ at a certain point in time, i.e. frame $I = i$ and for a certain *Bodypart bp*. This dependency is also reflected in the Bayesian-net of Fig. 4.5. 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 movement M is associated with the

Table 4.2: Space variables

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

concept space, while the *Vector Symbols* are part of both, the *Laban space* and the *physical space*. The frame I is given by the system as some kind of clock and thus regarded as a 'pure' low-level feature from the *physical space*. The model might be applied to any number of body parts bp which are treated as independent evidences and thus expressed through a product as shown in the joint distribution of (4.13).

$$
\begin{aligned}
&P(M,\, I,\, A,\, B,\, C) \\
={}& P(M)\quad P(I)\quad \prod_{bp} \{P(A_{bp} \mid M,\, I)\quad P(B_{bp} \mid M,\, I)\quad P(C_{bp} \mid M,\, I)\}
\end{aligned}
$$

$$(4.13)$$

Table 4.2 summarizes the variables used in this sub-model.

4.6 Effort Model

The *Effort model* describes the dynamic aspects of the movement. It relates the low-level features like speed (Vel), speed gain (Acc) and curvature (K) to *Effort* qualities like *Time* ($E.Ti$), *Space* ($E.Sp$), *Weight* ($E.We$) and *Flow* ($E.Fl$). In order to not confuse the *Space* component from the previous section with the *Space* quality of the *Effort* component all variable symbols of *Effort* have a leading $E.$ before the quality. Similar is true for the possible confusion of the *Flow* quality for the *Effort* component ($E.Fl$) and the *Shape*

component ($Sh.Fl$) in the following section. From the sets of variables presented in the *Global model* (see section 4.4) two sets of variables are used in this model:

$$\mathbf{LLF}_{Ef} \in \{Vel, Acc, K\}$$
$$\mathbf{Effort} \in \{E.Ti, E.Sp, E.We, E.Fl\}$$

(4.14)

The relation between the two sets has already been investigated, established and developed in Section 3.4. The *concept space* relates the *Effort* qualities to a specific movement M. This has been introduced in Section 2.3 and 2.7. The *phase Ph* which has been introduced in Section 3.5 can be seen as a concept of segmenting the performance of a human movement. Thus, it will be associated with the *concept space* rather than with the *physical space* like the *frame I* in the previous model. The *Effort model* is related with a specific plane and body part where the *Effort* qualities can be detected best. The variables and their sample space are shown in (4.15)

$$Vel \in \{slow, medium, fast\} \langle 3 \rangle$$
$$Acc \in \{no, low, medium, high\} \langle 4 \rangle$$
$$K \in \{no, small, medium, big\} \langle 4 \rangle$$
$$E.Sp \in \{direct, indirect\} \langle 2 \rangle$$
$$E.Ti \in \{sudden, sustained\} \langle 2 \rangle$$
$$E.We \in \{strong, light\} \langle 2 \rangle$$
$$E.Fl \in \{bound, free\} \langle 2 \rangle.$$

(4.15)

Each movement M will produce a certain set of *Effort* qualities during a certain *phase Ph*. Thus we have a conditional dependency of *Effort Space* $E.Sp$, *Effort Time* $E.Ti$, *Effort Weight* $E.We$ and *Effort Flow* $E.Fl$ from the movement M and the *phase Ph* as can be seen in Bayes-net of Fig. 4.8. The *Effort* variables can not be directly measured but observed through some low-level features (i.e. $\mathbf{LLF}_{Ef}$). Thus, there is a dependency of the non-observable variables from the *Effort* set and $\mathbf{LLF}_{Ef}$. The joint distribution
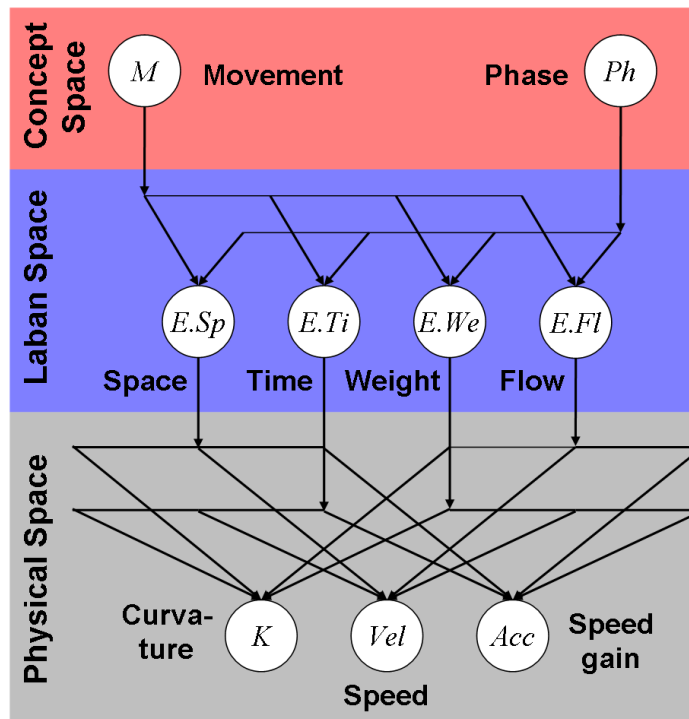
Figure 4.8: Bayes-Net of the *Effort* component of LMA. The movement M and the *phase Ph* belong to the *concept space*. The *Laban space* holds the four *Effort* qualities $E.Sp, E.Ti, E.We, E.Fl$. The *physical space* is based on *speed Vel*, *speed gain Acc* and *curvature K*.

Table 4.3: Effort variables

Variable	Symbol	Description
Movement	M	e.g. $byebye\_dab$
Phase	Ph	e.g. $pre-stroke$
Effort Space	$E.Sp$	e.g. $E.Sp = direct$
Effort Time	$E.Ti$	e.g. $E.Ti = sudden$
Effort Weight	$E.We$	e.g. $E.We = strong$
Effort Flow	$E.Fl$	e.g. $E.Fl = free$
Speed	Vel	e.g. $Vel = fast$
Speed gain	Acc	e.g. $Acc = high$
Curvature	K	e.g. $K = small$

can be expressed as

$$
\begin{aligned}
&P(M,\ Ph,\ E.Sp,\ E.Ti,\ E.We,\ E.Fl,\ Vel,\ Acc,\ K) \\
&= P(M)\quad P(Ph)\quad P(E.Sp \mid M,\ Ph)\quad P(E.Ti \mid M,\ Ph) \\
&\qquad P(E.We \mid M,\ Ph)\quad P(E.Fl \mid M,\ Ph) \\
&\qquad P(Vel \mid E.Sp,\ E.Ti,\ E.We,\ E.Fl) \\
&\qquad P(Acc \mid E.Sp,\ E.Ti,\ E.We,\ E.Fl) \\
&\qquad P(K \mid E.Sp,\ E.Ti,\ E.We,\ E.Fl).
\end{aligned}
\tag{4.16}
$$

Table 4.3 summarizes the variables used in this sub-model.

4.7   Shape Model

The *Shape model* describes the changes in the shape of the human body through the movements of the limbs. It relates the low-level features like the *direction signs* (Sx_bp, Sy_bp, Sz_bp) and the *triangle change* $(l_{Tr}, \Delta I_z)$ to *Shape* qualities like *Spatial Shaping* ($Sh.H$, $Sh.V$ and $Sh.S$) and *Shape Flow* ($Sh.Fl$). From the sets of variables presented in the *Global model* (see section

4.4) two sets of variables are used in this model:

$$\mathbf{LLF}_{Sh} \in \{Sx_i, Sy_i, Sz_i, l_{Tr}, \Delta I_z\}$$
$$\mathbf{Shape} \in \{Sh.V, Sh.H, Sh.S, Sh.Fl\} \tag{4.17}$$

The *Shape* component of LMA is closely related with the *Space* component, so are the low-level features *direction sign $Sx_{bp}, Sy_{bp}, Sz_{bp}$* with the *direction D*. Also similar to the *Space* component is that the three dimensions of the *direction sign $Sx_{bp}, Sy_{bp}, Sz_{bp}$* are used separately and related to the *Spatial Shaping* variables *Sh.S*, *Sh.H* and *Sh.V*. The *Shape model* differs from the *Space model* in two aspects. First, the *Shape model* fuses the *direction signs* for the tree body parts indicated by the index *bp*. Second, it uses a probability distribution over the whole sequence similar to the low-level features in the *Effort model*. The second level of the model relates the *Shape* qualities to a specific movement *M* as has been introduced in section 2.4. The set of *Shape* variables to be used by our Bayesian model are shown in (4.18).

$$Sx_{bp}, Sy_{bp}, Sz_{bp}, l_{Tr}, \Delta I_z \in \{neg, still, pos\} \langle 3 \rangle$$
$$Sh.Hor \in \{enclosing, still, spreading\} \langle 3 \rangle$$
$$Sh.Ver \in \{sinking, still, rising\} \langle 3 \rangle \tag{4.18}$$
$$Sh.Sag \in \{retreating, still, advancing\} \langle 3 \rangle$$
$$Sh.Flo \in \{shrinking, still, growing\} \langle 3 \rangle$$

As for the *Shape model* we assume that each movement *M* will produce a certain set of *Shape* qualities during a certain *phase Ph*. Thus, we have a conditional dependency of *Spatial Shaping Sh.V, Sh.H, Sh.S* and *Shape Flow Sh.Fl* from the movement *M* and the *phase Ph* as can be seen in Bayes-net of Fig. 4.9. As in the *Effort model* the movement *M* and the *phase Ph* are seen as being part of the *concept space*. The *Shape* variables of the *Laban space* can not be measured directly, but observed through some low-level features (i.e. $\mathbf{LLF}_{Sh}$) from the *physical Space*. Thus, there is a dependency
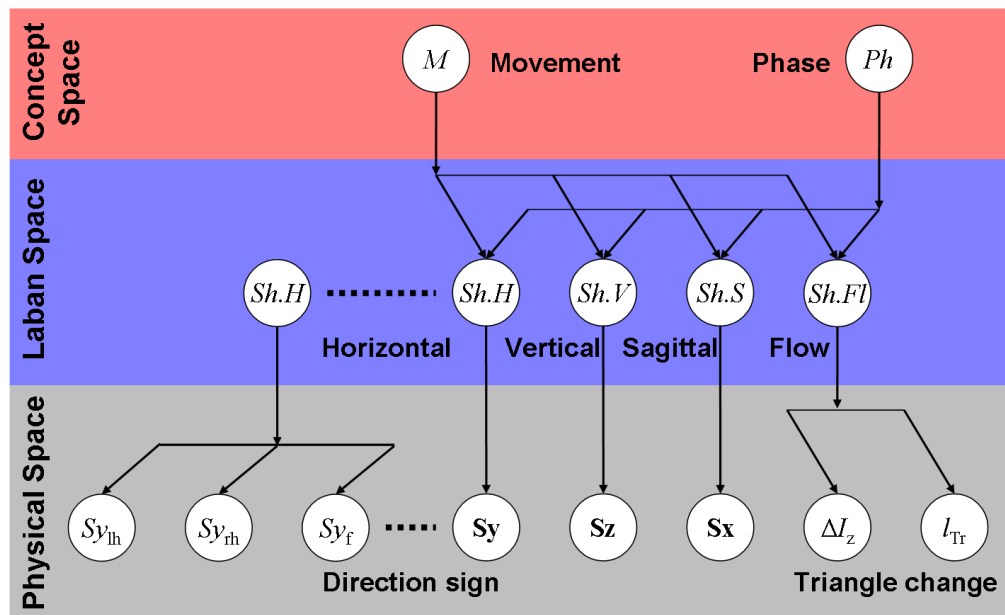
Figure 4.9: Bayes-Net of the *Shape* component of LMA. The movement M and the *phase Ph* belong to the *concept space*. The *Laban space* holds the four *Shape* qualities $Sh.V, Sh.H, Sh.S$ and $Sh.Fl$ of *Spatial Shaping* and *Shape Flow*. The *physical space* is based on the *direction signs* $Sx_{bp}, Sy_{bp}, Sz_{bp}$ and the *triangle change* $l_{Tr}, \Delta I_z$.

Table 4.4: Shape variables

Variable	Symbol	Description example
Movement	M	*byebye_dab*
Phase	Ph	$pre-stroke$
Body part	bp	*rh*
	$Sh.H$	*spreading*
Spatial Shaping	$Sh.V$	*rising*
	$Sh.S$	*advancing*
Shape Flow	$Sh.Flo$	*growing*
Direction sign	Sx_{bp}	*pos*
Triangle change	ΔI_z	*pos*
	l_{Tr}	*pos*

of the non-observable variables from the *Shape* set and $\mathbf{LLF}_{Sh}$. The joint distribution can be expressed as

$$
\begin{aligned}
&P(M,\ Ph,\ Sh.V,\ Sh.H,\ Sh.S,\ Sh.Fl,\ Sx_{bp},\ Sy_{bp},\ Sz_{bp},\ l_{Tr},\ \Delta I_z) \\
&= P(M)\quad P(Ph)\quad P(Sh.V \mid M,\ Ph)\quad P(Sh.H \mid M,\ Ph) \\
&\quad P(Sh.S \mid M,\ Ph)\quad P(Sh.Fl \mid M,\ Ph) \\
&\quad P(Sx_{bp} \mid Sh.S)\quad P(Sy_{bp} \mid Sh.H)\quad P(Sz_{bp} \mid Sh.V) \\
&\quad P(l_{Tr} \mid Sh.Fl)\quad P(\Delta I_z \mid Sh.Fl) \\
&\quad \prod_{bp} \{P(Sx_{bp} \mid Sh.S)\quad P(Sy_{bp} \mid Sh.H)\quad P(Sz_{bp} \mid Sh.V)\}
\end{aligned}
\tag{4.19}
$$

Table 4.4 summarizes the variables used in this sub-model.

4.8 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 $\bar{i}_{max}$ and variance
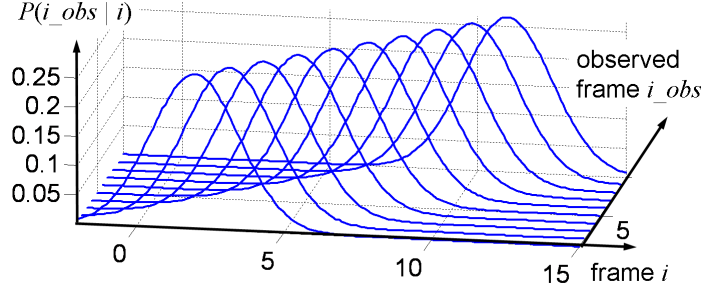
Figure 4.10: The observed (measured) variable *observed frame I_obs* relates to the value space of frame *I* with a certain probability. The relationship is modeled as a Gaussian distribution which will also regard evidences 'around' the currently measured frame. $P(i\_obs \mid i = i\_obs)$ is as a Gaussian distribution with 'sliding' mean.

σ_{i_max} 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

$$P(i\_obs \mid i) = N(i\_obs; \sigma_i) \qquad (4.20)$$

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 Fig. 4.10. One might notice that the standard deviation does not change, producing the relation of probabilities e.g. between $P(i\_obs \mid i = i\_obs)$ and $P(i\_obs \mid i = i\_obs + 1)$ for any observed frame in the interval. A different approach could be to increase also the standard deviation e.g. linearly with i_obs as shown in Fig. 4.11. This solution simply increases the uncertainty for evidences that arrive towards the end of
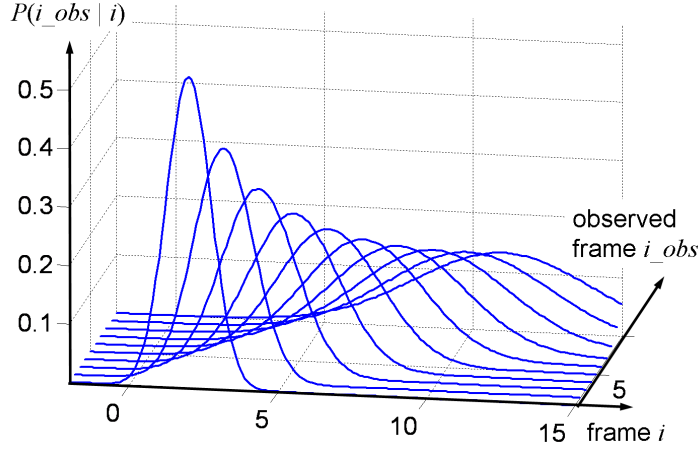
Figure 4.11: Additional to the previous figure 'later' frames will be regarded with a higher uncertainty. This results in an additional 'stretching' of the standard deviation of $P(i\_obs \mid i = i\_obs)$.

the performance. Evidences at the beginning of the movement are weighted higher.

The variables and their sample space are shown in (4.21).

$$I \in \{1, \dots, I_{max}\} \langle I_{max} \rangle$$
$$I\_obs \in \{1, \dots, I\_obs_{max}\} \langle I\_obs_{max} \rangle$$

(4.21)

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 Fig. 4.12. Comparing this figure with Fig. 4.5 shows that the additional variable observed frame I_obs belongs to the *physical space*. In this model I_obs is the variable that can be observed and I must be inferred from it. In other words: I_obs is measured directly as a *hard evidence* while I can be interpreted as a *soft evidence*. This *soft evidence* is used for the already known dependencies given by the previously discussed *Space model*. The joint distribution embedded *Space model* while omitting the body part index can be expressed
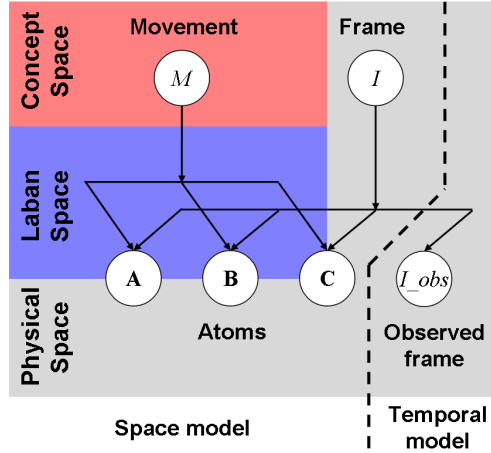
Figure 4.12: Bayes-Net of the *Temporal model* which connects to the *Space model*. The variable observed frame *I_obs* belongs to the *physical space* and is observable. The variable frame *I* also belongs to the *physical space* but will be inferred from *I_obs*. The remaining dependencies are given by the *Space model*.

Table 4.5: Variables used in the *Temporal model*

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

as

$$P(M, I\_obs, I, A, B, C)$$
$$= P(M) \quad P(I) \quad P(I\_obs \mid I) \quad P(A \mid M, I) \quad P(B \mid M, I) \quad P(C \mid M, I)$$

$$(4.22)$$

Table 4.5 summarizes the variables used in this model.

4.9   Joint Model using *Space* and *Effort*

With the previously discussed models many combinations can be built, tested and compared with each other. In this section the *Space, Effort* and *Temporal model* were combined as a *Joint model*. This type of *Joint model* is of

particular interest as it combines a kinematic (Space) and a non-kinematic (Effort) component of LMA. For this reason the model can be used to already prove the usefulness of computational LMA for automatic movement classification. From the sets of variables presented in the *Global model* (see section 4.4) four sets of variables are used in this model:

$$
\begin{aligned}
\mathbf{LLF}_{Sp} &\in \{A, B, C\} \\
\mathbf{Space} &= \mathbf{LLF}_{Sp} \\
\mathbf{LLF}_{Ef} &\in \{Vel, Acc, K\} \\
\mathbf{Effort} &\in \{E.Ti, E.Sp, E.We, E.Fl\}
\end{aligned}
\tag{4.23}
$$

Figure 4.13 shows the combination of the *Space*, *Effort* and *Temporal model*. The *Space* and *Effort model* are coupled through the movement variable M. This means that a certain movement $M = m$ will exhibit certain *atoms* $A = a$, $B = b$, $C = c$ and certain Effort variables $E.Ti = e.Ti$, $E.Sp = e.Sp$, $E.We = e.We$, $E.Fl = e.Fl$ along the time. The temporal dependency is modeled through the variables frame I, observed frame $I_{o}bs$ and *phase Ph*. A closer look a the connection points and the underlying equations can be seen in the annex F.

The joint distribution can be expressed as

$$
\begin{aligned}
&P(M, \ Ph, \ I\_obs, \ I, \ A, \ B, \ C, \ E.Sp, \ E.Ti, \ E.We, \ E.Fl, \ Vel, \ Acc, \ K) \\
&= P(M) \quad P(Ph) \quad P(I) \quad P(I\_obs \mid I) \\
&\qquad \prod_{bp} \{P(A_{bp} \mid M, \ I) \quad P(B_{bp} \mid M, \ I) \quad P(C_{bp} \mid M, \ I)\} \\
&\qquad P(E.Sp \mid M, \ Ph) \quad P(E.Ti \mid M, \ Ph) \\
&\qquad P(E.We \mid M, \ Ph) \quad P(E.Fl \mid M, \ Ph) \\
&\qquad P(Vel \mid E.Sp, \ E.Ti, \ E.We, \ E.Fl) \\
&\qquad P(Acc \mid E.Sp, \ E.Ti, \ E.We, \ E.Fl) \\
&\qquad P(K \mid E.Sp, \ E.Ti, \ E.We, \ E.Fl).
\end{aligned}
\tag{4.24}
$$
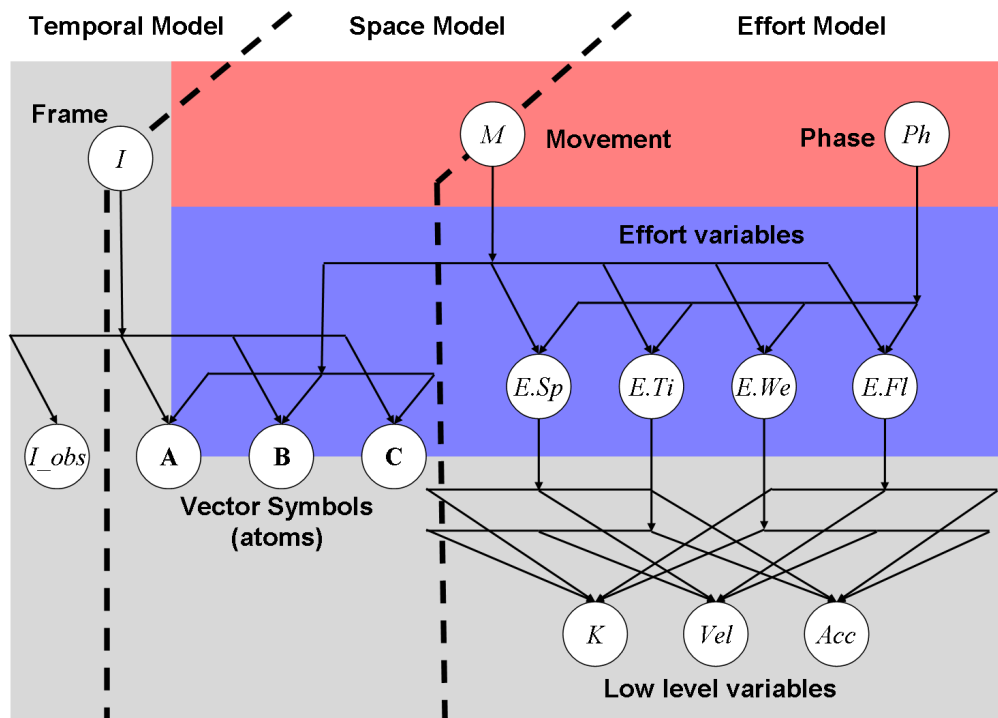
Figure 4.13: Joint model combining *Space*, *Effort* and *Temporal model*. The *Space* and *Effort model* are connected through the movement variable M in the *concept space*.

The tables that summarize the variables are found in the previous section (Table 4.2, Table 4.3 and Table 4.5).

4.10 Learning of probability tables

The previous section concluded with a joint distribution made of several distributions e.g. probability of *observed frame* given *frame* $P(I\_obs \mid I)$ and probability of *Vector Symbols* of *Door Plane* given *movement* and *frame* $P(B \mid M, I)$. All distributions belong to one of the following groups: i) Distributions that can be determined by 'expert wisdom' and ii) distributions that need to be 'learned'.

In the first group priors like $P(M)$, $P(Ph)$ and $P(I)$ can be found. Through 'expert wisdom' we state that all movements are equally likely to occur and thus a uniform distribution is assigned to $P(M)$. Similar is done for the priors $P(Ph)$ and $P(I)$. The probability of *observed frame* given *frame* $P(I\_obs \mid I)$ is also determined by 'expert wisdom'. In this case the parametric form of a *Gaussian distribution* chosen. The parameter *standard deviation* was obtained through the total length (in frames) of the different trials. The probabilities of *Effort* given *movement* and *phase* $P(E.Sp \mid M, Ph)$, $P(E.Ti \mid M, Ph)$, $P(E.We \mid M, Ph)$ and $P(E.Fl \mid M, Ph)$ were also defined by 'expert wisdom'. In this case it was assumed that a certain movement will not exhibit with 100 percent certainty the established *Effort* qualities. Through the chosen probability distribution e.g.

$$P(E.Sp = direct \mid M = pointing, Ph = pre - stroke) = 0.8$$
$$P(E.Sp = indirect \mid M = pointing, Ph = pre - stroke) = 0.2$$

(4.25)

the domination of the *Effort* component over the *Space* component was balanced.

All other distributions belong to the second group of distributions that need to be 'learned'. In our case 'learning' means that trials with a known label are fed into the system which in return identifies the parameters of a

chosen distribution. In the following the chosen type of distribution and the associated scheme for learning is discussed.

The two groups of distributions introduced above can also be found in the *Bayesian programming* framework (see Fig. 4.2). There, the distributions that can be determined by 'expert wisdom' (first group) are defined as *parametric forms* inside the *specification* part. The distributions that need to be 'learned' are defined through *identification* inside the *description* part.

Histogram Learning. One type of distribution that can be chosen is a histogram. The probability distribution can be learned by counting the observations a variable has a certain value. 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 by (4.26).

$$P\left(V = i\right) = \frac{n_i}{n} \tag{4.26}$$

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

The Laplace Succession Law. 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 (4.27) is based on the *Laplace succession law*.

$$P\left(V = i\right) = \frac{n_i + 1}{n + \lfloor V \rfloor} \tag{4.27}$$

Table 4.6: Learned probability tables

Type	Size
$P(Atom \mid M, I)$	$\lfloor M \rfloor \times 9 \times I_{max}$
$P(LLF \mid E.x)$	4×2
$P(E.x \mid M)$	$2 \times \lfloor M \rfloor$

The minimum probability which is produced when no observation has happened ($n_i = 0$) is shown in (4.28).

$$P_{min}\left(V = i\right) = \frac{1}{n + \lfloor V \rfloor} \qquad (4.28)$$

Taking the *atom* variable A_{rh} which has nine values $\lfloor V \rfloor = 9$ as an example, it can be seen that 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 $n_i = 0$.

Types of tables. The types of learned probability tables used in the previous models are summarized in Table 4.6. The learned table $P(Atom \mid M, I)$ holds the probability distribution of the variables *Atom* e.g. horizontal plane right hand *atom* A_{rh}. The variable has two conditions, the movement M and the frame I. Figure 4.14 represents this multidimensional table. 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.

Table 4.6 also indicates the size of the table that can be expected. Taking $P(Atom \mid M, I)$ as an example, the size 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.
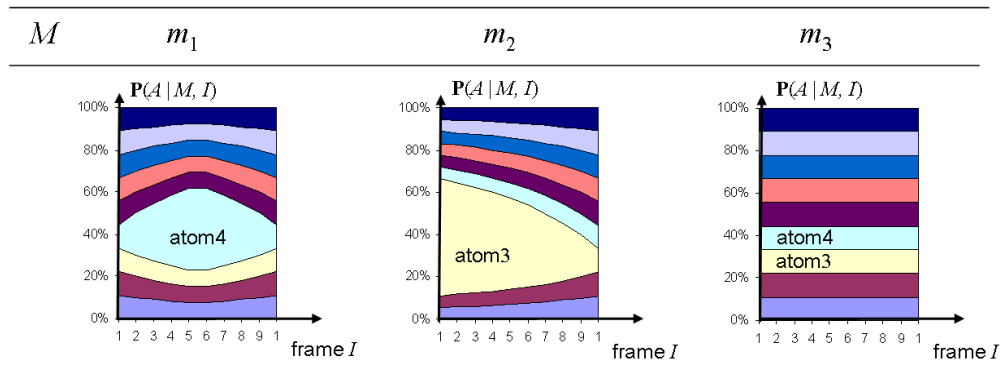
Figure 4.14:   Learned   table   for   generic   movements   of   the   type
$P(Atom \mid M, I)$. The movements m_1 and m_2 have a dominating *atom* (4 and
3) during certain phases (middle and beginning) while movement m_3 shows
no spatial pattern at all.

4.11   Questions for classification

Classification is the final step after the model has been established and the
tables have been learned.

> Classification accesses the knowledge of model and learning through
> inference.

In the *Bayesian programming* framework, classification is defined as *ques-
tions* inside the *program* part (see Fig. 4.2). In the following classification,
recognition and *question* will be used synonymous.

Given our joint distribution

$$P(M, Ph, I\_obs, I, A, B, C, E.Sp, E.Ti, E.We, E.Fl, Vel, Acc, K)$$
(4.29)

we need to formulate a *question*, i.e. what we want to classify and what we
can observe. The following set of equations shows particularly interesting
questions.

The first *question* (4.30) asks for the distribution of the variable movement M knowing the observed frame *I_obs* and the *atom B* from the *Door Plane* π_v.

$$P(M \mid I\_obs, B) \qquad \begin{cases} \text{Movement } M \text{ knowing} \\ \text{observed frame } I\_obs \\ \text{and atom } B \end{cases} \qquad (4.30)$$

This represents the classification of a movement taking into account the *Space* component of LMA from a frontal view.

The second *question* (4.31) asks for the distribution of the variable movement M knowing the observed frame *I_obs* and the *atoms A, B* and *C* from all *principal planes* π_h, π_v and π_s.

$$P(M \mid I\_obs, A, B, C) \qquad \begin{cases} \text{Movement } M \text{ knowing} \\ \text{observed frame } I\_obs \\ \text{and atoms } ABC \end{cases} \qquad (4.31)$$

This represents the classification of a movement taking into account the *Space* component of LMA from a top, frontal and side view.

The third *question* (4.32) asks for the distribution of the variable *Effort.Time E.Ti* knowing the speed gain *Acc*.

$$P(E.Ti \mid Acc) \qquad \begin{cases} \text{Effort.Time } E.Ti \\ \text{knowing speed gain } Acc \end{cases} \qquad (4.32)$$

This represents the classification of the *Effort* quality *Time* knowing the speed gain. With this *question* we are able to label a certain movement with the *Effort* quality *Time*.

The forth *question* asks for the distribution of the variable movement M knowing the two *Effort* variables *Effort.Time E.Ti* and *Effort.Space E.Sp*.

$$P(M \mid E.Sp, E.Ti) \quad \begin{cases} \text{Movement } M \text{ knowing} \\ \text{Effort Space } E.Sp \\ \text{and Effort Time } E.Ti \end{cases} \quad (4.33)$$

This represents the classification of a movement taking into account only the *Effort* component of LMA by using the two qualities *Space* and *Time*. With the latter *question* (4.33) and the former *question* (4.30) we are able to compare movement classifications based on components from the kinematic and the non-kinematic group (see Chapter 2).

The fifth *question* asks for the distribution of the variable movement M knowing the two *Effort* variables *Effort.Time E.Ti*, *Effort.Space E.Sp* and the observed frame *I_obs* and the *atom B* from the *Door Plane* π_v.

$$P(M \mid I\_obs, B, E.Sp, E.Ti) \quad \begin{cases} \text{Movement } M \text{ knowing} \\ \text{observed frame } I\_obs \\ \text{and atom } B \\ \text{and Effort Space } E.Sp \\ \text{and Effort Time } E.Ti \end{cases} \quad (4.34)$$

This can be seen as the 'ultimate' *question* asked in this work: The classification of a movement taking into account components from the kinematic and the non-kinematic group of LMA represented by *Space* and *Effort*.

4.12   Continuous classification of movements

After the interesting *questions* have been defined the problem of continuous update will be tackled. Continuous update of the believe is a desirable characteristic of Human-Machine Interaction. With this the system can continuously refine his classification results through the newly incoming evidences.

The concept will be developed be means of *question* (4.30) $P(M \mid I\_obs, B)$ but can be applied accordingly to all other questions.

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

Knowing the conditional probability $\mathbf{P}(B \mid M, I)$ together with the prior probabilities for the movements $\mathbf{P}(M)$ we are able to apply Bayes' rule and compute the probability distribution for the movements M given the frame I and the *atom B* with

$$P(M \mid I, B) \propto P(M)P(B \mid M, I) \tag{4.35}$$

It is possible to compute how likely it is that an observed sequence of n *atoms* was caused by a certain movement m. An example for this stream of *atoms* was shown in 3.17 for the movement $m = byebye$. To compute the *likelihood* we assume that the observed *atoms* are independently and identically distributed (i.i.d.). In (4.36) the sequence of n observed values for atom b is represented by $b_{1:n}$. For each movement m the joint probability will be the product of the probabilities from frame $i = 1$ to $i = n$, where the jth frame of the sequence is indicated by i_j.

$$P(b_{1:n} \mid m, i_{1:n}) = \prod_{j=1}^{n} P(b_j \mid m, i_j) \tag{4.36}$$

We can formulate (4.36) in a recursive way and for all movements M and get

$$P(b_{n+1} \mid M, i_{1:n+1}) = P(b_n \mid M, i_{1:n}) \, P(b_{n+1} \mid M, i_{n+1}) \tag{4.37}$$

The *likelihood computation* (4.37) can be plugged in our question (4.35). Assuming that each frame i a new observed direction symbol arrives we can

continuously (online) update our classification result.

$$P(M_{n+1} \mid i_{1:n+1}, b_{1:n+1}) \propto P(M_n) \, P(b_{n+1} \mid M, i_{n+1}) \qquad (4.38)$$

In the last step we introduce the *Temporal model* through the conditional probability $P(i\_obs_{n+1} \mid i_{n+1})$ and get

$$\begin{aligned} P(M_{n+1} \mid i\_obs_{1:n+1}, b_{1:n+1}) \propto \\ P(M_n) \, P(i\_obs_{n+1} \mid i_{n+1}) \, P(b_{n+1} \mid M, i_{n+1}) \end{aligned} \qquad (4.39)$$

which yields the recursive 'answer' to *question* (4.30).

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. The probability distribution of the initial prior ($n = 0$) is in this case not important. This holds true as long as non of the probabilities drop to zero. If a probability drops to zero the associated hypotheses will be ruled out once and for all. The final classification result is given by the *maximum a posteriori* (MAP) method.

For the sake of completeness the recursive 'answers' to the remaining four *questions* will be presented briefly. The recursive form for the second *question* (4.31) is

$$\begin{aligned} P(M_{n+1} \mid i\_obs_{1:n+1}, a_{1:n+1}, b_{1:n+1}, c_{1:n+1}) \propto \\ P(M_n) \, P(i\_obs_{n+1} \mid i_{n+1}) \, P(a_{n+1} \mid M, i_{n+1}) \\ P(b_{n+1} \mid M, i_{n+1}) \, P(c_{n+1} \mid M, i_{n+1}) \end{aligned} \qquad (4.40)$$

The recursive form for the third *question* (4.32) is

$$\begin{aligned} P(E.Ti_{n+1} \mid Acc_{1:n+1}) \propto \\ P(E.Ti_n) \, P(Acc_{n+1} \mid E.Ti) \end{aligned} \qquad (4.41)$$

The recursive form for the fourth *question* (4.33) is

$$P(M_{n+1} \mid E.Sp_{1:n+1},\ E.Ti_{1:n+1}) \propto$$
$$P(M_n)\ P(E.Sp_{n+1} \mid M)\ P(E.Ti_{n+1} \mid M) \tag{4.42}$$

The recursive form for the fifth *question* (4.34) is

$$P(M_{n+1} \mid i\_obs_{1:n+1},\ b_{1:n+1},\ E.Sp_{1:n+1},\ E.Ti_{1:n+1}) \propto$$
$$P(M_n)\ P(i\_obs_{n+1} \mid i_{n+1})\ P(b_{n+1} \mid M,\ i_{n+1}) \tag{4.43}$$
$$P(E.Sp_{n+1} \mid M)\ P(E.Ti_{n+1} \mid M)$$

4.13 Anticipation and Certainty

Three important observations can be made from Human-Human Interaction:

i) Starting from an initial belief (*prior*) the probabilities of this belief are continuously updated by newly incoming evidences.

ii) Depending on the distribution of a belief, a belief can be described by using labels like 'uncertain', 'quite certain' and 'certain'.

iii) Using an initial belief with few or no evidences in order to react without 'certainty' can be expressed as using the skill of 'anticipation'.

In order to improve Human-Robot Interaction the above stated observations should be taken into account. Figure 4.15 shows an exemplary situation that could occur during the interaction between a robot and a person. Assuming that the person starts to perform a movement (A) represented by timing diagram a) some evidences will be generated from the observation as shown in diagram b). There will be a delay given by the performance of the system (camera frame rate, computations, etc.). Taking the simple case of having two hypothesis (movement A and B) the probability of movement A will rise and the probability of movement B will fall as shown in diagram c). After the belief in movement A reaches a certain level a decision can be made as shown in diagram d). After the system has reached the 'certainty' that it observes movement A an arbitrary action can be triggered.
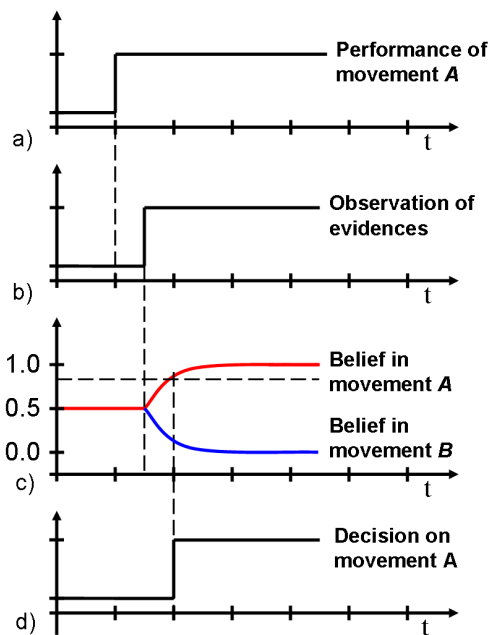
Figure 4.15: Timing diagram for observation, belief and decision. a) shows the timing of a performed movement (*A*). b) with some delay evidences will be generated by the observer. c) the continuous update of belief will increase the probability of observing movement *A*. After passing some level of 'certainty' a decision will be made in d).

The above described behavior will have some more implications. A prior that is not uniformly distributed like in Fig. 4.15 c) will speed up or slow down the recognition process depending if the prior was 'correct' or not. If the prior has a belief in a hypothesis higher than the level of certainty the system will decide immediately. The system can change its decision several times during the process whenever the belief passes the level of certainty. All these characteristics relate to the notion of anticipation and certainty.

A robot that offers the skill of anticipation inside a human-robot interaction scenario will be able to react faster at the cost of probably being wrong. This skill also effects the level of autonomy in which the robot is able to act. The higher autonomy will be reflected through the ability to cope with a situation which requires immediate action, though its outcome is yet uncertain. It is clear that the quality of anticipation is given by what has been learned. So far, this issue is only beginning to receive attention as in the European project Mind RACES[2]. The ultimate goal might be to establish anticipation as a performance measure for intelligent machines interacting with humans. In this section we propose a first method of evaluation by coining three factors of anticipation:

- the anticipation-speed (α-factor) as a measure of how fast a machine anticipates a social cue

- the anticipation-confidence (β-factor) as a measure of how strong a machine believes in the most likely hypothesis

- the anticipation-stability (γ-factor) as a measure of how stable the a machine believes in the most likely hypothesis

Given the evolution of probability distribution over time we first define three points in time: The start of the movement N_S, the end of the movement N_E and the anticipation point N_A where the entropy ratio $H(M)/H(M)_{max}$ drops and stays under a minimum threshold e.g. 0.1 (see Fig. 4.16. For the
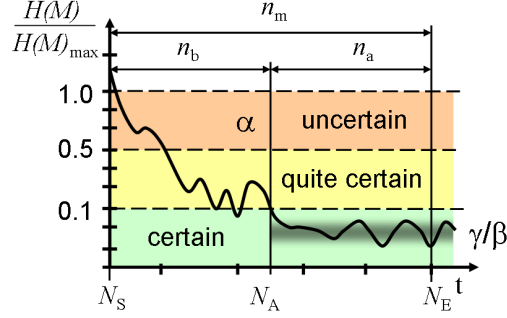
Figure 4.16: Performance measures for anticipation and levels of certainty. Given the entropy ratio $H(M)/H(M)_{max}$ three levels of certainty can be defined. When the entropy ration passes second level the anticipation point N_A can be determined. Using the start N_S and end N_E of the movement the anticipation factors can be determined.

sake of simplified expressions we name the duration of the whole movement n_m, the duration between N_S and N_A as n_b (b from build-up time), and the remaining duration between N_A and N_E as n_A (a from anticipation time).

We define the anticipation-speed (α-factor) as the ratio of n_m and n_b, the anticipation-confidence (β-factor) as the mean of the entropy during n_a, and the anticipation-stability (γ-factor) as the standard deviation from the β-factor during n_a. The three factors are shown in equation 4.44.

$$\alpha = 1 - \frac{n_a}{n_m} \qquad \beta = \frac{1}{N} \sum_{i=1}^{N} H_i \qquad \gamma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (H_i - \beta)^2} \qquad (4.44)$$

Related with the characteristic of anticipation is also the system's certainty in its belief. We defined three levels of certainty as shown in Fig. 4.16:

- *Uncertain*: $H(M)_{max} \geq H(M) \leq 0.5H(M)_{max}$.

- *Quite certain*: $0.5H(M)_{max} \geq H(M) \leq 0.1H(M)_{max}$.

- *Certain*: $H(M) \leq 0.1H(M)_{max}$.

4.14 Conclusions

This section presented the Bayesian models that describe the relationship of movements, LMA descriptors and low-level features. It introduced the basic Bayesian concepts, *Bayesian Programming* and entropy. A *Global model* was presented which encompasses the LMA descriptors for *Space*, *Effort*, *Shape*, *Body* and *Relationship*. The *Space* and *Effort* models were developed in detail as they represent the basic groups of LMA components, kinematic and non-kinematic respectively. Additionally a temporal model was developed to deal with the variance of movement duration. For the two basic components *Space* and *Effort* a joint model was developed. For each model the Bayes-net and the joint distribution was given. After the design of the models has been presented solutions for learning and continuous classification were shown. Bayesian nets offer the possibility to represent dependencies, parameters and their values intuitively understandable, which is a frequently expressed request from non-engineers [Loe01]. Using the property of conditional independence the dimensionality of the parameter space that describes the human movements can be reduced. Finally a connection between continuous update of a probabilistic belief and human anticipation was drawn. 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.

Chapter 5

Online Movement Recognition System

Contents

The previous chapters of this thesis reflect the five steps of designing a probabilistic model as shown in Fig. 5.1. The implementation of the processes and its results can also be organized in five steps. The first step a) is the extraction and computation of the low-level features and was already addressed in Chapter 3.

The second step b) is the definition of the *probabilistic variables* and *conditional probability tables*. Those definitions can be derived directly from

137
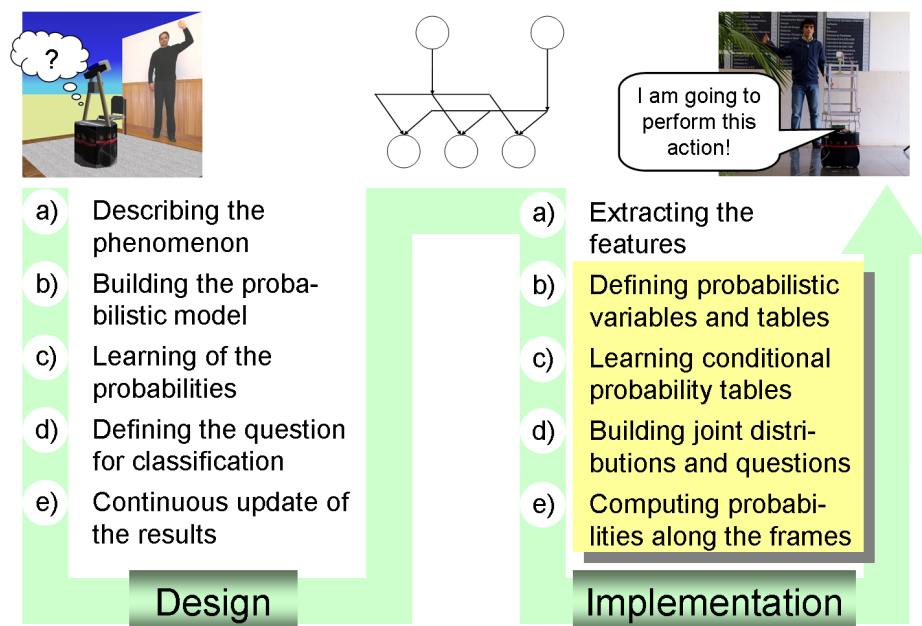
Figure 5.1: The five steps of designing a probabilistic model (shown in the previous chapters) relate to the five steps of implementing the processes (presented in this chapter).

the second step b) of the design process, i.e. building of the probabilistic model. Each step in *implementation* has, in fact, a corresponding step in *design*. For the second step b) of *implementation* the main work of defining the variables and tables is simply a matter of 'coding', thus it will not be discussed in an own section.

Instead the following Section 5.1 will present the third step c), i.e. the process of learning and the resulting *conditional probability tables*. The discussion will show that the Bayesian approach allows a comparison of those learned tables, even before entering the classification stage.

The fourth step d) in which the joint distributions and questions are built is similar to step b) just a matter of 'coding'. Thus, it is only shortly summarized in the beginning of Section 5.2 which has its focus on the fifth step e) of *implementation*, i.e. the computation of probabilities along the frames (time). The discussion on the evolution of probabilities is presented and give some insight on anticipation and certainty for a specific trial. With this the process of *implementation* is concluded. To emphasize the important characteristic of the system it is called *Online Movement Anticipation and Recognition (OMAR)* system.

The remaining sections will discuss the performance of the OMAR system when using different types of models for the classifier. The performance is evaluated through confusion tables for the two sets of movement data 'expressive movements' and 'bye-byes'. The set 'expressive movements' is used for models based on the *Space* component with different types of projections. The discussion can be seen in Section 5.3 to Section 5.5. The set 'bye-byes' is used to evaluate the influence of the *Effort* descriptor on the classification result. The discussion can be seen in Section 5.6 to Section 5.9. After this the implementation of the OMAR system inside the social robot, Nicole will be presented.
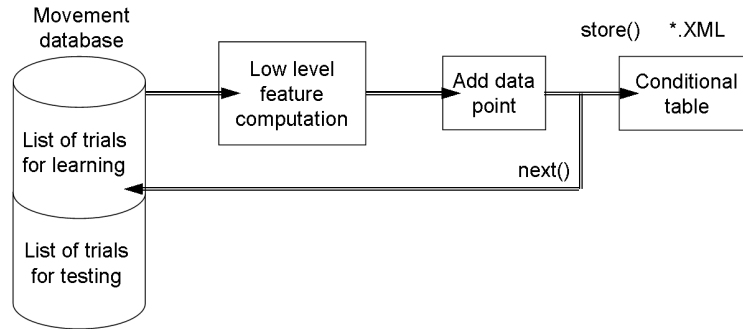
Figure 5.2: Learning process: Low level features are extracted each frame and 'add' points to the histogram. After all trials are processed the conditional probability tables are stored, e.g. in XML-format.

5.1   Learning conditional probability tables

The third step of implementation is the learning of the conditional probability tables. Figure 5.2 shows the flow chart of the learning process. 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.

The learned tables represent the 'fingerprint' and prototype of a movement. The table preserves the possibility to evaluate what has been learned. Figure 5.3 uses a stacked representation of the probabilities to show which *Vector Symbols* (*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 goes 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 and represents the phase $pre-stroke$. From the
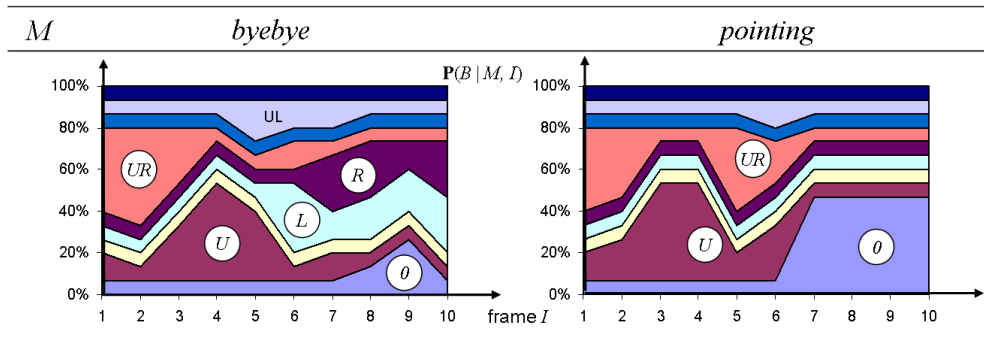
Figure 5.3: Learned table for the movements *byebye* and *point* with stacked probabilities. Broad stripes show dominant *atoms B* during certain phases. The thinnest region represents the minimum probability given by the Laplace assumption.

5th or 6th frame on, the gestures become distinct. The gesture *byebye* has mainly movements to the left and right (L and R) with some zero *atoms* 0 at the points of inflection. The gesture *pointing* mainly non-movement *atoms* (0) leaving the other probabilities at their minimum given by the Laplace assumption.

Figure 5.4 shows the learned table for the movement *circle* from the set 'gestures'. It can be seen that the pattern is quite distinct from the previous two movements as nearly all *atoms* appear along the time. The non-movement *atom* 0 is an exception as it stays at the minimum given by the Laplace assumption. This shows that for the gesture 'Circle' the hand is in continuous motion. 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.

The next set of movements under investigation is the 'byebye' set. Figure 5.5 shows the *Space* patterns for the movements *byebye_dab*, *byebye_glide*, *byebye_flick* and *byebye_float*. The movement patterns look much less distinct than the 'gesture' set. Specially *byebye_dab* and *byebye_glide* can be easily confused. Both have strong probabilities at left L and right *atoms*
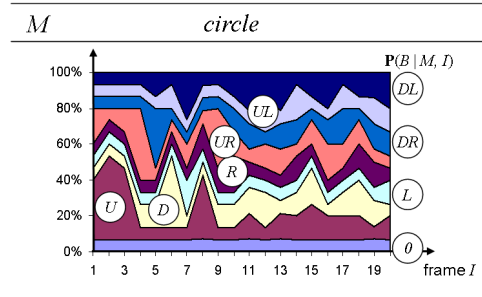
Figure 5.4: Learned table for the movement *circle* with stacked probabilities. All motion indicating *atoms B* have similar breadth indicating similar probability. The zero motion *atom* has the minimum given by the Laplace assumption.
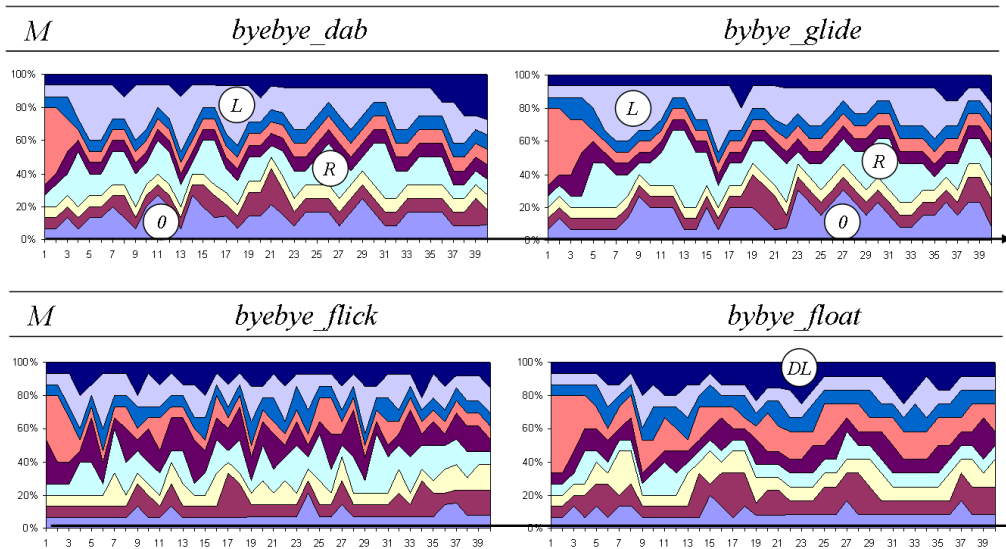
Figure 5.5: Learned table for movement set 'bye-bye' with stacked probabilities. The set was chosen to get a high similarity for the *Space* component. The similarity is reflected through the breadth of the stripes for the *atom B* probability.

R together with always apparent zero *atoms* 0. This is different to the *byebye_flick* and *byebye_float* movements which have zero *atoms* only at minimum probability. It also appears that the other *atoms* are distributed more uniformly.

5.2 Continuous classification, anticipation and certainty

The fourth step d) of implementation has been presented in chapter 4, where the joint distributions of interest were stated as:

- Movement classification using *Space* from a single projection.

- Movement classification using *Space* from triple projections.

- *Effort* classification

- Movement classification using *Effort.*

- Movement classification using *Space* and Effort.

The fifth and final step is the process of continuous classification which is presented as a flow chart in Fig. 5.6. 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 probability tables it possible to compute the desired probability distribution. This goes according to the defined joint distribution and the desired question. Through feeding in (replacing) the result of the computed probability distribution 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.
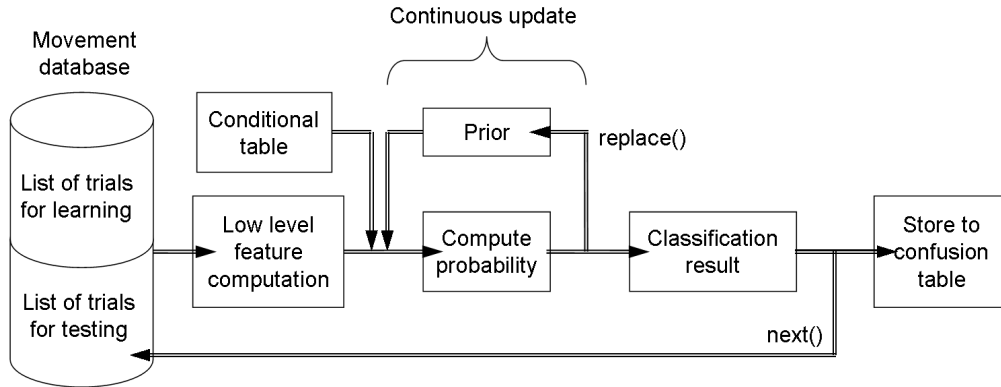
Figure 5.6: Classification process: The inner loop of continuous update produces the evolution of probabilities, the outer loop of next trial produces the confusion table.
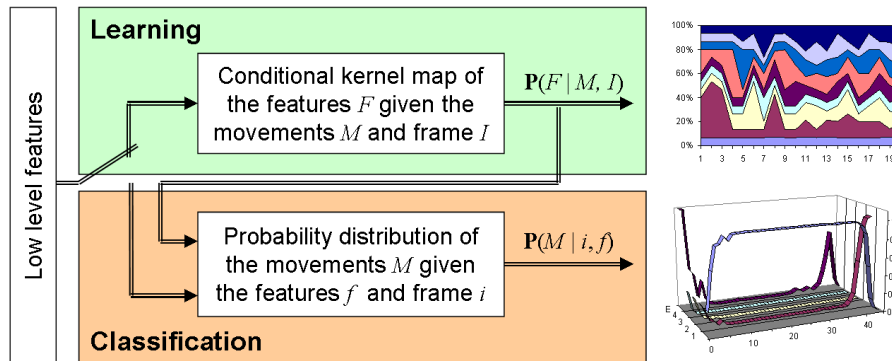
Figure 5.7: Switching between Bayesian learning and classification. The two processes share the same low-level features. First the probability of the features F is learned given the movements M. Later probability of a movement M is determined given the features F
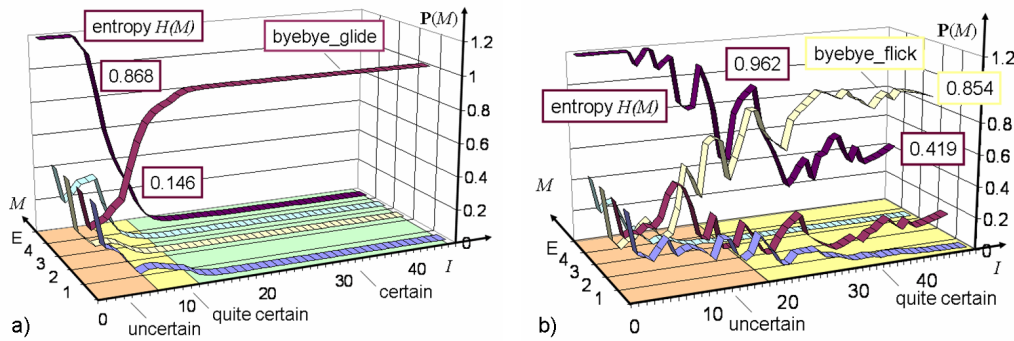
Figure 5.8: Evolution of the movement probabilities $P(M)$ along the time (i) for two trials of the 'byebye' set. The colored bottom signals the level of certainty.

We can conclude that the two processes of learning and classification are based on the same type of observations as shown in Fig. 5.7. The previously presented scheme starts by learning and, after the conditional probability tables have been build, continues with classification. An important feature of Bayesian histogram learning is that the OMAR system can 'switch back' at any time to learn new and more data. This opens the possibility to create a system that is able to continuously learn new data during its daily operation.

The OMAR system yields the qualities of anticipation and certainty through the Bayesian approach and the continuous classification. The evolution of the anticipated movements and the certainty of the belief is shown in Fig. 5.8. The two cases classify trials from the 'bye-bye' set. Figure 5.8 a) shows a case where the correct movement is anticipated fast and confident. The probability distribution for the four movements starts with a uniformly distributed prior, i.e. 25% and the maximum entropy, i.e. 2. In the eight frame the belief passes from *uncertain* to *quite certain* having an entropy of 0.868. In the twelfth frame the belief passes from *quite certain* to *certain* with an entropy of 0.146. The tree levels of certainty are also reflected as the color on the bottom of the diagram. The trial is finally correctly classified as *byebye_ glide* with a probability of 100%.

A different case is shown in Fig. 5.8 b) where the belief passes from *uncertain* to *quite certain* only in frame 18 having an entropy of 0.962. The system never reaches the *certain* state and concludes with the correct movement *byebye_flick* at an entropy of 0.419. It can be seen that the entropy was not monotonically decreasing and surpassed the minimum around the thirtieth frame. The behavior of the OMAR system to actually change its belief while producing a peak of entropy is a natural characteristic of this type of Bayesian classifier.

5.3 Expressive movement recognition using single and triple projections

The following three sections will present the performance of the OMAR system under different aspects of planar projections. These models consider only the *Space* component of LMA and a special set of movements, i.e. 'expressive movements' was assembled. These movements were not collected in anticipation of an optimal classification rate, but to the contrary to enforce some confusions depending on the perspective.

The loss of spatial information when using only one projection is investigated in this section. First the confusion table when using only *atoms B* from the *Door Plane* π_v is shown. Then the performance of the OMAR system when fusing *atoms A*, *B*, *C* from the tree planes is compared. The experiment uses the eight movements of the 'expressive movements' set, i.e.

1. Lunging: Lunging for a ball

2. Maestro: Conducting an orchestra

3. Stretch: Stretch to yawn

4. Ok: OK-sign gesture

5. Point: Pointing gesture

Table 5.1: Confusion table using only the 2-D (vertical) *atoms.*

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

6. Byebye: Waving bye-bye

7. Shake: Reach for someone's hand

8. Nthrow: Waving sagittally (approach sign)

Table 5.1 shows the results for using the *B atoms* of the *Door Plane* π_v. The sum of all numbers in each row usually adds up to thirteen, though some movements have fewer trials. The Tables G.1, G.2, G.3 and G.4 of the annex show additional information concerning the trials. There, sub-colums have been introduced to distinguish between trials and persons that have been used for learning (known trials, known person (kk)) and persons and trials that are unknown to the system (uu).

In the single-projection 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 Fig. 5.9) we can see that they are quite similar, though the *Door Plane* π_v appears as the most distinctive one.

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
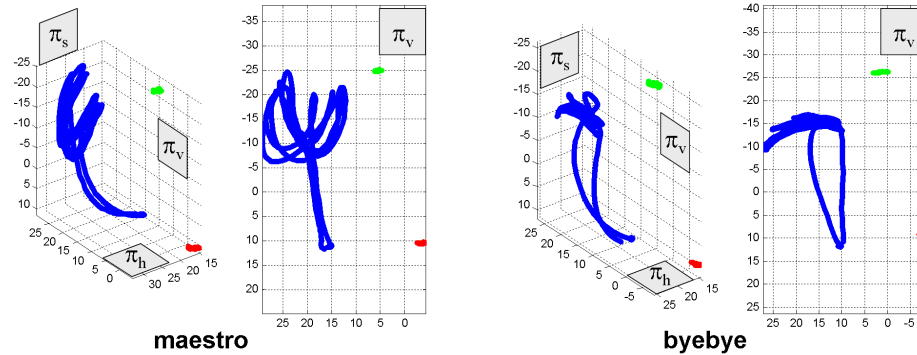
Figure 5.9: Comparing the traces of the movements *byebye* and *maestro* in 2-D and 3-D. The projections in the *Door Plane* π_v appear similar.

but also due to the model. From Fig. 5.10 it can be seen that for the 2-D projection the right hand traces (blue) are similar leading partly to the confusion. 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 Fig. 5.11 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. In this case the confusion goes in both directions as can be seen in the seventh row. The final confusion occurs in row eight. From Fig. 5.12 it can be seen that the 2-D projection does not convey the information on the sagittal oscillation the right hand is performing.

Table 5.2 shows the results for fusing the A, B and C *atoms* of the three planes. 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
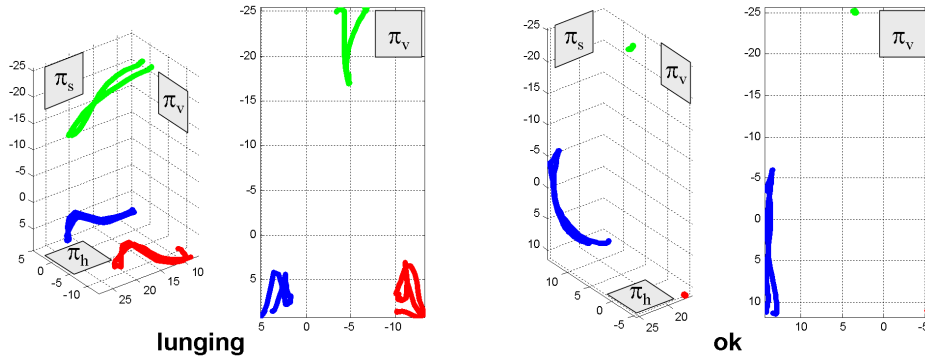
Figure 5.10: Comparing the traces of the movements *lunging* and *ok* in 2-D and 3-D. The 2-D projection for the right hand traces (blue) are similar.
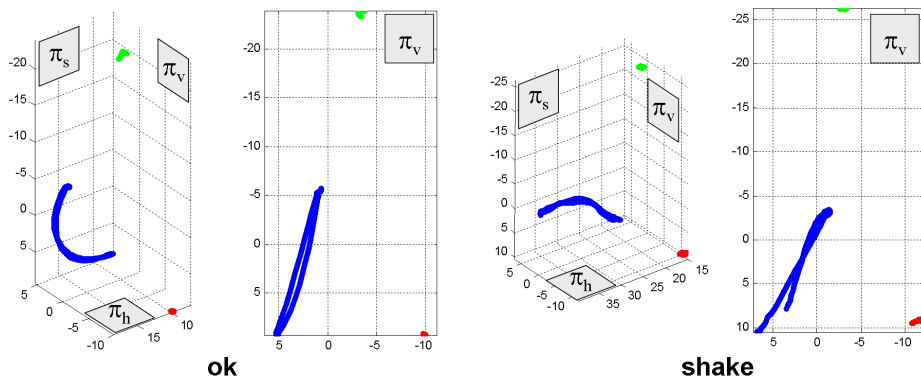
Figure 5.11: Comparing the traces of the movements *lunging* and *ok* in 2-D and 3-D. The 2-D projection can be confused easily.

Figure 5.12: Comparing the traces of the movements *nthrow* and *ok* in 2-D and 3-D. The 2-D projection does not convey the information on the sagittal oscillation.

Table 5.2: Confusion table for fusing *atoms* from three projections.

Movement	1	2	3	4	5	6	7	8	Σ_e
1 lunging	11			1			1		2
2 maestro		2				11			11
3 stretch			11	1			1		2
4 ok				9			4		4
5 pointing					10		2		2
6 byebye						13			0
7 shake				1			12		1
8 nthrow								5	0
									22

trials as the evidences of the sagittal waving are now processed. In the triple-projection case 22 of 95 trials are classified wrongly leaving a recognition rate of 77%. The *maestro* movement of the second row is significantly worse for triple-projections 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 the three planes (from 67% to 78%). Some movements can not be seen in certain planes, e.g. *nthrow* in the *Door 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.

5.4 Expressive movement recognition using distorted projections

The previous section showed that a certain recognition rate can be achieved by considering only information from a single projection, i.e. the *Door Plane* π_v. It was assumed that the projection used in the learning phase fits with the projection observed in the classification phase. This section investigates the effect of feeding in data from a perspective that differs from the one that has been learned. This might happen when a robot during the phase of positioning does not reach the expected orientation, or if the actor is not facing the camera. In this sense the difference in orientation will cause a 'distorted' projection.

For this experiment the perspective for classification was rotated $22.5 \, \mathrm{deg}$ and $45 \, \mathrm{deg}$ with respect to the learned fronto-parallel view of $0 \, \mathrm{deg}$. Table 5.3 shows the results for using the B *atoms* of a $22.5 \, \mathrm{deg}$ perspective towards the *Door Plane* π_v. In the case of a $22.5 \, \mathrm{deg}$ perspective 46 of 95 trials are classified wrongly leaving a recognition rate of 52%.

Table 5.4 shows the results for using the B *atoms* of a $45 \, \mathrm{deg}$ perspective

Table 5.3: Confusion table using *B atoms* from a 22.5 deg perspective.

Movement	1	2	3	4	5	6	7	8	Σ_e
1 lunging	8			3					3
2 maestro		3				10			10
3 stretch			11	2					2
4 ok				3	8		2		10
5 pointing				3	9				3
6 byebye						13			0
7 shake				5	8				13
8 nthrow				1		4			5
									46

Table 5.4: Confusion table using *B atoms* from a 45 deg perspective.

Movement	1	2	3	4	5	6	7	8	Σ_e
1 lunging	10			2	1				3
2 maestro		1			1	11			12
3 stretch			7	2	4				6
4 ok				4	9				9
5 pointing				2	10				2
6 byebye						13			0
7 shake				2	10		1		12
8 nthrow						5			5
									49

towards the *Door Plane* π_v. In the case of a 45 deg perspective 49 of 95 trials are classified wrongly leaving a recognition rate of 48%.

We can conclude that a changed perspective decreases greatly the recognition rate. This problem can be solved by calibrating the camera and the active sensor. When the transformation is known, the correct 2-D projection can be extracted from the 3-D data.

Table 5.5: Confusion table using a single-projection for 21 frames.

Movement	1	2	3	4	5	6	7	Σ_e
1 lunging	6			3		1		4
2 maestro		9		1				1
3 stretch			6	4				4
4 ok				5	4		1	5
5 pointing					10			0
6 byebye						10		0
7 shake					4		6	4
								18

5.5 Expressive movement recognition after partial observations

In this section the influence of the partial observations on the recognition results and thus on the confusion tables are shown. In this case the update of the classification of a single projection is stopped after a certain frame (i.e. 21).

Table 5.5 shows the results for using the *B atoms* and extracting the classification result after 21 frames. Table 5.5 shows the results for fusing the *A*, *B* and *C atoms* of the three planes and extracting the classification result after 21 frames.

In the single-projection case 18 of 70 trials are classified wrongly leaving a recognition rate of 74%. In the triple-projection case 10 of 70 trials are classified wrongly leaving a recognition rate of 86%. The recognition result improves for both cases (single- and triple-projection) compared with the results from observing the full sequence. It appears that the different lengths of the trials causes a higher confusion during the 'late' frames. This points towards a solution where a certain measure needs to found which triggers the conclusion of the classification process. Later it will be shown that entrophy is the correct choice in most of the cases. This section concludes the

Table 5.6: Confusion table for fusing the *A, B* and *C atoms* for 21 frames.

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

Table 5.7: Subset of the *Basic Effort Action Drives*

Action	Space	Time
Dab	Direct	Sudden
Flick	Indirect	Sudden
Glide	Direct	Sustained
Float	Indirect	Sustained

experiments concerned with *Space* models using the data set of 'expressive movements'.

5.6 Classification of bye-bye movements using only Space

This and the following sections will evaluate the effect on the performance of the OMAR system when the *Effort* model is introduced. The experiments will use the data set 'bye-byes' which is made of four movements that represent essentially a horizontal waving. Thus, no high spatial distinctiveness can be expected. Each of the four movements has a special 'flavor' in the way it is performed. This flavor is defined by *Basic Effort Actions* know from LMA (see Section 2.3). Table 5.7 shows a fragment of the *Basic Effort Action Drives* for the *Effort* qualities *Time* and *Space*.

Table 5.8: Confusion table for classification of movements using *Space*.

Movement	1	2	3	4	Σ_e
1 dab	3	12			12
2 glide		15			0
3 flick	2		12	1	3
4 float				15	0
					15

This section will conduct again the experiment of movement classification using the *Space* model from a single (*Door Plane* π_v) projection. The obtained results can then be compared to models using *Effort*, to models using *Space* and *Effort*, and so on. The results for all trials are shown in Table 5.8. By using *Space* 15 of 60 trials are classified wrongly leaving a recognition rate of 75%. Though the number of hypothesizes is low (only four movements) the recognition rate is also low. The confusion between *byebye_dab* and *byebye_glide* was already anticipated given the learned tables from the previous section. The 'byebye' set provides only few spatial distinctiveness.

5.7 Classification of Effort qualities for bye-bye movements

In order to get results on movement classification using *Effort*, first *Effort* itself needs to be classified. Thus, this section presents the performance of a classifier for *Effort* which will be used in the following section to classify movements.

For this the probabilities for the two *Effort* qualities will be calculated each frame a movement happens (non-zero velocity frame). Each *Effort* quality uses only one low-level descriptor: *Time E.Ti* is associated to *speed gain Acc* and *Space E.Ti* is associated to *curvature K*. The belief will be updated and converges to the hypothesis that explains the observations (*Acc* and *K*)
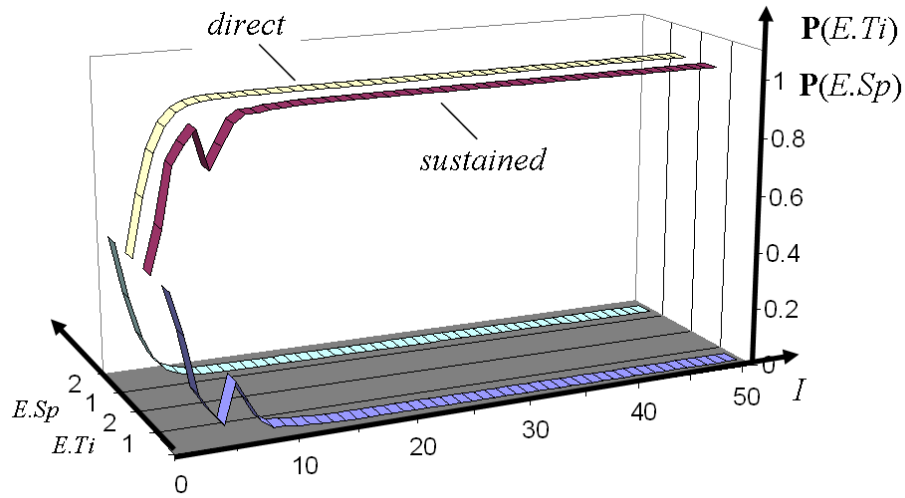
Figure 5.13: Evolution of the probabilities for *Effort Time* and *Space* quali-
ties by the sub-model of the trial *byebye_glide Jorg1*.

best. Figure 5.13 shows the probabilistic evolution of the two *Effort* qualities
for the trial *byebye_glide Jorg1*. In this case the qualities converge early to
the correct hypothesizes. The results for all trials are shown in Table 5.9. For
the *Effort* quality *Space* 9 results and for the *Effort* quality *Time* 3 results
are classified wrongly. This yields a recognition rate of 90% for the former
and 70% for the latter. This suggests that the low-level feature *speed gain
Acc* represents a stronger descriptor for *Time* than *curvature* for *Space*.

Table 5.9: Confusion table for classification of *Effort*.

Movement	1	2	Σ_e
1 direct	25	5	5
2 indirect	4	26	4
1 sudden	29	1	1
2 sustained	2	28	2
			12
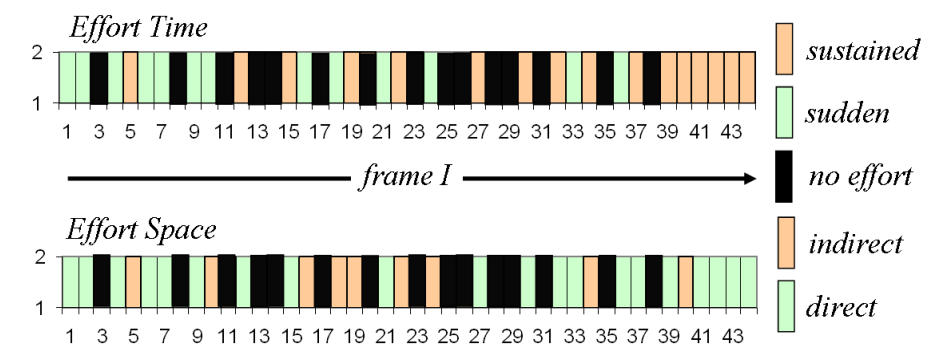
Figure 5.14: *Effort Time* and *Space* variables as a stream of hard evidences computed by the sub-model of the trial *byebye_ dab Diego2*.

5.8 Classification of bye-bye movements using only Effort

The next step is a model which classifies movements M given the results from the previous model as evidences. For this the OMAR system uses the function `best()` provided by the probabilistic library. The function returns the variable value with the highest probability. With this the resulting *Effort* values will be used as certain (hard) evidences.

Figure 5.14 shows the *Effort Time* and *Space* variables as a stream of hard evidences computed by the sub-model of the trial *byebye_ dab Diego2*. It can be seen that for the trial *byebye_ dab Diego2* not every frame the correct Effort qualities are calculated. The ideal stream should only consist of light green evidences, i.e. *Time.sudden* and *Space.direct*. The black rectangles represent frames with zero velocity where the probability remains constant. The computation of hard evidences using the function BEST() obscures the uncertainty of the (soft) Effort qualities. Figure 5.14 also shows that Effort Time becomes more sustained (light red) close to the end of the trial. This will finally also result in a change of belief concerning the type of movement. Figure 5.15 shows the probability evolution of the trial *byebye_ dab Diego2*. The system classifies correctly $M = byebye\_dab$ most of the time and only
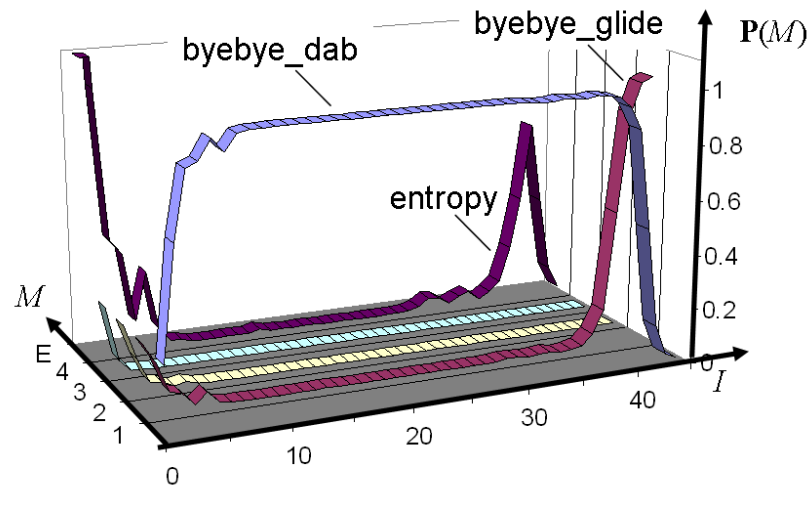
Figure 5.15: Evolution of the probabilities for movement M using Effort and Space of the trial *byebye_ dab Diego2*.

Table 5.10: Confusion table for classification of movements using *Effort*.

Movement	1	2	3	4	Σ_e
1 dab	13	3	2	2	7
2 glide		15			0
3 flick			12	3	3
4 float	4			11	4
					14

changes its belief in favor of $M = byebye\_glide$ in the last eight frames. This behavior can be observed for three more trials and seems to indicate the performers 'anticipation' of the end.

The results for all trials are shown in Table 5.10. By using *Effort* 14 of 60 trials are classified wrongly leaving a recognition rate of 77%. We can see that the recongition rate has improved slightly compared to using a *Space model*.

Table 5.11: Confusion table for classification of movements using *Space* and *Effort*.

Movement	1	2	3	4	Σ_e
1 dab	9	6			6
2 glide		15			0
3 flick			13	2	2
4 float		2		13	2
					10

5.9 Classification of bye-bye movements using Space, Effort and entropy

In this section the models for *Space* and *Effort* will be combined to evaluate the effect on the classification results. The *Joint model* which has been presented in section 4.9 fuses the evidences from the *Space* and *Effort* model. Like in the previous model each *Effort* quality uses only one low-level descriptor: *Time E.Ti* is associated to *Speed Gain Acc* and *Space E.Ti* is associated to *Curvature K*. The results are shown in Table 5.11. By using both, Space and Effort 10 of 60 trials are classified wrongly leaving a recognition rate of 83%. It appears that with a joint model of *Effort* and *Space* a better recognition rate can be obtained as with each of the single models.

Some of the trials that have been classified wrongly are of the type already shown in Fig. 5.15. The system classifies correctly most of the time and only changes its belief during last frames. A reasonable solution is to let the system decide when it is 'certain' about it's classification. The right measure for this certainty is the entropy. A value of 0.1 was chosen for the entropy to finish the classification. Table 5.12 shows the results for deciding based on a low entropy level (0.1). When deciding based on a low entropy level only 5 of 60 trials are classified wrongly leaving a recognition rate of 92%. The best recognition rate can be obtained by deciding at the first appearance of certainty.

Table 5.12: Confusion table for classification of movements using *Space* and *Effort* and by deciding based on low entropy.

Movement	1	2	3	4	Σ_e
1 dab	13	2			2
2 glide		15			0
3 flick			14	1	1
4 float		2		13	2
					5
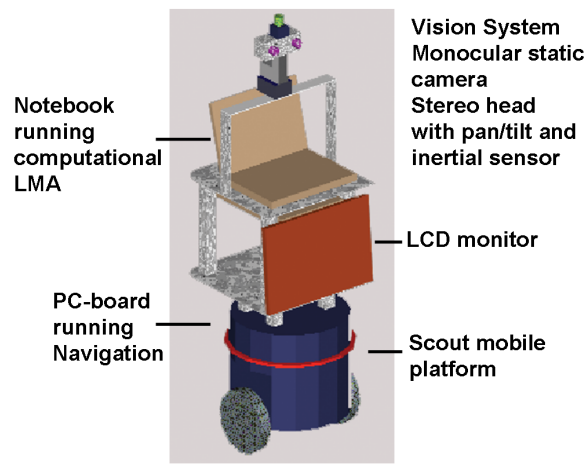
Figure 5.16: CAD drawing of the social robot Nicole.

5.10 Scenario Nicole@Play

The current target application for the continuous classification of movements is the social robot 'Nicole'. The social robot 'Nicole' is designed as a autonomous platform with which human-robot interaction can be investigated. Figure 5.16 shows the main parts of the platform. The vision system is mainly using a static monocular camera. The system for continuous classification of movements runs on a notebook where the process of tracking can be observed. An additional monitor can be used to present an avatar like an animated face or full body character. The navigation system is held by a PC
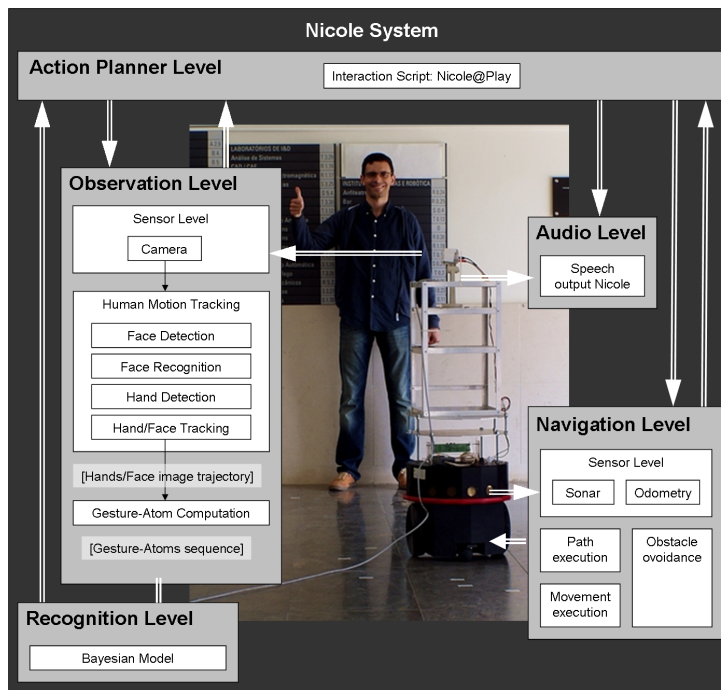
Figure 5.17: Architecture of the Nicole-System V1.0.

board inside the Scout platform.

The complete scenario 'Nicole@Play' includes also stationary PC which runs the script interaction. It holds way-points for navigation, asks for the presence of a person and the result of the movement recognition. The different processes exchange information by using 'sockets'. Figure 5.17 shows the functional order of our implemented modules inside the Nicole V1.0 architecture. Version 1.0 of the system architecture is a redefined version the gesture perception system (GP-System) presented in [RD05]. Apart from the already discussed modules for perception the system architecture also includes the *Action Planner* which controls the sequential execution of the tasks inside the interaction scenario. It holds the script that tells in which way the robot acts upon the perceptions. Table 5.13 shows how the recognized gestures are mapped to actions of the robot platform's navigation system. In our first trials Nicole was using audio outputs like asking for confirmation on

Table 5.13: Gesture-Action Mapping

	Recognized Gesture	Interpretation	Action
1	Draw Circle	Turn 360°	Rotation
2	Sagittal Waving	Come closer	Move forward
3	Horizontal Waving Left side	Step aside (left)	Move right
4	Horizontal Waving Right side	Step aside (right)	Move left
5	Waving Bye-Bye	Ignore last gesture	Wait
		Stop Interaction	Switch off system
6	Pointing gesture	Change godfather	Proceed to next way-point
		Acknowledgement	Perform Action

the recognized commands and robot movements. Movies of the trials can be downloaded from the project's web-page [RBD07]

The scenario was tested in a natural environment at the entrance hall of the University of Coimbra, Department of Electrical Engineering in June 2006. Figure 5.18 shows some of the states during the interaction. After Nicole has been called she navigates to the position where she expects the user (Phase 1: Long Distance Approach). She will then, look around in search for a person (Phase 2: User Search). The first person she detects will be approached (Phase 3: Short Distance Approach). After taking the optimal interaction position she will greet the user and ask for a gesture (Phase 4: Initiate Interaction). In the next phase Nicole will observe and anticipate the movement of the user's hand(s) (Phase 5: Tracking and Gesture Recognition). After being certain about the perceived gesture Nicole will perform a related action (e.g. turning around) (Phase 6: Action). After this Nicole will end up in phase two or three start all over again.

Following the taxonomy presented in [MG01] the features of the Nicole-system can be described as follows: The Initialization process establishes the probability distribution of 'skin' color and the initial position of head and hands. The Tracking process uses a figure-ground segmentation approach based on spatial data and blob statistics. A summary on the constraints

Figure 5.18: Nicole V1.0 interacting at the entrance of our department.

that are applied to the Nicole system are shown in Table H.1 of the annex.

Chapter 6

Discussion and Conclusions

Contents

6.1 Resume

This thesis started with the premise that the field of computational Human Movement Analysis is in need of an annotated database fur human movements. Low level features like 'velocity' can easily be extracted by machines and action descriptors like 'dabbing paint on a canvas' can easily be understood by humans. A good descriptive language needed to be chosen that provides the labels on a medium level in between features and actions. The second chapter of this thesis shows that Laban Movement Analysis (LMA) is a good choice for this descriptive language. The chapter presents a thorough overview over the properties and capabilities of each component and their relation to each other. This part concludes with examples from our Human Interaction Database (HID) to outline the applicability of LMA for an annotated database.

This work will to contribute to applications like 'social robots', 'smart

rooms' or 'rehabilitation'. The required technical solution brings together a monocular camera mounted on a mobile platform, monocular cameras mounted on the walls of a room and high precision data from an active sensor. The third chapter bases the computation of the low-level features on two very different sensor types, i.e. monocular camera and active sensor. The sensor data can be registered by calibrating the two devices which allows to work with a database of rich 3-D position data and sensory input from 2-D projections. The low-level features are extracted from trials of our database and the evaluation shows that these features are useful as evidences for LMA descriptors.

To extract the LMA descriptors automatically the Bayesian framework is used as presented in the fourth chapter. It i) presents the models as Bayesian nets which allows multidisciplinary evaluations, ii) can be designed in a modular fashion so the influence of each component can be studied and iii) takes into account that LMA is based on human observations where incompleteness and uncertainty are issues. It provides skills like anticipation and certainty to the system that gives the impression of being more 'human-like'.

The fifth chapter presents the implementation which proves the feasibility of this approach. The probabilistic approach provides the learned data in a way that allows its visual inspection and evaluation. With this, expected results for classification can be anticipated. The chosen histogram-based approach for learning provides a simple way of adding data points. The characteristics to add data at any time opens the possibility for a continuously learning artificial agent. The Human Interaction Database (HID) provides several sets of movements, e.g. 'gestures', 'expressive movements' and 'bye-byes'. The 'gestures' set 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 OMAR system results for movement classification can be presented and compared separately for single- and triple projection *Space*, *Effort* and joining *Space* and *Effort*. Also a 'stand-alone'

classification of the *Effort* qualities is possible. One observed characteristics of the implemented type of Bayesian classifier was that the system actually changed its belief while producing a peak of entropy. The recognition rate of the *Space* model improves in general when using evidences from all three planes (from 67% to 77%). Some movements can not be seen in certain planes, e.g. *nthrow* in the *Door Plane* π_v. The change of perspective decreases greatly the recognition rate. Which speakes in favour of our appoach to calibrate the monocular camera and the active sensor.

The evolution of the OMAR system for the 'bye-bye' set when using the different models is: i) A recognition rate of 75% when using only *Space*, ii) a recognition rate of 77% when using only *Effort*, iii) a recognition rate of 83% when using both and finally iv) a recognition rate of 92% when deciding based on a low entropy level, i.e. at the first appearance of certainty. To proof that the system could also perform well in a natural indoor environment a series of demonstration of the social robot 'Nicole' was conducted since Summer 2006 at various expositions.

6.2 Future Tasks

Through this work on computational LMA a new axis of research was created at the Mobile Robotics Lab (MLR) in Coimbra. The ongoing EU FP6 project Bayesian Approach to Cognitive Systems (BACS[1]) introduced since the second half of its duration research on computational LMA as part of the work package Human Body Motion Recognition and Behavior Analysis (WP 5.3.3). Ongoing works within these environments tackle the extension of the database of movements with annotated Laban Movement Analysis (LMA) descriptors by certain classes. A further extension is the implementation and evaluation of the *Shape model* and the estimation of 3-D trajectories from 2-D projections. The latter follows the line of research of computational LMA in multi-camera environments.

---

[1]BACS, `http://www.bacs.ethz.ch/`

The recently started EU FP7 project Prediction and Interpretation of Human Behavior Based on Probabilistic Structures and Heterogeneous Sensors (Prometheus[2]) was proposed with contributions from computational LMA. Especially the results from multi-camera systems can be transfered to the work packages of this project. These type of research projects aim toward applications like Smart Rooms which is in turn related to the recent FP7 effort to tackle the challenge of an aging society.

Computational LMA can be contributed whenever more advanced interfaces between human and machine are required. The research departments of some big enterprises have already shown interest to apply computational LMA to well-chosen problems. The cases of application range from systems which support medical surgery[3] to wearable computers for logistics and maintenance[4]. First contacts with institutions from the health sector (through IPN[5]), specialized in rehabilitation, point toward an employment of a socially assistive robot. There, the movements of a patient can be recorded and annotated with LMA descriptors. Through this the system will be able to present the progression of the rehabilitation process. This branch represents also a continuation of the successful collaboration between engineers and neuroscientists represented through the BACS project.

From the pool of ideas for future extensions and improvements the following issues present a short overview. Computational LMA will be applied to further movements and actions like manipulatory movements (e.g. drinking from a mug) and pedestrians walking. Additional sensors will be implemented and tested e.g. inertial sensor units [LD04]). Advanced Bayesian models will be designed and evaluated like a *Body model* including the geometric constraints of the *kinesphere*, a *Segmentation model* which uses the phases of a movement and a *Relationship model* that models bodily interaction between people. The question for classification will be extended toward the

---

[2]Prometheus,`http://paloma.isr.uc.pt/projects/prometheus/index.php`

[3]OrthoPilot, `http://www.orthopilot.com/index.cfm`

[4]SiWear, `http://www.siwear.de/`

[5]IPN-las, `http://www.las.ipn.pt/index.php`

recognition of persons. Advanced interfaces will show an animated character (avatar) to hold the user in the interaction process.

6.3 Future Technologies

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 comparisons 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. A successful research toward these goals must be based on the creation of a large database, the implementation of systems to collect this data, an interface design that appears 'natural' to the patient and an intensive multidisciplinary discussion and cooperation.

Appendix A

Computational Human Movement Analysis

The following Table A.1 shows constraints of assumptions, which are frequently found in the works contributing to area of 'Computational Human Movement Analysis' taken from [MG01].

Table A.1: Constraints frequently found in the area of 'Computational Human Movement Analysis'

Assumptions related to movements	Assumptions related to appearance
The subject remains inside the workspace	Environment
None or constant camera motion	Constant lighting
Only one person in the workspace at the time	Static background
The subject faces the camera at all time	Uniform background
Movements parallel to the camera-plane	Known camera parameters
No occlusion	Special hardware
Slow and continuous movements	
Only move one or a few limbs	Subject
The motion pattern of the subject is known	Known start pose
Subject moves on a flat ground plane	Known subject
	Markers placed on the subject
	Special colored clothes
	Tight-fitting clothes

Appendix B

Expressive movements

Taking into account the four *Drives* (i.e. *Action Drive*, **Spaceless**, *Weightless* and *Timeless*) of Laban Movement Analysis a number of $4 \times 8 = 32$ distinct movements is possible. The *Effort* qualities are given for the most expressive part of the movement, thus we have added a descriptor of the phase to the table. In our implementation the low-level features are defined for all three planes separately. This leads to different Effort estimates for the different planes. For this reason we have also included a descriptor for the principal plane in which the movement happens. We could have further distinguished that distinct body parts may have different *Effort* qualities. As in our system only two case occurred, whether both hands having similar qualities or one hand is not moving at all, we have neglected this fact. We have summarized prototypical movements we found in the literature [BL80], [Zha02] in the four Tables B.1, B.2, B.3, B.5 and B.4. Some of the examples are taken from Timothy Yontz [1].

We have only added those movements that produce curves in 3-D space with a certain length. This leaves out wrist-based movements like 'Wringing a towel' with *Indirect Space*, *Strong Weight* and *Sustained Time* or 'Squashing a fruit' with *Direct Space*, *Strong Weight* and *Sustained Time*. The tables

---

[1] www.timothyyontz.com

Table B.1: Expressive movements of *Action Drive* part 1 with *Effort*, principal plane π_{prin} and principal phase Ph_{prin}

Action	Space	Weight	Time	π_{prin}	Ph_{prin}
Punch	Direct	Strong	Sudden		
Forward punch				XY	Pre-Str
Pointing strong				XY	Pre-Str
Swatting a fly				YZ	Stroke
Slash	Indir.	Strong	Sudden		
Slash with a sword				YZ	Stroke
Crack with a heavy whip				YZ	Stroke
Swinging a baseball bat[1]				XZ	Stroke
Casting a fishing line[1]				XZ	Stroke
Golfing[1]				XZ	Stroke
Dab	Direct	Light	Sudden		
Pointing				XY	Pre-Str
Dabbing paint on a canvas				XY	Stroke
Lunging for a ball				XY	Pre-Str
Grabbing a child from a path of danger				XY	Pre-Str
Making a snap move				XY	Pre-Str
Typing[1]				YZ	Stroke
Playing darts[1]				XZ	Stroke
Flick	Indir.	Light	Sudden		
Cleaning with a brush				XY	Stroke
Removing an insect from a dress				XY	Stroke
Touching a hot stove[1]				XY	Stroke

Table B.2: Expressive movements of *Action Drive*, part 2 with *Effort*, principal plane π_{prin} and principal phase Ph_{prin}

Action	Space	Weight	Time	π_{prin}	Ph_{prin}
Press	Direct	Strong	Sust.		
Pushing a heavy object				XY	Stroke
Lifting a heavy object				YZ	Stroke
Pushing a child on a swing[1]				XY	Stroke
Wring	Indir.	Strong	Sust.		
Expressing a firmly held opinion				XY	Stroke
Stretch to yawn				YZ	Stroke
Wringing a towel[1]				YZ	Stroke
Heavy twisting off a bottle cap[1]				YZ	Stroke
Glide	Direct	Light	Sust.		
Ironing a shirt				XY	Stroke
Pointing light				XY	Stroke
Stroking a pet				XY	Stroke
Reaching to shake hands[1]				XY	Stroke
Erasing a blackboard[1]				YZ	Stroke
Float	Indir.	Light	Sust.		
Cradle a soap bubble				YZ	Stroke
Using a bubble wand [1]				YZ	Stroke
Spraying with air freshener [1]				YZ	Stroke

Table B.3: Expressive movements of *Spaceless* with *Effort*, principal plane π_{prin} and principal phase Ph_{prin}

Action	Flow	Weight	Time	π_{prin}	Ph_{prin}
Type A	Bound	Light	Sust.		
Threading a needle				YZ	Stroke
Pulling out a splinter				YZ	Stroke
Carrying a cup of hot tea				YZ	Stroke
Type B	Free	Light	Sudden		
Waving wildly				YZ	Stroke
Swing an object to fling it away				XY	Stroke
Type C	Free	Strong	Sudden		
Shaking of water				YZ	Stroke

Table B.4: Expressive movements of *Weightless* with *Effort*, principal plane π_{prin} and principal phase Ph_{prin}

Action	Flow	Space	Time	π_{prin}	Ph_{prin}
Type A	Free	Direct	Sudden		
Waving away bugs				YZ	Stroke
Type B	Bound	Direct	Sust.		
Carefully tracing a drawing				YZ	Stroke

for **Spaceless**, *Weightless* and *Timeless* are sparse as not all possible combinations could be filled with an appropriate action.

Table B.5: Expressive movements of *Timeless* with *Effort*, principal plane π_{prin} and principal phase Ph_{prin}

Action	Flow	Space	Weight	π_{prin}	Ph_{prin}
Type A	Free	Indirect	Light		
Conducting an orchestra				YZ	Stroke

Appendix C

Human Interaction Database

The following figures show some image sequences, which are part of the HID. The movement 'Byebye' is a one-hand, oscillating gesture. For each movement several trials (4 to 6) were performed by several persons (2 to 3) the following figure shows the trajectories for all trails of 'byebye'.

Figure C.1: All trials of Alberto (A) performing a Bye-bye gesture.

Figure C.2: All trials of Davim (D) performing a Bye-bye gesture.
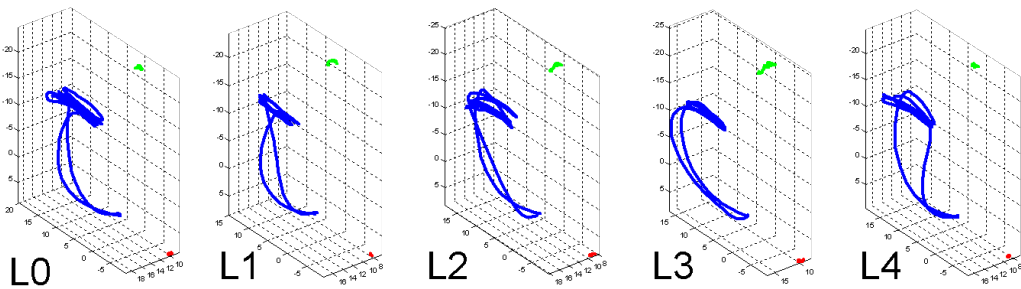
Figure C.3: All trials of Luis (L) performing a Bye-bye gesture.

The movement 'Maestro' is a one-hand, oscillating movement. For each movement several trials (4 to 6) were performed by several persons (2 to 3) the following figure shows the trajectories for all trails of 'Maestro'.
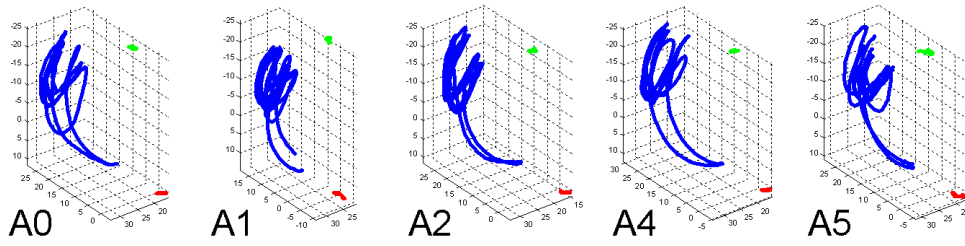
Figure C.4: All trials of Alberto (A) performing a Maestro gesture.
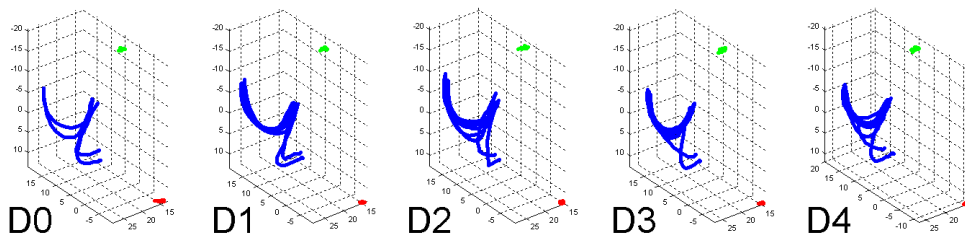
Figure C.5: All trials of Davim (D) performing a Maestro gesture.
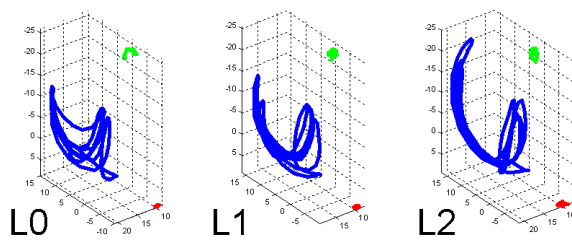
Figure C.6: All trials of Luis (L) performing a Maestro gesture.

The movement 'Ok' is a one-hand, non-oscillating gesture. For each movement several trials (4 to 6) were performed by several persons (2 to 3) the following figure shows the trajectories for all trails of 'Ok'.
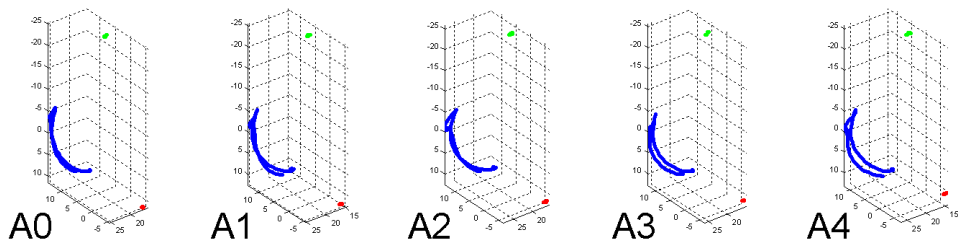
Figure C.7: All trials of Alberto (A) performing a Ok gesture.
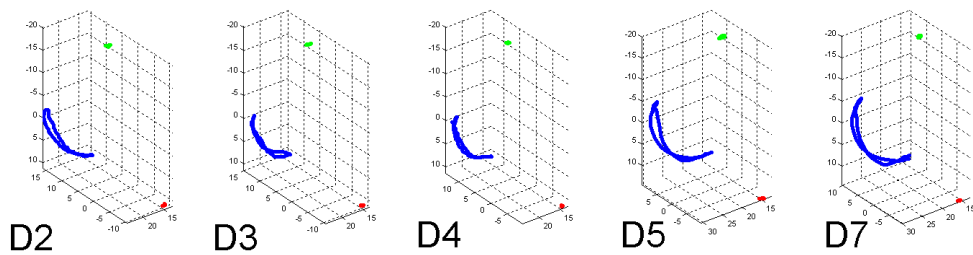
Figure C.8: All trials of Davim (D) performing a Ok gesture.
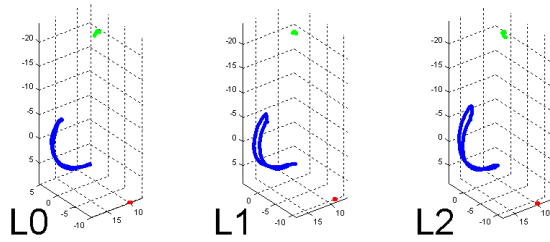
Figure C.9: All trials of Luis (L) performing a Ok gesture.

The movement 'Point' is a one-hand, non-oscillating movement. For each movement several trials (4 to 6) were performed by several persons (2 to 3) the following figure shows the trajectories for all trials of 'Point'.
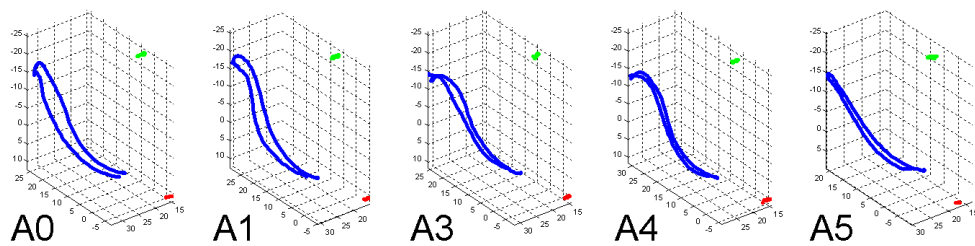
Figure C.10: All trials of Alberto (A) performing a Point gesture.
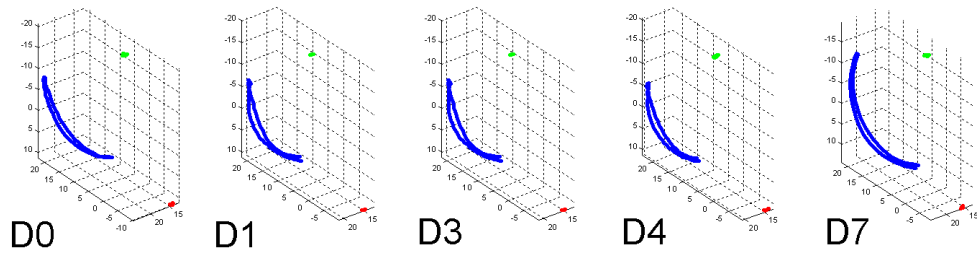
Figure C.11: All trials of Davim (D) performing a Point gesture.
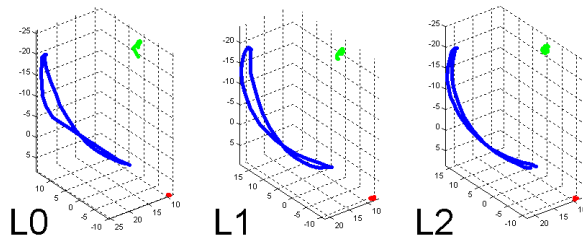
Figure C.12: All trials of Luis (L) performing a Point gesture.

The movement 'Stretch' is a two-hand, non-oscillating gesture. For each movement several trials (4 to 6) were performed by several persons (2 to 3) the following figure shows the trajectories for all trials of 'Stretch'.
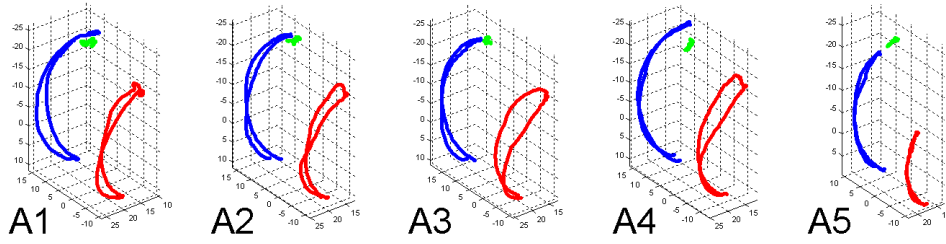
Figure C.13: All trials of Alberto (A) performing a Stretch gesture.
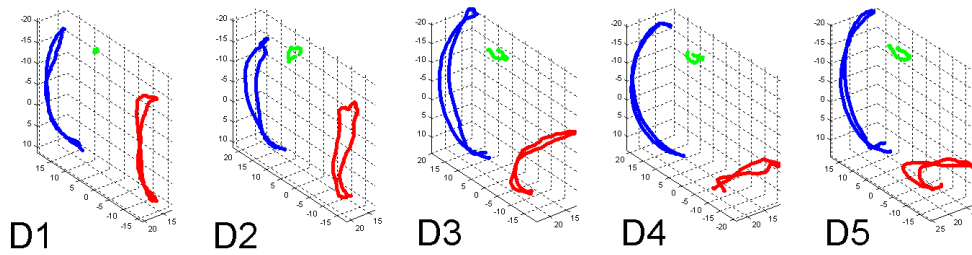
Figure C.14: All trials of Davim (D) performing a Stretch gesture.
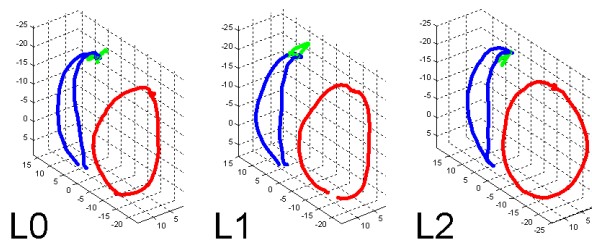
Figure C.15: All trials of Luis (L) performing a Stretch gesture.

The movement 'Lunging' is a two-hand, non-oscillating gesture. For each movement several trials (4 to 6) were performed by several persons (2 to 3) the following figure shows the trajectories for all trials of 'Lunging'.
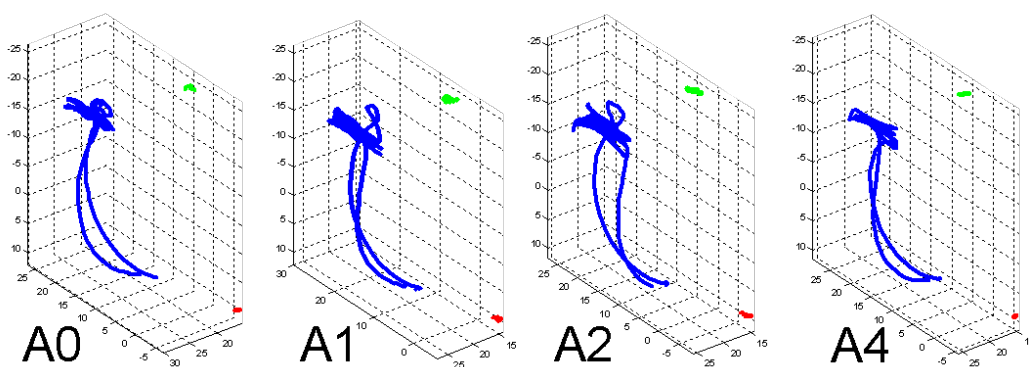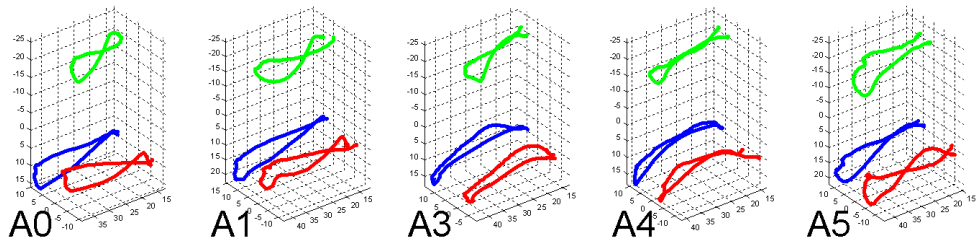

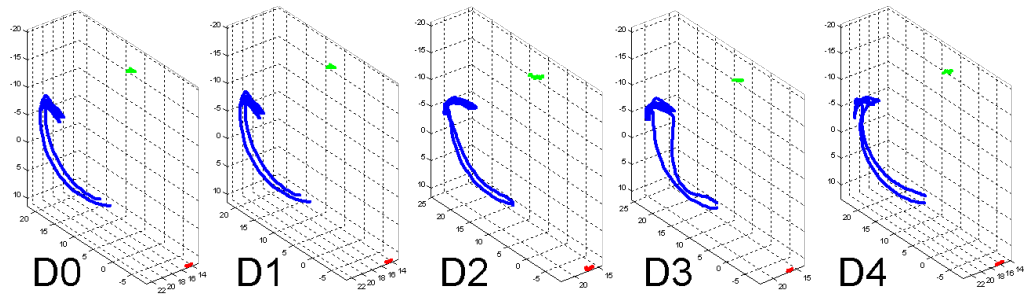Figure C.16: All trials of Alberto (A) performing a Lunging gesture.
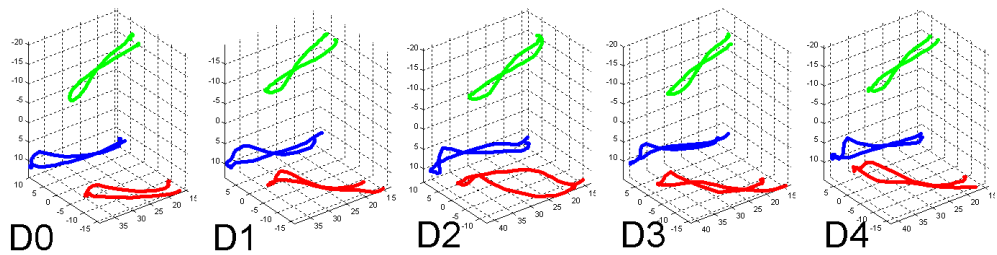
Figure C.17: All trials of Davim (D) performing a Lunging gesture.
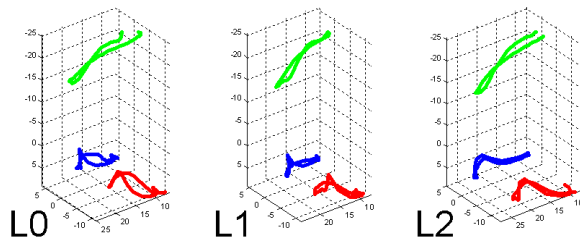
Figure C.18: All trials of Luis (L) performing a Lunging gesture.

The movement 'Nthrow' is a one-hand, oscillating gesture. For each movement several trials (5) were performed by one person the following figure shows the trajectories for all trials of 'Nthrow'.
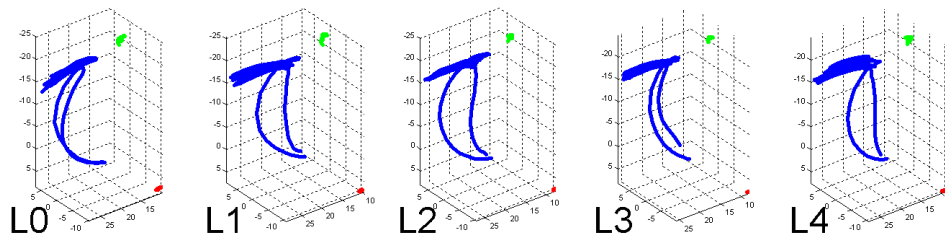
Figure C.19: All trials of Luis (L) performing a Nthrow gesture.

The movement 'Punch' is a two-hand, non-oscillating gesture. For each movement several trials (5) were performed by one person the following figure shows the trajectories for all trials of 'Punch'.
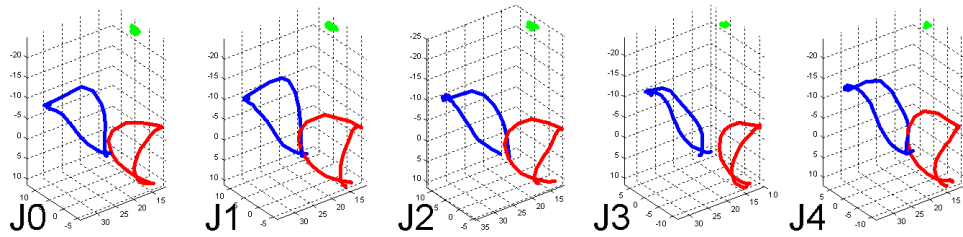
Figure C.20: All trials of myself (J) performing a punch.

The Bye-Bye movement was performed in different styles. For each style several trials (5) were performed by one person. Figure C.21 shows the trajectories for all trials of 'Bye-Bye Dab'.
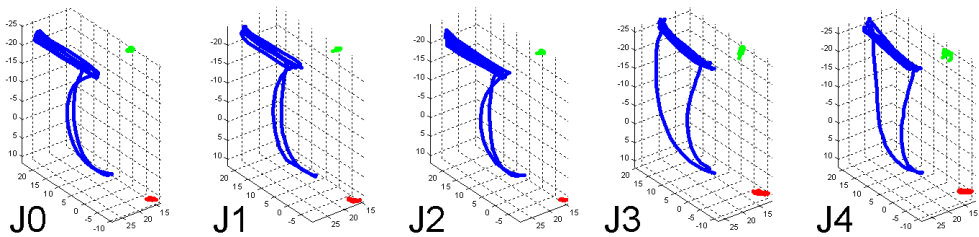
Figure C.21: All trials of myself (J) performing a Bye-Bye Dab movement.

Figure C.22 shows the trajectories for all trials of 'Bye-Bye Flick'.
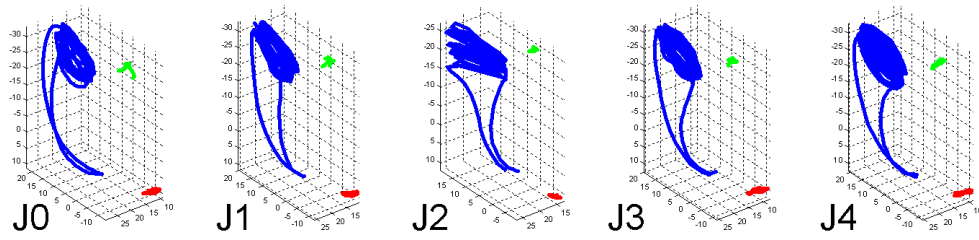
Figure C.22: All trials of myself (J) performing a Bye-Bye Flick movement.

Figure C.23 shows the trajectories for all trials of 'Bye-Bye Float'.
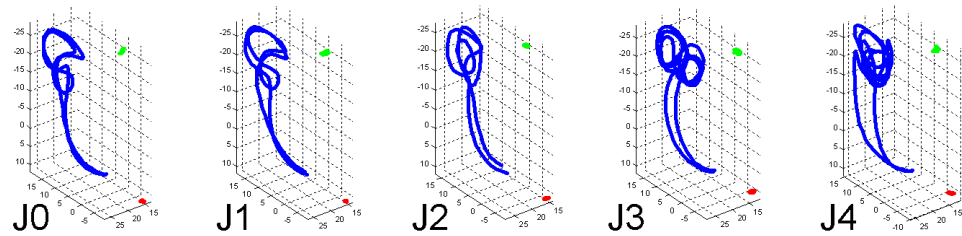
Figure C.23: All trials of myself (J) performing a Bye-Bye Float movement.

Figure C.24 shows the trajectories for all trials of 'Bye-Bye Glide'.
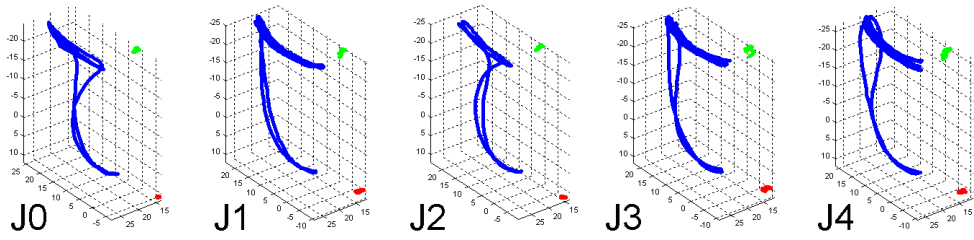
Figure C.24: All trials of myself (J) performing a Bye-Bye Glide movement.

The following figure shows representative trials to compare the traces of the 2-D and 3-D trajectories.

byebye-A2   maestro-A2

ok-A2   point-A3

stretch-A2   shake-A2

lunging-A3   nthrow-L2
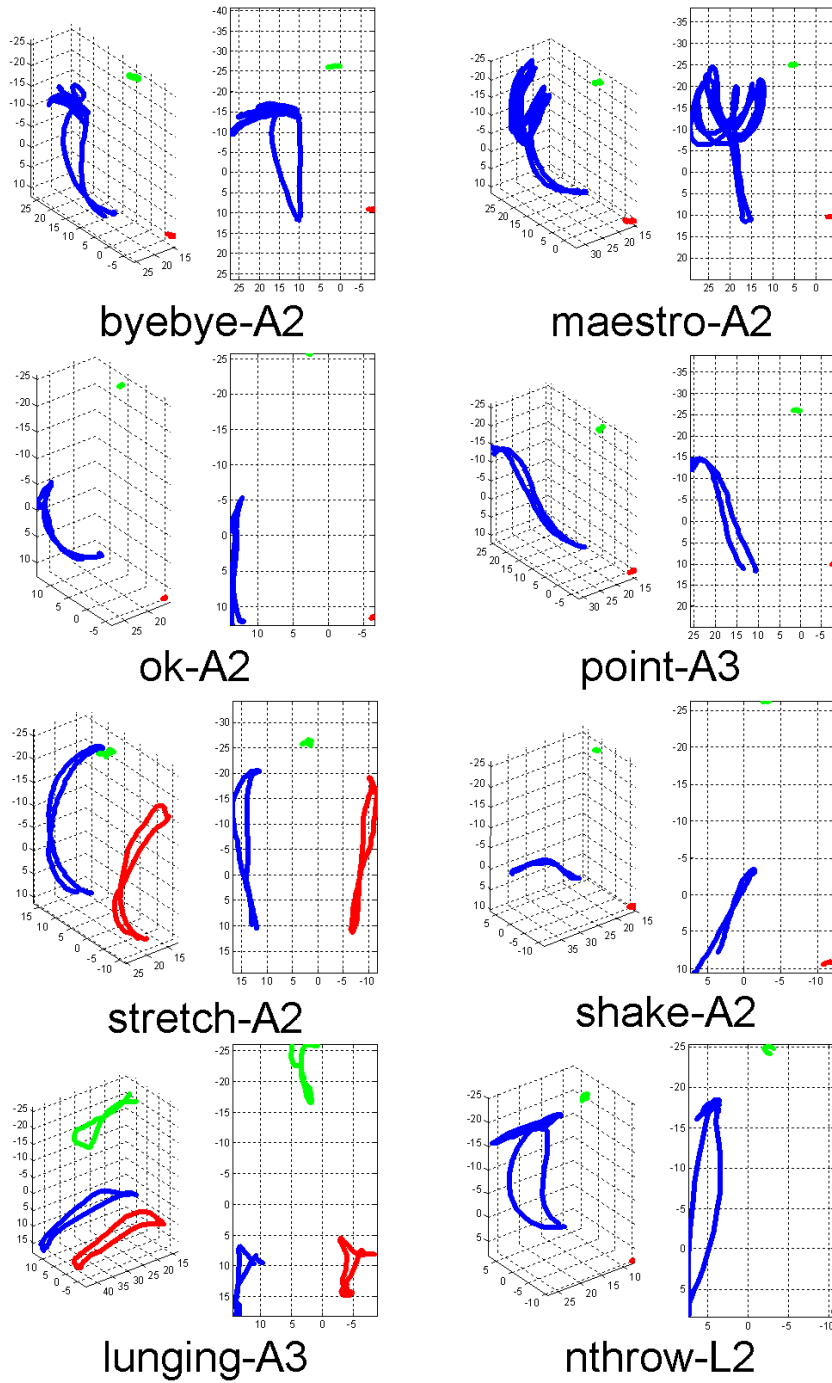
Figure C.25: 2D vs. 3D.

Table C.1: Characteristics of our gesture-set used to interact with the social
robot Nicole.

No.	Gesture	Hands	Level
1	Sagittal Waving	Two	High
2	Waving to Left	Two	Medium
3	Waving to Right	Two	Medium
4	Waving Bye-Bye	One	High
5	Pointing	One	High
6	Draw Circle	One	Medium

The set of gestures that has been used for interaction with our social
robot Nicole is shown in Table C.1.

Appendix D

Motion capture device

The orientation is expressed in Euler angles, i.e. three rotations stated in the inverse (clockwise) sense in each orthogonal axis which align $\{M_T\}$ with $\{M_R\}$. The first rotation α around the Z-axis is followed by a rotation β around Y and concluded with a rotation γ around the X-axis. Figure D.1 shows the relation between Euler angles and the 'Roll-Pitch-Yaw' convention. The position is given by of the sensor referential's origin measured relatively to $\{M_T\}$, $(x; y; z)$. The transformation between the transmitter $\{M_T\}$ and the sensor (receiver) $\{M_R\}$ can be determined by a rotation with the Euler angles starting with γ and ending with α and then translating the origin of $\{M_R\}$ to coincide with $\{M_T\}$. The rotations can be expressed through three
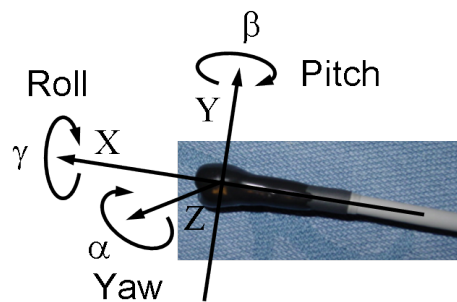
Figure D.1: Polhemus 'teardrop' sensor with rotations defined by 'Roll-Pitch-Yaw', and Euler angles.

matrices as shown in (D.1).

$$\mathbf{R}_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma & \cos\gamma \end{bmatrix} \mathbf{R}_Y = \begin{bmatrix} \cos\gamma & 0 & \sin\gamma \\ 0 & 1 & 0 \\ -\sin\gamma & 0 & \cos\gamma \end{bmatrix}$$

$$\mathbf{R}_Z = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{D.1}$$

The full transformation is then given by

$${}^{\{M_T\}}\mathbf{T}_{\{M_R\}} = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & x \\ r_{2,1} & r_{2,2} & r_{2,3} & y \\ r_{3,1} & r_{3,2} & r_{3,3} & z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{D.2}$$

with

$$
\begin{aligned}
r_{1,1} &= \cos(\alpha)\cos(\beta) \\
r_{1,2} &= \cos(\alpha)\sin(\beta)\sin(\gamma) - \sin(\alpha)\cos(\gamma) \\
r_{1,3} &= \cos(\alpha)\sin(\beta)\cos(\gamma) + \sin(\alpha)\sin(\gamma) \\
r_{2,1} &= \sin(\alpha)\cos(\beta) \\
r_{2,2} &= \sin(\alpha)\sin(\beta)\sin(\gamma) + \cos(\alpha)\cos(\gamma) \\
r_{2,3} &= \sin(\alpha)\sin(\beta)\cos(\gamma) - \cos(\alpha)\sin(\gamma) \\
r_{3,1} &= -\sin(\beta) \\
r_{3,2} &= \cos(\beta)\sin(\gamma) \\
r_{3,3} &= \cos(\beta)\cos(\gamma)
\end{aligned}
\tag{D.3}
$$

Appendix E

Discretization of LLF

Discretization of speed, speed gain and curvature. The discretization through thresholds is defined in the file 'learnCam.cpp'. Table E.1 shows the thresholds for speed, speed gain and curvature.

Vector Symbols (Atoms) A,B and C. The calculation of the direction symbols A,B and C are based on the displacements in X, Y and Z direction. For each of the three planes we calculate the displacement angle using the atan2 function. The angles are discretized in 8 sectors, resulting in a $45\,\mathrm{deg}$ resolution and coded with integers in the range $\{2\ldots 9\}$. the integer value 1 is reserved for a non-displacement. The integers are later decoded to symbols e.g. value 3 in the xy-plane relates to forward-left (FL).

Table E.1: Thresholds for speed, speed gain and curvature.

Feature	Value	Thresholds
1 Vel	zero	$\|Vel\| \leq 2.5$
	slow	$2.5 \leq \|Vel\| \leq 13$
	medium	$13 \leq \|Vel\| \leq 21$
	fast	$21 \leq \|Vel\|$
2 Acc	no	$\|Acc\| \leq 25$
	low	$25 \leq \|Acc\| \leq 50$
	medium	$50 \leq \|Acc\| \leq 100$
	high	$100 \leq \|Acc\|$
3 K	zero	$\|K\| \leq 7.5$
	small	$7.5 \leq \|K\| \leq 22.5$
	medium	$22.5 \leq \|K\| \leq 40$
	big	$40 \leq \|K\| \leq 180$
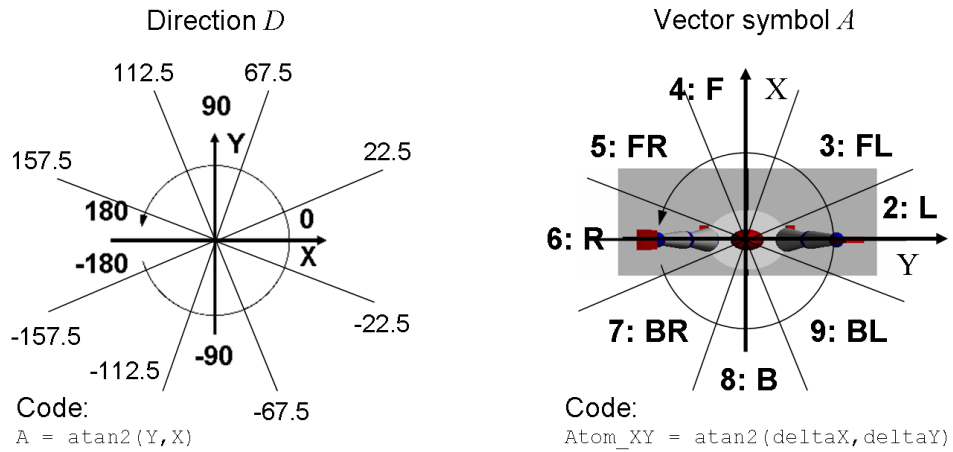
Figure E.1: The *Direction D* and *Vectors Symbols* (*Atoms*) *A* using the 'atan2' function.
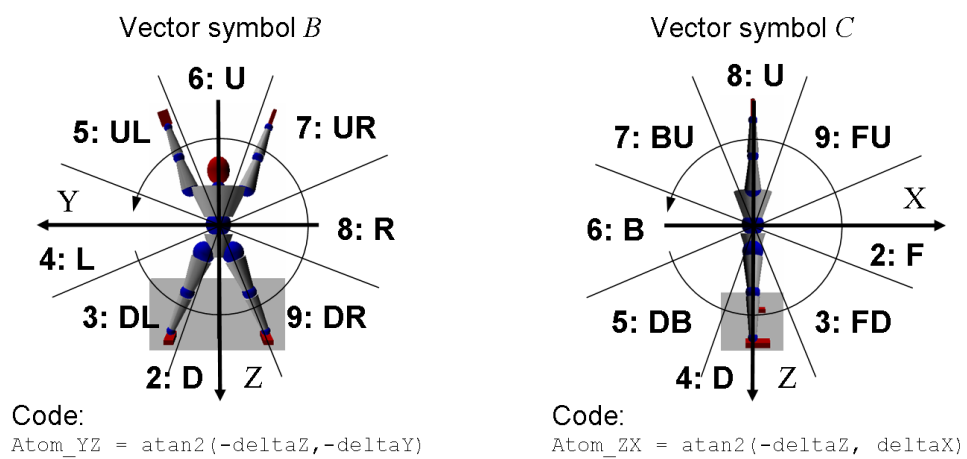
Figure E.2: *Vectors Symbols (Atoms) B and C.*

Appendix F

Bayesian Models for LMA

The following model includes the *Space* and *Effort model* presented in Chapter 4 plus a model for segmentation as shown in Fig. F.2. The model also regards the appearance of uncertain evidences through the function `best()` or a coherence variable c. Boxes above the arcs show the probabilistic terms for learning and recognition.

The upper part of the complete Bayesian model for Laban Movement Analysis is shown in Fig. F.1. The middle part of the complete Bayesian model is shown in Fig. F.2. The lower part of the complete Bayesian model is shown in Fig. F.3.
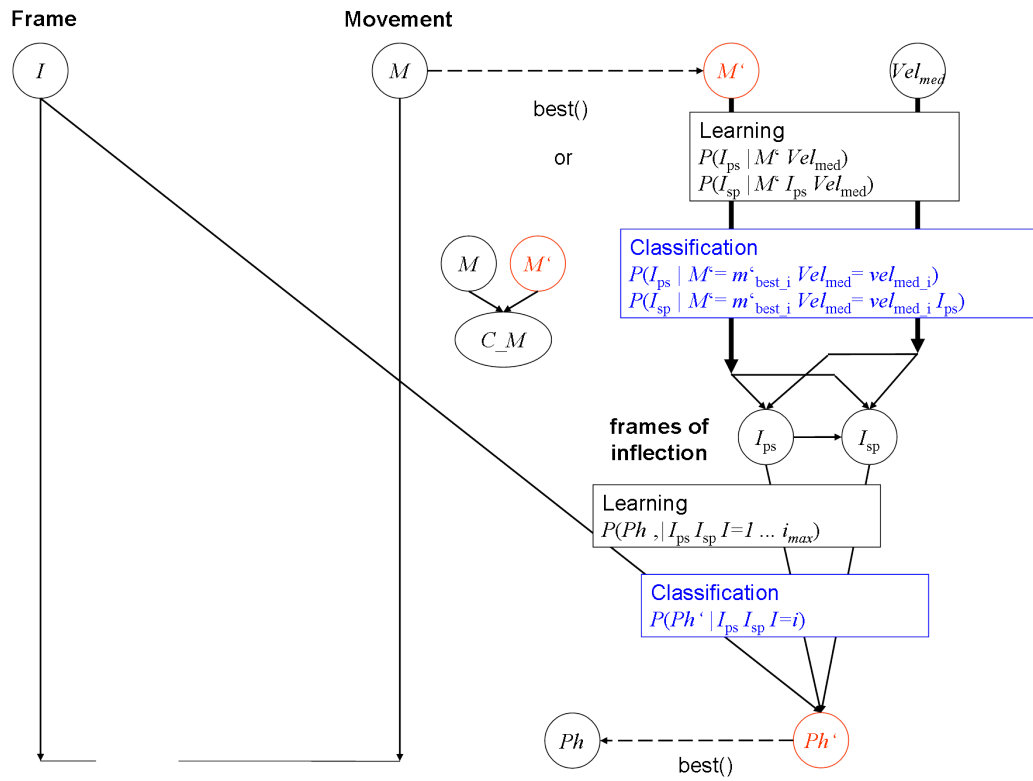
Figure F.1: Upper part of the full Bayesian model.
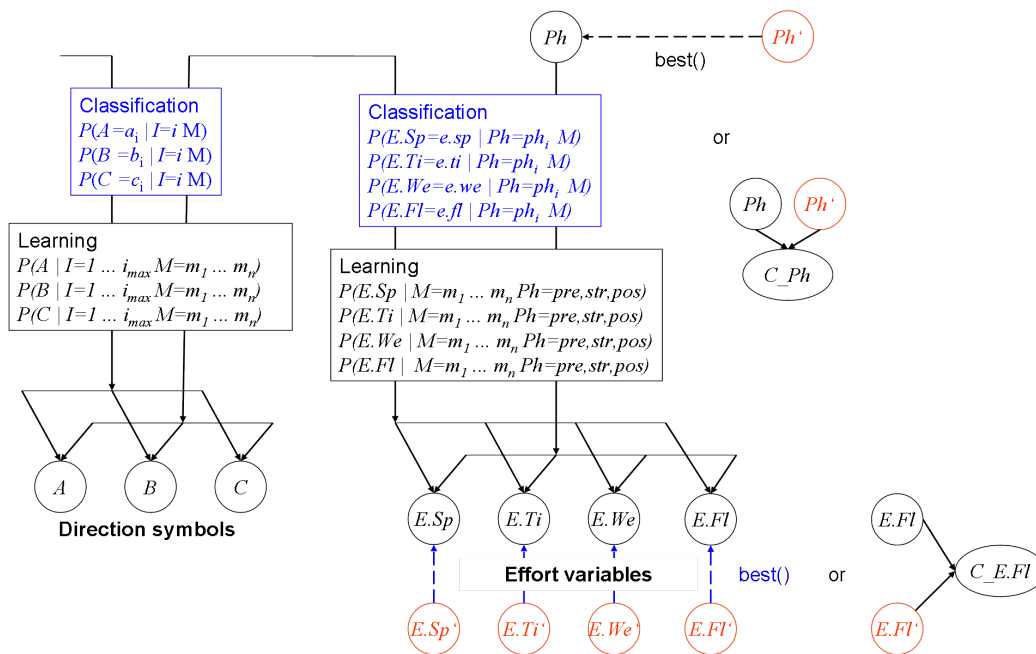
Figure F.2: Middle part of the full Bayesian model.
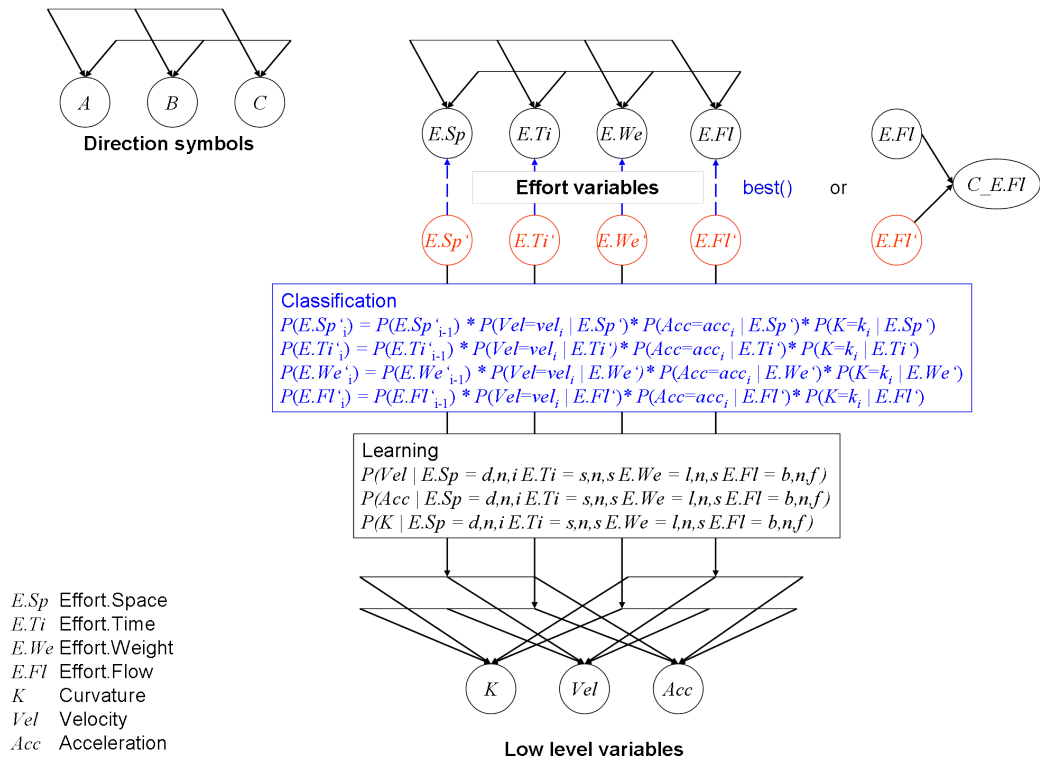
Figure F.3: Lower part of the full Bayesian model.

Appendix G

Classification results for 2-D and 3-D

Comparing the results of classification when using the displacements symbols (atoms) of all the tree planes A, B, C to the results gained when only the atom B of the vertical plane π_v is used.

Table of movements

1. Lunging: Lunging for a ball

2. Maestro: Conducting an orchestra

3. Stretch: Stretch to yawn

4. Ok: OK-sign gesture

5. Point: Pointing gesture

6. Byebye: Waving bye-bye

7. Shake: Reach for someone's hand

8. Nthrow: Waving sagittally (approach sign)

The Tables G.1 and G.2 shows the results for using the B atoms of the vertical plane π_v. The sum of all numbers in each row usually adds up

Table G.1: First part of the confusion table using only the 2-D (vertical) atoms with known trials and persons (kk), unknown trials but known person (uk) and unknown person and unknown trials (uu).

Movement	1			2			3			4			...
Person	k	u	u	k	u	u	k	u	u	k	u	u	...
Trial	k	k	u	k	k	u	k	k	u	k	k	u	...
1 lunging	5	2								1	1	3	...
2 maestro				3		2							...
3 stretch							6	4	2				...
4 ok										5	3		...
5 pointing												1	...
6 byebye													...
7 shake										1	2	1	...
8 nthrow										4			...
													...

Table G.2: Second part of the confusion table using only the 2-D (vertical) atoms with known trials and persons (kk), unknown trials but known person (uk) and unknown person and unknown trials (uu).

Movement	5			6			7			8			Σ_e
Person	k	u	u	k	u	u	k	u	u	k	u	u	
Trial	k	k	u	k	k	u	k	k	u	k	k	u	
1 lunging										1			6
2 maestro				3	4	1							8
3 stretch									1				1
4 ok		1					1		3				5
5 pointing	6	4							1				2
6 byebye				6	4	3							0
7 shake							5	2	2				4
8 nthrow							1						5
													31

Table G.3: First part of the confusion table using 3D (all planes) atoms with known trials and persons (kk), unknown trials but known person (uk) and unknown person and unknown trials (uu).

Movement	1			2			3			4			Σ_e
Person	k	u	u	k	u	u	k	u	u	k	u	u	
Trial	k	k	u	k	k	u	k	k	u	k	k	u	
1 lunging	6	4	1									1	2
2 maestro				2									11
3 stretch							6	4	2			1	1
4 ok										5	4		4
5 pointing													2
6 byebye													0
7 shake												1	1
8 nthrow												1	0
													21

to thirteen, though some movements have fewer trials. The sub column (kk) indicates trials that have been used for learning, which means that the trials itself and thus the person is known to the system (known, known = kk). A person is considered to be known if at least one movement pattern for a particular movement has been learned. A person might be known for one movement but unknown for another. The usual number for (kk) trials (training-sequences) is six, though some movements have less. The case in which the movement pattern of a person is known but the particular trial has not been used for learning is indicated by (uk). The usual number for (uk) trials (test-sequences) is four. The last case in which the person has never contributed a movement pattern for the learning process is called (uu). The usual number for (uu) trials (test-sequences) is three.

Table G.4: Second part of the confusion table using 3D (all planes) atoms with known trials and persons (kk), unknown trials but known person (uk) and unknown person and unknown trials (uu).

Movement	5			6			7			8			Σ_e
Person	k	u	u	k	u	u	k	u	u	k	u	u	
Trial	k	k	u	k	k	u	k	k	u	k	k	u	
1 lunging									1				2
2 maestro				4	4	3							11
3 stretch													1
4 ok							1		3				4
5 pointing	6	4							2				2
6 byebye				6	4	3							0
7 shake							6	4	2				1
8 nthrow										5			0
													21

Appendix H

Nicole system

Table H.1: Constraints that apply to the Nicole-system

Constraints	Nicole V1
Movement	
The subject remains inside the workspace	Yes
None or constant camera motion	Yes
Only one person in the workspace at the time	Yes
The subject faces the camera at all time	Yes
Movements parallel to the camera-plane	Yes
No occlusion	Yes
Speed of movements	Slow - Normal
Continuous movements	No
Only move one or a few limbs	One/Two hands
The motion pattern of the subject is known	No
Subject moves on a flat ground plane	Stands still
Environment	
Constant lighting	No
Static background	No[1]
Uniform background	No
Known camera parameters	No
Special hardware	No
Subject	
Known start pose	Yes
Known subject	No
Markers placed on the subject	No
Special colored clothes	No[1]
Tight-fitting clothes	No

Bibliography

[AC99]   J. K. Aggarwal and Q. Cai, *Human motion analysis: A review*, CVIU **73** (1999), no. 3, 428–440.

[AML04]   P. Ahrendt, A. Meng, and J. Larsen, *Decision time horizon for music genre classification using short time features*, EUSIPCO (Vienna, Austria), sep 2004, pp. 1293–1296.

[BB83]   R. Benesh and J. Benesh, *Reading dance: The birth of choreology.*, McGraw-Hill Book Company Ltd, 1983.

[BCF⁺98]   Wolfram Burgard, Armin B. Cremers, Dieter Fox, Dirk Hahnel, Gerhard Lakemeyer, Dirk Schulz, Walter Steiner, and Sebastian Thrun, *The interactive museum tour-guide robot*, AAAI/IAAI, 1998, pp. 11–18.

[BCF⁺99]   _____, *Experiences with an interactive museum tour-guide robot*, Artificial Intelligence **114** (1999), no. 1-2, 3–55.

[BG03]   Pierre Bessière and BIBA-INRIA Research Group, *Survey: Probabilistic methodology and techniques for artefact conception and development*, Tech. report, INRIA Rhône-Alpes, 2003.

[BL80]   I. Bartenieff and D. Lewis, *Body movement: Coping with the environment*, Gordon and Breach Science, New York, 1980.

[BPW93]   N. I. Badler, C. B. Phillips, and B. L. Webber, *Simulating humans: Computer graphics, animation, and control*, Oxford Univ. Press, 1993.

[Bra98]   Gary R. Bradski, *Computer vision face tracking for use in a perceptual user interface*, Intel Technology Journal **2** (1998), no. 2, 15.

[Bre97]   C. Bregler, *Learning and recognizing human dynamics in video sequences*, Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, 1997.

[Bre03]   Cynthia Breazeal, *Toward sociable robots*, Robotics and Autonomous Systems **42** (2003), 167–175.

[CCZB00]   Diane Chi, Monica Costa, Liwei Zhao, and Norman Badler, *The emote model for effort and shape*, SIGGRAPH 00, Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH, ACM Press, July 2000, pp. 173–182.

[CSJ07]   Yinpeng Chen, Hari Sundaram, and Jodi James, *A computational estimate of the physical effort in human poses*, Advances in Multimedia Modeling, Lecture Notes in Computer Science, vol. 4352, Springer Berlin, 2007, pp. 393–404.

[CT91]   T. Cover and J. Thomas, *Elements of information theory*, Wiley Series in Telecommunications, John Wiley & Sons, 1991.

[Dau98]   Kerstin Dautenhahn, *The art of designing socially intelligent agents: science, fiction and the human in the loop.*, Applied Artificial Intelligence **12** (1998), 573–617.

[DBM03]   Julien Diard, Pierre Bessière, and Emmanuel Mazer, *A survey of probabilistic models, using the bayesian programming methodology as a unifying framework*, Proc. of the Int. Conf.

on Computational Intelligence, Robotics and Autonomous Systems (Singapore (SG)), December 2003.

[Dia94]   J. Dias, *Reconstrução tridimensional utilizando visão dinâmica*, Ph.D. thesis, University of Coimbra, Portugal, 1994.

[ENUT06]   C. Eberst, H. Nohmayer, G. Umegher, and M. Takagi, *Towards programming robots by gestures, test-case: programming bore inspection for small lotsizes*, IEEE International Conference on Robotics and Automation, ICRA, 2006, pp. 333–338.

[EW58]   N. Eshkol and A. Wachmann, *Movement notation*, Weidenfield and Nicholson, 1958.

[FMH02]   M. Fabri, D. Moore, and D. Hobbs, *Expressive agents: Non-verbal communication in collaborative virtual environments*, Autonomous Agents and Multi-Agent Systems, 2002.

[FND03]   T. Fong, I. Nourbakhsh, and K. Dautenhahn, *A survey of socially interactive robots*, Robotics and Autonomous Systems **42** (2003), 143–166.

[Fri69]   N. Frijda, *Recognition of emotion*, Advances in Experimental Social Psychology **4** (1969), 167–223.

[FSM05]   D. Feil-Seifer and M. J. Mataric, *Defining socially assistive robotics*, IEEE 9th International Conference on Rehabilitation Robotics, 2005.

[FW06]   Afra Foroud and Ian Q. Whishaw, *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** (2006), 137–149.

[Gav99]   D. M. Gavrila, *The visual analysis of human movement: A survey*, CVIU **73** (1999), no. 1, pp. 82–98.

[Gol76]   I. Golani, *Homeostatic motor processes in mammalian interactions: a choreography of display*, Perspectives in Ethology, New York: Plenum Press. **2** (1976), 69–134.

[Gue89]   A. Hutchinson Guest, *Choreographics: a comparison of dance notation systems from the fifteenth century to the present.*, Routledge, 1989.

[HNT+01]   M. Hattori, S. Nishizawa, S. Tadokoro, T. Takamori, and K. Yamada, *The description of bunraku puppet's motion in computer based on labanotation*, 10th IEEE International Workshop on Robot and Human Interactive Communication, 2001, pp. 128–133.

[HS97]   J. Heikkila and O. Silven, *A four-step camera calibration procedure with implicit image correction*, CVPR97, 1997, pp. 1106–1112.

[Hut70]   A. Hutchinson, *Labanotation or kinetography laban*, Theatre Arts, New York, 1970.

[HZ00]   R.I. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, 2000.

[Int]   Intel, *Opencv reference manual.*

[Joh73]   G. Johansson, *Visual perception of biological motion and a model for its analysis*, Perception and Psychophysics **14** (1973), no. 2, 201–211.

[JT81]   Ollie Johnston and Frank Thomas, *The illusion of life: Disney animation*, Abbeville Press, 1981.

[Kal60]   R. E. Kalman, *A new approach to linear filtering and prediction problems*, Trans. ASME—J.Basic Eng. **82** (1960), 35–45.

[Ken04]   Adam Kendon, *Gesture: Visible action as utterance*, Cambridge University Press, 2004.

[KH04]   Sara Kiesler and Pamela Hinds, *Introduction to this special issue on human-robot interaction*, Human-Computer Interaction **19** (2004), no. 1/2, 1–8.

[KP04]   David C. Knill and Alexandre Pouget, *The bayesian brain: the role of uncertainty in neural coding and computation*, TRENDS in Neurosciences **27** (2004), 712–719.

[KSPF96]   R. E. Kahn, M. J. Swain, P. N. Prokopowicz, and R. J. Firby, *Gesture recognition using the perseus architecture*, IEEE International Conference on Computer Vision and Pattern Recognition, 1996.

[KYS02]   S. Kettebekov, M. Yeasin, and R. Sharma, *Prosody based co-analysis for continuous recognition of coverbal gestures*, International Conference on Multimodal Interfaces (ICMI'02) (Pittsburgh, USA), 2002, pp. 161–166.

[Lab66]   Rudolf Laban, *Choreutics*, MacDonald & Evans., London, 1966.

[LBDM04]   Olivier Lebeltel, Pierre Bessière, Julien Diard, and Emmanuel Mazer, *Bayesian robot programming*, Autonomous Robots **16** (2004), no. 1, 40 – 79.

[LD04]   Jorge Lobo and Jorge Dias, *Inertial sensed ego-motion for 3d vision*, Journal of Robotic Systems **21** (2004), 3–12.

[LDC06]   Vijay Laxmi, R.I. Damper, and J.N. Carter, *Human and machine perception of biological motion*, Cognitive Systems Research **7** (2006), 339–356.

[LM92]   Meredith Ellis Little and Carol G. Marsh, *La danse noble, an inventory of dances and sources*, Broude Brothers Ltd, 1992.

[LM02]   Rainer Lienhart and Jochen Maydt, *An extended set of haar-like features for rapid object detection.*, IEEE International Conference on Image Processing, vol. 1, 2002, pp. 900–903.

[Loe01]   Gerald E. Loeb, *Learning from the spinal cord*, Journal of Physiology **533.1** (2001), 111–117.

[Lon96]   J. S. Longstaff, *Cognitive structures of kinesthetic space; reevaluating rudolf laban's choreutics in the context of spatial cognition and motor control*, Ph.D. thesis, City University, London Human Movement Studies, Laban Centre, London, 1996.

[Lon01]   _____, *Translating vector symbols from laban's (1926) choreographie*, 26. Biennial Conference of the International Council of Kinetography Laban, ICKL, Ohio, USA, 2001, pp. 70–86.

[MEFSW07]   Maja J. Mataric, Jon Eriksson, David J. Feil-Seifer, and Carolee J. Winstein, *Socially assistive robotics for post-stroke rehabilitation*, Journal of NeuroEngineering and Rehabilitation **4** (2007), no. 5, 1–9.

[MG01]   Thomas B. Moeslund and Erik Granum, *A survey of computer vision-based human motion capture*, CVIU **81** (2001), no. 3, 231–268.

[MHK06]   T.B. Moeslund, A. Hilton, and V. Kruger, *A survey of advances in vision-based human motion capture and analysis*, CVIU **103** (2006), no. 2-3, 90–126.

[MLD06]   Paulo Menezes, Frédéric Lerasle, and Jorge Dias, *Data fusion for 3d gestures tracking using a camera mounted on a robot*, International Conference on Pattern Recognition, August 2006.

[MN03]   Thomas B. Moeslund and Lau Norgard, *A brief overview of hand gestures used in wearable human computer interfaces*, Tech. report, Computer Vision and Media Technology Lab., Aalborg University, DK, 2003.

[MST⁺97]   Hiroshi MIZOGUCHI, Tomomasa SAT0, Katsuyuki TAKAGI, Masayuki NAKAOI, and Yotaro HATAMURA, *Realization of expressive mobile robot*, IEEE International Conference on Robotics and Automation, ICRA, 1997.

[NBG⁺99]   Illah Nourbakhsh, Judith Bobenage, Sebastien Grange, Ron Lutz, Roland Meyer, and Alvaro Soto, *An affective mobile educator with a full-time job*, Artificial Intelligence **114** (1999), no. 1-2, 95–124.

[NKW03]   I. Nourbakhsh, C. Kunz, and T. Willeke, *The mobot museum robot installations: A five year experiment*, IROS 2003, 2003.

[NMS02]   T. Nakata, T. Mori, and T. Sato, *Analysis of impression of robot bodily expression*, Journal of Robotics and Mechatronics **14** (2002), 27–36.

[OKN⁺06]   N. Otero, S. Knoop, C.L. Nehaniv, D. Syrda, K. Dautenhahn, and R. Dillmann, *Distribution and recognition of gestures in human-robot interaction*, The 15th IEEE International Symposium on Robot and Human Interactive Communication, 2006. ROMAN 2006., September 2006, pp. 103–110.

[Pav99]   Vladimir Ivan Pavlovic, *Dynamic bayesian networks for information fusion with applications to human-computer interfaces*, Ph.D. thesis, Graduate College of the University of Illinois, 1999.

[PdW07]   Amy S. Pollick and Frans B. M. de Waal, *Ape gestures and language evolution*, Proceedings of the National Academy of

Sciences of the United States of America (PNAS) **104** (2007), no. 19, 8184–8189.

[Pea01]   K. Pearson, *On lines and planes of closest fit to systems of points in space*, Philosophical Magazine **2** (1901), no. 6, 559–572.

[Pen00]   Alex Pentland, *Looking at people: Sensing for ubiquitous and wearable computing*, IEEE Transactions on PAMI **22** (2000), no. 1, 107–119.

[PSH97]   Vladimir Pavlovic, Rajeev Sharma, and Thomas S. Huang, *Visual interpretation of hand gestures for human-computer interaction: A review*, IEEE Transactions on Pattern Analysis and Machine Intelligence **19** (1997), no. 7, 677–695.

[RBD07]   J. Rett, B.Sousa, and J. Dias, *Social robot nicole: http://paloma.isr.uc.pt/nicole/*, 2007.

[RD04]   Joerg Rett and Jorge Dias, *Autonomous robot navigation-a study using optical flow and log-polar image representation*, Proceedings of the Colloquium of Automation, Salzhausen 2003/2004, 2004.

[RD05]   _____, *Visual based human motion analysis: Mapping gestures using a puppet model*, EPIA 05 (C. Bento, A. Cardoso, and G. Dias, eds.), LNCS (LNAI), vol. 3808, Springer, Heidelberg, 2005.

[RD06]   _____, *Gesture recognition using a marionette model and dynamic bayesian networks (dbns)*, ICIAR 2006 (A. Campilho and M. Kamel, eds.), LNCS, vol. 4141, Springer, Heidelberg, 2006, pp. 69–80.

[RD07a]   _____, *Human robot interaction based on bayesian analysis of human movements*, EPIA 07 (J. Neves, M. Santos, and

J. Machado, eds.), LNAI, vol. 4874, Springer, Berlin, 2007, pp. 530–541.

[RD07b] _____, *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, Noordwijk, The Netherlands, June 2007.

[RDA08a] Joerg Rett, Jorge Dias, and Juan Manuel Ahuactzin, *Bayesian reasoning for laban movement analysis used in human machine interaction*, International Journal of Reasoning-based Intelligent Systems **4** (2008).

[RDA08b] _____, *Brain, vision and ai*, ch. Laban Movement Analysis using a Bayesian model and perspective projections, pp. 183–210, InTech Education and Publishing, Vienna, August 2008.

[Ros04] N. Rossini, *The analysis of gesture: Establishing a set of parameters*, GW 2003 (A. Camurri and G. Volpe, eds.), LNCS (LNAI), vol. 2915, Springer, Heidelberg, 2004, pp. 124–131.

[RS00] R. Rosales and S. Sclaroff, *Learning and synthesizing human body motion and posture*, Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 506–511.

[RSD08] Joerg Rett, Luis Santos, and Jorge Dias, *Laban movement analysis for multi-ocular systems*, IEEE/RSJ 2008 International Conference on Intelligent Robots and Systems, 2008.

[Sea03] Roland Siegwart and et al., *Robox at expo.02: A large-scale installation of personal robots*, Robotics and Autonomous Systems **42 No. 3-4** (2003), 203–222.

[Sha49]   C.E. Shannon, *The mathematical theory of communication*, University of Illinois Press, 1949.

[SNM96]   Tomomasa Sato, Yoshifumi Nishida, and Hiroshi Mizoguchi, *Robotic room: Symbiosis with human through behavior media*, Robotics and Autonomous Systems **18** (1996), 185–194.

[SP95]   T. Starner and A. Pentland, *Visual recognition of american sign language using hidden markov models*, In International Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland, 1995, pp. 189–194.

[Sta95]   Thad Starner, *Visual recognition of american sign language using hidden markov models*, Master's thesis, MIT, Feb 1995.

[Sum84]   S. Sumi, *Upside-down presentation of the johansson moving lightspot pattern*, Perception **13** (1984), no. 3, 283–286.

[Sut82]   Valerie Sutton, *Dancewriting shorthand for modern and jazz dance*, Center Sutton Movement Writing, 1982.

[SWP98]   Thad Starner, Joshua Weaver, and Alex Pentland, *Real-time american sign language recognition using desk and wearable computer based video*, IEEE Transactions on Pattern Analysis and Machine Intelligence **20** (1998), no. 12, 1371–1375.

[UF04]   Raquel Urtasun and Pascal Fua, *3d tracking for gait characterization and recognition.*, FGR, 2004, pp. 17–22.

[VJ01]   Paul Viola and Michael J. Jones, *Rapid object detection using a boosted cascade of simple features*, IEEE International Conference on Computer Vision and Pattern Recognition, vol. 1, 2001, p. 511.

[ZB05]   Liwei Zhao and Norman I. Badler, *Acquiring and validating motion qualities from live limb gestures*, Graphical Models **67** (2005), no. 1, 1–16.

[Zha99]   Z.Y. Zhang, *Flexible camera calibration by viewing a plane from unknown orientations*, ICCV99, 1999, pp. 666–673.

[Zha02]   Liwei Zhao, *Synthesis and acquisition of laban movement analysis qualitative parameters for communicative gestures*, Ph.D. thesis, University of Pennsylvania, 2002.