Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Int. J. Human-Computer Studies 59 (2003) 213–225
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques

Antonio Camurri[a], Ingrid Lagerlöf[b], Gualtiero Volpe[a]

[a] *InfoMus Lab (Laboratorio di Informatica Musicale), DIST, University of Genova, Viale Causa 13, I-16145 Genova, Italy*

[b] *Department of Psychology, University of Uppsala, Sweden*

Received 13 November 2002; accepted 25 December 2002

Abstract

This paper illustrates our recent work on analysis and classification of expressive gesture in human full-body movement and in particular in dance performances. An experiment is presented which is the result of a joint work carried out at the DIST–InfoMus Lab, University of Genova, Italy, and at the Department of Psychology of the University of Uppsala, Sweden, in the framework of the EU-IST project MEGA (Multisensory Expressive Gesture Applications, www.megaproject.org). The experiment aims at (i) individuating which motion cues are mostly involved in conveying the dancer's expressive intentions to the audience during a dance performance, (ii) measuring and analyzing them in order to classify dance gestures in term of basic emotions, (iii) testing a collection of developed models and algorithms for analysis of such expressive content by comparing their performances with spectators' ratings of the same dance fragments. The paper discusses the experiment in detail with reference to related conceptual issues, developed techniques, and obtained results.

Keywords: Expressive gesture; Expressive multisensory interfaces; Analysis of human full-body movement; Performing arts

1. Introduction

Recent studies show that adults as well as children show great skill in their ability to decode emotions from full body movements (Boone and Cunningham, 1998, Dittrich et al., 1996; Lagerlöf and Djerf, 2002a,b, Van Meel et al., 1993). However, the question of what is guiding human perception of emotion has received very little

1071-5819/03/$ - see front matter
doi:10.1016/S1071-5819(03)00050-8

attention in empirical research. Body motion contains a high degree of flexibility that makes it a challenging task to uncover cues that are conveying emotional content. The purpose of this study is to identify cues that are important for emotion recognition and to show how these cues can be tracked by automated recognition techniques. A comparison of results obtained from spectators' recognition with those obtained from automated recognition techniques is presented. A complete description of this research can be found in Camurri et al. (in preparation).

In the search of decisive movement cues it is useful to distinguish between propositional and nonpropositional aspects (Buck cited in Boone and Cunningham, 1998). Propositional movements are established signs to transmit meaning such as a raised hand to indicate stop. These movements give distinct and scripted meaning. We maintain that also specific movements of emotions are propositional like a clenched fist to show anger or raised arms to demonstrate joy. In contrast nonpropositional movements are embodied in the direct and natural emotional expression of body movement based on fundamental elements such as tempo and force that could be combined in a vast range of movement possibilities. In this meaning nonpropositional movements do not rely on specific movements, but build on the quality of movements i.e., how movements are carried through, for instance whether it is with lightness or heaviness. In this study, we focus on the nonpropositional style of movements.

To find a nonpropositional style of movements we approach the genre of modern dance. A main characteristic of modern dance is to explore how emotional experiences could be expressed in body movements. The basis of natural body expression is further developed in terms of dance movements. This close linkage between modern dance and human emotions, has led to the proposal that modern dance contains cues from the underlying principle of natural emotional movement expression. Therefore, it has been suggested that emotions expressed in dance movements are a unique way to extract cues for emotions in natural bodily expressions (Boone and Cunningham, 1998; Stevens et al., 2002).

2. Background

In a previous study by Lagerlöf and Djerf (2002c), spectators showed high recognition of emotional content of the same dance that was varied in four emotions; anger, fear, grief and joy. Qualitative inspections of how movements were varied in the four different emotion expressions showed that these descriptions could be classified in the main dimensions of Laban movement analysis such as time, space, flow and weight. From this point of view, our work is similar to the one described in Zhao (2001), with the difference that we not only aim at classifying motion with respect to Laban's dimensions, but also with respect to the emotions it communicates (Camurri et al., 2001).

In this study, the Laban dimensions are operationalized into measurable elements, roughly described as follows: the time dimension is divided into two parts, overall duration of time and tempo changes. Further, tempo changes are also elaborated as

the underlying structure of rhythm or flow in the movement. The space dimension is considered in its aspects related to 'personal space', e.g., to what extent limbs are contracted or expanded in relation to the body centre. The flow component is considered in terms of analysis of shapes of speed and energy curves, and frequency/rhythm of motion and pause phases. The weight component is viewed in terms of amount of tension and dynamics in movement (vertical component of acceleration). These cues were predicted to be associated in different combinations for each emotion category as in the following configuration (Lagerlöf and Djerf, 2002c):

Anger	short duration of time
	frequent tempo changes, short stops between changes
	movements reaching out from body centre
	dynamic and high tension in the movement; tension builds up and then 'explodes'
Fear	frequent tempo changes, long stops between changes
	movements kept close to body centre
	sustained high tension in movements
Grief	long duration of time
	few tempo changes, 'smooth tempo'
	continuously low tension in the movements
Joy	frequent tempo changes, longer stops between changes
	movements reaching out from body centre
	dynamic tension in movements; changes between high and low tension

3. Method for experiments on spectators rating

Five dancers performed the same dance with four different emotional expressions: anger, fear, grief and joy. Each dancer performed all four emotions. The dancers' performances were video-recorded and then judged with regard to perceived emotion by 32 observers.

3.1. Participants

3.1.1. Dancers

An experienced dancer of modern dance was recruited to compose a short piece of dance, designed such that it excluded any propositional gesture or posture to avoid stereotyped emotions. The choreographer taught the 2 min neutral dance to experienced dancers of modern dance. The dancers then worked on the same dance and made separate variations. There were five dancers and each dancer performed the same dance four times to express anger, fear, grief and joy. This gave a total number of twenty dance performances. The dancers were instructed that they could not change or remove any sequence of movement. Neither did the dancers receive any instruction how to express the emotions.

3.1.2. Observers

There were 32 observers, divided in two different groups with 16 participants in each group, equally divided on women and men. All were adult students enrolled in different courses (psychology, social sciences, law and natural sciences) at Uppsala University. They were all naive observers, e.g., they had no previous dance training, or any experiences of previous dance performances.

3.2. Recording of dances

Each dance performance was recorded by two digital videocameras (DV recording format) standing fixed in the same frontal view of the dance (a spectator view). One camera obtained recordings to be used as stimuli for spectators' ratings. The second videocamera was placed in the same position but with specific recording conditions and hardware settings to simplify and optimize automated recognition of movement cues (e.g., manual shutter). Dancers' clothes were similar (dark), contrasting with the white background, in an empty performance space without any scenery. Digitized fading eliminated facial information and the dancers appeared as dark and distant figures against a white background.

3.2.1. Procedure

A total number of 20 video-recorded dance performances (five dancers each performing the same choreography four times, one for each basic emotion) were presented in a randomized order to two groups of spectators. In one group ratings were collected by 'forced choice' (chose one emotion category and rate its intensity) for each performance and the other group were instructed to use a multiple-choice schemata, i.e., to rate intensity of emotion on all four emotions scales for each performance.

4. Automated recognition of movement cues

A layered approach (Camurri et al., 2001) has been adopted to model human movement and gesture, from low-level physical measures (e.g., position, speed, acceleration of body parts) toward descriptors of overall motion features (e.g., motion fluency, directness, impulsiveness).

If, from the one hand, such high-level descriptors are grounded within the consolidated tradition of biomechanics, on the other hand, they are inspired to studies by psychologists (e.g., Wallbott, 1980) and researchers on human movement coming from the fields of performing arts and humanities, e.g., Rudolf Laban and his Theory of Effort (Laban, 1947, 1963).

Such a layered approach is suitable both for modelling (that is, generating) movement (for example of avatars, virtual characters, robots in Mixed Reality scenarios) and for recognizing movement qualities, as it is discussed in this paper.

Fig. 1 sketches the layered model as it was used in this experiment: for each layer inputs and outputs are displayed, as well as the kind of employed techniques. The

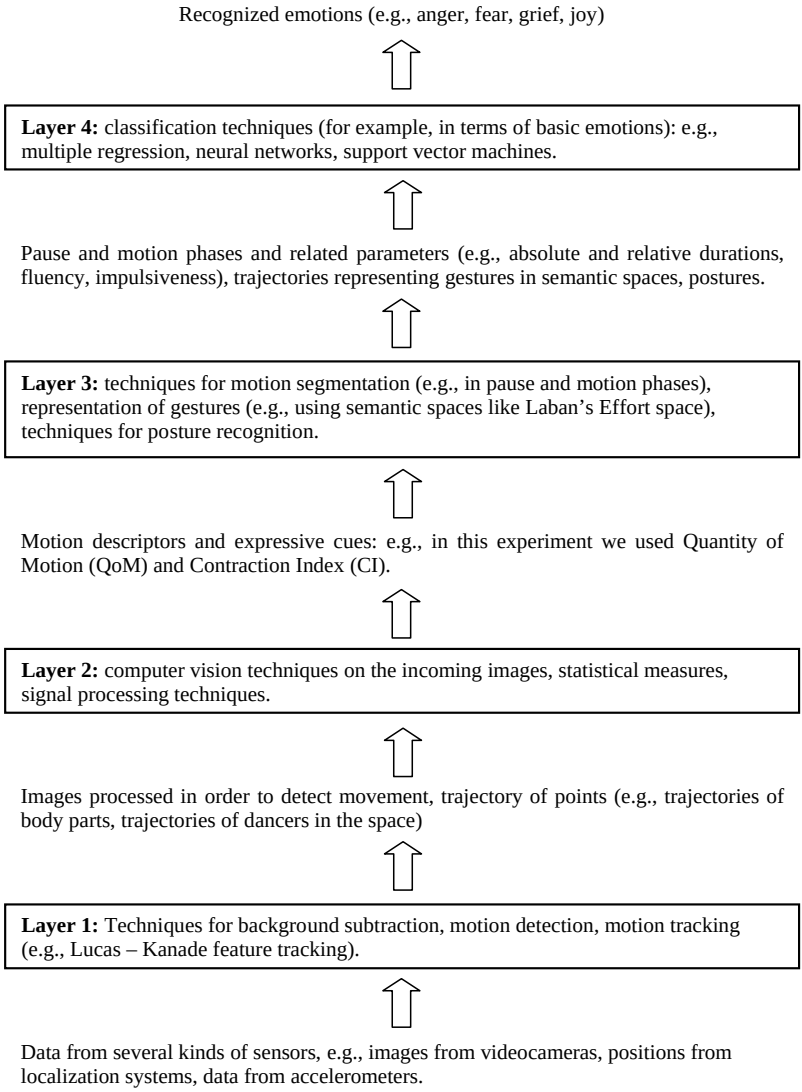

Fig. 1. The layered approach.

layered model is also motivated by an integrated multimodal representation of different channels of information (visual, acoustic, etc). This is however beyond the scope of this paper.

4.1. Layer 1

Layer 1 is responsible of the processing of the incoming video frames in order to detect and obtain information about the motion that is actually occurring. It receives

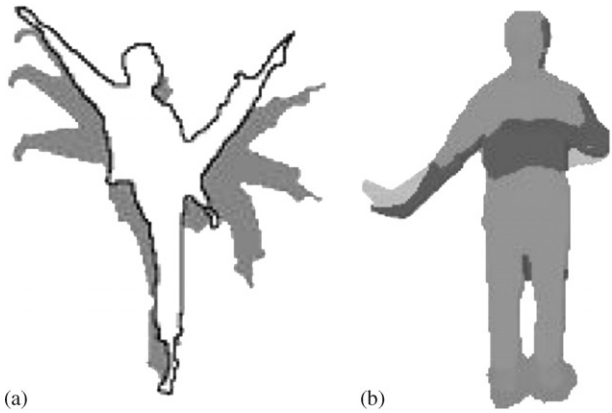

Fig. 2. (a) An example of SMI with time window of four frames. (b) Measure of internal motion in SMIs.

as input images from one or more videocameras and, possibly, information from other sensors (e.g., accelerometers). Two types of output are generated: processed images (e.g., see Fig. 2) and trajectories of body parts. Its (Layer 1) task is accomplished by means of consolidated computer vision techniques usually employed for real-time analysis and recognition of human motion and activity: see for example the temporal templates technique for representation and recognition of human movement described in Bobick and Davis (2001). It should be noticed that in contrast to Bobick and Davis research, we do not aim at detecting or recognizing a specific kind of motion or activity. The techniques we use include feature tracking based on the Lucas–Kanade algorithm (Lucas and Kanade, 1981), skin colour tracking to extract positions and trajectories of hands and head, and Silhouette Motion Images. A Silhouette Motion Image (SMI) is an image carrying information about variations of the silhouette shape and position in the last few frames. SMIs are inspired to motion-energy images (MEI) and motion-history images (MHI) (Bradsky and Davis, 2002, Bobick and Davis, 2001). They differ from MEIs in the fact that the silhouette in the last (more recent) frame is removed from the output image: in such a way only motion is considered while the current posture is skipped. Thus, SMIs can be considered as carrying information about the "amount of motion" occurred in the last N frames. Information about time is implicit in SMI and is not explicitly recorded. We also use an extension of SMIs, which takes into account the internal motion in silhouettes (see Fig. 2). In such a way, we are able to distinguish between global movements of the whole body in the General Space and internal movements of body limbs inside the Kinesphere.

The techniques described above have been developed in our EyesWeb open software platform. Free download of technical documentation and full software environment available from www.eyesweb.org. The *Expressive Gesture Processing Library* includes these and other processing modules.

Information motion detection and tracking provides to the upper levels is actually encoded in two different forms: positions and trajectories of points on the body (possibly related to specific body parts, e.g., hands, head, feet), and images directly resulting from the processing of the input frames (e.g., human silhouettes, SMIs).

4.2. Layer 2

The second layer is responsible of the extraction of a set of motion cues from the data coming from low-level motion tracking. Its inputs are the processed images and the trajectories of points on body coming from Layer 1. Its output is a collection of motion cues describing movement and its qualities. For the particular aim of this study two of such motion cues have been measured: Quantity of Motion (QoM) and Contraction Index (CI). Layer 2 employs computer vision, statistical, and signal processing techniques.

QoM is computed as the area (i.e., number of pixels) of a SMI (e.g., the number of pixels in the grey area in Fig. 2a). It can be considered as an overall measure of the amount of detected motion, involving velocity and force. QoM can be thought as a first rough approximation of the physical momentum, i.e., $q = mv$, where m is the mass of the moving body and v stands for its velocity. The shape of the QoM graph is close to the shape of the graphs of velocity of a marker put on a limb. QoM has two problems: (i) the measure depends on the distance from the camera; (ii) difficulties emerge when comparing measures from different dancers. We solved these problems by scaling the SMI area by the area of the most recent silhouette:

$$\text{Movement} = \text{Area}\,(\text{SMI}[t,n])/\text{Area}\,(\text{Silhouette}\,[t])$$

In this way, the measure becomes relative, i.e., independent from the camera's distance (in a range depending on the resolution of the videocamera), 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.

The CI is a measure, ranging from 0 to 1, of how the dancer's body uses the space surrounding it. It is related to Laban's "personal space".

The algorithm to compute the CI combines two different techniques: the individuation of an ellipse approximating the body silhouette and computations based on the bounding region. The former is based on an analogy between the image moments and mechanical moments: in this perspective, the three central moments of second order build the components of the inertial tensor of the rotation of the silhouette around its centre of gravity: this allows to compute the axes (corresponding to the main inertial axes of the silhouette) of an ellipse that can be considered as an approximation of the silhouette: eccentricity of such an ellipse is related to contraction/expansion; orientation of the axes is related to the orientation of the body (Kilian, 2001). The second technique used to compute CI is related to the bounding region, i.e., the minimum rectangle surrounding the dancer's body. The

algorithm compares 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, this component of the CI will be low, while, if the limbs are kept tightly nearby the body, it will be high (near to 1). While the dancer is moving, the CI varies continuously. Even if it is used with data from only one camera, its information is still reliable, being almost independent from the distance of the dancer from the camera. A use of this cue consists of sampling its values at the end and the beginning of a stretch movement, in order to classify that movement as a contraction or expansion.

4.3. Layer 3

A third layer of analysis consists in segmenting motion in order to individuate motion and non-motion (pause) phases. The temporal duration of such phases is then measured and compared with the total duration of the dance performance. The QoM measure has been used to perform the segmentation between pause and motion phases. QoM is related to the overall amount of motion and its evolution in time can be seen as a sequence of bell-shaped curves (*motion bells*). In order to segment motion, a list of these motion bells has been extracted and their features (e.g., peak value and duration) computed. An empirical threshold has been defined for these experiments: the dancer is considered to be moving if the area of the motion image (i.e., the QoM) is greater than 2.5% of the total area of the silhouette. Fig. 3 shows motion bells after automated segmentation: a motion bell characterizes each motion phase.

Motion segmentation can be considered as a first step toward the analysis of the rhythmic aspects of the dance. Analysis of the sequence of pause and motion phases and their relative time durations can lead to a first evaluation of dance tempo and its evolution in time, i.e., tempo changes, articulation (the analogous to music legato/staccato). Parameters from pause phases are also extracted to individuate real still standing positions from active pauses involving low-motion (hesitating or oscillation movements).

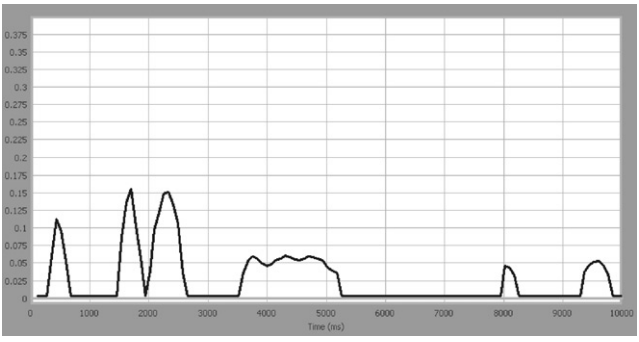

Fig. 3. Motion segmentation.

Furthermore, motion fluency and impulsiveness are evaluated. They are related to Laban's Flow and Time axes. Fluency can be estimated starting from an analysis of the temporal sequence of motion bells. A dance fragment performed with frequent stops and restarts (i.e., characterized by a high number of short pause and motion phases) will result less fluent than the same movement performed in a continuous, "harmonic" way (i.e., with a few long motion phases). The hesitating, bounded performance will be characterized by a higher percentage of acceleration and deceleration in the time unit (due to the frequent stops and restarts), a parameter that has been demonstrated of relevant importance in motion flow evaluation (see, for example, Zhao, 2001, where a neural network is used to evaluate Laban's flow dimension).

A first measure of impulsiveness can be obtained from the shape of a motion bell. In fact, since QoM is directly related to the amount of detected movement, a short motion bell having a high peak value will be the result of an impulsive movement (i.e., a movement in which speed rapidly moves from a value near or equal to zero, to a peak and back to zero). On the other hand, a sustained, continuous movement will show a motion bell characterized by a relatively long time period in which the QoM values have little fluctuations around the average value (i.e., the speed is more or less constant during the movement).

4.4. Layer 4

In the experiment described in this paper, Layer 4 collects inputs from Layers 2 (QoM, CI) and 3 (duration of pause and motion phases, fluency) and tries to classify dance fragments in term of the four basic emotions anger, fear, grief and joy. A number of techniques is available for this task: statistical methods like multiple regression, neural networks (e.g., classical backpropagation networks, Kohonen networks), support vector machines, methods based on fuzzy sets.

As a first step, statistical techniques have been used: a one-way ANOVA has been computed for each motion cue and a regression analysis has been carried out. A first set of results is reported in the following section. In a further paper (Camurri et al., in preparation) a detailed comparison between spectators' rating and the output of the classification process will be discussed. We are also working at a comparison between different classification techniques in this type of movement analysis task.

5. Results and conclusions

Overall the two groups of spectators (forced vs. multiple alternative) show similar results for emotion recognition. The accuracy rate for the spectators with the forced choice alternative was above chance level for all but one performance of grief. With the multiple-choice alternative spectators' ratings show that all the intended emotion of grief received significant higher mean ratings than the ratings of the other emotions. Ratings of the intended emotions of anger and joy received significant

higher means in six performances, although it was a mix-up between anger and joy in four cases (of ten). The recognition of fear was comparatively lower than for the other emotions, with both forced and multiple-choice alternatives. Although, the fear recognition was above chance level in all but one case, with the forced choice alternative.

These results show that spectators were able to detect the intended emotion that was transmitted by the performances of the same dance movements. The highest recognition rate was received by the grief performances followed by anger and then joy. Performances of fear received a comparatively lower recognition rate, although exceeding chance level in all performances but one. Table 1 summarizes spectators' recognition rate for each dancer and for each intended emotion.

The results obtained from the application of algorithms reveal different mean values for each of the emotion categories. A one-way ANOVA ($n = 5$), with repeated measures of emotion was computed for each movement cue (Overall Duration of Time, CI, QoM and Motion Fluency) followed by Tukey's HSD test. The result revealed a significant main effect for Duration of Time showing that duration for grief performances is significant longer than for the other emotions. A significant main effect was also found for CI, revealing that fear and grief receive significant higher mean value for CI compared to joy. Another significant main effect was revealed for QoM meaning that the performances of anger and joy received significant higher mean score than grief.

As an example of some results obtained in this research, Figs. 4 and 5 show the mean values computed for each motion phase of QoM and CI, respectively. In each figure the four graphs refer to four performances by the same dancer in which the dancer tried to express the four basic emotions. In the figures line types are associated to emotions as follows: anger—solid line; fear—dashed line; joy—dash–dot line; grief—dotted line.

It can be noticed, for example, that curves representing the average QoM for anger (solid line) and fear (dashed line) have a similar trend: i.e., they starts with low values and slow increase at the beginning, then they continuously increase with increasing steepness. Fear, however, have much more motion phases than anger indicating a less fluent motion.

Table 1
Spectators' recognition rate for each dancer and each intended emotion

Dancer	Spectators agreement (%)			
	Anger	Joy	Fear	Grief
1	40	33	40	67
2	93	40	56	81
3	53	75	47	75
4	73	67	31	76
5	44	60	25	53

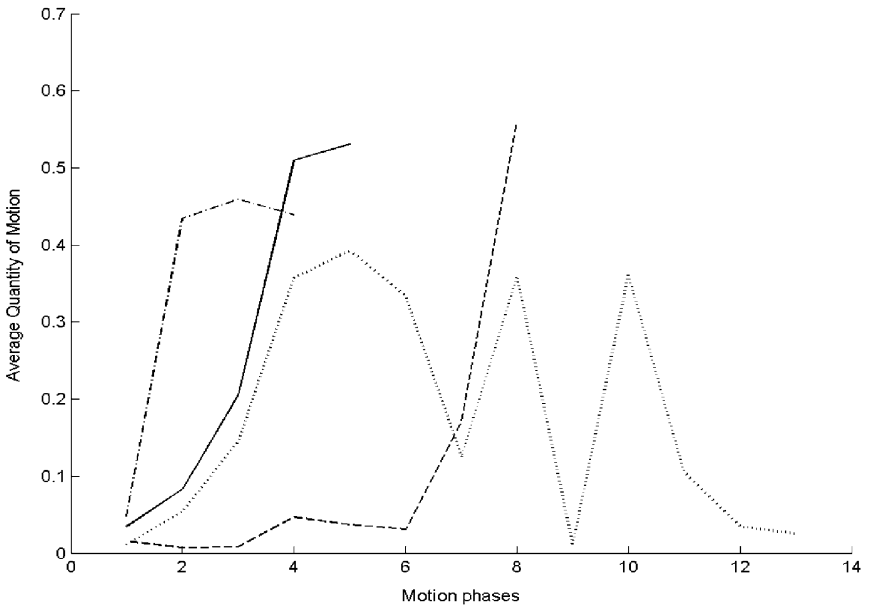

Fig. 4. Mean values of the QoM computed for each motion phase (the four graphs refer to four performances by the same dancer, each one expressing a different basic emotion: anger—solid line; fear—dashed line; joy—dash–dot line; grief—dotted line). The *X*-axis is the index of the motion phase in which the movement has been segmented (therefore, X is not the time axis).

CI for joy (dash–dot line) has quite low values with respect to the other emotions, while fear (dashed line) has quite high values, meaning that the body is often contracted (i.e., limbs are often close to the centre of gravity).

Grief (dotted line) always has a high number of motion phases and a high variance of the average values of QoM, meaning frequent transitions between motion and pause phases and very low fluency. Joy (dash–dot line), instead, has only four long motion phases indicating a very fluent motion.

We also noticed that, while from the one hand each of the four dancers has a particular trend allowing us to distinguish between them, on the other hand what we observed above holds for all the four dancers, i.e., they expressed the four emotions by acting on the expressive cues in the same way.

More in general, the analysis performed allowed us to formulate the hypothesis that the emotion categories can be differentiated by the automatically obtained movement cues and that they are in line with the main predicted associations between emotion and movement cues described in Background section. Further, the extracted cues are psychologically validated by spectators' recognition of the different emotions in movement.

In a further paper (Camurri et al., in preparation) we describe all the obtained results on automatic classification with detailed tables and comparisons of human

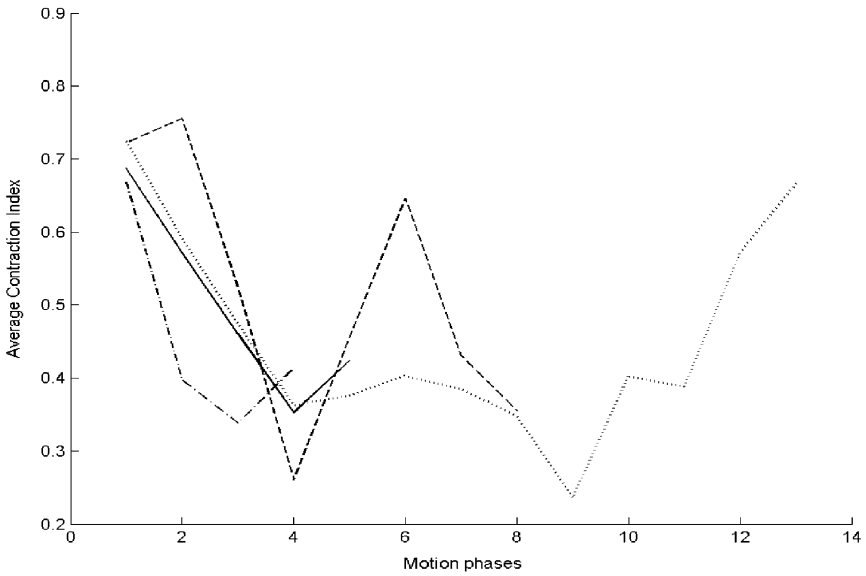

Fig. 5. Mean values of the CI computed for each motion phase (the four graphs refer to four performances by the same dancer, each one expressing a different basic emotion: anger—solid line; fear—dashed line; joy—dash–dot line; grief—dotted line). The X-axis is the index of the motion phase in which the movement has been segmented (therefore, X is not the time axis).

ratings with automated recognition of emotions, as well as details related to the automated recognition techniques.

Acknowledgements

The automated recognition techniques described in this paper are included in the Motion Analysis library of the EyesWeb open software platform (www.eyesweb.org).

We thank Marie Djerf, Barbara Mazzarino, Matteo Ricchetti for discussions and their concrete contributes to this research project. We thank also the other members of the EyesWeb staff. The research described in this paper has been implemented as part of the Expressive Gesture Processing Library for the EyesWeb open software platform (www.eyesweb.org).

This work has been partially supported by the EU—IST Project MEGA (Multisensory Expressive Gesture Applications).

References

Boone, R.T., Cunningham, J.G., 1998. Children's decoding of emotion in expressive body movement: the development of cue attunement. Developmental Psychology 34, 1007–1016.

Bobick, A.F., Davis, J., 2001. The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (3), 257–267.

Bradsky, G., Davis, J., 2002. Motion segmentation and pose recognition with motion history gradients. Machine Vision and Applications 13, 174–184.

Camurri, A., De Poli, G., Leman, M., 2001. MEGASE—A multisensory expressive gesture applications system environment for artistic performances. Proceedings of the International Conference CAST01, GMD, St Augustin-Bonn, pp. 59–62.

Dittrich, W.H., Troscianko, T., Lea, S.E.G., Dawn, M., 1996. Perception of emotion from dynamic point-light displays represented in dance. Perception 25, 727–738.

Kilian, J., 2001. Simple Image Analysis by Moments. OpenCV library documentation.

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

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

Lagerlöf, I., Djerf, M., 2002a. Communicating emotions in dance performance. Department of psychology, Uppsala (manuscript under revision).

Lagerlöf, I., Djerf, M., 2002b. Children's understanding of emotion in dance. Department of psychology, Uppsala (manuscript under revision).

Lagerlöf, I., Djerf, M., 2002c. On cue utilization for emotion expression in dance movements. Department of psychology, Uppsala (in preparation).

Lucas B. , Kanade, T., 1981. An iterative image registration technique with an application to stereo vision. In: Proceedings of the International Joint Conference on Artificial Intelligence, 1981.

Stevens, C., Malloch, S., Hazsard-Morris, R., McKechnie, S., 2002. Shaped time: a dynamical systems analysis of contemporary dance. Paper presented at ICMP7.

Van Meel, J., Verburgh, H., De Meijer, M., 1993. Children's interpretation of dance expressions. Empirical Studies of the arts 11 (2), 117–133.

Wallbott, H.G., 1980. The measurement of human expressions, In: Walbunga von Rallfer-Engel, Aspects of communications, pp. 203–228.

Zhao, L., 2001. Synthesis and acquisition of Laban movement analysis qualitative parameters for communicative gestures. Ph.D. Dissertation, University of Pennsylvania.