**Proceedings of the 14th IASTED International Conference
Robotics and Applications (RA 2009)
November 2 - 4, 2009 Cambridge, MA, USA**

# HUMAN-ROBOT INTERACTION: INVARIANT 3-D FEATURES FOR LABAN MOVEMENT ANALYSIS SHAPE COMPONENT

Luís Santos and Jorge Dias
Institute of Systems and Robotics, Department of Electrical Engeneering
University of Coimbra
Pólo II
3000 Coimbra, Portugal
email: luis,jorge@isr.uc.pt

**ABSTRACT**

In the field of human-machine interaction, there are still lacking efficient tools within visual perception of human non-verbal comunication. In this context, some investigation has been conducted within the paradigm of Laban Movement Analysis (LMA) [1–5]. This work, will explore how visual signals can be processed in order to retrieve useful features which will allow the characterization of movements in a semantic/intuitive way (e.g. reaching). The LMA shape component will be the main focus of this work, and the implementation will follow the guidelines of previous work. The head and both hands will be tracked within the image, and stereo vision model will be used to retrieve 3-D information of the performer's pose. From this 3-D data, invariant features will be generated, and used as evidences in a Bayesian framework, which is the selected tool for Laban Movement Analysis implementation. The current work intents to further extend the LMA Bayesian models, towards a full robust descriptor of non-verbal cues for machine interpretation of human behaviour. Results show that the more complete the global Laban Movement Analysis model becomes, better results are achieved, leading to the thought that Laban can provide a good computational movement classifier.

**KEY WORDS**

Human-Machine Interaction, Laban Movement Analysis, Bayesian Models, Gesture Recognition, Signal Processing.

## 1 Introduction

The scientific community is still lacking a robust tool for movement classification, similar to what is already available in speech recognition. This work, is a follow up of Rett J. [5], where a descriptive language for movements is studied, Laban Movement Analysis (LMA). LMA is used to describe, notate and interpret human movements. The underlying concept of Laban is to translate movements in a unique semantic, coding them in symbols. Laban provides descriptors such as *fast, sudden, retreating, strong, etc* to classify movements which is very intuitive to read. It is divided in components, each of them dealing with different aspects of a movement, turning it a very complete movement dictionary. One of the strong points of LMA is the ability to describe expressive content of movements, which makes a prospective powerful tool for emotion/behavior analysis.

However, LMA lacks full computational implemention. Rett J. et al [5–7] proposed a Bayesian framework as a support for its implementation. The Bayesian theory gives us the possibility to deal with incompleteness and uncertainty, make predictions on future events and, most important, provides an embedded scheme for learning. It also allows the implementation of the components in an hierarquic fashion, i.e. we can have a global model that is divided in several submodels, which is useful providing the components can be implemented separatelly. The bayesian models are formulated based on [8].

The lowest level of the bayesian framework are evidences, which are generally features extracted from signals.

This study emerges from a natural consequence of previously developed Space and Effort Models. For contextualization purposes, the Effort component and obtained results will be presented. This work presents a study on Laban's *shape* component through the analysis of body geometry. First step is to compute the 3-D points correspoding to the tracked key points (head and both hands). Using a stereo vision system and homography, one can build a disparity and depth map. From this information, we can use the disparity value to introduce a concept, *virtual depth*. A triangle based on these three sets of 3-D coordinates can be found, along with geometrical inherent characteristics/features. These features, upon calculated will generate probabilistic distributions which can be evaluated and analyzed to detect patterns, hence dominant features.

The study of this component, has the objective of augmenting the global model, with the purpose to add strenght to LMA as a movement classifier. There are inherent advantages, as one may instantly perceive what a movement classifier capable of giving information regarding human expressiveness and emotion, can give to the study of human-machine interaction. Fields like social robotics pursue the goal of *humanizing* machines, i.e. build robots capable of smoothly and intuitively interact with humans. By perceiving and/or predicting human expressive content, a robot can improve its response capability. This work intents to give one more step towards this direction.

# 2 Laban Movement Analysis

## 2.1 Theory

Laban Movement Analysis (LMA) is a method to observe, describe, notate and interpret human movements within the context of music coreography. The general framework is already applied in physical and menthal therapy [1] and is beggining to find application within the engineering domain. Neuroscience researches (e.g. [4]) started to investigate LMA as a tool to describe certain effects on the movements on animals and humans. Other recent studies on LMA by Rett J. et al [5, 6], explored how LMA can be used to classify human expressive movements within human-machine interaction.

Theory divides LMA in components, however there is no standard regarding their cardinality. Most notably, Norman Badler's group [2, 3] mentions five major components in their work.

Briefly, these are divided in:

- **Non kinematic** components: *Body* specifies which body parts are moving and their relation to the body center; *Space* deals directly with the trajectory executed by the body parts while performing a movement.

- **Kinematic** components: *Effort* which deals with the dynamic qualities of the movement, and the inner attitude towards the use of energy; *Shape* (emerging from *Body* and *Space*) is focused on the body itself.

- **Relantionship**: does not belong to any of the previous groups, and appears as the less explored component, and describes the interaction with oneself, others and the environment.

Some literature considers only the first four components [4]. The *Space* and *Effort* components have already been investigated by Rett et. al. [5, 6]. In this work, the focus will be on the *Shape* component, however the Effort component will also be presented to strengthen Laban Movement Analysis as a robust computational movement descriptor.

## 2.2 Effort component

Effort is the component that deals with the 'expressiveness' that accompanies the spatial trajectory (*Space* component). By selecting a set of suitable features from the trajectories described by hands and head of the performer, the *Effort* component can be seen as the key descriptor to solve the task of analyzing 'expressive movements'.

*Effort* describes the dynamic qualities of the movement and the inner attitude towards using energy. It is divided in four *Effort* qualities: *Space*, *Weight*, *Time*, and *Flow*.

Each quality is bipolar and lies between two extremes. For instance, Effort Time, in extreme cases can be either Sudden or Sustained. The values for the *Effort* qualities are shown in (1)

$$\begin{aligned} Space &\in \{direct, neutral, indirect\} \\ Time &\in \{sudden, neutral, sustained\} \\ Weight &\in \{strong, neutral, light\} \\ Flow &\in \{bound, neutral, free\} \end{aligned} \tag{1}$$

Movements are described and distinguished by those qualities close to an extreme, e.g. a *Punch* has *Strong Weight*, *Sudden Time* and *Direct Space*.

Despite having four qualities, combinations of three with the fourth considered to be neutral, appear as the most natural way to perform an action. Literature presents four and single quality combinations as rare, as they produce extreme movements (e.g. tearing something apart) [1, 3].

Since *Effort* deals with dynamic characteristics of movements, previous studies [5] concluded that the most natural way to characterize it would be using dynamic characteristics of the trajectories. Hence, from the generated trajectories, parameters like Acceleration, Velocity and Curvature were chosen as reliable features for the Effort Bayesian Model.

## 2.3 Shape component

Bartenieff and Lewis [1] do not define Shape a component of its own but rather a set of qualities emerging from the Body and Space components. Two Shape qualities were mentioned particularly: Shape Flow describes movements that are focused on the body itself, going towards or away from the body center and using descriptors like shrinking and growing, bulging and hollowing (also including breathing). The term Spatial Shaping is used for movements that are going towards a goal in space (e.g. reaching).

It is usually described in a Euclidean frame of reference that is aligned with an initial position of the egocentric frame of reference. Due to this, movements can be described by using the vertical, horizontal and sagittal axes and relating them to bipolar descriptors like sinking and rising, enclosing and spreading, and retreating and advancing. Fig.1 shows the descriptors (left) and some exemplary movements (right).

Body movement is mainly described using head, hands and torso movements. Yet, to give our work consistency in its approach, only hand and head movements will be used to generate features, has sucessfuly been done in previous work. To approach shape, it was decided to study the triangle connecting these previously mentioned body parts. Relations between Shape descriptors and this triangle become intuitive, e.g. if the area of the triangle

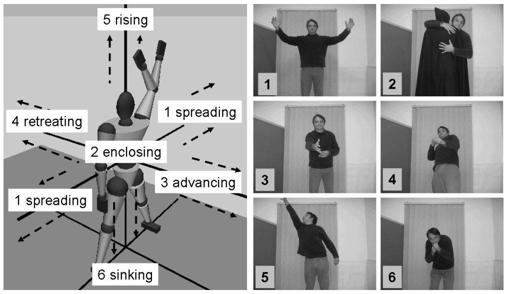


Figure 1: The Shape component with its spatial qualities and some exemplary movements: 1. embracing, 2. hugging, 3. shake, 4. retreating 5.reaching and 6.ducking.

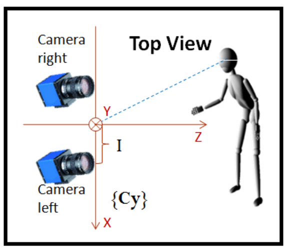


Figure 2: The cyclopean referential $\{Cy\}$ defined at the center of the baseline connecting the two cameras.

grows, then it is probably that the body shape can be described as growing. The mathematical characterization of this triangle will be address in further sections.

## 3 Data and Signal Processing

To study the *Shape* component, this work proposes the geometrical behavior analysis of a triangle in the three dimensional (3-D) space whose vertices are defined by the head and hands coordinates.

The chosen sensor type that will originate the signal is vision sensor through image processing, however the processing technique itself is not the focus of this work, and will not be described. The reader will assume that a camera tracks the performer's head and both hands in the image, through commonly known techniques (color histograms and camshift). This section will address the signal processing from that point forward.

In order to allow the vision system to find coordinates in the 3-D space through, a stereo vision approach is used. These systems use two cameras, however a *cyclopean* referential $\{Cy\}$ needs to be established. It is common to define a *cyclopean* referential in the middle of the baseline[1] connecting both cameras as seen in Fig.2.

[1]Baseline is commonly defined as the line that connects both sensors of a vision stereo system

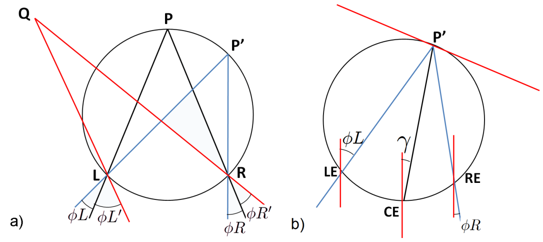


Figure 3: a) Left (L) and Right (R) cameras with focus on points (P and P') over the ViethMuller Circle; An example is given (point Q) lying outside the circle, where the disparity value is negative b) stereo system angle representation.

### 3.1 Disparity and Depth Maps

Disparity maps represent the difference distance between points of a pair of images, whilst depth maps represent the expected depth/distance of specific region to the camera's referential. Videre [9] library first constructs a disparity space image from a pair of stereo images, and then calculates temporary disparity maps using the SAD method [10]. The applied algorithm reduces both the blurred errors at depth discontinuities and the mismatched errors at half occluded areas. Then a median filter is used to interpolate the dense disparity map. Once one has calibrated the cameras and the disparity map calculated, it is trivial to get the depth map.

To further understading, a definition is briefly introduced. The *Viethmuller Circle* defines the region where disparity is zero. The disparity values are positive inside the circle, and negative outside. The circle can be observed in Fig.3

The disparity value however is an interesting characteristic to be explored. Consider the head to be on the ViethMuller Circle, hence its disparity is equal to zero; the disparity value for the hands can be related to the depth of the hands, relative to the head position. Having the disparity value computed using commonly know algorithms, the $z$ coordinate for both hands can be found.

In this article, a definition is proposed, *virtual depth*. As mentioned, the value for disparity is zero along the circumference of the ViethMuller circle (it is considered that the head is the focus point, hence, its disparity is always zero). To have hand depth relative to the head, the value of disparity can be used directly. Whilst the value of disparity does not reflect directly the depth relative to the cyclopean referential, it has implicit depth associated. If the disparity is positive for one hand, that hand is inside the circle, however if it negative, it is outside. We can define a unidimensional *virtual depth* axis as seen in Fig.4.

The axis is coincident with the line that connects the origin of the cyclopean referential and the focus point. The origin of the *virtual depth* axis, lays in the intersection of the axis with the ViethMuller Circle, hence, it can be de-

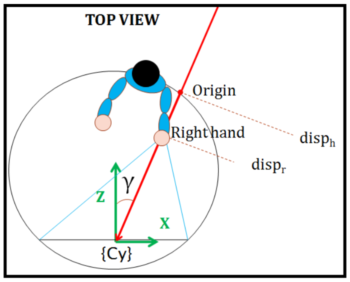


Figure 4: In red, the proposed *virtual depth* axis, increasing towards the center of the baseline with origin in the intersection between the axis and the ViethMuller circle.

fined, without loss of generalization as seen in (2), where $d$ defines the distance between the center of the baseline and the target object and the disparity can be depicted in Fig.4.

$$virtual\ depth \quad \approx \quad d - disparity\_value \qquad (2)$$

The result, regarding the Z coordinate in the camera referential $C$ is given by $z_h = \cos(\gamma) * disp_h$, and the right and left hand are respectively $z_r = \cos(\gamma) * (d - disp_r)$ and $z_l = \cos(\gamma) * (disp - d_l)$. The minus signals are due to the fact that the value for disparity is positive inside the circle, thus $depth < d \rightarrow depth = d - disp_r$. The angle $\gamma$ can be seen in Fig.3.

This tackles the depth problem. For the remaining 2 coordinates, $x$ and $y$, the homography [11] method is applied. Briefly, we can define homography as a geometrical method which allows a linear transformation (using the homographic matrix) of coordinates between two planes. Thus, by having the image coordinates, the resulting world coordinates are given by (3)

$$[x \quad y \quad 1]^t = H\,[u \quad v \quad 1]^t \qquad (3)$$

where H is the homographic matrix, calculated upon calibration method [11]

### 3.2 Features

Once the 3-D coordinates for head and hands are acquired, they are processed in order to obtain feasible features related with LMA *Shape* component.

This work proposes the definition of a triangle in the 3 dimensional space (Fig.5), whose vertices are defined by the head and hands coordinates, and study its behaviour in geometrical space. Some characteristics can be computed, e.g. we can retrieve its area $A$, as well as the *normal* vector $\vec{n}$. The relative movement between each of its vertices is implict within the orientation changes of the vector $\vec{n}$ within a period of time.

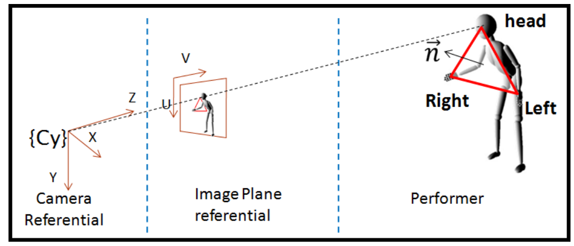


Figure 5: Triangle defined by head and hands defined in the Image and Cyclopean referentials.

Based on the three points, a plane can be defined and the normal vector $\vec{n}$ can be computed. Consider the points defined by the coordinates of the head, left and right hands as H,L and R respectively. Define two vectors $\vec{HR}$ and $\vec{HL}$ as $H - R$ and $H - L$.

By computing the cross product $\vec{HL} \times \vec{HR}$ we obtain the normal vector $\vec{n}$ which after normalization gives

$$\widehat{n} = (a, b, c) = \frac{\vec{n}}{\|\vec{n}\|} \qquad (4)$$

It was decided to have more information about the triangle to this vector. This encompassed the inclusion of the area $A$ information in the normal vector coordinates, hence the newly defined vector $\vec{V}$ is defined as

$$\vec{V} = (v_1, v_2, v_3) = A * \widehat{n} \qquad (5)$$

### 3.3 Spherical Space of Analysis

Vector $\vec{V}$ 3 dimensional and its values are unlimited. Having this fact in mind, it was decided to study this vector in a different space of analysis, the spherical. This invariant geometric space will allow to instantly retrieve orientation and size information.

$$\begin{aligned} r &= \sqrt{v_1^2 + v_2^2 + v_3^3} = A \\ \varphi &= atan2(v_2, v_1) \\ \theta &= \arccos \frac{v_3}{r} \end{aligned} \qquad (6)$$

As it can be seen in Fig.1, we have different descriptors depending on which of the 3 planes that compose the 3-D space we are processing. The sugested hypothesis tries to explore these invariant features in the spherical space over the conviction that the 3-D orientation and size will provide enough information to classify the overall *Shape Space* component. The next figures will allow to visualize the coordinate transformation.

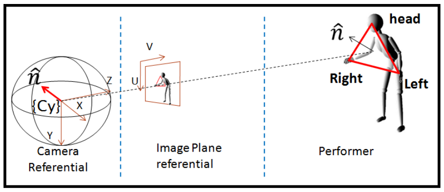


Figure 6: Normal vector to the triangle, projected in the camera referential in spherical coordinates.

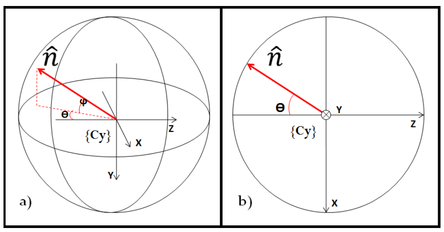

Figure 7: a) Projection of the normal vector $\widehat{n}$ in detail. b) Top view of the projection of $\widehat{n}$ to camera referential.

As it can be seen in Fig.7, if one hand is more active than other, this will originate dominance of one of the octants of the spherical space. Also the movement itself, will cause the area of the triangle to shift. For example the resulting signal for the area $A$ in a bye-bye movement, will result in a sine wave like, due to the fact the area will be increasing and descreasing in a somewhat periodic way. The spherical coordinates, hence compose the basic set of features for *Shape* component for LMA. These can be related to semantically describe *Shape* as seen in Fig.1 (right). The 'a-priori' proposed relations are sugested in Tab.1.

However, it will be shown in Section 4 an extensive study on how each of these features behaves in each of the qualities for *shape*, i.e. movements are labeled with *shape* qualities and a probabilistic study (in the form of histograms) will show the dominant features as well as expected patterns.

Table 1: Initial feature hypothesis for *Shape* component.

| Shape | Extremes | Related Physical feature |
|---|---|---|
| Flow | Shriking | Area $A$ decreasing |
| | Growing | Area $A$ increasing |
| Space | Retreating | back orientation |
| | Reaching | $A$ increasing, frontal orientation |
| | Spreading | $A$ increasing, side orientation |
| | Enclosing | $A$ decreasing, back orientation |

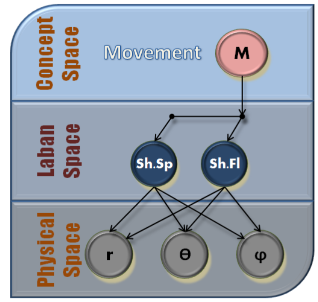


Figure 8: Bayesian model for the *Shape* component; The highest level is the Concept space, where the movement is defined; This node depends on the previous nodes within the Laban Space, which are dependent from the Physical Space, where the low-level features (LLF) are computed.

## 4 Probabilistic Modeling

### 4.1 Shape Model

The *Shape* model describes the changes in the shape of the human body through the movements of the limbs. It relates the low-level features like triangle changes $(\theta, \varphi, r)$ to Shape qualities like Shape Space (Sh.sp) and Shape Flow (Sh.Fl). The sets of variables are presented in (7).

$$\begin{aligned} Feat.Sh &\in \{\theta, \varphi, r\} \\ Shape &\in \{\ Sh.Sp, Sh.Fl\} \end{aligned} \tag{7}$$

The second level of the model relates the Shape qualities to a specific movement M as has been introduced in the previous section. The set of Shape variables to be used by our Bayesian model are shown

$$\begin{aligned} Sh.Fl &\in \{Shrink, still, growing\} \\ Sh.Sp &\in \{Enclosing, Spreading, Retreating\} \end{aligned} \tag{8}$$

The simplified model for the Shape component can be seen in the following figure.

Each of the nodes in the model correspond to a probabilistic distribution. For the movement we have

$$P(M|Sh.Fl \quad Sh.Sp) = P(M)P(Sh.Fl|M)P(Sh.Sp|M) \tag{9}$$

and the nodes corresponding to Shape itself are defined as

$$\begin{aligned} &P(Sh.Fl|r \quad \theta \quad \varphi) \\ &P(Sh.Sp|r \quad \theta \quad \varphi) \end{aligned} \tag{10}$$

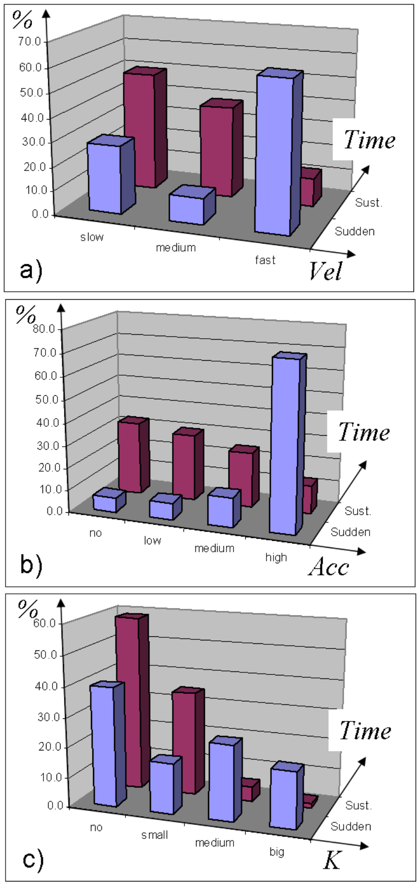


Figure 9: Behaviour of acceleration within the Time quality of *Effort* component

The lowest level distributions for the features, are represented by normal distributions, i.e. we compute for a number of trials of one determined movement, the mean and variance values for r, $\theta$ and $\varphi$, hence defining the probabilistic distribution (e.g. $P(r) = N(\mu_r, \sigma_r)$ where $\mu_r$ and $\sigma_r$ correspond to the average and variance for the value of r).

### 4.2 Feature evaluation

In order to prove the usability of these features, a probabilistic evaluation has to be made. As mentioned before, this is a follow up of the work taken by Rett J. et al. In [6] the authors implemented the *Space* component, which yelded a classification rate of 66%. In [7], one other component was implemented, the *Effort*. This component is theoretically taken as the most relevant when evaluanting emotion. It uses descriptors such as *sudden* and *sustained* to characterize movements. The standalone implementation had a database designed specifically for this component's study. Movements where 'spaciatilly' similar, but had different *effort* qualities, in a away to make them unique between each other. The database is composed by four different bye bye gestures and sequences of random actions, all performed with different *Emph* characteristics, yielding a data set whose movement were discretized mainly through their *Emph* parameters. The features used were extracted from the trajectories performed by the body parts in the period relative to the movement itself, and can be seen in (11)

Table 2: Results for *Space* and *Effort* components.

| | Space | Effort | Global |
|---|---|---|---|
| Classification rate | 63% | 77% | 83% |

$$LLF_{ef} \in \{velocity, acceleration, curvature\} \tag{11}$$

As an example, results from the feature evaluation can be quoted, for better understading of the analysis made. As it can be seen in Fig.9, for the LLF acceleration regarding quality *Time* of component *Effort*, results show that *sudden* movements have a predominance of high accelerations, whilst *sustained* ones have *low* or *no* acceleration as dominant. From this fact a pattern can be extracted and modeled, thus yielding *strong* characterization for the qualities of a component, in this specific case, *Effort*.

The results for *Effort* were considered good on their own, a the bayesian model implemented for its study, reached a classification of positive results of nearly 80%. Despite results were better than for *Space* component, Bayesian framework allowed further improvement. By fusing the two component models, *Effort* and *Space* into one global model, where the movement presents the following probabilistic distribution

$$P(M|Effort \quad Space) \tag{12}$$

Results increased the positive results from 66% and 77% to 83% using the global model, which could guide to the conclusion that Laban Movement Analysis has a potential to become a robust computational movement classifier, if more components were added to the global model. Table 2 summarizes these results, conducted in previous works, and supporting ongoing work.

Feature evaluation is assumed to be understood at this point, and the same study is being employed to study the *Shape* features. The invariant features chosen, in the spherical space of analysis, are thought as enough to characterize this component. It is intuitive that the area of the triangle, will be a strong index on whether the movement is *growing* or *shrinking*, and the relative movement between body parts can be implicit in the normal vector pose. Lets assume that the angle $\theta$ shifts from positive to negative values. This fact alone leads one to know that the hands are moving in different directions, relative to each other. To better visualize this fact, visual support is given with the help of Fig.10

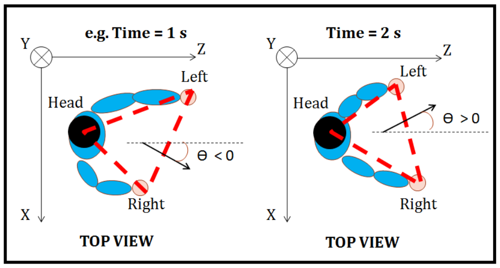


Figure 10: to consecutive instants of time, where the hands exhibit inverse relative movement, which is reflected by the angle $\theta$; In t = 1s, $Z_{left} > Z_{right}$ and in t = 2s $Z_{left} < Z_{right}$.

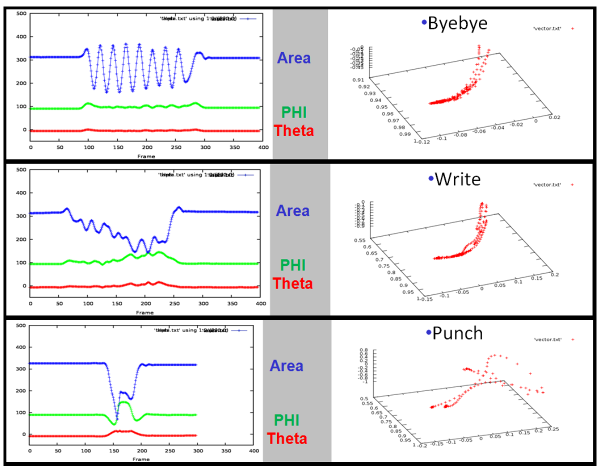


Figure 11: The movement trajectories: Bye-bye, Write and Punch and the behavior of $\phi$, $\theta$ (degrees) and the Area along the time of execution of each.

To evaluate the *Shape* features, a wider database was used, with movements like *Punch*, *Point*, *Shaking hands*, etc, and several trials per movements. An example of the results yielded by the feature behavior is presented in Fig.11.

As it can be seen, different movements present different patterns for the chosen features. It is felt that these features can be strengthen, and work on their study and development continues. This is a first approach to invariant features to evaluate the *Shape* component.

## 5 Conclusion

From this ongoing work, some conclusions arise. The main one is that there is a need to implement the four basic Laban Movement Analysis components in order to fully evaluate its potential as a standard computational movement classifier. It has been proved that, the more components are implemented, the better results are achieve. Also, it is possible to extract using only vision sensing, simple mathematical features from the trajectories generated by tracking a small number of body parts instead of full body geometric models. These simple evidences, are so far, considered to be enough to provide good results, within a probabilistic approach, based on Bayesian frameworks. The approach, allows one to implement components separately, and to fuse them in upper layers of the model. Related to the *Shape* component, it was felt the need to add invariant features, because the shape taken by the body along some movements, can generate very complex, high dimensional data for analysis, which is the inverse of what our models pretend to achieve. Results already show that chosen patterns yield different behavior for different movements. Steps are being taken to complete all models within LMA, and thus generating, what is thought to be a very capable and robust movement classifier, with expressiveness content within resulting analysis.

## Acknowledgement

This work is partially supported by the BACS-project-6th Framework Programme of the European Commission contract number: FP6-IST-027140.

## References

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

[2] D. Chi, M. Costa, L. Zhao, and N. Badler, "The emote model for effort and shape," in *SIGGRAPH 00, Computer Graphics Proceedings, Annual Conference Series*, ser. ACM SIGGRAPH. ACM Press, July 2000, pp. 173–182.

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

[4] 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.

[5] J. Rett, "Robot-human interface using laban movement analysis inside a bayesian framework," Ph.D. dissertation, University of Coimbra, 2009.

[6] J.Rett, L. Santos, and J. Dias, "Laban movement analysis for multi-ocular systems," in *IEEE/RSJ International Conferece on Intelligent Robots and Systems (IROS08)*, 2008.

[7] J. Rett, J. Ahuactzin, and J. Dias, *International Journal of Reasoning-based Intelligent Systems (IJRIS)*. ., 2008, ch. Bayesian reasoning for Laban Movement Analysis used in Human Machine Interaction.

[8] P. Bessire, “Survey: Probabilistic methodology and techniques for artefact conception and development,” INRIA Rhne-Alpes, Tech. Rep., 2003.

[9] Videre. Videre design. http://www.videredesign.com/.

[10] O. Faugeras, B. Hotz, H. Mathieu, T. V. ville, Z. Zhang, P. F. E. T. ron, L. Moll, G. Berry, J. Vuillemin, P. Bertin, and C. Proy, “Real time correlation-based stereo algorithm, implementations and application,” INRIA, Technical Report 2013, Tech. Rep., 1993.

[11] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.