Communicating emotions and mental states to robots in a real time parallel framework using Laban movement analysis

Tino Lourens[a], Roos van Berkel[a], Emilia Barakova[a]

[a]*Eindhoven University of Technology, P.O. Box 513, Eindhoven, The Netherlands*

Abstract

This paper presents a parallel real time framework for emotions and mental states extraction and recognition from video fragments of human movements. In the experimental setup human hands are tracked by evaluation of moving skin-colored objects. The tracking analysis demonstrates that acceleration and frequency characteristics of the traced objects are relevant for classification of the emotional expressiveness of human movements. The outcomes of the emotional and mental states recognition are cross validated with the analysis of two independent certified movement analysts (CMA's) who use Laban movement analysis (LMA) method. We argue that LMA based computer analysis can serve as a common language for expressing and interpreting emotional movements between robots and humans, and in that way it resembles the common coding principle between action and perception by humans and primates that is embodied by the mirror neuron system. The solution is part of a larger project on interaction between a human and a humanoid robot with the aim of training social behavioral skills to autistic children with robots acting in a natural environment.

Keywords: emotion recognition from body movements, real time parallel processing, framework for motion analysis and synthesis, Laban movement analysis.

1. Introduction

Sociable humanoid robots pose a dramatic and intriguing shift in the way one thinks about control of autonomous robots. The introduction of robots that have to demonstrate different degrees of autonomy imposed different requirements than industrial robots that have been pre-programmed to work in a fully controlled environment. Biologically inspired and cognitive approaches to robotics have been successfully applied to achieve a degree of autonomy that is determined on the level of an individual organism. By sociable robots it is necessary to go beyond sensory-motor interaction by also taking into account the interplay between constitutive and interactive aspects of autonomy. This implies that sensory-motor interaction has to be enriched with intentional, reward, and emotional features [1, 2].

Most of the existing sociable robots are mainly used in tele-operated or "wizzard of Oz" manner [3, 4], although there are a number of exciting developments in robots that interact autonomously in a meaningful for the human way [5, 6, 7] or use brain like mechanisms or neural mechanisms [8, 9, 10].

Brain inspired robots have been used for investigating animal locomotion and motor control, [11, 12, 13, 14, 9], to learn to avoid obstacles [15, 16], produce accurate vision functions [17, 18, 19] generate adaptive arm movements [17, 20, 13, 9], perform (rat-like) learning and memory tasks [21, 14, 22, 23, 24], or emulate the human or rodent reward and value systems [25, 26].

Many of these functionalities are building blocks for social behavior. Social neuro-robots have been based mainly on the theories of the mirror neuron system. Mirror neurons are visuo-motor neurons that fire both when an action is performed , and when a similar or identical action is passively observed [27]. Using this notion of mirror neurons, it has been shown that observing complex movements such as reaching and locomotion [7, 28, 29, 9] as well as postures, non-goal directed movements, and facial expressions, interacts with the preparation of similar actions [30], due to common motor encoding of the observed and the performed actions. Many studies identify that the mirror neuron system for motor imitation is comprised of three main cor-

Email addresses: `t.lourens@tue.nl` (Tino Lourens), `r.e.a.v.berkel@tue.nl` (Roos van Berkel), `e.i.barakova@tue.nl` (Emilia Barakova)

tical areas: the premotor cortex (F5), the inferior parietal lobule (PF), and the superior temporal sulcus (STS). It creates an affordance to first implement goal-oriented tasks, such as grasping. This system has further evolved to dynamically represent the sensory and the motor correlates of an action. This way perceiving and performing an action exploits the same representation i.e. the action of the conspecifics is represented in the acting agent. Therefore the mirror neuron system can facilitate imitation of action, as one of the earliest forms of reciprocal interaction observed between infant and caregiver [31].

Motor imitation is fundamental for infant's emerging ability to detect the correspondence between self and others [31]. The early opportunity for an infant to detect similarities with others leads to later understanding of other's intentional behavior and the development of theory of mind. More general system for social interaction facilitates the emotional aspects of the imitative or other reciprocal behaviors. The studies of human emotion that are related to observable human behavior in terms of postures and movement, such as [32], place the amygdala at the core of a network of emotional brain structures. The amygdala and STS are directly connected and involved in recognition of emotional body language. It decodes the affective relevance of sensory inputs and initiates adaptive behaviors via its connections to the motor mirror system [33].

To model the cascade of brain areas or mechanisms that relate to imitation of motor and emotional actions is a challenging task. Instead, we model motor imitation and other reciprocal behaviors on a functional level, by describing the movement in the framework of Laban movement analysis. LMA is a formal 'language' for movement description [34, 35] and emphasizes on how internal states, feelings and intentions govern the patterning of movement throughout the whole body. Because of that LMA makes it possible to implement the common coding principle (i.e. the functional principle of the mirror neuron system) on a behavioral level. Similar to common coding/mirroring paradigm LMA is useful to describe the interaction in the physical world, through using the same description for expressing and interpreting of emotional movements [36].

LMA captures both the kinematic as well as the non-kinematic features of movement. Non-kinematic features of movements are the qualitative aspects of movement, characterized by changes in its intensity, shape, force, flow and rhythm. These changes are more expressive than changes in the spatial-temporal body relations. LMA detects the changes in several body parts and axes.

Using robots for social training and interaction is concordant with the interest for nonverbal social communication, expressed via movement, postures and facial expressions. The robots are required to produce and perceive nonverbal behavior in real time, and in parallel (i.e. 2 or more simultaneous movement, mimic or postural behaviors). In this paper we will focus on perceiving and interpreting nonverbal behavior for the purpose of modeling reciprocal social behavior on a the robot. The interpretation of the behavior also through LMA is rooted in the dynamic and qualitative characteristics of the movement in real time, and it can be concluded from observation of parallel movement behaviors.

Real time parallelism on a PC has been strongly facilitated by recent developments of graphical processing units (GPUs), not only have these GPUs become fast[1] but they also can be used as general processing units [37].

We use a standardized humanoid robot NAO (Aldebran) and the GPU based parallel processing for more realistic behavioral applications and have adopted the approach of functional brain modeling [38]. To realize and integrate these functional models we use graphical software environment TiViPE [39], as suggested in [40, 18, 41, 19].

More specifically we are interested in scenarios involving multiple simultaneous actions performed by different body parts of a human or a robot. We assume realistic imitation scenarios, i.e., scenarios where a human freely behaves and a robot tracks its actions with the intend to act upon the meaningful ones, for instance by imitation [42] or by encoding and recollecting episodes of memories [43].

The focus in this paper will be to define the technical aspects of this architecture, and show detecting and tracking of multiple behavioral trajectories of different body parts with a further aim to detect incongruent, atypical, or emotional behavior. We demonstrate the results for emotional and neutral hand waving behavior and guitar playing and demonstrate that both visible skin colored body parts can be tracked in real time. The paper also describes Laban movement analysis (LMA) and elaborates on a method for detection and recognition of emotional movements using this LMA by evaluating hand waving and guitar playing experiments.

The paper is organized as follows: Section 2 describes the experimental setup and gives the implementation of marking a hand in an image using skin color. For the sequence of images such region is marked to construct a stream that is used to analyze hand waving

[1] A single GPU card is able to process up to 5 tera (10^{12}) floating point operations per second (TFLOPS).

behavior. This section provides some preliminary experimental results. Section 3 provides an initial setup how these data streams can be used to extract social behavior for interaction with a robot. Section 4 and 5 describe Laban movement analysis and the hand waving experiment from this perspective. Section 6 elaborates on human emotion recognition while playing guitar and describes analogies with the hand waving experiment. The paper finishes with a discussion and future research.

2. Experimental setup

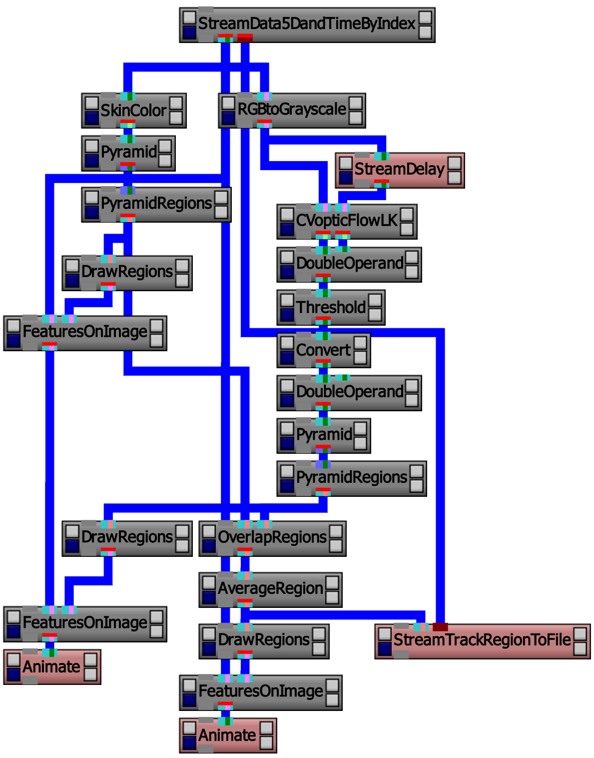

Figure 1: TiViPE (www.tivipe.com) implementation of hand waving. The icons from top to bottom at the left-side process skin areas, while motion sensitivity is processed by the functional blocks at the right-side.

We have been conducting hand waving experiments within scenarios where different emotions and mental states has been enacted. Figure 3 depicts four (happy, angry, sad, and polite) different emotional waving patterns. A camera records images that are processed using a combination of skin color and motion detection, with the aim of tracking a single area. This area is associated with the waving movement. A device or robot should be able to extract a simple motion pattern and interpret the intend, the emotion, or the mental state conveyed by this movement behavior, imitate this movement pattern or eventually adjust its own behavior. The aim of the overall project is to teach or to influence behavior of the human in order to improve his or her social skills, that is necessary, for instance, for people with autism.

The implementation of detection and tracking a moving hand is given in Figure 1. It is used for both hand waving and guitar playing experiment and comprises the following steps:

1. acquiring data from a camera or reading a stored image sequence
2. binarizing an image by marking a pixel either as skin color or other color and in parallel binarizing an image by marking pixels either as observed motion or as static element
3. marking skin and motion regions by a pyramid decomposition
4. selection of these regions that are both skin and motion region
5. averaging skin-in-motion regions to a single region
6. tracking an averaged skin-in-motion region
7. visualization of regions in an image
8. visualization of a waving hand
9. classification of waving profiles

The theoretical background and the computational details behind the implemented signal processing is described in more detail in the following subsections.

2.1. *Skin color detection*

An image is defined as a two-dimensional matrix where a pixel at position (x, y) has an intensity $I_c(x, y) = (r, g, b)$, where r, g, and $b \in [0, \ldots, 255]$ are the red, green, and blue component. Segmentation using skin color can be made independent of differences in race when processing image pixels in Y-Cr-Cb color space [44]. The following (r, g, b) to (Y, Cr, Cb) conversion is used

$$\begin{aligned} Y &= 0.2989r + 0.5866g + 0.1145b \\ Cr &= 0.7132(r - Y) \\ Cb &= 0.5647(b - Y), \end{aligned}$$

where threshold values $77 < Cb < 127$ and $133 < Cr < 173$, see [44] yield good results for classifying pixels belonging to the class of skin tones.

In our experiments we also excluded the "white area". Formally an element belongs to the "white area" if it satisfies the following:

$$\frac{|r - g|}{m} < 0.1 \wedge \frac{|r - b|}{m} < 0.1 \wedge \frac{|g - b|}{m} < 0.1, \qquad (1)$$

where $m = \min(r, g, b)$, $r > 0.3$, $g > 0.3$, and $b > 0.3$. Its implementation given by 'the "SkinColor" icon in Figure 1 yields a binary image.

The functional concept as described above contains similarities with how the brain processes visual data, especially in the way the primary visual cortex area V4 provides a substantial role in processing color [45].

2.2. *Motion detection*

A pixel at location (x, y) with intensity $I(x, y) = (r + g + b)/3$ will have moved by $(\delta x, \delta y)$ over a time span δt, hence the following image constraint equation can be given:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t). \qquad (2)$$

This is known as the brightness constancy assumption. The optical flow constraint equation is obtained using a Taylor expansion on (2), dropping it linear terms, and is expressed as

$$I_x V_x + I_y V_y = -I_t, \qquad (3)$$

where $(V_x, V_y) = \left(\frac{\partial x}{\partial t}, \frac{\partial y}{\partial t}\right)$ denotes the flow, I_x, I_y, and I_t the derivatives of I at horizontal, vertical, and temporal direction, respectively. Equation 3 is applied to all locations (x, y) in image I. The Lucas-Kanade operator [46] is used, it is a two-frame differential method for optical flow estimation where the derivatives are obtained by using a Sobel filter kernel [47]. Instead of using Sobel kernels the more biologically plausible Gabor kernels can be used as well [48, 49]. Receptive fields of a cat's primary visual cortex area V1 and V2, that show striking similarities with these Gabor kernels, have been found in the 1950's by neuroscientists and Nobel laureates Hubel and Wiesel [50, 51]. It is plausible that a similar activity flow map might be expected in the middle temporal are MT, also known as primary visual cortex area V5 [45].

From (V_x, V_y) the L-2 norm (top right "Double-Operand" icon of Figure 1) is taken and thresholded at 5 to obtain a binary "motion classified" image.

2.3. *Rectangular region marking*

The next stage is marking a cluster of "skin tone classified" or "motion classified" pixels by a rectangular window. This is performed by decomposing the image into a pyramid, where every pixel in the next level of the pyramid is computed as follows:

$$I_{i+1}(x, y) = \frac{1}{4} \sum_{x'=0}^{1} \sum_{y'=0}^{1} I_i(2x + x', 2y + y') \qquad (4)$$

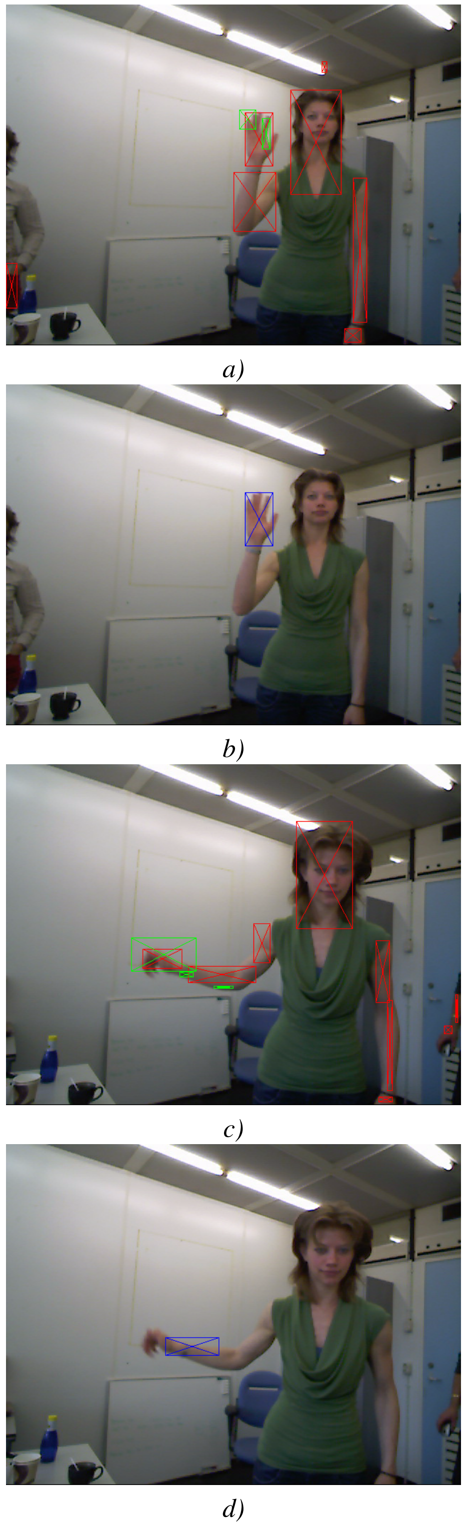

Figure 2: Marked regions of interest. Red and green areas denote skin and motion area, respectively. A blue area is the combined moving skin area that has been averaged over all skin areas that are moving.

where (x, y) is the position in image I_i, i denotes the level in the pyramid. Base level 0 contains the original image I_0. The construction of a pyramid using (4) provides a strongly reduced search space, since if in level $i+1$ a pixel $I_{i+1}(x, y)$ is found to belong to the desired region then in level i of the pyramid a cluster of 2x2 pixels ($I_i(2x, 2y)$, $I_i(2x+1, 2y)$, $I_i(2x, 2y+1)$, and $I_i(2x+1, 2y+1)$) belong to the same region.

The search for regions of interest starts at the highest level, and decreases until an a-priori known minimum level has been reached. It is therefore possible that no regions of interest are found. Taking into consideration that if a pixel is marked as "skin tone" or "motion" it has value 1, and 0 otherwise. We define a pixel to belong to a unique region j if it satisfies the following:

$$R_i^j(x, x+1, y, y+1) = I_i(x, y) \equiv 1. \qquad (5)$$

Regions R_i^j in their initial setting are bound by a single pixel $I_i(x, y)$, and a region growing algorithm is applied to determine the proper size of the rectangular region. Lets assume that the initial size of the rectangle is $R_i^j(x_l, x_r, y_u, y_d)$ and that the possible growing areas are left ($R_i^{jl} = R_i^j(x_l - 1, x_l, y_u, y_d)$), right ($R_i^{jr} = R_i^j(x_r, x_r + 1, y_u, y_d)$), above ($R_i^{ju} = R_i^j(x_l, x_r, y_u - 1, y_u)$), and below ($R_i^{jd} = R_i^j(x_l, x_r, y_d, y_d + 1)$ this region. The average value of all four growing areas is taken, where the maximum value determines the direction of growing. The following procedure

$$\begin{aligned} A_i^{jx} &= \operatorname{avg}\left(R_i^{jx}\right), x \in \{l, r, u, d\} \\ M_i^{jx} &= \max_x \left(A_i^{jx}\right) \\ R_i^j &= R_i^j \cup R_i^{jx}, \text{if } M_i^{jx} \geq T_{rg} \end{aligned}$$

is repeated until $M_i^{jx} < T_{rg}$. From experiments $T_{rg} = 0.67$ provides a rectangle that corresponds roughly to a skin area in the original image and 0.5 gives a sufficiently large motion area, see also Figure 2.

The method described above is able to find all uniform skin color and motion regions in an image in real time.

2.4. Tracking

Two examples of the waving experiment using color images of 640x480 pixels at a speed of 29 frames per second are provided in Figure 2. Creating a single region in the image rather than multiple regions of interest is accomplished in order to unambiguously track an object of interest. These tracked objects are stored as file, see also icon "StreamTrackRegionToFile" of Figure 1, and processed further.

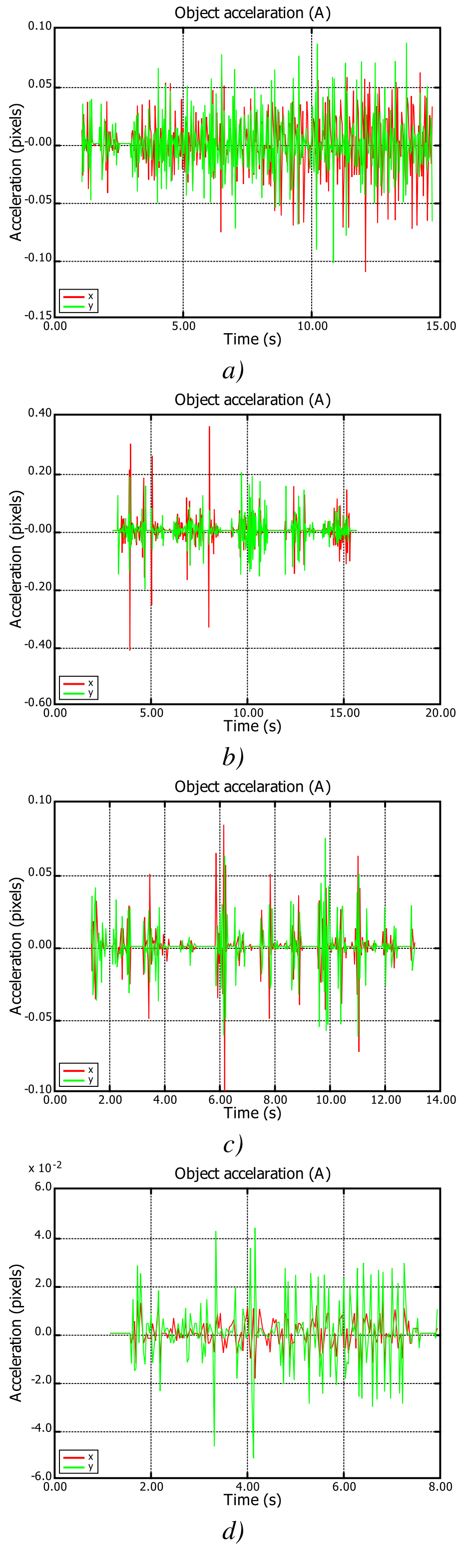

Figure 3: Waving patterns given by acceleration profiles for (a) happiness, (b) anger, (c) sadness, and (d) politeness.

Fifteen recordings of 20 seconds have been made where a performer was asked to demonstrate happiness, anger, sadness, or politeness. In each plot the acceleration profile has been obtained by taking the second derivative of the central point of the tracked object. Examples for all four different emotional states are depicted in Figure 3. From this figure the following can be observed:

1. *happy* waving provides a regular waving pattern with a relatively high frequency.
2. *anger* demonstrates bursts with tremendous acceleration
3. *sadness* demonstrates a profile of low acceleration, its frequency is relatively low and appears to have a lower frequency compared to the other three emotions.
4. *politeness* that demonstrates a queen type of waving profile is a regular pattern with a high frequency that is obtained by using minimal energy.

In an average acceleration-frequency plot of the recorded movements four distinctive clusters are formed (Figure 4). In one of the image sequences the actor was instructed to perform polite waving, but in the sequence she seemed to be happy, indicating that there might be a smooth boundary between these emotional states. The average energy in one of the five bursts in Figure 3b shows an average acceleration score of more than 0.07 and gives an indication of the upper bound of used energy by performing these emotions by the actor.

3. Behavioral Primitives

Understanding motion from waving patterns requires a mechanism that is able to learn to interpret and classify these sequences, and ideally able to extract the observations provided in Section 2.4. In the following section we are attempting to classify motion by so-called Laban primitives. Using these primitives we classify the intend and the mental state of the person that perform movement behavioral patterns.

The current method is developed to enable a robot to interact with a human in realistic scenarios. If a robot is able to track in parallel regions of interest, a considerable number of interacting scenarios are possible even without interpreting of the meaning of an object. Moreover, using an earlier developed technique [52] the robot recognizes and learns repeating patterns of behavior, which it considers important, and discards occasional movements which most often are not important. For instance, if during waving of the hand a head movement takes place because at this time somebody enters the room, this movement will be ignored. However, if a person that interacts with the robot performs repeatedly a movement with his/her head while waving, this will be learned and eventually included in the imitation behavior of the robot.

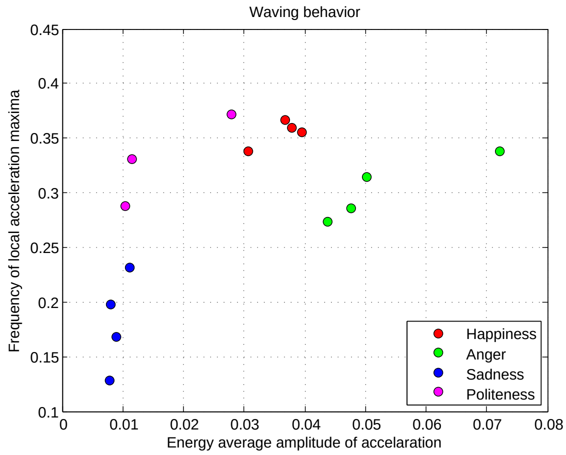

Figure 4: Distinct emotion profiles are revealed by average frequency and acceleration.

4. Laban movement analysis

Rudolf Von Laban, a dance theorist, created a practical method for recording all forms of human motion. While Laban first refers to his notation, as well as to other systems as choreography, for its final form he coined the term kinetography[2] and initially published it as “Schrifttanz” (written dance or script-dance) [53, 54], which also appeared as the Schrifttanz Magazine form 1928-1931. The Laban notation system [34] became known a Kinetography Laban [55, 56] and Labanotion [57].

Bartenieff [35], a student of Laban, applied developmental principles and Laban’s theories to her work with polio patients and dancers, her system is known as Laban/Bartenieff movement analysis or Laban movement studies and comprises:

- Laban movement analysis,
- Anatomy and kinesiology,
- Bartenieff fundamentals (BF), and
- Labanotation.

[2]Kinetography is derived from the Greek words kinesis (movement) and graphein (to write).

LMA emphasizes the processes underlying motor actions rather than the resultant motor action. It records how the four movement components 'Body, Effort, Shape, and Space (BESS)' are integrated, or not, throughout the observed movements. LMA was used to evaluate fighting behaviors of rats [58], to diagnose autistic patients [59], to explain the differences in sexual behavior in Japanese macaques [60] and to analyze the quality of movement by recovery of stroke patients [61]. We use LMA for description of the kinematic and non-kinematic movements made by human subjects that perform emotional actions. The reliability of the non-kinematic measures in LMA has been validated in previous studies [62, 63, 60]. In this paper the focus is on *effort* [64] or dynamics, it is a system for understanding the more subtle characteristics about the way a movement is done with respect to inner intention.

For instance, the difference between punching or reaching for an object is small in terms of body organization, since both rely on extension of the arm. The attention to strength, control, and timing of the movement are different. *Effort* has four sub-categories, each of which has two opposite polarities:

- *Space* (*direct* or *indirect*). Space effort constitutes a single-focused or multi-focused approach to the environment.

- *Weight* (*light* and *strong*). Weight effort determines how I use the impact of my body weight during a movement, ranging from delicate to more forceful.

- *Time* (*sustained* or *quick*). Time effort reveals a deceleration or acceleration within movement.

- *Flow* (*free* or *bound*). Flow effort is responsible for the continuousness or ongoingness of motions, varying from uncontrolled to more controlled use of flow within movement.

Laban considers movement expressions to be almost always a combination of two or three atoms of movement, i.e. dominant effort factors. A combination of *two effort factors* constitutes a *state* and a combination of *three effort factors* constitutes a *drive*. States and drives are movement structures, to be considered as maps in which different motion factors are combined. These *states* are

- Awake = space + time
- Dream = weight + flow
- Stable = weight + space
- Mobile = time + flow
- Remote = flow + space
- Rhythm = weight + time

and its distinct *drives* are

- Action = space + weight + time
- Vision = time + space + flow
- Spell = weight + space + flow
- Passion = time + weight + flow

Laban described the *effort actions* or *action drive* using the subdivision of time, weight, and space, as provided in Table 1. While, flow, either free or bound, is responsible for the continuousness or ongoingness of motions. A detailed description of action drive is found in [65].

Action	Time	Weight	Space
Punch	quick	strong	direct
Dab	quick	light	direct
Slash	quick	strong	indirect
Flick	quick	light	indirect
Press	sustained	strong	direct
Glide	sustained	light	direct
Wring	sustained	strong	indirect
Float	sustained	light	indirect

Table 1: Laban action drives, given by combinations of effort factors time, weight, and space.

All effort factors are visualized in an effort graph, as given in Figure 5. The graph should be evaluated by interpreting four factors each having two polarities, yielding 8 different line segments. The horizontal axis is given by free flow on the left and bound flow on the right. Its vertical axis by light weight on top and strong weight at the bottom, a secondary plane is given by direct space on the horizontal axis and indirect space on the vertical axis. The left and right, non-connected, horizontal lines represent sustained and quick time respectively. In this paper a red segment denotes that the respective effort polarity has been found, while black denotes that it is not. The diagonal line segment is used to connect the secondary 'space' plane.

5. Laban analysis of hand waving experiment

Analysis of videos with a *happy* waving scenario displays a clear standard use of accelerated movement that

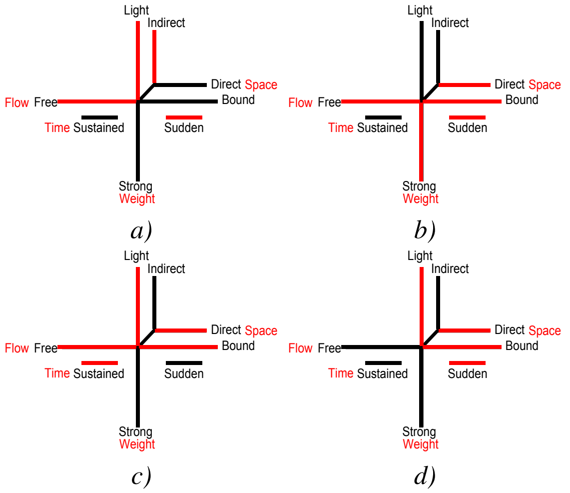

Figure 5: Laban effort graphs for (a) happiness, (b) anger, (c) politeness, and (d) sadness. These results have been obtained by taking the results of the first CMA as provided in Table tab:laban.

travels up along the vertical axis of the body, being a combination of Quick Time and Light Weight. Most emotions that are qualified as positive have an affinity with movement that goes upward in relation to the spine. The first recording reveals a Passion Drive because the Free Flow Effort factor is dominant, meaning the enthusiasm comes out strong and the mover is -taken over-by emotion. During the second and third recording a shift from Free Flow into Indirect Space has been observed. With Indirect Space Effort, the mover uses a spatial awareness that is multi-focused instead of focusing on one point directly. During the second and third recording, movement that is more controlled as well as more connected to the surrounding space has been observed. This shift from Free Flow into Indirect Space means that the Effort Drive, which is a combination of three Effort factors, changes from a Passion Drive into an Action Drive.

Analysis of videos[3] with an *angry* waving scenario reveals a clear standard use of Strong Weight with a tendency towards Direct Space and Quick Time. In terms of accelerated, less controlled movement (Quick Time and Free Flow), the first movement is similar to recording 1 of the happy movement. The difference between the recording of the happy movement expression and the angry movement expression is the Weight Effort factor. The negative impulse of the emotion causes an accelerating impulse that travels downward in relation to the spine. The Effort factors that usually accompany negative emotions are shown quite clearly in the Effort graph. One can see that the emphasis in the movement lies on the use of Strong Weight, Direct Space and Quick Time. The Flow Effort factor diminishes in the second and fourth example, relating the movement to an Action Drive instead of the Passion Drive. The third recording does not show acceleration or deceleration, meaning the Time Effort factor is not sufficiently dominant to maintain the Action Drive. In this case the Time Effort is replaced by use of Bound Flow Effort as the movement is more controlled.

Video analysis of *sad* waving scenarios show a clear standard use of Sustained Time Effort and a frequent use of Bound Flow. The movement recordings contain Passion Drives and Vision Drives. The difference between the Passion Drive and the Vision Drive is that the Passion Drive is less connected to the surrounding space and the horizontal axis of the mover, whereas the Vision Drive is less connected to the sense of gravity and the vertical axis of the mover. Taking into account that this movement example of sadness does not use a strong impulse, the combination of Sustained Time and Bound Flow seems logical. The movement slows down within Sustained Time Effort and due to the control that is related to the Bound Flow Effort, the beginning of the movement is relatively unaccented. One might be surprised that this movement example is related to Light Weight instead of Strong Weight as sadness fits within the realm of negative emotions. In this case the mover approaches sadness in a very gentle, melancholic way. These characteristics establish Light Weight Effort as these movement expressions are more delicate than brusque.

Video analysis of *polite* waving behavior displays a clear standard use of the Light Weight Effort factor and the Direct Space Effort factor. As noted with the previous movement example, the mover has a soft, delicate approach to politeness and therefore the use of Light Weight Effort is dominant. The movements are strongly connected to specific points in the surrounding environment, establishing a clear use of Direct Space Effort. The first two movement examples do not contain specific Time Effort factors, giving the movement a very stable character. During the third movement example, the mover does use Time Effort with slight acceleration, creating a shift from the stable Spell Drive into a more impulsive Action Drive. In this case the consequence of using Time Effort means the initial control within the movement is diminished, completing the sense of Action Drive.

The four distinct emotion profiles illustrated in Fig-

[3]The numbering of these videos is given in Table 2.

	Weight				Space				Flow				Time				Drive/ State	
	Strong		Light		Direct		Indirect		Bound		Free		Quick		Sustained			
Happy 1			*			x					x	*	*	*			Passion	Vision
Happy 2		x	x				x			x			*	*			Action	Action
Happy 3		x	*				x					x	*	*			Action	Passion
Happy 4		x	*								*	*	*	*			Passion	Passion
Angry 5	*	*				x					x		x	*			Passion	Action
Angry 6	*	*			x					x			x	*			Action	Passion
Angry 7	*	*			*				*	*				x			Spell	Passion
Angry 8	*	*			*	*							x	*			Action	Action
Sad 9			x	*						x	x				x	*	Passion	Passion
Sad 10					x	*			x	*					x	*	Vision	Vision
Sad 11			x	*		x			x						x	*	Passion	Action
Sad 12				x	x			x	x	*					x		Vision	Spell
Polite 13			x	*	x			x	x								Spell	Stable
Polite 14			x		x			x	x	*							Spell	Remote
Polite 15			x		x	*				x			x				Action	Remote

Table 2: Laban table of 15 evaluated waving patterns. An asterisk (*) denotes strongly available while a cross (x) denotes available. Marks at the left and right denote evaluations of two different CMA's.

ure 4 are described by an energy measure by means of the average absolute acceleration, and clearly correlates with the weight effort, as illustrated in Table 2. While the frequency of hand waving correlates with time effort and space effort. The Laban movement analysis of the waving patterns has been performed by two different certified movement analysts. Their main difference was in the interpretation of weight for the four happy patterns. The first considered happiness in combination with strong effort still light weight, while the latter considered the substantial effort as strong weight. When comparing the this conclusion with the emotion profiles of Figure 4 we conclude that this interpretation is a slight shift in the average amplitude of acceleration.

In the evaluation of video files that expresses the mental state of politeness was analyzed differently by both analysts. While the first analyst defined it as a drive, i.e. containing 3 active atoms of movement, which is typical for emotions, the second analyst defined it as a state, since she could identify only two active atoms. The evaluation of video file denoted in Table 2 with "Polite" 13, 14, 15 the second analyst stated that patterns 13 and 14 are in indirect space, and pattern 15 shows direct space, which explains the higher frequency of the acceleration pattern compared to the other two Polite waving patterns.

6. Emotion retrieval from guitar playing

Human emotions are, apart from other ways, expressed by our way of moving. Playing guitar, or making music for that matter, is a way of expressing your emotions. The way how these emotions are expressed and perceived could be different for each guitar player, but it is likely that there are some general characteristics of emotions into the movement of guitar players.

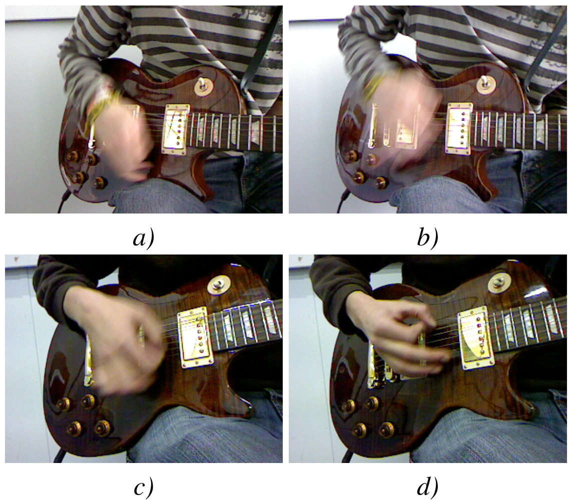

Figure 6: Recordings of two guitar players 1 and 2, at top and bottom row, respectively. On the left happy on the right angry behavior.

The experimental setup is similar to the hand waving experiment, see also Figure 1, i.e., a camera has been used to record images at a resolution of 320x240 pixels at 30 frames per second. These images have been firstly thresholded at a gray value level of $T = 64$ to avoid skin color detection on the dark brown guitar:

$$O(x,y) = \begin{cases} 0 & \text{if } \sum_{c\in\{r,g,b\}} I_c(x,y) < 3T \\ I(x,y) & \text{otherwise} \end{cases}, \quad (6)$$

where O and I denote the color output and input image, respectively. Further processing is exactly the same as described in Figure 1.

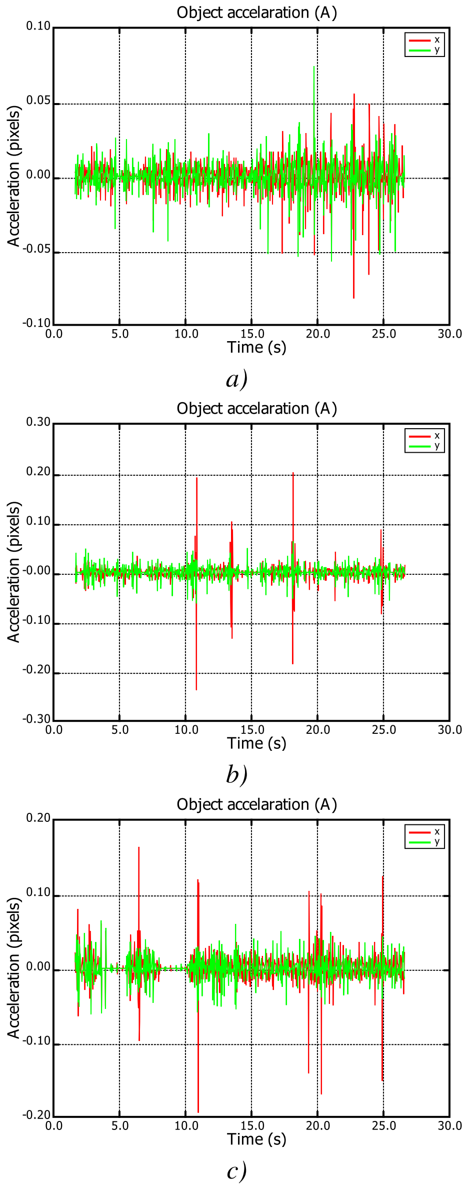

Figure 7: Acceleration profile of player 1 for (a) sad, (b) happy, and (c) angry behavior.

All participants were asked to play 3 songs on an electric guitar. One song which they perceived to be happy, one to be sad, and one to be angry. They were completely free in which songs they wanted to play for each emotion. Single images of some of the videos are illustrated in Figure 6.

Guitar players make repetitive movements with their hands while playing and are therefore suitable to compare with hand waving. However the frequency component depends on the type of songs that are played and do not necessarily reveal characteristic information about the emotion. The acceleration profiles illustrated in Figure 7 show that the acceleration profiles in vertical direction look similar in different songs. However, when computing the average amplitude of acceleration there was a difference between sad, happy, and angry, the ratio in energy was 1:1.4:1.7 for the profiles given in Figure 7, yielding a similar energy ratio profile compared to hand waving emotions as given in Figure 4.

Most striking differences are seen in horizontal direction, and due to the posture the backward swing is the only degree of freedom where substantial higher acceleration of the hand can be observed. The main difference between the emotions is the number of occurrences of such a backward swing. The number of times this happens is 0, 4, and 7 times respectively for sad, happy, and angry, over a period of 25 to 30 seconds, as illustrated in Figure 7.

7. Discussion and Future work

In this paper we have shown that a robot or other device with a simple camera is able to detect and track a waving hand in real time by using a combination of skin tone and motion. We have shown that the display in frequency-amplitude domain of the tracked moving hand of a person who enacts emotional or mental states provides a clear indication of the expressed emotion. The movements that express emotions performed by a single person are unambiguously clustered and segregated in this domain making the current approach suitable for emotion recognition. The distinctive acceleration patterns for the different emotions are independent from the trajectory of the movement. For instance for angry movement the performer was enacting angry mother calling for her child or a person who wanted to push away an annoying fly. Despite of the very different trajectories the acceleration patterns were looking similar and were clustered in the same class. That makes possible to extract emotional primitives that are implemented in robot movements so the robot can socially interact with a human.

It is obvious that we have barely touched the surface of the overall research that we would like to conduct. The experiments with hand waving were partially confirmed with guitar experiments, but still the number of tested subjects is not representative in order to conclude that the extracted emotional primitives are valid for most people and cultures. We have some success in expressing emotional behaviors by robots [2] but our aim is to create reciprocal interactive behaviors that appear natural and are not restricted to simple imitation. Questions, such as: Do the simulation of emotional primitives lead to adaptive or predictive behavior; How does the design of simple interaction behavior look like; and How to design interaction, such that it appears natural and distinctive on a humanoid robot; remain to be answered.

It is clear that the observed results for guitar playing are merely an indication what to expect rather than a proof. The experiment conducted involved 5 different players, and each player was asked to play a song with a sad, happy, and angry emotion setting. An extensive user experiment is required to evaluate a single player over a longer period of time. It is a fact that emotions can not be invoked on command. When the participants were asked to play 3 songs linked with a certain emotion, they were not feeling those emotions at that moment, but they were performing a song which they believe expresses that emotion. To validate whether the emotion is really felt by the player a feedback mechanism provided for instance by the heart-rate variability measurement might be useful. Furthermore, another aspect that influences the validity of our results is the diversity of the character of the players. For example, an introvert person could express his/her emotions differently then an extrovert person. Therefore it is important to look between the relative difference and define a range of the characteristic of the emotions, rather then absolute patterns of emotions.

An important design aspect of each humanoid robot is how closely human movement can be emulated on it. The emulation is restricted by the understanding the physical limitations of the robots, and the mechanism that cause the movement behavior. Moreover, the emotion is not expressed by an individual body part. We assume that by negative emotions the coordination of the movement between different body parts becomes worse. Our parallel architecture makes possible to conduct such experiments and robot emulations.

Acknowledgment

The analysis of CMA Katya Bloom, of the different waving sequences is greatly acknowledged.

[1] C. Breazeal, Emotion and sociable humanoid robots, International Journal of Human-Computer Studies 59 (2003) 119–155.

[2] E. I. Barakova, T. Lourens, Expressing and interpreting emotional movements in social games with robots, Personal and Ubiquitous Computing.

[3] B. Robins, P. Dickerson, K. Dautenhahn, Robots as embodied beings - interactionally sensitive body movements in interactions among autistic children and a robot, ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication (2005) 54–59.

[4] B. Scassellati, How social robots will help us to diagnose, treat, and understand autism, International Symposium of Robotics Research (ISSR'05) (2005) 552–563.

[5] J. Nadel, Early imitation and the emergence of a sense of agency, in: L. Berthouze, H. Kozima, C. G. Prince, G. Sandini, G. Stojanov, G. Metta, C. Balkenius (Eds.), Proc. 4th Intl. Workshop on Epigenetic Robots, Vol. 117, Lund University Cognitive Studies, 2004, pp. 15–16.

[6] H. Kozima, C. Nakagawa, Y. Yasuda, Interactive robots for communication-care: a case-study in autism therapy, ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005. (2005) 341–346doi:10.1109/ROMAN.2005.1513802.

[7] A. Billard, M. J. Mataric, Learning human arm movements by imitation: Evaluation of a biologically inspired connectionist architecture, Robotics and Autonomous Systems 37 (2-3) (2001) 145–160. doi:10.1016/S0921-8890(01)00155-5.

[8] E. Sauser, A. Billard, Biologically Inspired Multimodal Integration: Interferences in a Human-Robot Interaction Game, in: E. are here (Ed.), In Proceedings of IROS'2006, Beijing, China, 2006, pp. 5619–5624.

[9] E. I. Barakova, W. Chonnaparamutt, Timing sensory integration for robot simulation of autistic behavior, IEEE Robotics and Automation Magazine 16 (3) (2009) 51–58.

[10] J. L. Krichmar, A. K. Seth, D. A. Nitz, J. G. Fleischer, G. M. Edelman, Spatial navigation and causal analysis in a brain-based device modeling cortical-hippocampal interactions, Neuroinformatics 3 (3) (2005) 197 – 221. doi:DOI: 10.1385/NI:3:3:197.

[11] A. J. Ijspeert, A. Crespi, D. Ryczko, J.-M. Cabelguen, From Swimming to Walking with a Salamander Robot Driven by a Spinal Cord Model, Science 315 (5817) (2007) 1416–1420. doi:10.1126/science.1138353.

[12] H. Kimura, Y. Fukuoka, A. H. Cohen, Biologically inspired adaptive walking of a quadruped robot, Philosophical Transactions of the Royal Society A 365 (2007) 153–170.

[13] E. Chinellato, A. Morales, E. Cervera, A. P. d. Pobil, Symbol grounding through robotic manipulation in cognitive systems, Robotics and Autonomous Systems 55 (12) (2007) 851 – 859, robotics and Autonomous Systems in the 50th Anniversary of Artificial Intelligence, Campus Multidisciplinary in Perception and Intelligence. doi:DOI: 10.1016/j.robot.2007.07.011.

[14] E. Barakova, T. Lourens, Efficient episode encoding for spatial navigation, International Journal of Systems Science 36 (14) (2005) 887 – 895.

[15] J. L. McKinstry, G. M. Edelman, J. L. Krichmar, A cerebellar model for predictive motor control tested in a brain-based device, Proceedings of the National Academy of Sciences of the United States of America 103 (9) (2006) 3387–3392. doi:10.1073/pnas.0511281103.

[16] B. Porr, F. Wörgötter, Isotropic sequence order learning, Neural Comput. 15 (4) (2003) 831–864. doi:http://dx.doi.org/10.1162/08997660360581921.

[17] P. Dean, J. E. W. Mayhew, N. Thacker, P. M. Langdon, Saccade control in a simulated robot camera-head system: neural net architectures for efficient learning of inverse kinemat-

ics, Biological Cybernetics 66 (6) (1991) 27 – 36. doi:DOI: 10.1007/BF00196450.
[18] T. Lourens, E. I. Barakova, H. G. Okuno, H. Tsujino, A computational model of monkey cortical grating cells, Biological Cybernetics 92 (1) (2005) 61–70, dOI: 10.1007/s00422-004-0522-2.
[19] T. Lourens, E. I. Barakova, Orientation contrast sensitive cells in primate v1 –a computational model, Natural Computing 6 (3) (2007) 241–252.
[20] S. Eskiizmirliler, N. Forestier, B. Tondu, C. Darlot, A model of the cerebellar pathways applied to the control of a single-joint robot arm actuated by mckibben artificial muscles, Biological Cybernetics 86 (5) (2002) 379–394.
[21] A. Arleo, W. Gerstner, Modeling rodent head-direction cells and place cells for spatial learning in bio-mimetic robotics, ABC journal (2000) 236–245.
[22] N. Burgess, J. G. Donnett, K. J. Jeffery, J. O'Keefe, Robotic and neuronal simulation of the hippocampus and rat navigation, Philosophical Transactions: Biological Sciences 352 (1360) (1997) 1535–1543.
[23] M. Milford, R. Schulz, D. Prasser, G. Wyeth, J. Wiles, Learning spatial concepts from ratslam representations, Robot. Auton. Syst. 55 (5) (2007) 403–410. doi:http://dx.doi.org/10.1016/j.robot.2006.12.006.
[24] H. Wagatsuma, Y. Yamaguchi, Context-dependent adaptive behavior generated in the theta phase coding network, Neural Information Processing (2008) 177–184.
[25] A. Arleo, F. Smeraldi, W. Gerstner, Cognitive navigation based on nonuniform gabor space sampling, unsupervised growing networks, and reinforcement learning, ABC journal.
[26] O. Sporns, W. H. Alexander, Neuromodulation and plasticity in an autonomous robot, Neural Netw. 15 (4) (2002) 761–774. doi:http://dx.doi.org/10.1016/S0893-6080(02)00062-X.
[27] G. Rizzolatti, M. A. Arbib, Language within our grasp, Trends in Neurosciences 21 (5) (1998) 188 – 194. doi:DOI: 10.1016/S0166-2236(98)01260-0.
[28] S. Schaal, Is imitation learning the route to humanoid robots?, Trends in Cognitive Sciences 3 (1999) 233–242.
[29] S. Schaal, N. Schweighofer, Computational motor control in humans and robots, Current Opinion in Neurobiology 15 (6) (2005) 675 – 682, motor sytems / Neurobiology of behaviour. doi:DOI: 10.1016/j.conb.2005.10.009.
[30] K. J. Montgomery, N. Isenberg, J. V. Haxby, Communicative hand gestures and object-directed hand movements activated the mirror neuron system, Social Cognitive and Affective Neuroscience 2 (2) (2007) 114–122. doi:10.1093/scan/nsm004.
[31] A. N. Meltzoff, M. K. Moore, Explaining facial imitation: a theoretical model, Early Development and Parenting 6 (1997) 179–192.
[32] B. de Gelder, Towards the neurobiology of emotional body language, Nat Rev Neurosci 7 (3) (—2006—) 242–249, 10.1038/nrn1872.
[33] Baron-Cohen, Social intelligence in the normal and autistic brain: an fmri study, European Journal of Neuroscience 11 (1 June 1999) 1891–1898(8).
[34] R. Laban, Principles of Dance and Movement Notation, Macdonald & Evans, New York, 1956.
[35] I. Bartenieff, D. Lewis, Body Movement –Coping with the environment, Gordon and Breach Scinece Publishers, 1980.
[36] E. I. Barakova, J. Gillesen, L. Feijs, Social training of autistic children with interactive intelligent agents, Journal of Integrative Neuroscience 8 (1) (2009) 23–34.
[37] T. R. Halfhill, Parallel processing with cuda, IN-STAT Microprocessor Report (2008) 1–8.
[38] T. Lourens, E. I. Barakova, H. Tsujino, Interacting modalities through functional brain modeling, in: J. Mira, J. R. Álvarez (Eds.), Proceedings of the International Work-Conference on Artificial and Natural Neural Networks, IWANN 2003, Vol. 2686 of Lecture Notes in Computer Science, Springer-Verlag, Menorca, Spain, 2003, pp. 102–109.
[39] T. Lourens, Tivipe –tino's visual programming environment, in: The 28th Annual International Computer Software & Applications Conference, IEEE COMPSAC 2004, 2004, pp. 10–15.
[40] R. P. Würtz, T. Lourens, Corner detection in color images through a multiscale combination of end-stopped cortical cells, Image and Vision Computing 18 (6-7) (2000) 531–541.
[41] T. Lourens, E. I. Barakova, Tivipe simulation of a cortical crossing cell model, in: J. Cabastany, A. Prieto, D. F. Sandoval (Eds.), IWANN 2005, no. 3512 in Lecture Notes in Computer Science, Springer-verlag, Barcelona, Spain, 2005, pp. 122–129.
[42] E. I. Barakova, T. Lourens, Mirror neuron framework yields representations for robot interaction, Neurocomputing 72 (4-6) (2009) 895–900.
[43] E. I. Barakova, T. Lourens, Efficient episode encoding for spatial navigation, International Journal of Systems Science 36 (14) (2005) 877–885.
[44] D. Chai, K. N. Ngan, Face segmentation using skin-color map in videophone applications, IEEE Transactions on Circuits and Systems for Video Technology 9 (4) (1999) 551–564.
[45] S. Zeki, A Vision of the Brain, Blackwell science Ltd., London, 1993.
[46] B. D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: Proceedings of Imaging understanding workshop, 1981, pp. 121–130.
[47] I. E. Sobel, Camera models and machine perception, Ph.D. thesis, Electrical Engineering Department, Stanford University, Stanford, CA (1970).
[48] J. G. Daugman, Complete discrete 2-d Gabor transforms by neural networks for image analysis and compression, IEEE Transactions on Acoustics, Speech and Signal Processing 36 (7) (1988) 1169–1179.
[49] T. Lourens, A Biologically Plausible Model for Corner-based Object Recognition from Color Images, Shaker Publishing B.V., Maastricht, The Netherlands, 1998.
[50] D. H. Hubel, T. N. Wiesel, Receptive fields of single neurones in the cat's striate cortex, J. Physiol. 148 (1959) 574–591.
[51] D. H. Hubel, T. N. Wiesel, Ferrier Lecture functional architecture of macaque visual cortex, Proc. R. Soc. Lond. B. 198 (1977) 1–59.
[52] E. I. Barakova, D. Vanderelst, From spreading of behavior to dyadic interaction -a robot learns what to imitate, International Journal of Intelligent SystemsIn press.
[53] R. Laban, Schrifttanz, Universal Edition, Wien, 1928.
[54] R. Laban, Schrifttanz –Kleine Tänze mit Vorübungen, Universal Edition, Wien, 1930.
[55] A. Knust, Kinetographie laban, parts A - O. Eight volume unpublished manuscript in German (1953).
[56] A. Knust, Handbook of Kinetography Laban, Das Tanzarchiv, Hamburg, 1958.
[57] A. Hutchinson, Labonation –The System for Recording Movement, New Directions, New York, 1956.
[58] A. Foroud, S. M. Pellis, The development of ldquoroughnessrdquo in the play fighting of rats: A laban movement analysis perspective, Developmental Psychobiology 42 (1) (—2003—) 35–43, 10.1002/dev.10088.
[59] L. P. Dott, Aesthetic listening: Contributions of dance/movement therapy to the psychic understanding of motor stereotypes and distortions in autism and psychosis in childhood and adolescence, The Arts in Psychotherapy 22 (3) (1995) 241 – 247, european Consortium for Arts Therapy Edu-

cation (ECArTE). doi:DOI: 10.1016/0197-4556(95)00033-2.
[60] P. Vasey, A. Foroud, N. Duckworth, S. Kovacovsky, Male–female and female–female mounting in japanese macaques: A comparative study of posture and movement, Archives of Sexual Behavior 35 (2) (2006) 116–128.
[61] A. Foroud, 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 158 (1) (2006) 137 – 149. doi:DOI: 10.1016/j.jneumeth.2006.05.007.
[62] F. R., C. J., K. E., Observing behavioral qualities, International Journal of Comparative Psychology 10 (1997) 167–179.
[63] A. Foroud, I. Q. Whishaw, S. M. Pellis, Experience and cortical control over the pubertal transition to rougher play fighting in rats, Behavioural Brain Research 149 (1) (2004) 69 – 76. doi:DOI: 10.1016/S0166-4328(03)00230-4.
[64] R. Laban, F. C. Lawrence, Effort, MAcdonald & Evans, 1947.
[65] J. Newlove, J. Dalby, Laban for all, Nick Hern, London, 2004.