Interactive Systems Design: A KANSEI-based Approach[1]

Antonio Camurri **Riccardo Trocca** **Gualtiero Volpe**
InfoMus - Laboratory on Musical Informatics (http://infomus.dist.unige.it)
DIST - University of Genoa, Viale Causa 13, I-16145 Genoa, Italy

Abstract
This paper presents some our recent research on computational models and algorithms for real-time analysis of full-body human movement. The focus here is on techniques to extract in real-time expressive cues relevant to KANSEI and emotional content in human *expressive gesture*, e.g., in dance and music performances. Expressive gesture can contribute to new perspectives for the design of interactive systems. The EyesWeb open software platform is a main concrete result from our research work. EyesWeb is used in interactive applications, including music and other artistic productions, museum interactive exhibits, therapy and rehabilitation, based on the paradigm of expressive gesture. EyesWeb is freely available from `www.eyesweb.org`.

1. INTRODUCTION

The integration of gesture, music, and visual languages, to enable artists with conceptual as well as practical tools for creating novel participative, shared, interactive, mixed-media performance spaces is a challenge faced by several researchers (Battier and Wanderley 2000; Camurri and Trocca 2000; Machover 1989; Rowe 1993, 2001; Winkler 1998). Hyper-instruments, interactive dance, museum exhibits, therapy and rehabilitation are examples of possible concrete applications. In a broader perspective, the goal is to enhance man-machine communication and interfaces in multimedia systems.

We deem that the process of design of interactive systems can benefit by adding a new channel: expressiveness. Computational models, analysis, and synthesis of expressive content are scarcely considered in current state of the art of interactive systems and interfaces. In particular, we refer to the inner qualities of *gesture*: expressive cues, emotional and KANSEI (Hashimoto 1997) content.

The modeling of *Expressive Gesture* is a central issue of our work (Camurri, De Poli, Leman and Volpe 2001). The concept of expressive gesture includes musical, human movement, visual (e.g. computer animated) gesture. This paper focuses on the problems of real-time analysis of expressive gesture in human movement, and on how the introduction of expressive gesture can affect the design of interactive systems. The research presented in this paper is part of a broader project on real-time analysis of expressive gesture in artistic contexts (with particular reference to non-verbal communication through human movement and music signals), and on multimodal/cross-modal mapping.

Gestures carry what Cowie (2001) calls implicit messages: the same action can be performed in several different ways, stressing different qualities of the movement, expressing feelings, moods, intentions. It is also possible to recognize a person from the way he/she moves. The attention is focused in this paper on dance, an artistic expression of human gesture with strong potential of emotional communication and arousal on spectators.

The analysis methods and computational models for analyzing expressive gesture are inspired to several sources, including KANSEI Information Processing research (Hashimoto 1997), humanistic theories of non-verbal communication, such as Rudolf Laban's Theory of Effort (Laban and Lawrence, 1947; Laban, 1963), developed for dance and choreography and Shaeffer's Morphology (Schaeffer, 1977).

KANSEI Information Processing has been proposed as the third target of information processing. Hashimoto (1997) argued that the first target of information processing is the physical signal (i.e. sound, light, force), the second is language (i.e., the field of logic, of symbolic knowledge), the third is KANSEI, that refers to feelings, intuition, sympathy.

Our research can be considered as an attempt to shed some light on aspects related to the KANSEI communication in artistic performance, and to formulate hypotheses on "KANSEI rating" by developing and experimenting analysis and mapping techniques.

Consolidated research results are transferred into software libraries for our EyesWeb open software platform (www.eyesweb.org), also widely used in public concerts and a number of applications.

2. METHODOLOGY

A set of evoked basic emotions, the arousal or engagement on spectators by an artistic performance, or other components of KANSEI or emotion communication are usually described in terms of emotional spaces. The model of a KANSEI evaluation system may consist of the following components:

1. The *KANSEI function* maps features of the studied phenomenon to a space (e.g., an emotional space). This function models the interaction between the physical world and the emotional space, emulating the effects that certain physical features of the studied phenomenon would have on the evoked emotional response.

[1] This research is partially funded by the EU IST Projects CARE HERE (Creating Aesthetically Resonant Environments for the Handicapped, Elderly and Rehabilitation) no. IST-2001-32729, MEGA (Multisensory Expressive Gesture Applications) no. IST-1999-20410 (www.megaproject.org), and by National Projects COFIN2000 and CNR project CNRG0024AF "Metodi di analisi dell'espressività nel movimento umano per applicazioni in Virtual Environment".

2. The *Interpretation Function* of a point in the emotional space. For example, a function expressing the distance of a point from a set of labelled emotions in that space (e.g. in the well-known circumplex model, valence/arousal).

The definition of the KANSEI and emotional spaces and the labelling of relevant points, e.g. in terms of basic emotions, is a crucial issue, widely faced by psychologists (see e.g. the survey in Cowie 2001). Another crucial issue concerns the modeling of the interpretation function. This can be based on different approaches, for example neural networks or clustering algorithms. In this way, neural networks are used to find non-linear relations between physical measures and the KANSEI space. An example can be found in (Suzuki and Hashimoto, 1997), focusing on sound perception, where a neural network is trained to place its output in a sort of KANSEI space.

Another family of approaches involves the creation of an explicit description of the phenomenon. For example, starting from a signal on human movement, this approach tries to reconstruct a description in terms of expressive cues, such as fluentness, directness, energy, rhythm, of shapes, phrases, etc. As an analogy, in music this would be equivalent to build a perceptual representation of the music signal in terms of objects, for example based on Shaeffer's morphology (1977). Such symbolic description would allow a system to detect perceptual patterns in the studied event (or series of events), to study each object's relavant cues and invariants. Then, the next step consists of capturing more complex structural relations that might require the use of logics.

So, in short, we might say that a KANSEI-oriented type of approach the focus is on an implicit, emerging model of KANSEI space, while in another (we might say affective or emotional type of approach) the focus is on explicit models of the emotional space.

The important issue here is that KANSEI or emotional spaces can play an important role in the design of interaction and mapping in interactive systems.

2.1 Conceptual Architecture

G*esture* is the carrier of a set of temporal/spatial features responsible of conveying expressiveness. In this perspective, this work adopts the general guidelines of the layered conceptual framework for expressive gesture applications described in (Camurri, De Poli, Leman, Volpe, 2001), and discusses the relations to the previously mentioned KANSEI approach, on the special case of human movement analysis.

Analysis of movement is performed through different layers/steps following a bottom-up approach:

Layer 1 – Physical Signals: This is the information that is captured by the sensors of a computer system. Physical signals may have different formats strongly dependent on the kind of sensors used to study movement. For example, sampled signals from tactile, haptic, IR or US sensors, or low-level data frames in video. The term "sensor" is related to both the physical sensors employed and the algorithm used to extract a given set of low level data. We can therefore speak of "virtual" or "emulated" sensors. A CCD camera can be an example of a physical sensor, while the *optical flow* or the *motion templates* are examples of data extracted from "virtual sensors" implemented by the cited algorithms.

Layer 2 – Low-level features and statistical parameters: A collection of motion cues describing the movements being performed. Motion cues are usually processed by statistical techniques. Examples include the amounts of contraction/expansion, the stability, the "rotation" in movement, the equilibrium. An important set of cues are those inspired to the effort dimensions described in Laban's Theory of Effort (1947, 1963): space, time, weight, and flow. For example: straightness (i.e., how much a movement is direct or flexible), impulsiveness (a movement can be quick or sustained), fluency (bounded or unbounded movements).

Layer 3 – Mid-level features and maps: "In this layer, the purpose is to represent expression in gestures by modelling the low-level features in such a way that they give an account of expressiveness in terms of events, shapes, patterns or as trajectories in spaces or maps." (Camurri, De Poli, Leman, Volpe, 2001). Data from several different physical and virtual sensors are likely to be integrated in order to perform such a step. A movement sequence is divided in gestures. Each gesture is characterized by specific cues (layer 2: e.g., speed, impulsiveness, straightness). The problem here is to identify relevant strokes in movement stream and associate to them the qualities/cues deemed important for expressive communication. For example in dance analysis, a fragment of a performance might be segmented into a sequence of gestures where gesture's boundaries are detected by studying velocity and direction variations, thus identifying a sequence of trajectories in a semantic space. A gesture is a trajectory in such a map, representing categories of semantic features related to emotion and expression on a pre-defined grid. A sequence of gestures is associated to a sequence of trajectories in the map.

Layer 4 – Concepts and structures: for example, emotional content and KANSEI concepts: basic emotions (fear, grief…), arousal or emotional engagement measures in spectators of the gesture, intentional gestures such as Laban's types of effort ("pushing", "gliding", …). This high-level information is built from the other layers, using various analysis techniques (e.g., statistical, time series).

Following the scheme depicted in the previous section, the *KANSEI function* lies in the first three layers, while the *Interpretation Function* is a main concern of Layer 4.

Music gesture follows the same conceptual architecture, as described in (Camurri, De Poli, Leman, Volpe, 2001): from musical low-level signals (audio, midi etc) at Layer 1, to general concepts of Layer 4.

The approach to the design of interactive systems is grounded on this proposal of unified gesture architecture for movement and music gesture.

2.2 Interactive systems design: some examples

The inclusion of a layered model of expressive and emotional features can contribute to improve the design of interactive systems.

An example is the demo of "expressive hi-fi system" presented at the EU-IST-E3 Booth at IBC2001 (Intl Broadcasting Convention, Amsterdam, Sept 2001) that we developed in collaboration with CSC-DEI University of Padova in the framework of the EU-IST MEGA Project. A music piece (Chopin and Mozart pianoforte pieces were used in this demo) is played, allowing the user to control not only the usual “volume”, “balance” etc. knobs of a standard hi-fi system, but also “knobs” related to the interpretation of the piece. User’s full-body movement is analysed and a set of expressive cues are extracted in real-time (layer 2). This cues array is then compressed (the previously mentioned Kansei function) to a 2-dimensional space (the expressive space, see figure 4). In this example, the two axes of the expressive space (layer 3) correspond to the features *fluentness* and *quantity of movement*. Then, another function maps a point on this space into an array of music cues (e.g. IOI, dynamics, …) which are used in real time to modulate the performance of the music piece. In this demo we developed an intuitive mapping, at different time scales, of expression in movement into coherent music interpretation parameters: e.g., a “light” style of movement causes a tendency toward a “light” (*leggero*) interpretation of the piano piece.

We participated to several artistic productions and workshops with artists, where we explored the role of expressive gesture in interactive systems. Several useful directives matured and some lessons were learned from such work with artists. An example: since expressive data varies slowly, the perception of the interaction process (awareness in the performers as well as in spectators) can be lost if only expressive information is used in the loop. Direct immediate cause-effect reactive mappings (the usual musical instrument metaphor adopted in interactive systems) can be integrated to consolidate this perception of dialogue and interaction in a performance. The metaphor of interaction goes beyond the “musical instrument”. It is not (only) the dance or the body that performs the music as a musical instrument, but rather there is a counterpoint, a dialogue mechanism between humans and virtual (acoustic/musical as well as visual or robotic) participants, where different modalities including music, human movement (dance), visual media (eg animated figures or synthetised images, see e.g. Fels and Mase 1999), mobile scenery on stage (e.g. mobile robots), in a mixed-media distributed scenario. See figures 5 and 7 for simple examples on visual mapping of expressive cues.

2.3 Analysing human movement: Microdances

A reference archive of microdances has being created and studied. We call “microdance” a short fragment of choreography, of typical duration of 15-90s. A microdance is conceived as a potential carrier of expressive information. Several performances of the same microdance can convey different expressive or emotional content to spectators: e.g. light/heavy, fluent/rigid, happy/sad dimensions. A microdance is the analogous of a musical score fragment, which can performed with different interpretation. Humans (e.g. spectators) evaluate each microdance performance. The outputs of developed algorithms for analysis of expressive cues are compared with corresponding spectators’ rating of the same microdance performance. Microdances can be useful to isolate factors related to KANSEI and expressiveness, to help to provide an experimental evidence with respect to the cues that choreographs and psychologists identified.

2.4 Subtractive analysis approach

One of the main challenges is to identify basic factors of KANSEI and of deep emotional arousal or engagement in spectators observing a dance performance. The same applies to spectators of a music performance. In general, our aim is to unify the approaches to movement and music.

To this aim, we are developing an approach based on the live observation of genuinely artistic performances, and corresponding video (audio) recordings. A reference archive of artistic performances has to be carefully defined for this method, chosen after a strict interaction with artists. Image (audio) processing techniques are utilized to gradually subtract information from the recordings. For example, parts of the dancer’s body could be progressively hidden until only a set of moving points remain, deforming filters could be applied (e.g., blur), the frame rate could be slowed down, etc. Each time information is reduced, spectators are asked to rate the intensity of their “arousal” in a scale ranging from negative to positive values (a negative value meaning that the video fragment would rise some feeling in the spectator but such feeling is a negative one). The transitions between positive and negatives rates and a rate of zero (i.e. no expressiveness was found by the spectator in the analyzed video sequence) would help to identify what are the movement features carrying expressive information. A deep interaction is needed between the image processing phase (i.e. the decisions on what information has to be subtracted) and the rating phase.

This subtractive approach is different from the previous studies by Johansson (1973) and more recently by Cowie (2001), where it is demonstrated that a limited number of visible points on human joints allow an observer to recognise information of movement, including certain emotional content.

Our subtractive method is currently subject to investigations and experiments. The feedback from these experiments provides information on which movement cues our research should focus on. The cues described in the following sections are also motivated by the results of these preliminary experiments.

2.5 Analysis Perspectives: Kinesphere and General Space

According to Laban, a main distinction exists between the analysis of movement in the *Personal Space*, referred also as *Kinesphere*, and the analysis of movement in the *General Space*. In “Modern Educational Dance” (Laban 1963, p. 85) Laban writes: “Whenever the body moves or stands, it is surrounded by space. Around the body is the sphere of movement, or Kinesphere, the circumference of which can be reached by normally extended limbs without changing one’s stance, that is, the place of support. The imaginary inner edge of this sphere can be touched by hands and feet, and all points of it can be reached. Outside this immediate sphere lies the “general” space, in which the human can enter only by moving away from his/her original stance. He/she has to step outside the borders of his immediate sphere and create a new one from the new stance, or, in other words, he transfers what might be called his "personal" sphere to another place in the general space. Thus, in actual fact, he never goes outside his personal sphere of movement, but carries it around with him like a shell."

Movement is therefore considered from two different perspectives:

1. Detailed movement of a single human, e.g., the movement of the centre of gravity or the joints of a dancer, in his own "Kinesphere" or "Personal Space";
2. Movement of one or more humans in a wider space, the "General Space" (e.g., a group of dancers moving on a stage, a group of visitors in a museum exhibit).

3. EXAMPLES OF ANALYSIS IN THE PERSONAL SPACE

The methodology sketched above has been applied to analyse movement of dancers both in the Personal Space and in the General Space. Our General Space methods are described in (Camurri Mazzarino Trocca Volpe 2001). Here some examples of analyses in the Personal Space are presented, organized on the different levels according to the described layered approach:

(i) Processing of low-level data coming from a camera (Layer 1): background subtraction techniques are used in order to extract the dancer's silhouette. The resulting images are then used to calculate *Silhouette Motion Images* (SMI).

(ii) Extraction of low level features and parameters (Layer 2): in particular, the quantity of motion and the contraction index are here presented as examples of cues at this level.

(iii) Segmentation of movement in motion and pause phases (Layer 3) by using the quantity of motion calculated in the previous layer 2.

(iv) Examples of gesture representations by means of suitable feature spaces and/or symbolical descriptions are also shortly sketched.

Analysis is here performed in real-time on a set of recorded (or live) microdances using a collection of software modules implemented in the framework of the EyesWeb open architecture for expressive gesture processing (Camurri et al, 2000).

3.1. Layer 1: Silhouette Motion Images (SMI)

A Silhouette Motion Image is an image carrying information about variations of the silhouette shape and position in the last few frames. SMIs can be seen as a special case of Motion Templates (see OpenCV Reference Manual: www.intel.com/research/mrl/research/opencv), where information about time is implicit in the image and not explicitly recorded. The SMI is generated by the following formula:

$$motion\_image[t] = \sum_i silhouette[t-i] - silhouette[t]$$

The motion image at frame t is generated adding images of the silhouette in the previous N frames and then subtracting the silhouette at frame t. The resulting image contains just variations happened in the previous frames. If N is the number of frames in which the SMI is calculated and N=1, then the SMI carries information about the instantaneous variations of the silhouette. Working with $N>1$ allows capturing more information about the shape of motion and results are smoother, because the effect is similar to filtering. Figure 1 shows a SMI with n=4. In the figure, the SMI is the grey area, while the darker contour shows the current silhouette.

Figure 1. SMI with n=4.

3.2. Layer 2: Quantity of Motion

The simplest use of a SMI is calculating its area. The result can be thought as a rough approximation of the quantity of motion, i.e. $q=m * v$, where m is the mass and v stands for velocity. Of course the area of a SMI is not q, but the behavior is similar: actually the shape of the graph is close to the shape of the graphs of velocity of a marker put on a limb.

However the SMI area alone is not a very reliable measure of movement, first because it suffers the same limitations of silhouette ("internal" motion is not detected, but more sophisticated techniques can overcome this problem); second, it is strongly dependent on the dancer's distance from the camera; third it is difficult to compare results of different dancers. It is possible to partially overcome the last problem scaling the SMI area by the area of the most recent silhouette.

$$Movement=Area(SMI[t,n])/Area(Silhouette[t])$$

In this way the measure becomes almost independent from the camera's distance and it is expressed in terms of fractions of the body area that moved. For example it is possible to say that at instant t a movement corresponding to the 2.5% of the total area covered by the silhouette happened.

3.3. Layer 2: Contraction Index

The contraction index is a measure, ranging from 0 to 1, of how the dancer's body uses the space surrounding it. We define a bounding box that surrounds the dancer's whole body and compare the area covered by this rectangle with the area actually covered by the silhouette. Intuitively, if the limbs are fully stretched and not lying along the body, the contraction index will be low, while, if the limbs are kept tightly nearby the body, the contraction index will be high (near to 1). While the dancer is moving, the contraction index varies continuously. Even if it is used with data from only one camera, its information are still reliable, being almost independent from the distance of the dancer from the camera. Of course, in case of too long distance, image quantization problems appear.

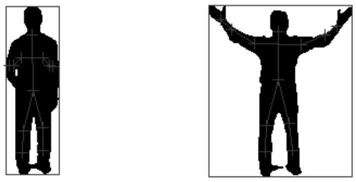


Figure 2. Silhouettes and their bounding boxes The leftmost one has high contraction index while the other has low contraction index.

Figure 2 shows two examples of silhouette, displayed with their bounding boxes, with high and low, respectively, contraction indexes. A possible use of this parameter consists of sampling its values at the end and beginning of a stretch of movement, in order to classify that movement as a contraction or expansion.

3.4. Layer 3: Motion segmentation and gesture representation

The SMI has interesting properties: the evolution in time of its (normalized) area (what we called quantity of motion) resembles the evolution of velocity of biological motion, which can be roughly described as a sequence of bell-shaped curves (*motion bells*). In order to segment motion by identifying the component gestures, it is interesting to extract a list of these motion bells and their features, e.g. peak value and duration. This can be also useful to obtain a first simple symbolical description of motion. One of the problems with the SMI approach, even dividing it in two vertical halves, is that several different movements may result super-imposed to each other, resulting in several motion bells to be overlapped. It is necessary to separate those motion bells in order to have a better description of motion. A first attempt at this consists in recognizing phases during which the dancer is moving (motion phases) and phases during which the dancer *does not appear (i.e, movement is not perceived by a spectator) to move* (pause phases). Actually, even if the dancer seems not to be moving, very small movements occur and they are detected by the motion image (together with some noise). An empirical threshold has been defined: the dancer is considered to be moving if the area of the motion image is greater than 2.5% of the total area of the silhouette. Figure 3 shows motion bells after segmentation: a motion bell characterizes each motion phase.

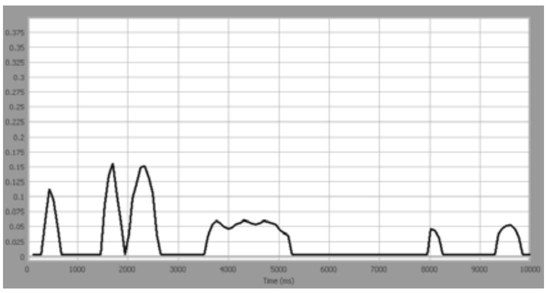
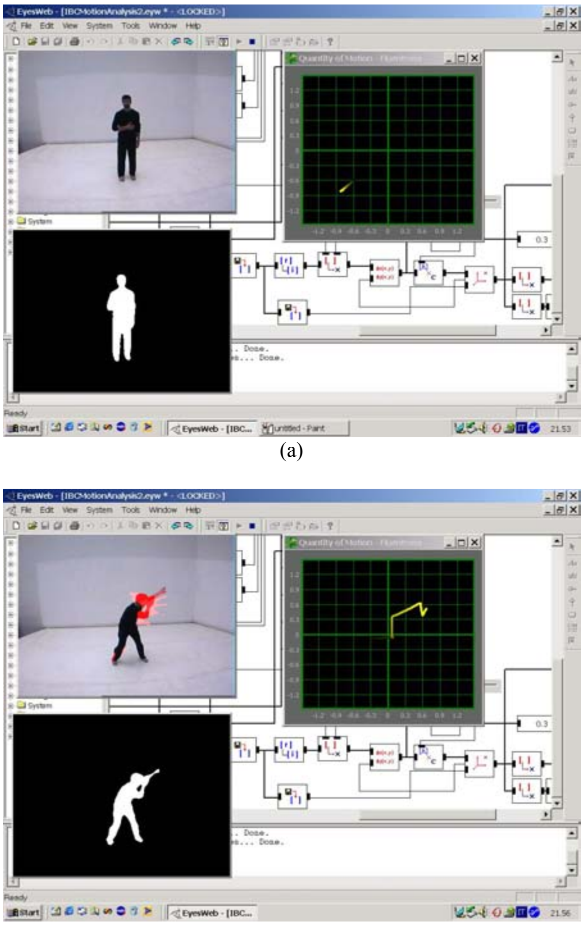


Figure 3. Motion segmentation

Motion bells may also be represented as trajectories in suitable spaces. For example, figures 4a and 4b show a running EyesWeb patch in which features are represented as trajectories in a 2D space. The dimensions are quantity of motion and fluentness.



(a)

(b)

Figure 4a and 4b. A running EyesWeb patch: gestures are represented as trajectories in a 2D space (window on the right: X axis is fluentness, Y axis is quantity of movement).

In figure 4a the dancer is not moving: the current position in the space (window in the right) is moving toward the bottom left parts of the 2D space (yellow stripe), a position characterized by low quantity of motion and low fluentness (i.e., the amount of pause phases is dominating the amount of motion phases). In figure 4b, a high-energy fluent gesture is displayed. The red shadow around the dancer in the upper-left window of figure 4b is is the SMI. The yellow point in the right window is then moving toward the top-right region in that window, characterized by high quantity of motion and high fluentness.

It is interesting to notice that the *motion bells* approach can be successfully applied also to sound signal analysis.

3.5. Toward a symbolic description of human movement

An output of Layer 3 is a symbolic description of the movement, which can be further analyzed to produce inferences on the underlying emotional content (e.g., in terms of basic emotions expressed by the dancer). The following example is obtained by using the "quantity of motion" to operate segmentation and the "contraction index" to distinguish between contraction and expansion phases of movement:

```
pause(9, 32).
expansion( 41, 4, 0.032, 0.023, 2, 0.034,
```

```
-0.087, 0.40).
pause(45, 6).
contraction(51, 12, 0.035, 0.035, 9, 0.080,
0.092, 0.15).
pause(63, 64).
contraction(127, 5, 0.036, 0.024, 2, 0.039,
0.76, 0.83).
```

The meaning of each term in the symbolic description above is the following:

- *Pause(start_frame, length)* corresponds to a phase of stillness that started at *start_frame* and lasted *length* frames.
- *Contraction(start_frame,length,start_value,end_value, peak_value_offset,peak_value,contraction_index_delta, start_contraction_index)* corresponds to a phase of motion, started at frame *start_frame* and lasted *length* frames. First and last SMI area values are *start_value* and *end_value,* while *peak_value_offset* says after how many frames the peak value *peak_value* was found. Finally *contraction_index_delta* shows the variation of the contraction index between the start and the end of the phase, and *start_contraction_index* says what was the value of the contraction index at the first frame of this phase.
- *Expansion(...)* the same has before, but while *contraction(...)* has a *contraction_index_delta* greater than 0, for an expansion it is less than 0.

4. IMPLEMENTATION: THE EYESWEB MOTION ANALYSIS LIBRARY

The work described in the previous sections is part of a collection of software modules for the EyesWeb open architecture for expressive gesture processing (Camurri et al, 2000; www.eyesweb.org). These and other modules are grouped as a separate library of EyesWeb for real-time analysis of expressive cues: the *EyesWeb Motion Analysis Library*. It includes:

(i) Blocks and patches for extraction and pre-processing of physical signals (typically, video frames from videocameras): e.g., feature tracking using the Lucas and Kanade algorithm (see figure 6b);

(ii) Blocks and patches for extraction and processing of low-level features and statistical parameters: e.g., contraction index, directness index, stability index, quantity of motion, fluentness, pause and motion durations (see the examples of cues in figures 5 and 7);

(iii) Blocks and patches for posture recognition using various techniques, e.g., Hu moments (Hu, 1962). For example, posture recognition enables to associate postures to pause phases. Body postures and postural attitudes can have an important role in conveying expressive intentions. Argyle (Argyle, 1980) discusses the importance of postural attitudes in non-verbal communication: postures are used to express interpersonal attitudes, emotions, and personality traits;

(iv) Blocks and patches for analysis of movement in Laban's General Space: e.g., position in the General Space, expressive potential fields, occupation rates (Camurri, Mazzarino, Trocca, Volpe, 2001);

(v) Patches for segmentation of movement in pause and motion phases.

5. CONCLUSIONS AND FUTURE WORKS

This paper illustrates a methodology to face the problem of KANSEI analysis in human movement. This methodology is the application to human movement of a wider conceptual framework dealing with expressive gestures especially in their artistic manifestations, i.e., in interactive systems involving music, dance, visual media. In this perspective, this work should be considered as part of a broader research context in which KANSEI is seen as a level enabling a deeper integration of interaction and mapping strategies in interactive systems. Our aim is to contribute to better support cross-language interactions, and enhance human-computer communication. As a consequence, if from one hand future research will aim to better understand KANSEI by widening the set of cues to measure and developing cross-modal cues (e.g., based on comparisons of Schaeffer morphology with Laban's Effort theories), building suitable representations for expressive gestures, developing algorithms correlating concepts (Layer 4) to measured values of lower level cues. In particular, cross-modal integration (i.e., to relate these findings with similar analysis in other domains such as music and visual media), and to mapping strategies (i.e., the possibility for an automatic system to synthesize suitable expressive outputs depending on the analyzed KANSEI) is still needed. Aesthetic and artistic implications still require further work, and composers and artists in general still need to metabolize these results in order to produce consolidated artistic outputs.

Figures 5 and 7 show simple but effective examples of mapping of movement cues on perceptually relevant visual cues. (We use these techniques to speed-up the study of microdances with dancers and choreographers). Similar examples are used with therapists for improving therapy and rehabilitation in Parkinson disease and severely handicapped children in the EU-IST project CARE HERE.

The EyesWeb Motion Analysis library have been used in several public events, including interactive concerts (e.g. Roberto Doati work opening La Biennale concert season, Sept 2001), and in museum installations (e.g. permanent interactive exhibits at Città della Scienza science center, Napoli).

ACNOWLEDGEMENTS

We thank our collegues of the EyesWeb s/w development staff (Paolo Coletta, Massimiliano Peri, Andrea Ricci), the student Barbara Mazzarino, Eidomedia and NumenSoft.

REFERENCES

Argyle M, Bodily Communication, Methuen&Co Ltd, London, 1980.

Battier M. and Wanderley M. (2000). Trends in Gestural Control of Music. Ircam Publ (CDROM)

Camurri A., De Poli G., Leman M., and Volpe G., A multi-layered conceptual framework for expressive gesture applications, in Proc. Workshop on Current Research Directions in Computer Music, Barcelona, November 2001.

Camurri A., Coletta P., Peri M., Ricchetti M., Ricci A., Trocca R., Volpe G., A real-time platform for interactive dance and music systems, in Proc. Intl. Conf. ICMC2000, Berlin, 2000.

A.Camurri, R.Trocca (2000). Movement and gesture in intelligent interactive music systems. In Battier and Wanderley (Eds.). Trends in Gestural Control of Music. Ircam Publ.

Camurri, A., Mazzarino, B., Trocca, R., Volpe, G. (2001). Real-Time Analysis of Expressive Cues in Human Movement. Proc. CAST01, GMD, St.Augustin-Bonn, Sept 2001.

Cowie R., Douglas-Cowie E., Tsapatsoulis N., Votsis G., Kollias S., Fellenz W. and Taylor J., Emotion Recognition in Human-Computer Interaction, IEEE Signal Processing Magazine, no. 1, January 2001.

Fels S., Mase K. 1999. Iamascope: A Graphical Musical Instrument. Computers and Graphics, 23(2), 277-286, Elsevier.

Hashimoto S. (1997) KANSEI as the Third Target of Information Processing and Related Topics in Japan, in Camurri A. (Ed.) "Proceedings of the International Workshop on KANSEI: The technology of emotion", AIMI (Italian Computer Music Association) and DIST-University of Genova, pp101-104

Hu, M.K. Visual pattern recognition by moment invariants. IRE Trans. Information Theory, vol. IT-8, pp. 179—187, Feb. 1962.

Johansson G (1973) *Visual perception of biological motion and a model for its analysis.* Perception & Psychophysics, 14, 201-211.

Johnson-Laird P. M., Mental models and probabilistic thinking. Cognition, 50, 1994.

Laban R., Lawrence F.C., Effort, Macdonald&Evans Ltd. London, 1947.

Laban R., Modern Educational Dance, Macdonald & Evans Ltd. London, 1963.

LeDoux J.E., The Emotional Brain, New York: Simon & Schuster, 1996.

Machover, T., J.Chung. 1989. Hyperinstruments: Musically intelligent and interactive performance and creativity systems. Proc. ICMC1989. San Francisco: ICMA, pp.186-190.

Rowe, R. 1993. Interactive Music Systems. MIT Press.

Rowe, R. 2001. Machine Musichianship. MIT Press.

Schaeffer, P., Traité des Objets Musicaux. Second Edition. Paris. Editions du Seuil, 1977.

Suzuki K., Hashimoto S., Modeling of Emotional Sound Space Using Neural Networks, in Camurri, A. (Ed.) Proc. Intl Workshop on KANSEI: The technology of emotion, AIMI and DIST-University of Genova, Genova, 1997, pp 116-121.

Wallbott H.G., The measurement of Human Expressions, in Walbunga von Rallfer-Engel, "Aspects of communications", pp. 203-228, 1980.

Winkler, T. 1998. "Composing Interactive Music." MIT Press, Cambridge, MA.

Open Computer Vision Library, Reference Manual. Available at: http://www.intel.com/research/mrl/research/opencv/ and http://sourceforge.net/projects/opencvlibrary/

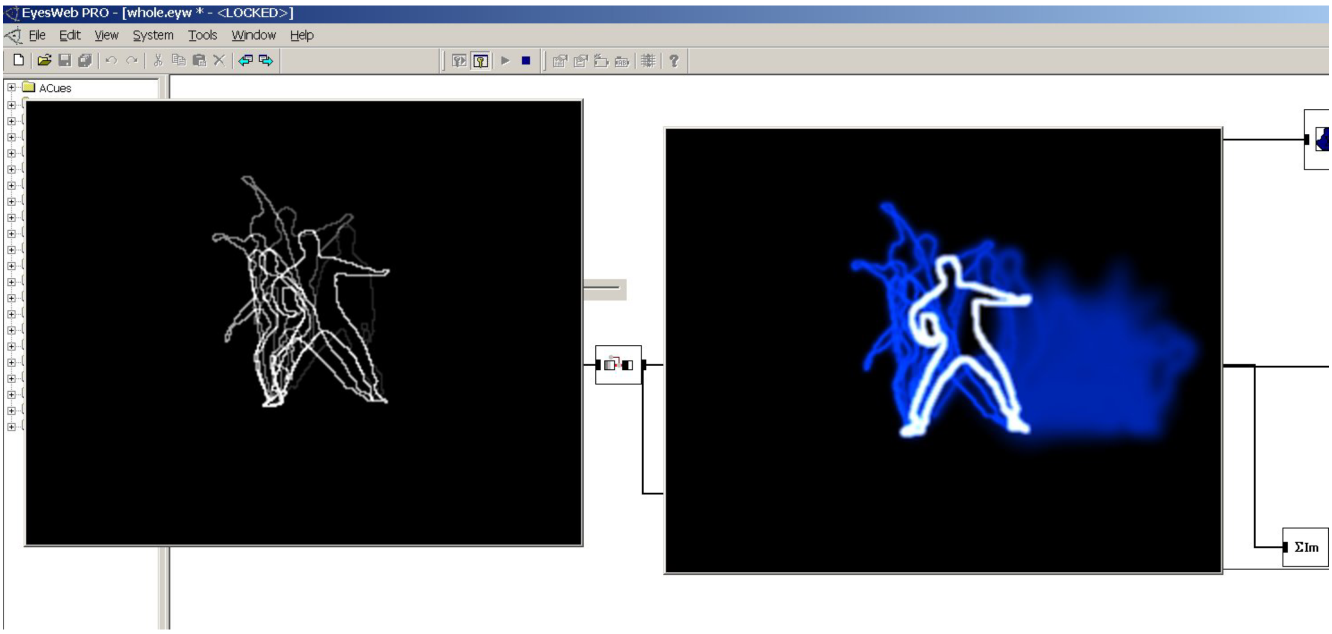


Figure 5: Two examples of real-time extraction of cues from dance (using a single videocamera). Left window: extraction of silhouettes only in points where velocity has a peak. Right window: visualisation of "how the dancer occupies (or sculpt) space"; the blue zone shows visually this cue.

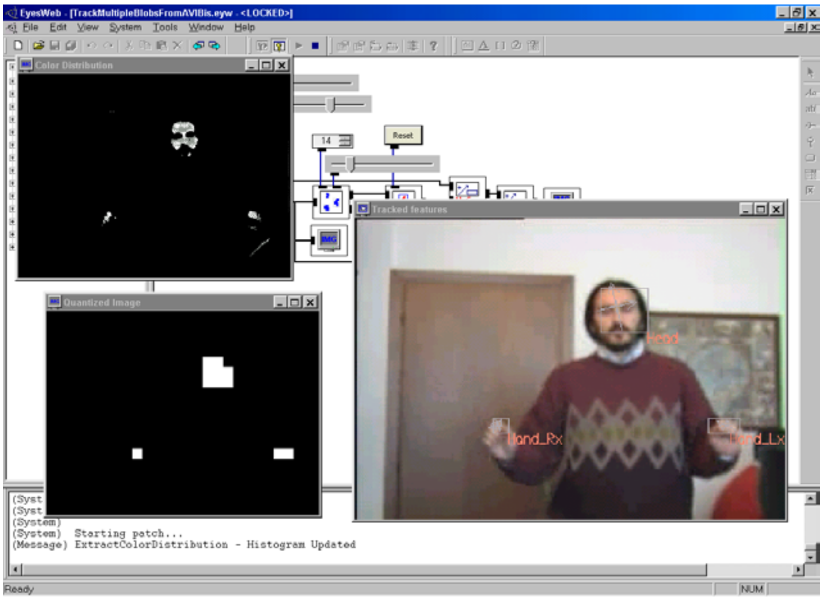


Figure 6a: An EyesWeb patch showing the color blob tracker (dev by M.Peri and A.Ricci).

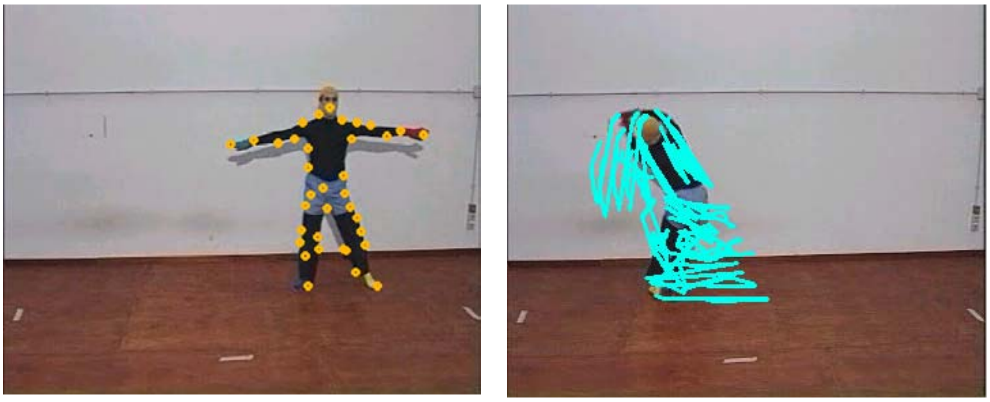


Figure 6b: LK Trackers.

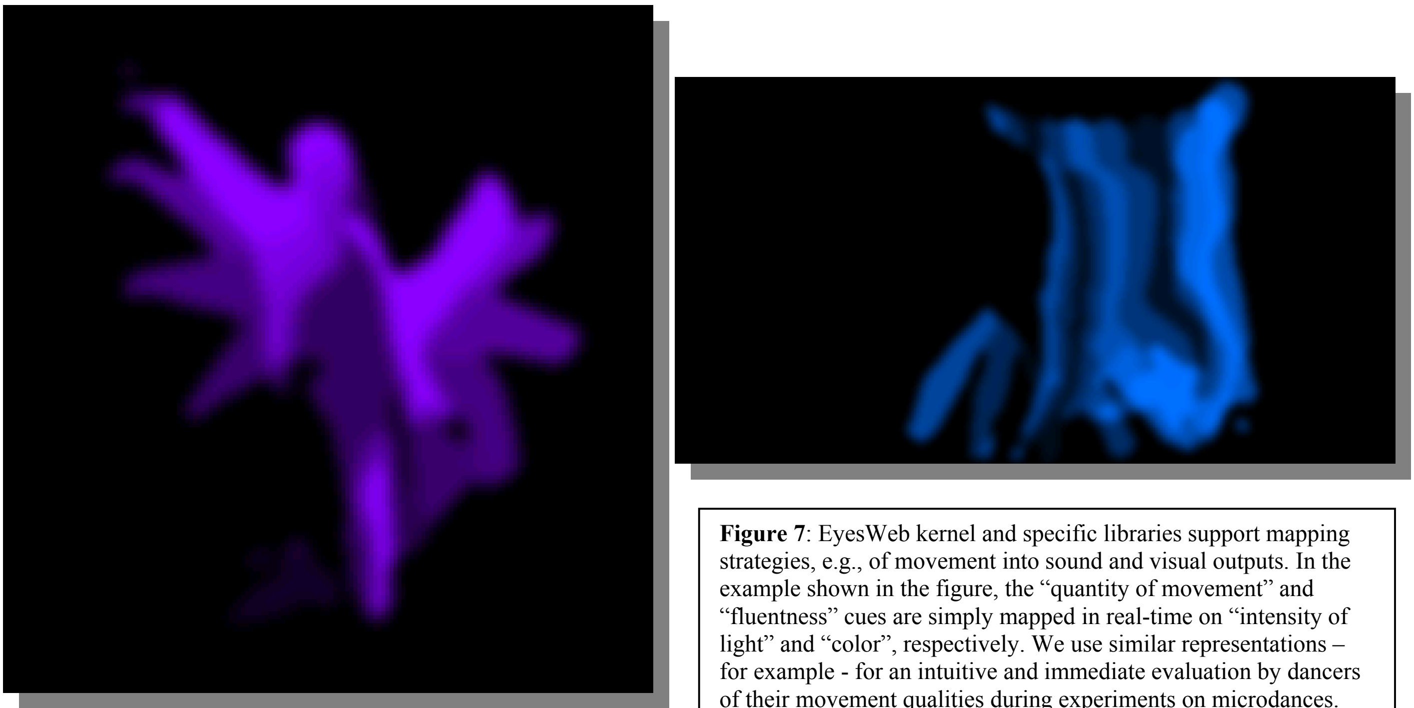


Figure 7: EyesWeb kernel and specific libraries support mapping strategies, e.g., of movement into sound and visual outputs. In the example shown in the figure, the "quantity of movement" and "fluentness" cues are simply mapped in real-time on "intensity of light" and "color", respectively. We use similar representations – for example - for an intuitive and immediate evaluation by dancers of their movement qualities during experiments on microdances.