Human-robot interface with anticipatory characteristics based on Laban Movement Analysis and Bayesian models

Jörg Rett and Jorge Dias[‡]

Abstract

In this work we contribute to the field of human-machine interaction with a system that anticipates human movements using the concept of Laban Movement Analysis (LMA). The implementation uses a Bayesian model for learning and classification and results are presented for the application to online gesture recognition. The merging of assistive robotics and socially interactive robotics has recently led to the definition of socially assistive robotics. What is necessary and we found still missing are socially interactive robots with a higher level cognitive system which analyzes deeply the observed human movement. In this article we provide a framework for cognitive processes to be implemented in human-machine-interfaces based on nowadays technologies. We present LMA as a concept that helps to identify useful low-level features, defines a framework of mid-level descriptors for movement-properties and helps to develop a classifier of expressive actions. Our interface anticipates a performed action observed from a stream of monocular camera images by using a Bayesian framework. With this work we define the required qualities and characteristics of future embodied agents in terms of social interaction with humans. This article searches for human qualities like anticipation and empathy and presents possible ways towards implementation in the cognitive system of a social robot. We present results through its embodiment in the social robot 'Nicole' in the context of a person performing gestures and 'Nicole' reacting by means of audio output and robot movement.

[*]This work is partially supported by FCT-Fundação para a Ciência e a Tecnologia Grant #12956/2003 to J. Rett.

[†]Jörg Rett is with the Institute of Systems and Robotics, University of Coimbra, Polo II, 3030-290 Coimbra, Portugal jrett@isr.uc.pt

[‡]Jorge Dias is with the Institute of Systems and Robotics, University of Coimbra, Polo II, 3030-290 Coimbra, Portugal jorge@isr.uc.pt

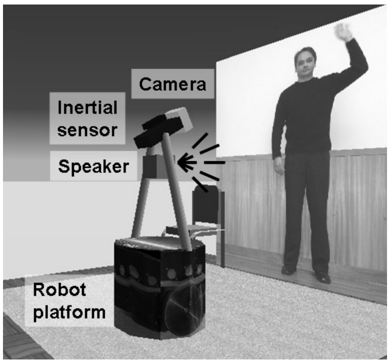

Figure 1. Nicole in position to interact.

1. Introduction

The merging of *Assistive Robotics* and *Socially Interactive Robotics* has recently led to the definition of *Socially Assistive Robotics* [6]. 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)[6]. In this example, the observation and analysis of the user's movements become a key mode of interaction. We find more examples for the importance of this modality in tasks like aiding the wheelchair navigation and controlling manipulator arms through pointing gestures.

What is necessary and we found still missing are socially interactive robots with a higher level cognitive system which analyzes deeply the observed human movement. One can think of the problem as a scenario where a robot is observing the movement of a human, analyzes the movement pattern and acts according to the extracted information (see fig. 1).

Our ultimate goal is to provide the robot with a cognitive system that mimics human perception in two aspects. One aspect involves anticipation, the ability to predict future situations in a dynamic world, physically in motion. The other is empathy, the ability to recognize, perceive and directly experientially feel the emotion of another person.

Towards the latter requirement this article will present the concept of Laban Movement Analysis (LMA) [3] as a way to describe intentional content and expressiveness of a human body movement. Three major components of LMA (i.e. *Space*, *Effort* and *Shape*) are described in detail. The *Space* component presents the different concepts to describe the pathways of human movements inside a frame of reference. The most suitable concept, that of *Vector Symbols* [11] will be adopted by our system. The components *Effort* and *Shape* will be presented with exemplary actions.

We show the technical realization of LMA for the cognitive system of the embodied agent which is based on a probabilistic (Bayesian) framework and a human-movement-tracking system [20]. The tracker extracts the movement-features of a human actor from a series of images taken by a single camera. The hands and the face of the actor are detected and tracked automatically without using a special device (markers). Face recognition is incorporated to take advantage of a person's "individual" movement pattern.

This work presents the Bayesian approach to LMA through the problem of learning and classification, also treating the system's online characteristic of anticipation. The probabilistic model anticipates the gesture given the observed features using the Bayesian framework. 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[9]. The system has been implemented in our social robot, 'Nicole' to test several human-robot interaction scenarios (e.g. playing).

Meanwhile distinct groups of social robots have been successfully developed. Some examples from the group of guide robots are a family of robot guides serving at the Carnegie Museum of Natural History as docents for five years [15], the autonomous tour-guide/tutor robot *RHINO* which was deployed in the "Deutsches Museum Bonn" in 1997 [5] and the mobile robot *Robox* which operated at the Swiss National Exhibition Expo02 [24]. The group of socially assistive robots is younger in age. The T-WREX system [23] uses a modified passive antigravity arm othosis as 3-D input device for measuring arm movements. The Computer interface (Java Therapy) poses exercises to the patient (e.g. shopping), stores and displays the patient's progress. Bandit [26], is a prototype hands-off therapist robot that assists, encourages and socially interacts with users in the process of cardiac convalescence, stroke rehabilitation, and education.

If the perceptual system of a robot is based on vision, interaction will involve *visual human motion analysis*. 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 [8]. Several surveys on visual analysis of human movement have already presented a general framework to tackle this problem [1], [8], [18] and [13]. Aggarwal and Cai point out in their survey [1] that one (of three) major areas related to the interpretation of human motion is motion analysis of the human body structure involving human body parts. The general framework consists of: 1. Feature Extraction, 2. Feature Correspondence and 3. High Level Processing. Meanwhile there is a large amount of work on gesture recognition mainly applied to control some sort of devices. In [17] DBNs were used to recognize a set of eleven hand gestures to manipulate a virtual display shown on a projection screen . Surveys specialized on gesture interfaces along the last ten years reflect the development and achievements [16, 14]. The most recent survey [12] is once more included in the broader context of human motion analysis. It reminds us, 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 (like social robots) it is preferable to use small and light devices for the sensory input as well as for the means of computation. For social robots it is also desired to have a fast classification result to be able to react accordingly.

Section 2 presents the concept of LMA, the three major components of LMA (i.e. *Space*, *Effort* and *Shape*) and the concept of *Vector Symbols*. Section 3 presents the tracking of human movements using vision through its implementation in the system for the social robot Nicole. Section 4 describes the Bayesian framework that is used to classify human movements and provides results for a performed gesture. Section 5 presents the learning of gestures, how the knowledge is represented and how it can be assessed. Section 6 introduces anticipation as a quality for human-robot interaction and defines a concept of evaluation. Section 7 presents the results on anticipation on two performed gestures. Section 9 closes with a discussion and an outlook for future works.

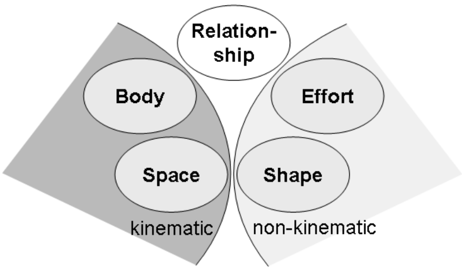

Figure 2. Major components of LMA

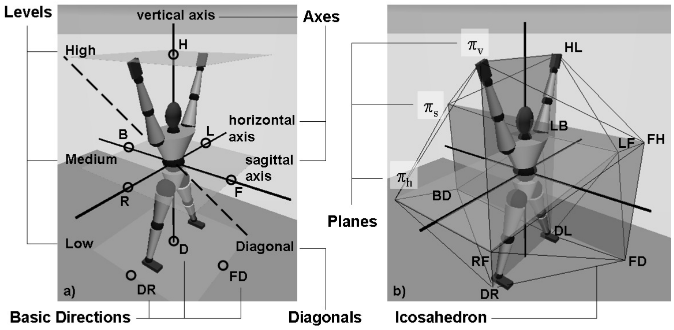

Figure 3. The concepts of a) Levels of Space, Basic Directions, Three Axes, and b) Three Planes and Icosahedron

2. Laban Movement Analysis (LMA)

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 [29]. While being widely applied to studies of dance and application to physical and mental therapy [3], it has found little application in the engineering domain. Most notably the group of Norman Badler, who already started in 1993 to re-formulate Labanotation in computational models [2]. More recently a computational model of gesture acquisition and synthesis to learn motion qualities from live performance has been proposed in [30]. Also recently but independently, researchers from neuroscience started to investigate the usefulness of LMA to describe certain effects on the movements of animals and humans. Foround and Whishaw adapted LMA to capture the kinematic and non-kinematic aspects of movement in a reach-for-food task by human patients whose movements had been affected by stroke [7]. 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 treats five major components shown in fig. 2 of which we adopted three. *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. Like suggested in [7] 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.

2.1. Space

The *Space* component addresses what form is being revealed by the spatial pathways of the movement. The actor is actually "carving shapes in space" [3]. *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 most important ones shown in fig. 3 are: 1) The *Levels of Space* - referring to the height of a position, 2) The *Basic Directions* - 26 target points where the movement is aiming at, 3) The *Three Axes* - Vertical, horizontal and sagittal axis, 4) The *Three Planes* - *Door Plane* π_v, *Table plane* π_h, and the *Wheel Plane* π_s each one lying in two of the axes, and 5) 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 [3].

Labanotation direction symbols encode a position-based concept of space. Recently, Longstaff [11] has translated an earlier concept of Laban which is based on lines of motion rather than points in space into modern-day Labanotation. Longstaff coined the expression *Vector Symbols* to emphasize that they are not attached to a certain point in space. It was suggested that the collection of *Vector Symbols* provides a heuristic for the perception and memory of spatial orientation of body movements. The 38 *Vector Symbols* are organized according to *Prototypes* and *Deflections*. The 14 *Prototypes* divide the Cartesian coordinate system into movements along only one dimension (*Pure Dimensional*

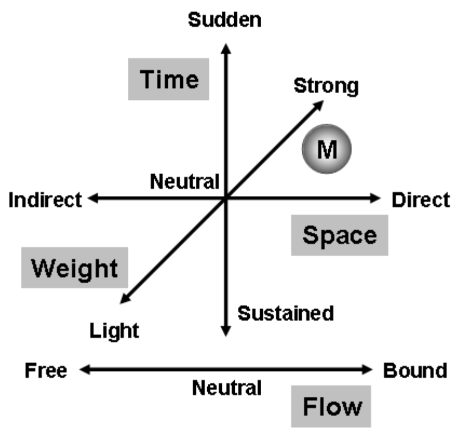

Figure 4. The bipolar *Effort* factors represented as a 4-D space containing a movement (M)

Movements) and movements along lines that are equally stressed in all three dimensions (*Pure Diagonal Movements*) as shown in fig. 3 a). Longstaff suggests that the *Prototypes* give idealized concepts for labeling and remembering spatial orientations. The 24 *Deflections* are mentally conceived according to their relation to the prototype concepts. The infinite number of possible deflecting orientations are conceptualized in a system based on 8 *Diagonal Directions*, each deflecting along 3 possible *Dimensions*.

2.2. Effort

The *Effort* component consists of four motion factors: *Space*, *Weight*, *Time*, and *Flow*. As each factor is bipolar and can have values between two extremities one can think of the *Effort* component as a 4-D space as shown in fig. 4. A movement (M) can be described by its location in the *Effort*-space. Exemplary movements where a certain *Effort*-value is predominant are given in table 1. It is important to remember, that a movement blends during each phase all four Effort-value. Most of the human movements have two or three *Effort*-values prominently high. In fact, it seems difficult even for a trained Laban performer (i.e. Laban notator) to perform single-quality movements [29].

2.3. Shape

LMA defines four *Shape* qualities: *Shape Flow* deals with the movers body shape within itself, the increasing or decreasing the body volume and the movement toward or away from the body center. *Directional Movement* treats the bridging of the action to a point in the environment, such as arc-like and spoke-like movements to reach an object . *Shaping* is concerned with the carving or molding as the body interacts with the environment, e.g. when adapting the body to move through a crowd. *Flow-Reach Space* deals with the reach space in the *Kinesphere*, e.g. the distance of the limbs away from the body center. Like suggested in [29] we summarize the first three *Shape* qualities and express it in terms of spatial directions. By using a major and a minor direction we are able to express the Shape in the concept of the *Three Planes* (π_{vert}, π_{horz}, π_{sag}).

It is important to keep in mind, that the reference system needs to be attached to the movers body rather than being static somewhere in the world. An example taken from [7] showing that the quality of *Sinking*, is independent from the direction of the movement in the world reference frame: A person can walk up a staircase while sinking in the torso, as if someone is pulling down an imaginary string tied to the tip of the tailbone.

3. Tracking of human movements using vision

Using cameras as the basic input modality for a robot provides the highest degree of freedom to the human actor. Though being the most natural way of interaction it also poses the biggest challenge to the functionality of detecting and tracking of human movements.

Figure 5 shows the functional order of our implemented modules inside the Nicole architecture. The current state of the system architecture is a redefined version the gesture perception system (GP-System) presented in [19]. The image data is used by the *Human (Motion) Tracking* module to perform face detection, face recognition, skin-color detection and object tracking and has been described in [19]. We use a face detection module based on Haar-like features as described in

Effort	Movement
Space Direct	Pointing gesture
- Indirect	Waving away bugs
Weight Strong	Punching,
- Light	Dabbing paint on a canvas
Time Sudden	Swatting a fly
- Sustained	Stretching to yawn
Flow Bound	Moving in slow motion
- Free	Waving wildly

Table 1. *Effort* qualities and exemplary movements

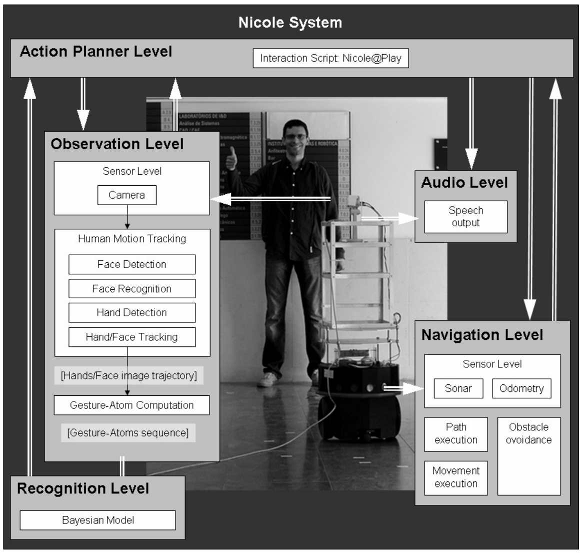

Figure 5. Architecture of the Nicole-System.

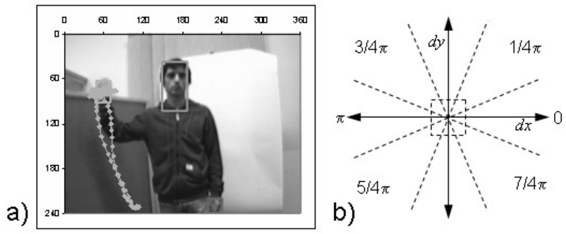

Figure 6. Tracking of a hand movement. a) Trajectory of a Bye-Bye gesture b) Discretization into Vector Atoms according to the concept of Vectors Symbols.

[28] and a face recognition based on eigen-objects and PCA [27]. For skin detection and segmentation we use the CAMshift algorithm presented in [4].

Figure 6 a) shows the tracking of a hand movement with the completed trajectory superimposed over a particular snapshot image. From the resulting trajectories we calculate the relative displacement between each frame and the absolute displacement from the initial position. The latter triggers the starting and end of the gesture. The former undergoes a further discretization according to the concept of *Vectors Symbols*. As we will actually calculate segments of the *Vector Symbols* we will refer to them as *Vector Atoms* or simply *Atoms* A.

Shape	Direction	Plane
Enclosing	Major: Sideward	horizontal
Spreading	Minor: For-/Backward	plane π_h
Sinking	Major: Up-/Downward	vertical
Rising	Minor: sideward	plane π_v
Retreating	Major: For-/Backward	sagittal
Advancing	Minor: Up-/Downward	plane π_s
Shrinking	-	Reach
Growing	-	Space

Table 2. *Shape* qualities

4. Bayesian Framework for Movement Classification

The classification of human movements is done with a probabilistic model using a Bayesian framework. The Bayesian framework can offer combinations of the whole family of probabilistic tools like Hidden Markov Models (HMMs), Kalman Filters and Particle Filters and their various modifications. Though, the Bayesian framework can be used for all kind of system modeling (e.g. navigation, speech recognition, etc.) they are specially suited for 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 [9].

The process of prediction and update represents an intrinsic implementation of the mental concept of anticipation. In general, modeling offers the opportunity to reach a modest dimensionality of the parameter space that describes the human motion. Bayesian models in particular also maintain an intuitive approach which can also be understood by non-engineers [10]. Furthermore these methods have already proven their usability in gesture recognition [25, 17].

In the following we give a short description of equation 1 representing our Bayesian model. A more detailed description can be found in [21]. Our solution assumes that the probability distribution for all possible values of atom A given all possible gestures G and frames I ,which is $\mathbf{P}(A|G,I)$ can be determined.

Applying Bayes rule we can compute the probability distribution for the gestures G given the frame I and the atom A expressed as $\mathbf{P}(G|I,A)$, which is the question the classification is based upon. $\mathbf{P}(G)$ represents the prior probabilities for the gestures.

Assuming the observed atoms are independently and identically distributed (i.i.d.) we can compute the probability that a certain gesture has caused the whole sequence of atoms $P(a_{1:n}|g, i_{1:n})$ by the product of the probabilities for each frame. Where $a_{1:n}$ represents the sequence of n observed values for atom and g a certain gesture from all gestures G. The jth frame of a sequence of n frames is represented by i_j.

We are able to express the probability of a gesture g that might have caused the observed sequence of atoms $a_{1:n}$ in a recursive way. As each frame a value for a is observed we can express the online behavior by using the index t.

Each temporal segmentation yields the problem that no two gestures are exactly the same in speed and duration. A simple solution is to stretch or compress the temporal axis of the observations according to an average. For this we use the learning samples to calculate the average duration (i.e. total number of frames) and variance. This produces prototypes of gestures with a certain length and segmented by a average frame i_avg. The real observation frame i_obs is then modeled by a Gaussian distribution $N(i\_obs, \sigma)$. This distribution represents the probability $P(i\_obs|i\_avg)$ of mapping the observation frame i_obs to an average frame i_avg.

Our Bayesian model is shown in equation 1. We see that the probability distribution of the gestures G at time $t+1$ knowing the observed atoms a until $t+1$ is equal to the probability distribution of G at time t times the probabilities of the current observed atom given the gestures G and frame i at $t+1$. The probability distribution of G for $t=0$ is the prior.

$$\begin{aligned}&\mathbf{P}(G_{t+1}|i_{1:t+1}, a_{1:t+1})\\&\quad= \mathbf{P}(G_t)P(i\_obs_{t+1}|i\_avg_{t+1})\mathbf{P}(a_{t+1}|G, i\_avg_{t+1})\end{aligned} \tag{1}$$

One characteristic of this Bayesian approach is that, as more observed atoms arrive, the probability distribution of the gestures will converge to the correct gesture even if the prior was wrong. The false gesture will vanish simply because the probability of generating "uncharacteristic" atoms for a long time is small. This will happen for any fixed prior, as long as it does not rule out the correct gesture by assigning zero probability to it.

We can likewise express our model in a *Bayesian Net* shown in fig. 7. It shows the dependencies of the above mentioned variables including the displacement dP from the previous section. The rule for classification is based on the highest probability value with a minimum threshold.

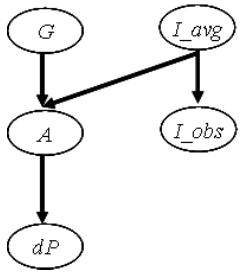

Figure 7. Bayesian Net for the gesture model.

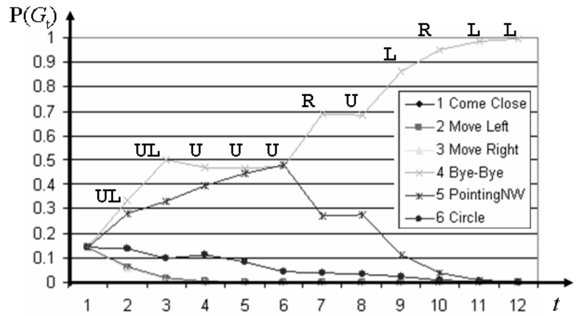

Figure 8. Probability evolution for a Bye-Bye gesture input.

4.1. Experimental results

For the experiment we have used 15 video sequences from each human actor for each of 6 distinct gestures as shown in table 3 and fig. 9. The sequences taken by the camera are stored in a database for future replications. We compute the image trajectories of hands and head, the sequence of *Vector Atoms* and the probability values of the gestures. To provide a ground truth for our image tracker we also collect data from a magnetic tracker. The contents of this database are publicly available at the project's web page (http://paloma.isr.uc.pt/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

Table 3. Characteristics of out gesture-set

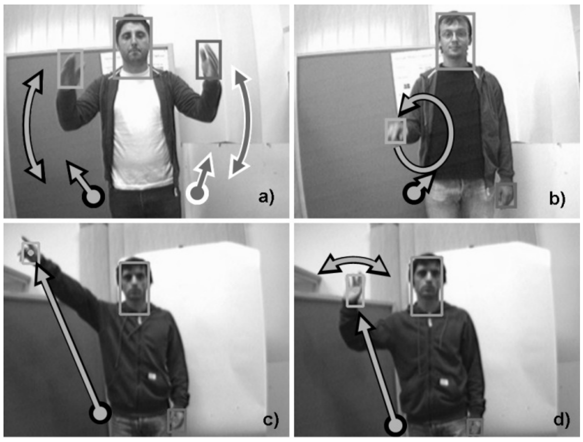

Figure 9. Four exemplary gestures: a)Sagittal Waving b)Draw Circle c)Pointing d)Waving Bye-Bye

Figure 8 illustrates how the gesture-hypothesizes, evolve as new evidences (atoms) arrive taken from the performance of a Bye-Bye gesture. After twelve frames the probabilities have converged to the correct gesture-hypothesis (No. 4). After four frames the probabilities of the two-hand gesture-hypothesis have reached nearly zero. (No. 1, 2, and 3). Until the sixth frame the probabilities of both *High-Level* gestures grow (No. 4 and 5) indicating what is called pre-stroke phase in gesture analysis [22]. Conversely the probability of the *Medium-Level* gesture (No. 6) drops slowly towards zero. After the sixth frame the oscillating left-right movement (and its associated atoms) makes the probability of the Bye-Bye-gesture hypothesis rise and the Pointing-NW-gesture hypothesis drop. A similar behavior was revealed when the remaining five gestures were performed. An unknown gesture, i.e. an unknown sequence of atoms produced more than one gesture-hypothesizes with a significant probability.

5. Bayesian Learning

As both, the gestures and the frame index are discrete values we can express $P(A|GI_{avg})$ in form of a conditional probability table. The probabilities can be learned from training data using a certain number of atom-sequences for each gesture. A simple approach is the one known as Histogram-learning. It counts the number of different atom-values that appear for a gestures along the frames. To overcome the problem of assigning zero probabilities to events that have not yet been observed an enhanced version often uses learning of a family of Laplace-distributions.

5.1. Experimental results

Currently we are using a table that is of size 18 x 31 x 6, that is 18 discrete values for the atom (9 for each hand), 31 frames and 6 gestures. Here we have used the value U for indicating atoms in the up direction rather than H (High) that was used in Laban *Space*. Figure 10 shows a fraction of the table which is the 9 atoms of the right hand for the first 11 frames and the Bye-Bye gesture. It represents the 'fingerprint' of

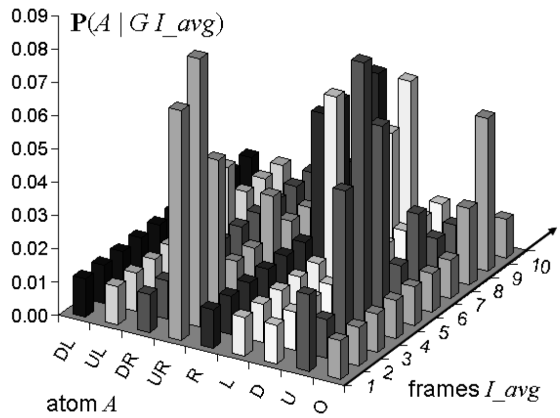

Figure 10. Learned Table $P(A|GI\_avg)$ for gesture 'Bye-Bye'.

the gesture prototype for waving Bye-Bye. Knowing the gesture we assume this sequence of distributions of the random variable atom to be extracted. The table represents an intuitive way to distinguish two gestures from each other. For the Bye-Bye gesture (see fig. 10) we can see, that during the first frames the most likely atom to be expected is the one that goes Up-Right (UR). This is similar for the Pointing gesture (see fig. 11) reflecting the already mentioned *Pre-Stroke* phase. The number of atoms during *Pre-Stroke* also reflect the *Levels of Space* in which the following *Stroke* [22] will take place. In our example we can distinguish the two gestures during *Stroke* as the Bye-Bye gesture has a roughly equal distribution along the line of oscillation (e.g. left-right), while the Pointing gesture produces mainly zero-motion atoms (O).

6. Anticipation

A robot that offers the skill of anticipation inside a human-robot interaction scenario will be able to re-

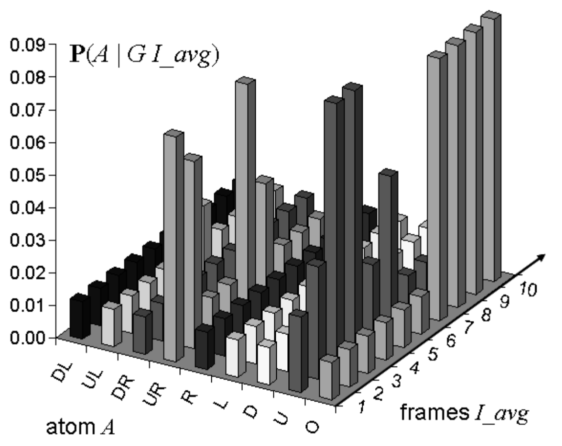

Figure 11. Learned Table $P(A|GI\_avg)$ for gesture 'Pointing NW'.

act faster at the cost of probably being wrong. It is clear that the quality of anticipation is given by what has been learned. So far, this issue has not received enough attention to act 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 how fast a machine anticipates a social cue
- the anticipation-confidence (β-factor) as a measure how strong a machine believes in the most likely hypothesis
- the anticipation-stability (γ-factor) as a measure 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 gesture N_S, the end of the gesture N_E and the anticipation point N_A where the probability of specific hypothesis (e.g. $P(g|i,a)$) reaches and then stays within a threshold of 0.63 (see fig 12). For the sake of simplified expressions we name the duration of the whole gesture n_g, 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_g and n_b, the anticipation-confidence (β-factor) as the mean of the probability values during n_a, and the anticipation-confidence (γ-factor) as the standard deviation from the β-factor during n_a. The three

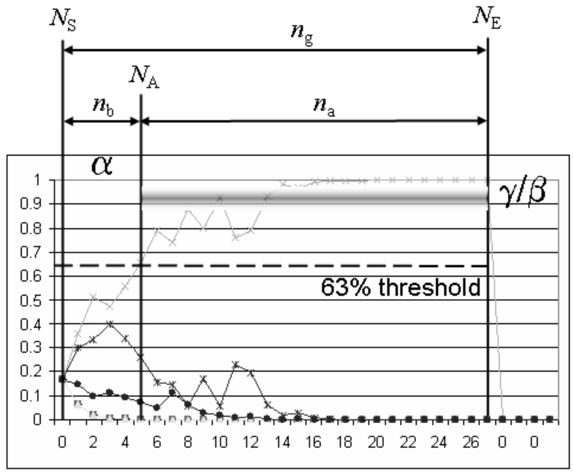

Figure 12. Example of a performed Bye-Bye gesture with anticipation-factors.

factors are shown in equation 2.

$$\alpha = 1 - \frac{n_a}{n_g} \qquad \beta = \frac{1}{N}\sum_{i=1}^{N} P_i \qquad \gamma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - \beta)^2} \tag{2}$$

Figure 12 shows an example of a performed 'Bye-Bye' gesture. As a sequence we chose one that produced a relatively high variance along the probability curve. Using 2 we calculate the value of $\alpha = 93.9\%$, $\beta = 91.1\%$, and $\gamma = 10.65\%$.

7. Results and Discussion

We present sequences taken from our experiments (see fig. 13 and fig. 14) that show the images of the performed gestures and the Laban Vector Atoms that are estimated from the relative displacement (top-right corner of the images). Under the images the probability distribution of the three highest valued gesture-hypothesizes is shown (1: Come Close; 2: Move Left; 3: Move Right; 4 ByeBye; 5: Pointing; 6 Circle). The vertical line in the bar-diagram indicates the threshold for the α-factor. The first sequence (see fig. 13) shows the performance of a bye-bye-gesture for the first eight frames. The system starts anticipating a circle-gesture in the first frame as the up-atom (U) evidence was being learned as belonging most probably to a circle-gesture. The evidences arriving in the second and third frame equalize the probability distribution. In frame four and five the agent starts believing that he perceives whether a pointing-gesture or a bye-bye-gesture. This is due to the fact for the circle-gesture hypothesis to be true the agent would expect to perceive some downward-atoms

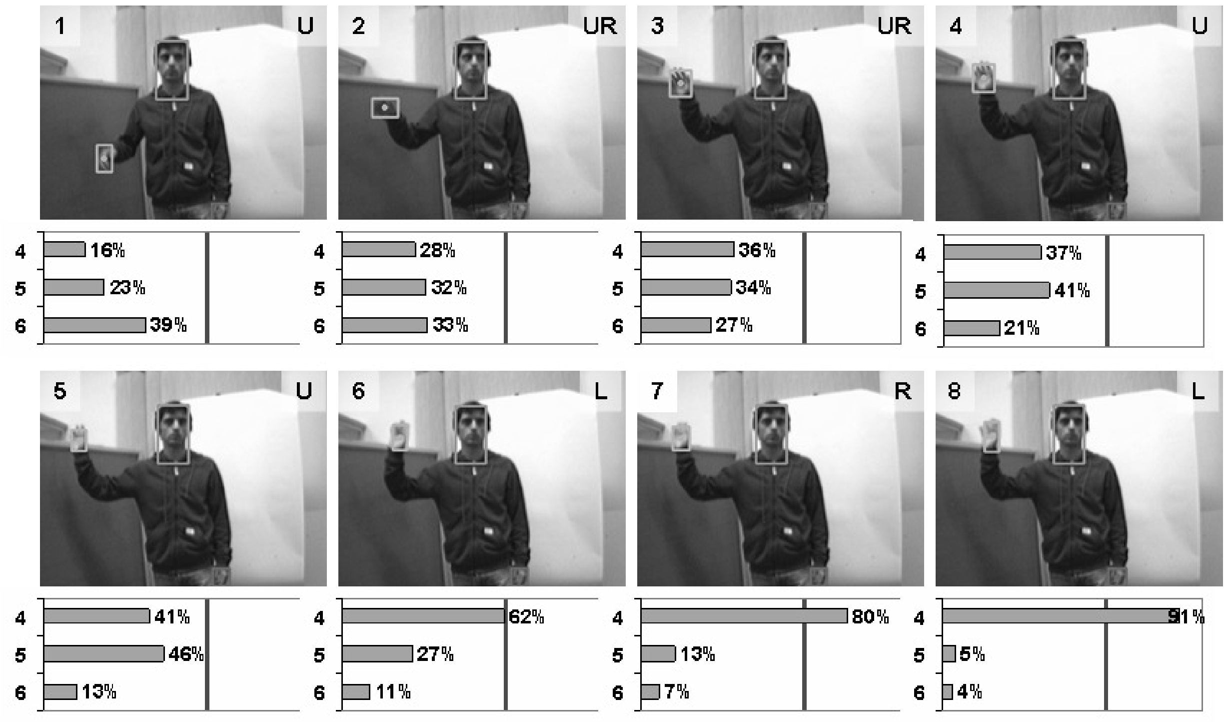

Figure 13. Bye-bye-gesture performance with probability distribution for the three highest valued gesture-hypothesizes (4 ByeBye; 5: Pointing; 6 Circle).

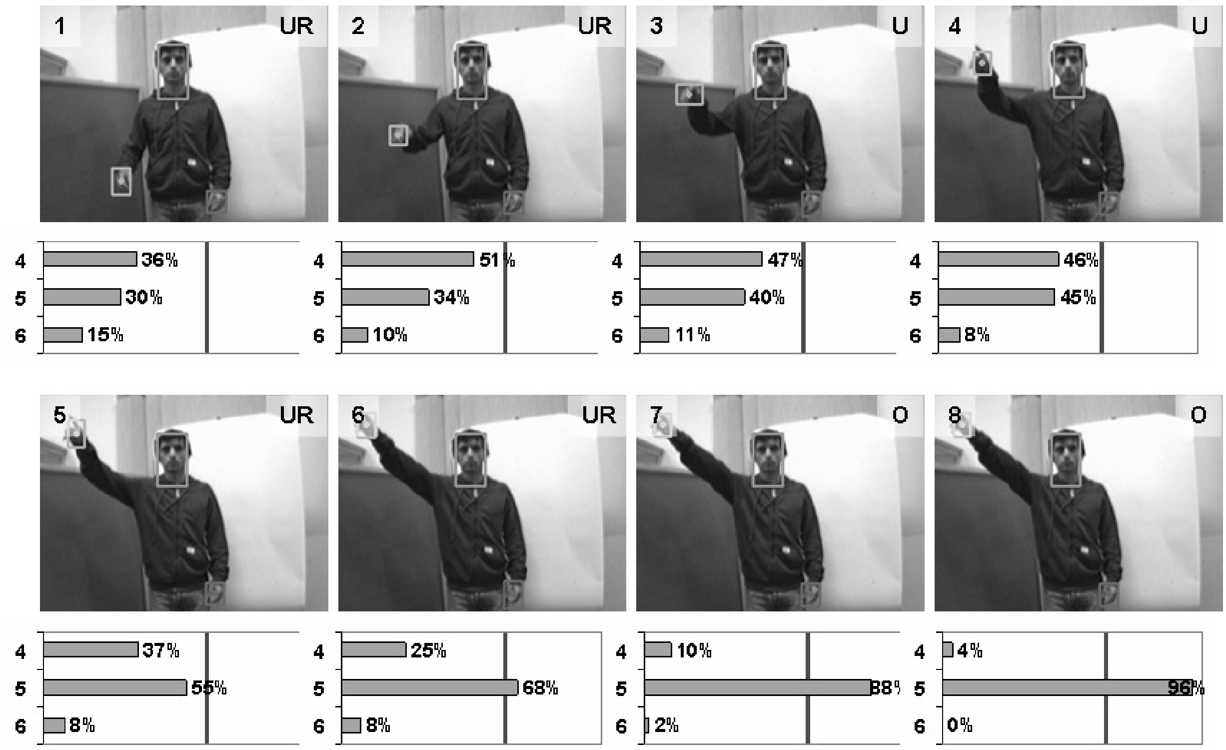

Figure 14. Pointing-gesture performance with probability distribution for the three highest valued gesture-hypothesizes (4 ByeBye; 5: Pointing; 6 Circle).

evidence (D, DL or DR). This is due to the arrival of the left-atom (L). Still the certainty is slightly under the threshold. From frame seven on, the agent is quite certain about the gesture which is due to the arrival of more and more left-right-atoms (R, L) while performing the waving of the hand. The whole performance continued for more 20 frames where the probabilities converged towards 100% for the bye-bye-gesture. Using our definition of anticipation-speed we got an α-factor) of 74%.

The second sequence (see fig. 13) starts already with a low probability for the circle-gesture as the upright-atom (UR) evidence was being learned as belonging most probably to a pointing-gesture or bye-bye-gesture. Until the fourth frame the agent's belief is similar to the one in the previous example (i.e. we see a nearly equal distribution). The fact that no evidences of a directional change arrive (in this case towards left) makes the agent believe stronger in the pointing-gesture hypothesis during frames five and six. The arrival of zero-atoms (O) in frames seven and eight makes the agent belief in the pointing-gesture without much doubts. The whole performance continued for more 15 frames where the probabilities converged towards 100% for the pointing-gesture. Using our definition of anticipation-speed we got an α-factor) of 74%.

8. The Social Robot Nicole

The tested scenario with the social robot Nicole is named Nicole@Play as shown in fig. 15. 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.

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. In our first trials Nicole was using audio outputs like asking for confirmation on the recognized commands and robot movements. Movies of the trials can be downloaded from the project's web-page: http://paloma.isr.uc.pt/nicole/

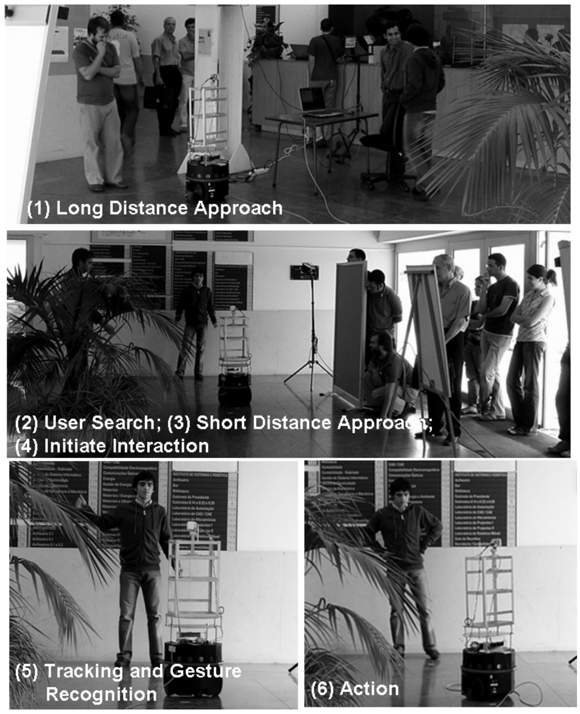

Figure 15. Nicole interacting at the entrance of our department.

9. Conclusions and Future Works

This work presented the application of the *Space* component of Laban Movement Analysis (LMA) to the Human-Robot Interface of the social robot, Nicole. It showed that trajectories of human movements can be learned and recognized using the concept of *Vector Symbols*.

This work demonstrates that the *Bayesian approach for movement classification* provides a robust and reliable way to classify gestures in real-time. Using naive Bayesian classification we are able to anticipate a gesture from its beginning and can take decisions long before the performance has ended.

We have shown that through *Bayesian Learning* the system memorizes learned data in an intuitive way which gives the possibility to draw conclusions directly from the look-up tables. The discussion on the learned probability tables could also be turned toward the distinctions of different people (e.g. person A performs the gesture like this, person B like that).

The scenario Nicole@Play was already tested outside the laboratory environment (e.g. the entrance hall of the department) with people that had never before interacted with any robot. The discussion with people from a local health institute for possible rehabilitation scenarios has just started.

The next step will be the application of the *Effort* and *Shape* component of the LMA to Nicole. Incorporating the dynamic qualities we hope to classify also the expressiveness of a human movement. We are also investigating the usefulness of LMA for other human movements like pedestrian walking and manipulatory movements like reaching, grasping and placing.

10. ACKNOWLEDGMENTS

The authors would like to thank Juan-Manuel Ahuatzin from ProBayes, Grenoble for his support on the implementation using the ProBT-software, while the work was partially sponsored by by BACS-project-6th Framework Programme of the European Commission contract number: FP6-IST-027140, Action line: Cognitive Systems

We also thank Bernardo Carvalho Torres de Sousa from ISR, Coimbra and Pedro Nuñes from ISIS, Málaga for their implementations, while their contributions have be partially sponsored by the VISOR-project, European Robotics Research Network (EURON), call:2, Topical Studies (TS), June 2005

References

[1] J. K. Aggarwal and Q. Cai, “Human motion analysis: A review,” *CVIU*, vol. 73, no. 3, pp. 428–440, 1999. [Online]. Available: citeseer.ist.psu.edu/aggarwal99human.html

[2] N. I. Badler, C. B. Phillips, and B. L. Webber, *Simulating Humans: Computer Graphics, Animation, and Control*. Oxford Univ. Press, 1993.

[3] I. Bartenieff and D. Lewis, *Body Movement: Coping with the Environment*. New York: Gordon and Breach Science, 1980.

[4] G. R. Bradski, “Computer vision face tracking for use in a perceptual user interface,” *Intel Technology Journal*, vol. 2, no. Q2, p. 15, 1998. [Online]. Available: citeseer.ist.psu.edu/bradski98computer.html

[5] W. Burgard, A. B. Cremers, D. Fox, D. Hahnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, “Experiences with an interactive museum tour-guide robot,” *Artificial Intelligence*, vol. 114, no. 1-2, pp. 3–55, 1999. [Online]. Available: citeseer.ist.psu.edu/article/burgard98experience.html

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

[7] A. Foroud and I. 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*, vol. 158, pp. 137–149, 2006.

[8] D. M. Gavrila, “The visual analysis of human movement: A survey,” *CVIU*, vol. 73, no. 1, pp. pp. 82–98, 1999.

[9] D. C. Knill and A. Pouget, “The bayesian brain: the role of uncertainty in neural coding and computation,” *TRENDS in Neurosciences*, vol. 27, pp. 712–719, 2004.

[10] G. E. Loeb, “Learning from the spinal cord,” *Journal of Physiology*, vol. 533.1, pp. 111–117, 2001.

[11] J. S. Longstaff, “Translating ”vector symbols” from labans (1926) choreographie,” in *26. Biennial Conference of the International Council of Kinetography Laban, ICKL, Ohio, USA*, 2001, pp. 70–86. [Online]. Available: http://www.laban-analyses.org/jeffrey/2001a_laban_vector_symbols/laban_vector_symbols.pdf

[12] T. Moeslund, A. Hilton, and V. Kruger, “A survey of advances in vision-based human motion capture and analysis,” *CVIU*, vol. 103, no. 2-3, pp. 90–126, November 2006.

[13] T. B. Moeslund and E. Granum, “A survey of computer vision-based human motion capture,” *CVIU*, vol. 81, no. 3, pp. 231–268, 2001. [Online]. Available: citeseer.ist.psu.edu/moeslund01survey.html

[14] T. B. Moeslund and L. Norgard, “A brief overview of hand gestures used in wearable human computer interfaces,” Computer Vision and Media Technology Lab., Aalborg University, DK, Tech. Rep., 2003. [Online]. Available: citeseer.ist.psu.edu/moeslund03brief.html

[15] I. Nourbakhsh, J. Bobenage, S. Grange, R. Lutz, R. Meyer, and A. Soto, “An affective mobile educator with a full-time job,” *Artificial Intelligence*, vol. 114, no. 1-2, pp. 95–124, 1999.

[16] V. Pavlovic, R. Sharma, and T. S. Huang, “Visual interpretation of hand gestures for human-computer interaction: A review,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 677–695, 1997. [Online]. Available: citeseer.ist.psu.edu/pavlovic97visual.html

[17] V. I. Pavlovic, “Dynamic bayesian networks for information fusion with applications to human-computer interfaces,” Ph.D. dissertation, Graduate College of the University of Illinois, 1999.

[18] A. Pentland, “Looking at people: Sensing for ubiquitous and wearable computing,” *IEEE Transactions on PAMI*, vol. 22, no. 1, pp. 107–119, January 2000.

[19] J. Rett and J. Dias, “Visual based human motion analysis: Mapping gestures using a puppet model,” in *Proceedings of EPIA 05, Lecture Notes in AI, Springer Verlag, Berlin*, 2005.

[20] ——, “Gesture recognition using a marionette model and dynamic bayesian networks (dbns),” in *Proceedings of ICIAR 2006, Lecture Notes in CS, Springer Verlag, Berlin*, vol. 4142/2006, 2006, pp. 69–80. [Online]. Available: http://dx.doi.org/10.1007/3-540-33316-9

[21] J. Rett, “Anticipation measures from bayesian classification and learning,” Institute of Systems and Robtics (ISR), Coimbra, Portugal, Tech. Rep., 2006, http://www.isr.uc.pt/ jrett/Pub-01.htm.

[22] N. Rossini, “The analysis of gesture: Establishing a

set of parameters," in *Gesture-based Communication in Human-Computer Interaction, LNAI 2915, Springer Verlag*, 2003, pp. 124–131.

[23] R. J. Sanchez, J. Liu, S. Rao, P. Shah, R. Smith, T. Rahman, S. C. Cramer, J. E. Bobrow, and D. J. Reinkenmeyer, "Automating arm movement training following severe stroke: functional exercises with quantitative feedback in a gravity-reduced environment," *IEEE Transactions on Neural and Rehabilitation Engineering*, vol. 14, no. 3, pp. 378–389, 2006.

[24] R. Siegwart and et al., "Robox at expo.02: A large-scale installation of personal robots," *Robotics and Autonomous Systems*, vol. 42 No. 3-4, pp. 203–222, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/B6V16-47X20PK-6/2/c0a0f251fac05ba67b7a0fcfd34b17b

[25] T. Starner, "Visual recognition of american sign language using hidden markov models," Master's thesis, MIT, Feb 1995.

[26] A. Tapus and M. J. Mataric, "Towards socially assistive robotics," *International Journal of the Robotics Society Of Japan*, vol. 24, no. 5, July 2006.

[27] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.

[28] P. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 1, 2001, p. 511. [Online]. Available: http://www.ai.mit.edu/people/viola/

[29] L. Zhao, "Synthesis and acquisition of laban movement analysis qualitative parameters for communicative gestures," Ph.D. dissertation, University of Pennsylvania, 2002.

[30] L. Zhao and N. I. Badler, "Acquiring and validating motion qualities from live limb gestures," *Graphical Models*, vol. 67, no. 1, pp. 1–16, January 2005.