Toward a Minimal Representation of Affective Gestures

Donald Glowinski, *Member*, *IEEE*, Nele Dael, Antonio Camurri, Gualtiero Volpe, Marcello Mortillaro, and Klaus Scherer

Abstract—This paper presents a framework for analysis of affective behavior starting with a reduced amount of visual information related to human upper-body movements. The main goal is to individuate a minimal representation of emotional displays based on nonverbal gesture features. The GEMEP (Geneva multimodal emotion portrayals) corpus was used to validate this framework. Twelve emotions expressed by 10 actors form the selected data set of emotion portrayals. Visual tracking of trajectories of head and hands were performed from a frontal and a lateral view. Postural/shape and dynamic expressive gesture features were identified and analyzed. A feature reduction procedure was carried out, resulting in a 4D model of emotion expression that effectively classified/grouped emotions according to their valence (positive, negative) and arousal (high, low). These results show that emotionally relevant information can be detected/measured/obtained from the dynamic qualities of gesture. The framework was implemented as software modules (plug-ins) extending the EyesWeb XMI Expressive Gesture Processing Library and is going to be used in user centric, networked media applications, including future mobiles, characterized by low computational resources, and limited sensor systems.

Index Terms—Emotion, expressive gesture, automatic features extraction.

1 Introduction

The analysis of human behavior, and in particular of the human ability to communicate through body movement and gesture, has been widely studied in psychology and is of growing importance for information and communication technologies (e.g., social media, online and mobile communities, and Web 2.0 technologies). Multimodal interfaces able to capture nonverbal expressive and affective features are needed to support novel multimedia applications, Future Internet, and user-centric media. For example, in networked user-centric media the expressive, emotional dimension is becoming particularly significant and the need is emerging for extraction, analysis, encoding, and communication of a small, but yet highly informative, amount of information characterizing the expressive-emotional behavior of people using or wearing mobile devices (e.g., mobile phones). This information is used for many different purposes, for example, in affective mobile applications (e.g., enabling users to control the expressive performance of a music piece by the expressive movement of their mobile phone), in the Future Internet for enhancing complex 3D reconstructions of the remote participants in an Internet-mediated distributed and embodied social interaction, by adding the expressive-emotional channel [48]. Anthropomorphic interfaces (e.g., a robot or an avatar) would also greatly improve their interactivity by recognizing and interpreting users' subtle emotional processes and by communicating expressive-emotional information [57], [41]. Computer vision and movement analysis techniques to extract and classify information related to the emotions of individuals or groups are another important research direction.

Following the work of Ekman and Friesen [32], [33], many studies focused on emotion recognition from facial expression. Recent neuroscientific and psychological studies have revealed that body movement is an additional important modality of emotion communication [51], [25], [67]. Nevertheless, very few contributions are available on the analysis of the dynamics of body movement to extract expressive and affective information [15], [40].

Upper-body movements are of particular interest. Head and hands movements are actually most often employed to express one's affect and human observer's attention spontaneously focuses around this body region when tempting to infer others emotions [36], [49]. This predominance of upper-body movements in emotional communication is reinforced by current computer interfaces and practice.

This paper presents a framework for analysis of affective behavior starting with a reduced amount of visual information related to human upper-body movements. The motivation is to embed it in user-centric, networked media applications, including future mobiles, characterized by low computational resource and limited sensor systems. We start from the hypothesis that even only with a reduced amount and quality of visual information it is still possible to classify a significant amount of the affective behavior expressed by the human. A small set of visual features are extracted from two consumer video cameras (25 fps), based on head and hands position and velocity. Neither facial

- *D. Glowinski, A. Camurri, and G. Volpe are with the Department of Communication Computer and System Sciences, InfoMus Lab/Casa Paganini (DIST), University of Genoa, 16145 Genova, Italy. E-mail: {Donald.Glowinski, Antonio.Camurri, Gualtiero.Volpe}@unige.it.*
- *N. Dael, M. Mortillaro, and K. Scherer are with the Swiss Center for Affective Sciences, 7, Rue des Battoirs, CH-1205, Geneva, Switzerland. E-mail: {Nele.Dael, Klaus.Scherer}@unige.ch.*

Manuscript received 30 Sept. 2010; revised 19 Jan. 2011; accepted 22 Feb. 2011; published online 28 Jan. 2011.
Recommended for acceptance by A.A. Salah, T. Gevers, and A. Vinciarelli.
For information on obtaining reprints of this article, please send e-mail to: taffc@computer.org, and reference IEEECS Log Number TAFFCSI-2010-09-0075.
Digital Object Identifier no. 10.1109/T-AFFC.2011.7.

expressions are considered nor fine-grain hand gestures (e.g., finger movements). Starting from behavioral studies, postural and dynamic expressive motor features, considered relevant for the communication of emotional content, are extracted from the low-level kinematic features of head and hands.

The GEMEP corpus of emotion portrayals [4], [5], [1] has been chosen for validating the framework. Our analysis focuses on 120 portrayals that have been systematically rated and selected among the original 7,000 audio-video emotion portrayals. These portrayals refer to 10 actors performing 12 emotions: elation, amusement, pride, hot anger (rage), fear, despair, pleasure, relief, interest, cold anger (irritation), anxiety, and sadness. Research results also include algorithms for real-time analysis of expressive gesture features, implemented as software modules (plug-ins) extending the EyesWeb XMI Expressive Gesture Processing Library [17], [15].

This paper is organized as follows: Section 2 reviews the existing approaches in the literature that attempt to recognize emotions through body postures and movements. Section 3 presents the developed experimental framework. The steps dedicated to motion tracking, expressive features extraction, dimension reduction, and clustering processes are detailed. Section 4 describes the obtained affect classification and compares the results with the emotion categories of the original GEMEP corpus. Section 5 outlines an implementation of the framework obtained by extending the EyesWeb XMI expressive gesture library and concludes by addressing future research directions.

2 Background

The body is an important source of information for affect recognition. A large trend of research concerns facial expression, gesture, and activity recognition (see, for example, the IEEE Face and Gesture Conferences). Many techniques have been suggested to analyze hand and head gestures over the years (Hidden Markov Models (HMMs), CRFs, neural networks, and DBN). However, our approach is different since we focus on the nonverbal expressive content rather than on the specific sign or gesture being performed. This section reviews the state of the art, considering the analysis of full-body features. We then detail recent interests for the analysis of upper-body movements. The section ends with a presentation of a conceptual framework to investigate higher-level, qualitative aspects of gesture performance (e.g., fluid versus impulsive) to describe behavior expressivity.

2.1 From Face to Body

Compared to the facial expression literature, attempts for recognizing affective body movements are few. Pioneering work by Ekman suggested that people make greater use of the face than the body for judgments of emotion in others [30], [31]. However, results from psychology suggest that body movements do constitute a significant source of affective information [2], [27], [74]. Yet, body gesture presents more degrees of freedom compared to facial expression. An unlimited number of body postures and gestures with combinations of movements of various body parts and postural attitudes are possible. No standard coding scheme equivalent to the FACS for facial expressions exists to decompose any possible bodily movement expression into elementary components. Various systems were proposed by psychologists [9], [13], [10], [37] or inspired by dance notation and theories [47], but none of them reach the consensus achieved by the Ekman's system for analysis of facial expression.

2.2 Full-Body Coarse Postures

Existing studies on full-body movement use coarse-grained posture features (e.g., leaning forward, slumping back) or low-level physical features of movements (kinematics). Bianchi-Berthouze and Kleinsmith [8] formalized a general description of posture based on angles and distances between body joints and used it to create an affective posture recognition system that maps the set of postural features into affective categories using an associative neural network. Mota and Picard [53] showed how sequences of postures can be predictive of affective states related to a child's interest level during a learning task on a computer. Naturally occurring postures data from 10 children were collected through pressure sensors mounted on a chair. HMMs were used to analyze temporal patterns among nine posture sequences to characterize three affective states (high and low interest and behavior of taking a break). A similar study was conducted by Kapoor et al. to detect prefrustration behavior [44] using multiple nonverbal channels of information.

2.3 Full-Body Kinematics

Other approaches have exploited the dynamics of gestures, referring to a few psychological studies reporting that temporal dynamics play an important role for interpreting emotional displays [25], [2], [59]. Kapur et al. [45] used full-body skeletal movement data (obtained with the VICON motion capture system on five participants) to distinguish automatically between four basic emotional states (sad, joy, anger, and fear). 3D positions of 14 body joints were recorded over time to identify the movements performed by actors for each of the selected emotion. The authors showed that very simple statistical measures of motion dynamics (e.g., velocity and acceleration) are sufficient for training successfully automatic classifiers (e.g., SVMs and decision trees classifier). The role of kinematic features has been further established by the recent study of Bernhardt and Robinson [6]. Further developing the motion-captured knocking motion from Pollick et al. [59], they developed a computational approach to extract affect-related dynamic features. Velocity, acceleration, and jerk measured for each joint composing the skeletal structure of the arm proved successful in the automatic recognition of the expressed affects (neutral, anger, happy, and sad).

2.4 Upper-Body Movements

Another trend of studies focused on upper-body movements. Two-handed gestures accompanying speech in particular have been mainly investigated and revealed the role of hand illustrators, i.e., hand movements accompanying, illustrating, and accentuating the verbal content of utterances [49]. As pointed out by Freedman, besides gait, hand illustrators are the most frequent, regularly occurring quantifiable bits of overt behavior available for objective study on expressivity [36]. The sign language production revealed that small detailed motions are performed in and around the face and upper body region, where the receiver (looking at the signer's face) can naturally observe gesture in high acuity [64]. Balomenos et al. combined facial expressions and hand gestures for the recognition of six prototypical emotions [3]. On the basis of hands position and trajectory patterns, they individuated four classes of gesture using HMMs: hands clapping, hands over the head, "Italianate" gestures (follows a repetitive sinusoidal pattern), and lift of the hand, which combinations proved successful for the classification of the six emotions. Gunes and Piccardi [40] proposed a bimodal system that recognizes twelve affective states. Starting from the frontal view of sitting participants, their motion was estimated by using optical flow, hands, head, and shoulders regions were tracked and analyzed (e.g., how the centroid, rotation, length, width, and area of the region changed over time), and classification performed based on SVMs.

2.5 Expressive Gesture Analysis

Camurri et al. developed a qualitative approach to human full-body movement for affect recognition [15]. Starting from low-level physical measures of the video-tracked whole-body silhouette, they identified motion features, such as the overall amount of motion computed with silhouette motion images, the degree of contraction and expansion of the body computed using its bounding region, or the motion fluency computed on the basis of the changes magnitude of the overall amount of motion over time. On the basis of these motion features, they were able to distinguish between four emotional performances of a dance sequence (anger, fear, grief, and joy) by four dancers. This framework of vision-based bodily expression analysis was used for a number of multimodal interactive applications, in particular in performing arts and mainly include SVM techniques for real-time affective classification [17], [20].

Most of these listed works attempt to recognize a small set of prototypic expressions of basic emotions like happiness and anger. The few exceptions include a tentative effort to detect more complex states such as frustration [44] or puzzlement [39]. Our work, on the one hand, reinforces the results obtained in previous research by showing that a limited number of gesture features, chosen among those already identified in psychological studies, seem sufficient for accurate automatic recognition of emotion. On the other hand, it 1) extends the set of emotions that are taken into account (a set including 12 emotions is considered, whereas most studies in the literature usually refer to the six emotions commonly referred as basic emotions), 2) introduces features at a higher-level with respect to the kinematical features used in most studies, and 3) considers a larger data set of emotion portrayals (120 portrayals) from an internationally validated corpus of emotion portrayals, the GEMEP corpus.

3 Experimental Framework

Our framework aims at individuating a minimal and efficient representation of emotional displays based on nonverbal gesture features, analyzing affective behavior starting from low-resolution visual information related to human upper-body movement. It grounds on the methodology adopted and results obtained in our previous research on expressive gesture.

Our approach starts from the multilayered framework developed by Camurri et al. to analyze and model expressive gesture [17]. Basing on the Kurtenbach and Hulteen's definition of gesture as "a movement of the body that contains information," a gesture can be said to be expressive since the information it carries is an expressive content, i.e., an "implicit message" [28] or KANSEI, in the Japanese culture a word indicating a complex process of coding and decoding of information concerning the affective, emotional sphere [41]. That is, expressive gesture is responsible for the communication of a kind of information (addressed as expressive content) that is different and independent, even if often superimposed, to a possible denotative meaning, and that concerns aspects related to affects. In the present analysis, we considered gestures accompanying verbal utterances. However, the concept of expressive gesture implies that behavioral features that support emotion expression can apply to a large range of gestures and are not restricted to specific types of gestures.

According to this framework, analysis is accomplished by three subsequent layers of processing, ranging from low-level physical measures (e.g., position, speed, acceleration of body parts) toward overall gesture features (e.g., motion fluency, impulsiveness) and high-level information describing semantic properties of gestures (affect, emotion, and attitudes).

The modules composing the framework refer to the conceptual layers presented above (see Fig. 1):

- Module 1 computes low-level motion features, i.e., the 3D positions and kinematics of the video-tracked hands and head.
- Module 2 computes a vector of higher-level expressive gesture features, including five sets of features aiming at explaining the qualitative aspects of movement. Three sets of features are inspired to Wallbott [74], [73]:
    - energy (passive versus animated),
    - spatial extent (expanded versus contracted),
    - smoothness and continuity of movement (gradual versus jerky). Two further components are considered here:
    - forward-backward leaning of the head and
    - spatial symmetry and asymmetry of the hands with respect to the horizontal and vertical axis [46];
- module 3 reduces the dimensionality of the data, while highlighting salient patterns in the data set.

3.1 Expressive Features Extraction Module

Considering literature mentioned above, our analysis is based on the position and the dynamics of the hands and the head. Extraction of expressive features from human movement is carried out using the EyesWeb XMI Expressive Gesture Processing Library, extended by the modules described in this and following sections.

A skin color tracking algorithm is used to extract the blobs of the head and of the two hands (see Fig. 2). A 3D

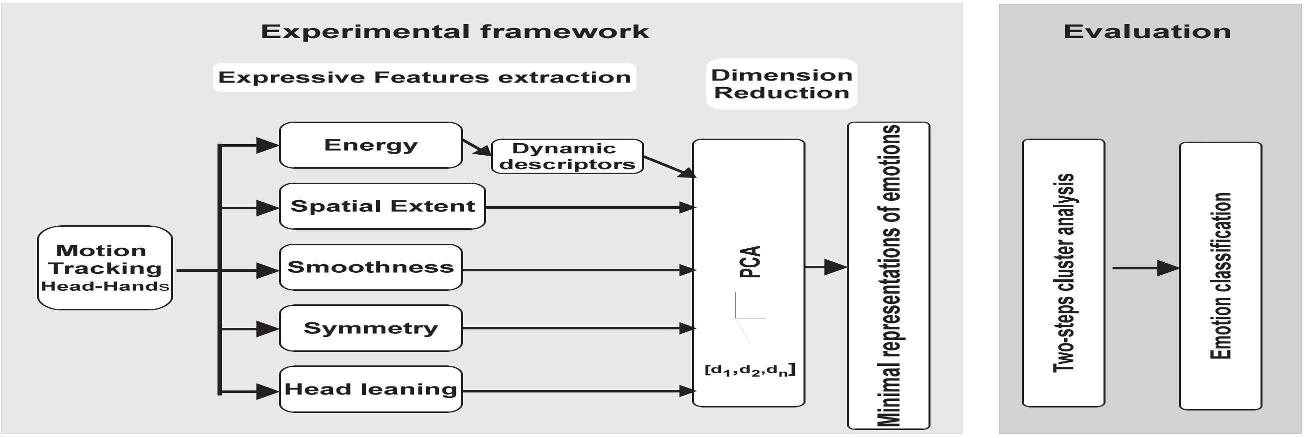


Fig. 1. The experimental framework and the evaluation component.

position of the blobs' center of mass (centroid) is obtained from the two synchronized videocameras (frontal and lateral). Velocity and acceleration (using the Savitzky-Golay smoothing filter) of the head and the hand's movements are also computed from the coordinates (x, y, z) of the blob's centroids. Most of the processing is performed on the x and y coordinates, whereas the z coordinate is used to get information about the depth. Thus, the system performs a 2D and half analysis, rather than a real 3D analysis. The movies from the two videocameras are synchronized manually in the same way audio and video are synchronized by means of a clapperboard.

3.1.1 Expressive Features

Starting from the 3D position, velocity, and acceleration of the head and the hands, as described in the previous section, five categories of expressive features are automatically extracted. They are described in the following sections:

Energy. The first expressive feature concerns the overall energy spent by the user for his performance, approximated by the total amount of displacement in all of the three extremities. A study by Wallbott revealed that the judged amount of movement activity is an important factor in differentiating emotions [74]. Following his seminal study [75], Wallbott considered that movements and postural activity as a whole may enable the distinction between fourteen emotions [74]. Specifically, the largest differences were obtained for the judgments of what he called *dynamics/energy/power* of movements: The highest values related to hot anger, elated joy and terror emotions, the lowest values corresponded to sadness and boredom [74]. Camurri et al. [15] showed that movement activity (in that case, computed from the analysis of silhouette variations and coined *quantity of motion*) is a relevant feature in recognizing emotion from the full-body movement of dancers. Results showed that the Quantity of Motion (QoM) in the anger and joy performances were significantly higher than in grief one. For these reasons, we include in our set of expressive features an approximated measure of the overall motion energy (activation) at time frame f.

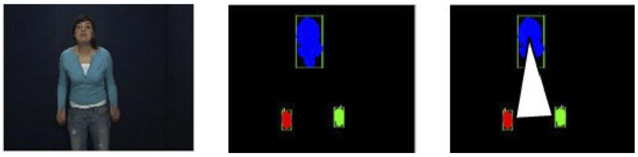


Fig. 2. From left to right: The original image, the three blobs (extracted by color skin tracking algorithms), and the bounding triangle (used to extract contraction/expansion and symmetry features). Two cameras (frontal and lateral view) are used to record emotion displays: A single time-of-flight camera can provide the same information.

Let $v_l(f)$ denote the module of velocity of each limb l (head and two hands in our case) at time frame f (1):

$$v_l(f) = \sqrt{\dot{x}_l(f)^2 + \dot{y}_l(f)^2 + \dot{z}_l(f)^2}. \tag{1}$$

We define E_{tot} as an approximation of the body kinematic energy, as the weighted sum of the limbs' kinetic energy (2):

$$E_{tot}(f) = \frac{1}{2}\sum_{l=1}^{n} m_l v_l(f),^2 \tag{2}$$

where m_i are the approximations of the mass of head and hands, and their value is computed starting from the biometrics anthropometric tables [26].

Spatial extent: the bounding triangle. We further consider a *bounding triangle* related to the three blobs' centroids of hands and head (see Fig. 3). The dynamics of the perimeter of such a bounding triangle approximates the space occupied by the head and the hands from the frontal view. The use of space in terms of judged expansiveness or spatial extension of movements have been regarded by Wallbott as another relevant indicator for distinguishing between active and passive emotions [74]. Mehrabian pointed out how the degree of openness in the arm arrangement (ranging from close-arm position to moderate or extreme open-arm position) characterizes the accessibility and the liking of a communicator's posture from the receiver's point of view [50]. De Meijer underlined that an open arm arrangement is generally associated with emotional warmth and empathy, thus related to emotions that are positively connoted [51].

Smoothness/jerkiness. In general, "smoothness" is synonymous to "having small values of high-order derivatives." Since Flash and Hogan, *movement jerk*—the third derivative of movement position—is frequently used as a descriptor of the smoothness of a movement [35], [43], [77].

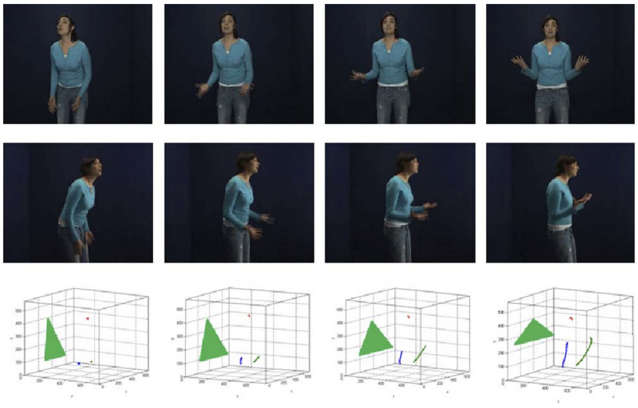


Fig. 3. From left to right, four key-frames of a portrayal representing elation; from top to bottom, the frontal view, the profile view, the 3D representation of the displacement of the head, and the two hands in the space with the *bounding triangle* relating the three blobs.

A minimum-jerk model was successfully applied by Pollick et al., together with other motion features to specify the arousal level communicated by a moving arm expressing basic emotion categories [59].

Wallbott, in his analysis of qualitative aspects of psychiatric patients' hand movements, noticed that movements judged as *smooth "are characterized distally by large circumference, long wavelength, high mean velocity, but not abrupt changes in velocity or acceleration (standard deviations of velocity and acceleration). Thus, smooth movements seem to be large in terms of space and exhibit a high but even velocity"* ([73], pp. 140). This relation between kinematical (i.e., velocity) and geometric (i.e., curvature) properties of movement is further interpreted as an epiphenomenon of smooth movement generation by studies on human motor control, e.g., the Two-Third Power Law [72] or smoothness maximization [66]. In this last case, however, the relationship between the path and the speed of the observed trajectory is defined primarily for specific situations (e.g., straight reaching movements or extemporaneous drawing movement). In the case of spontaneous or freely portraying emotions, predetermined trajectories of hands cannot be imposed. We therefore have adapted Wallbott's statements on the qualitative dimensions of underconstrained arm movements and we have computed hands trajectories curvature to identify trajectories' smoothness. Curvature (k) measures the rate at which a tangent vector turns as a trajectory bends (e.g., a hand trajectory following the contour of a small circle will bend sharply, and hence will have higher curvature; by contrast, a hand trajectory following a straight line will have zero curvature). Curvature is computed for each hand trajectory by (3):

$$k = \frac{\dot{x} \cdot \ddot{y} - \dot{y} \cdot \ddot{x}}{(\dot{x}^2 + \dot{y}^2)^{3/2}}, \tag{3}$$

where $\dot{x}, \ddot{x}$ and $\dot{y}, \ddot{y}$ are, respectively, the first and second order derivatives of the trajectory of one hand in its horizontal and vertical components.

Symmetry. Lateral asymmetry of emotion expression has long been studied in face expressions resulting in valuable insights about a general hemisphere dominance in the control of emotional expression. An established example is the expressive advantage of the left hemiface that has been demonstrated with *chimeric face stimuli*, static pictures of emotional expressions with one side of the face replaced by the mirror image of the other [11]. A recent study by Roether et al. on human gait demonstrated pronounced lateral asymmetries also in human emotional full-body movement [60]. Twenty-four actors (with an equal number of right and left-handed subjects) were recorded by using a motion capture system during neutral walking and emotionally expressive walking (anger, happiness, sadness). For all three emotions, the left body side moves with significantly higher amplitude and energy. Perceptual validation of the results were conducted through the creation of *chimeric walkers* using the joint-angle trajectories of one body half to animate completely symmetric puppets.

Apart from face and full-body, some studies investigated the symmetry of related upper-body movements, but a few accounted for its relation with expressivity. Merhabian showed in particular that arm-position asymmetry was a relevant behavioral feature to identify "relax" attitude and relative high social status of a person within a group [50].

Considering that literature pointed out the relevance of symmetry as behavioral and affective features, we address the symmetry of two handed-gestures and its relation with emotional expression. Hand symmetry is measured in two ways. We first compute spatial hands symmetry with respect to the vertical axis and with respect to the horizontal axis. Horizontal asymmetry ($SI_{horizontal}$) is computed from the position of the barycenter and the left and right edges of the bounding triangle that relate the head and the two hands (4):

$$SI_{horizontal} = \frac{\| x_B - x_L| - |x_B - x_R \|}{|x_R - x_L|}, \tag{4}$$

where x_B is the x coordinate of the barycentre, x_L is the x coordinate of the left edge of the bounding triangle, and x_R is the x coordinate of the right edge of the bounding triangle. Vertical asymmetry ($SI_{vertical}$) is computed by the difference between the y coordinates of hands. A first measure related to spatial symmetry ($SI_{spatial}$) results from the ratio of the measures of horizontal and vertical symmetries (5):

$$SI_{spatial} = \frac{SI_{horizontal}}{SI_{vertical}}. \tag{5}$$

A second measure of symmetry is based on the ratio between the measures of geometric entropy of each hand trajectory. The measure of geometric entropy provides information on how much the trajectory followed by a gesture is spread/dispersed over the available space [21]. Pijpers et al. [58] showed in a study on climbing a clear-cut correlation between such geometric entropy and the level of anxiety of the person. The more anxious the climber is (due to vertigo for example), the higher the geometric entropy index is. A high value of the index means that the climber's center of mass trajectory is dispersed with respect to the available space used to reach its final destination. Similar measures of geometric shape aspects of the trajectory were considered by Camurri et al., like *Directness Index* and *Spatial Allure*, which inform on how flexible or direct the trajectory is [17]. In line with such motion features, the *Geometric Entropy Index* further informs on how the available space is explored even in very restrained spaces, including the extreme condition of close trajectory. The hands' trajectories exhibited in the emotion portrayals frequently fell into such cases. The geometric entropy (H) associated to the hand's oscillation is computed by taking the natural logarithm of two times the length of the pattern traveled by the hand's center of mass (LP) divided by the perimeter of the convex hull around that path (6). H is computed on the frontal plane (XY) relative to the interlocutor's point of view:

$$H = \ln \frac{2{*}LP}{c}, \tag{6}$$

where LP is the path length and c is the perimeter of the convex hull around LP. The second measure of symmetry (SI_{spread}) thus considers how similar the hands' trajectories spreads were. SI_{spread} is computed as the ratio between H_{left_hand} and H_{right_hand} (7):

$$SI_{spread} = \frac{H_{left\_hand}}{H_{right\_hand}}. \tag{7}$$

Forward-backward leaning of the head. Head movement is relied on as an important feature for distinguishing between various emotional expressions. El Kaliouby and Robinson proposed a vision-based computational model to infer mental states from head movements and facial expressions [34]. Recent studies from the music field highlighted the importance of head movements in the communication of the emotional intentions of the player [23], [65]. The amount of head forward and backward leaning for each portrayal is measured by the velocity of the head displacement along its z component (depth) (8):

$$Head_{leaning}(f) = \dot{z}_{head}. \tag{8}$$

3.1.2 Dynamic-Features

The expressive features presented above are computed either at each video frame (shape features, e.g., bounding triangle) or on a sliding time window (dynamic features), thus providing a time series of values at the same sampling rate as the input video. A further processing step consisted of analyzing the temporal profiles of these time series in order to get information on their temporal dynamics. A growing body of psychological research on facial expression argues that temporal dynamics of human behavior (i.e., timing and duration of behavioral features) is a critical factor for interpretation of the observed behaviors [55]. For example, it has been shown that facial expression temporal dynamics are essential for categorization of complex psychological states like various types of pain and mood [76] and for interpretation of social behaviors, like social inhibition, embarrassment, amusement, and shame [22]. Again, much less attention has been devoted to temporal dynamics of expression through full-body movement and gesture, which is the subject of our research.

Following previous results by Castellano et al. [19], we defined a set of dynamics features, derived from the temporal shape of the curves of the expressive features [38]:

1. Maximum, Mean, Max/Mean, Standard Deviation, Coefficient of dispersion, Gesture duration: They are the maximum and mean values of the samples in the time series along a gesture, their ratio, the standard deviation (std) of the samples in the time series, the coefficient of dispersion (std/mean), and the gesture duration, respectively. These features summarize how much the variance in the time-series is localized around a specific time instant or it is spread along the whole time interval of the gesture under study.
2. Maximum/Main peak duration: The ratio between the maximum value in the time series and the time duration of its largest peak. This is an index approximating the overall impulsiveness of the movement: An impulsive movement is characterized by a short peak duration with a high absolute maximum, whereas a sustained movement is characterized by a longer peak duration with a low absolute maximum.
3. Main Peak Duration/Gesture duration, Max Index/Gesture duration: The ratio between the main peak duration and the total gesture duration, the ratio between the frame index of the maximum value and the total gesture duration (in frames), respectively. These features measure the temporal location and proportion of the main value with respect to the whole gesture.
4. Number of Maxima, Number of Maxima preceding and following the Main Maximum: Number of local maxima in the time series, number of relative maxima preceding and following the absolute one. These features assess how the magnitude of the motion feature evolves over time. These 10 dynamic features were computed for *energy*.

Each emotion portrayal can thus be analyzed and (automatically) annotated with a 25-features vector, including the standard statistical values (mean, std) of the above listed expressive features plus the dynamic features of energy (see Table 2). Automatic extraction of the features are performed using new software components developed in EyesWeb XMI (www.eyesweb.org [16]).

3.2 Dimension Reduction Module

The third module performs a dimension reduction procedure Principal Component Analysis (PCA) to reduce the dimensionality of the data and to obtain an appropriate set of independent features to be used as minimal representation of emotion portrayals. PCA is defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance of the data comes to lie on the first axis (i.e., first Principal Component, PC), the second greatest variance on the second axis, and so on.

In order to determine the number of components to retain, the module can apply to the features two complementary stopping rules [56]: parallel analysis (PA) (randomly permutated data, critical value 95th percentile) [42] and minimum average partial correlation statistics [69]. The rationale underlying parallel analysis is that nontrivial components (i.e., significant ones) from real data should have larger eigenvalues than parallel components derived from random data having the same sample size and number of variables. In Horn's original description of this procedure, the mean eigenvalues from the random data served as the comparison baseline, whereas a currently recommended practice is to use the eigenvalues that correspond to the desired percentile (typically the 95th) of the distribution of random data eigenvalues [54].

The second method, Velicer's minimum average partial (MAP) method, calculates the average of squared partial correlations after each component is partialled out. When the minimum average squared partial correlation is reached, the residual matrix resembles an identity matrix, and no further components are extracted.

TABLE 1
Selection of Emotional States Portrayed

		Valence	
		Positive	Negative
Arousal	High	Elation Amusement Pride	Hot anger (rage) Panic fear Despair
	Low	Pleasure Relief Interest	Cold anger (irritation) Anxiety (worry) Sadness (depression)

4 Evaluation of the Minimal Representation of Emotion Displays

To evaluate the efficiency of the proposed minimal representation of emotion displays, the experimental framework was applied to the GEMEP corpus.

4.1 GEMEP Corpus of Emotion Portrayals

We have chosen a corpus of emotion portrayals (the GEMEP corpus) to tune and validate the framework architecture. The Geneva Multimodal Emotion Portrayal corpus (GEMEP [5], [4], [1]) is general in the sense of number of portrayals and variety of emotions.

The GEMEP corpus was developed at the University of Geneva. It consists of more than 7,000 audio-video emotion portrayals, representing 18 emotions being portrayed by 10 actors in an interactive setting with a professional theatre director. All recordings are standardized in terms of technical and environmental features, such as light, distance, and video-camera settings, except for clothing. They also have similar time durations. These aspects make this database particularly suited for developing automatic analysis procedures. Our analyses are based on a selected subset of 120 portrayals representing a full within-subjects 12 (emotions) $\times$ 10 (actors) design. These portrayals have been validated by extensive ratings that ensured high believability (assessed as perceived capacity of the actor to communicate a natural emotion impression), reliability (interrater agreement), and recognizability (accuracy scores) of the encoded emotion [5]. It should be noted that the shared recognition of a particular emotion in an expression is the only possible way to validate a material set of this nature.

During the recording procedure, the actors were requested to express an emotion in interaction with a professional theater director on the basis of three short scenarios with a definition of every intended emotion (e.g., anger may result from a violent dissatisfaction caused by a malevolent person, like an owner subleasing his apartment and discovering that his flat was damaged during his absence, elation may ensue from being transported by a splendid thing which arrives to us in an unexpected way like gaining a fabulous sum with the lotto and announcing to one's family). Acted emotions were privileged with respect to the use of natural expression of emotions for they offer the possibility to record variable expressions for the same individuals. When working with natural expressions, a few emotional reactions can be recorded for a given individual and considering the variability of naturally occurring situations and events, the compatibility of emotional reactions across individuals tends to be reduced. Acted emotions permit acquiring sufficient variability on the level of the emotional states encoded in the nonverbal expressions, hence leading to more informative outcomes resulting from within—and between subjects analysis. On the one hand, emotional expressions produced by actors have been criticized for they may in part reflect cultural stereotypes or stereotypes pertaining to acting traditions, thus weakening the generalizability to spontaneous expressions occurring in real-life [28]. On the other hand, this kind of research could not possibly be done with real-life spontaneous items given practical and ethical constrains with regard to the induction and replicability, as well as the systematic selection of such expressive items.

The selection of affective states in this database was guided by those states that are frequently studied in the literature [61], [62], but less frequently examined states were also included on the basis of theoretical considerations (e.g., to study the possibility of differentiating positive emotional expressions or of disentangling the influence of arousal level with emotion family). Table 1 shows the set of affective states theoretically ordered according to two main emotion dimensions, valence and arousal. A complete description of the GEMEP material, procedure, and rationale for using actors can be found in [5], [4].

Two digital cameras with constant shutter, manual gain, and focus at 25 fps with a 720×576 pixel resolution were used to record the body movements of the actors from both the frontal and lateral view. With regard to body movement restrictions, the actors were only instructed not to move away from the focus of the camera.

The GEMEP was previously employed in two studies on emotion recognition. In [18], a subset of six portrayals depicting anger, joy, and sadness was used to model bidirectional communication between a user and an agent. Real-time analysis of actors' expressivity was based on the automatic extraction of expressive motion features related to the hand and hands in the frontal view (interlocutor's one). Kinematic data, degree of body expansion, and fluidity, referring to the uniformity of motion, were linearly mapped to animate an human-like autonomous agent (avatar) and generate expressive copying behavior. In [70], a subset of 40 portrayals designed to portray happiness, pleasure, anger, and irritation served to study the abilities of children and adults to categorize and label dynamic bodily/facial expressions. Our study pursues the objectives of [18] to automatize the extraction of expressive motion features related to a reduced amount of visual and dynamic information. In contrast, however, to both [18] and [70], our

TABLE 2
Rotated Component Matrix of the 25 Expressive Features (Rotation Converged in 10 Iterations) Extraction Method: Principal Component Analysis

	Component			
	1	2	3	4
Energy				
Max	.936			
Max/mean	.327	.611		
Coefficient of dispersion	.361	.612		
Max/peak (duration)	.935			
Max(frame) / gesture (duration)		.519		.385
Peaks number				.836
Peaks evolution				.872
Gesture[1] (mean)	.919			
Gesture (std)	.957			
Gesture (duration)				

	Component			
	1	2	3	4
Head				
Head velocity (mean)	.538	.370		
Head velocity (std)	.703	.344		
Bounding triangle				
Perimeter (mean)			.861	
Perimeter (std)	.703	.435		
Hands				
Hands distance (mean)			.861	
Hands distance (std)	.429	.608	.440	
Hands vertical component (%)	.491	-.319	-.364	
Hands upward movement				
Global symmetry[2] (mean)			.612	.371
Global symmetry (std)		.504	-.367	

	Component			
	1	2	3	4
Symmetry index[3]		-.428		
Horizontal symmetry (%)				
Symmetry (entropy)[4]				.346
Curvature (left hand)	.476			.428
Curvature (right hand)	.437			.376

[1] Gesture refers to the entire sequence of movements, [2] The Global symmetry refers to the sum of the horizontal and vertical symmetries [3] The symmetry index indicates whether the Global symmetry is related to vertical or horizontal symmetries. This last component is computed in details in the following variable horizontal symmetry (%). [4] Symmetry entropy is the difference between the geometry index of entropy of each hand trajectory (see section on *motion descriptors*)

Rotation method: varimax with kaiser normalization.

study stands out by 1) formulating a systematic approach to produce a gesture-based description of emotion expressions based on head and hands movement, 2) by considering a larger subset of 120 portrayals representing a much wider spectrum of emotions (12), and 3) by including both frontal and side views in the analysis.

4.2 Dimension Reduction

Results from tests on the GEMEP corpus are indicated in Table 2. From the application of the dimension reduction module (PCA, Velicer's, and Parallel analysis methods) to the features extracted from the GEMEP corpus, it appeared that the first four eigenvalues from the actual data were larger than the corresponding first four 95th percentile (and mean) random data eigenvalues (see Fig. 4). This indicates that four components had to be retained.

The four components accounted for 56 percent of the variance. Specifically, the first two components explained 29.4 percent and 10.3 percent of the variance, respectively, the third component explained 9.4 percent of the variance, while the fourth component explained 6.9 percent of the variance. In considering the component loadings of the motion features (Fig. 4), the first component roughly corresponds to *motion activity* (the amount of energy), the second component to the *temporal and spatial excursion of movement* (e.g., does energy concentrate at the beginning or at the end of the motion sequence? is the corresponding magnitude of the energy high or low with respect to the entire gestural sequence?), the third component corresponds to the *spatial extent and postural symmetry*, and the fourth component to the *motion discontinuity and jerkiness* (is the movement irregular?).

These results are in line with previous research findings [74], [50], [51]. The complex original set of expressive gesture features is thus best summarized by proposing a 4D scheme to characterize affective-related movement.

4.3 Two-Step Clustering

In order to validate the effectiveness of the features in representing emotion portrayals, we applied clustering techniques, an unsupervised learning method, on the PCA-based data to classify a portrayal as belonging to a

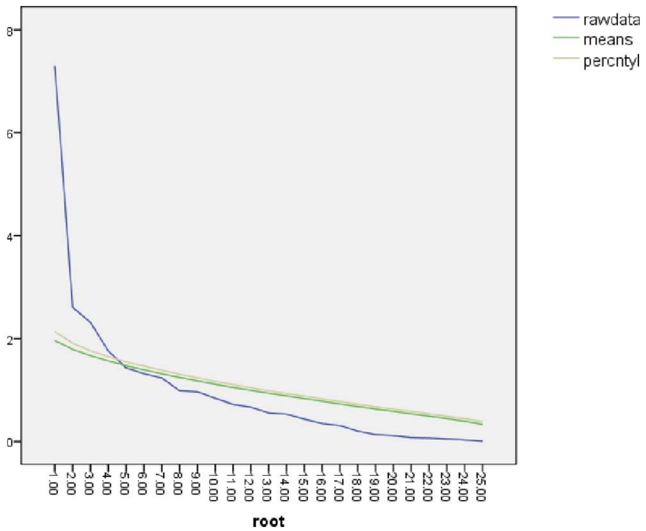
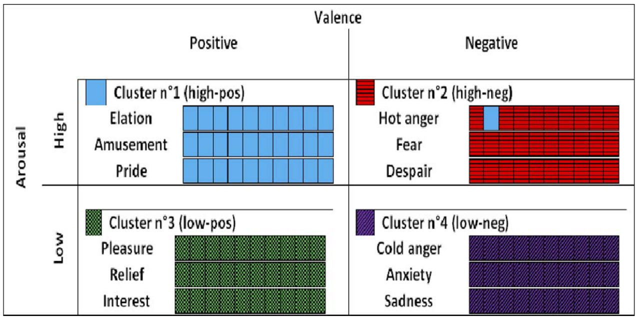


Fig. 4. Scree plot of the principal components: The parallel analysis indicates that four components had to be retained. The green line for parallel analysis in the graph crosses the solid PCA line before reaching the fifth component.

TABLE 3
Cluster Membership of the 12 Emotion Classes, Each Cell Denotes a Portrayal



		Valence			
Arousal		Positive		Negative	
	High	🟦 Cluster n°1 (high-pos)		🟥 Cluster n°2 (high-neg)	
		Elation	🟦🟦🟦🟦🟦🟦🟦🟦🟦🟦	Hot anger	🟥🟦🟥🟥🟥🟥🟥🟥🟥🟥
		Amusement	🟦🟦🟦🟦🟦🟦🟦🟦🟦🟦	Fear	🟥🟥🟥🟥🟥🟥🟥🟥🟥🟥
		Pride	🟦🟦🟦🟦🟦🟦🟦🟦🟦🟦	Despair	🟥🟥🟥🟥🟥🟥🟥🟥🟥🟥
	Low	🟩 Cluster n°3 (low-pos)		🟪 Cluster n°4 (low-neg)	
		Pleasure	🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩	Cold anger	🟪🟪🟪🟪🟪🟪🟪🟪🟪🟪
		Relief	🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩	Anxiety	🟪🟪🟪🟪🟪🟪🟪🟪🟪🟪
		Interest	🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩	Sadness	🟪🟪🟪🟪🟪🟪🟪🟪🟪🟪

single cluster. Two categorical variables corresponding to the theoretically derived value of the emotion on the arousal and valence continuum were also included for the classification (see Table 1). The number of clusters and their association with emotional categories were obtained by applying the clustering techniques to the GEMEP corpus. Given the number of available portrayals in the GEMEP corpus (10 portrayals for each of the 12 emotions), we selected unsupervised learning techniques. To automatically determine the optimal number of clusters and to avoid capitalizing on chance, we applied bootstrapping techniques [29], [24]. We generated a thousand resamples of the original data set, each of which was obtained by random sampling with replacement from the original data set. On each new sample we performed a two-step cluster analysis, a procedure implemented in spss (www.spss.com). The two-step cluster analysis procedure is a scalable cluster analysis algorithm that first preclusters the data into small subclusters using a sequential clustering approach and then groups these subclusters in the selected number of clusters through hierarchical clustering method. The most frequent solution resulting from the procedure applied on the GEMEP data set was a four-cluster solution (73.6 percent of bootstrapped samples).

4.4 Results

This section details the results and compare with the emotion categories of the GEMEP corpus. We investigate how much information is retained in the minimal representation obtained from the application of the experimental framework to the GEMEP corpus.

The four obtained clusters were compared with the original twelve emotion sets (*ground truth*) by creating a $12\ (\mathrm{emotions}) \times 4\ (\mathrm{clusters})$ contingency table.

The contingency coefficient $(\mathrm{C} = 0.864, p < 0.001)$ and the effect size measure $(\mathrm{w} = 0.99)$ reveals a very strong association between the four clusters and the 12 emotion categories. Clusters 1, 2, 3, and 4 contain, respectively, 25.8 percent, 24.2 percent, 25.0 percent, and 25.0 percent of the emotion portrayals.

As detailed in Table 3, results showed that the majority of emotion portrayals are clearly associated with one of the four clusters. Specifically, the valence quality of an emotion, usually referred to as the pleasantness-unpleasantness dimension, and arousal were equally appropriate to explain the cluster outcomes. Each of the four clusters represents one quadrant of the 2D structure proposed by [62]. Only one portrayal out of the 120 (actor 2 portraying hot anger) was classified under cluster 1 instead of cluster 2. The following discussion of the cluster profiles stem from these data-confirmed groupings. We show that the minimal representation can distinguish between the two theoretical concepts of arousal and valence loosing the fine-grain distinction among the 12 original emotions of the GEMEP corpus.

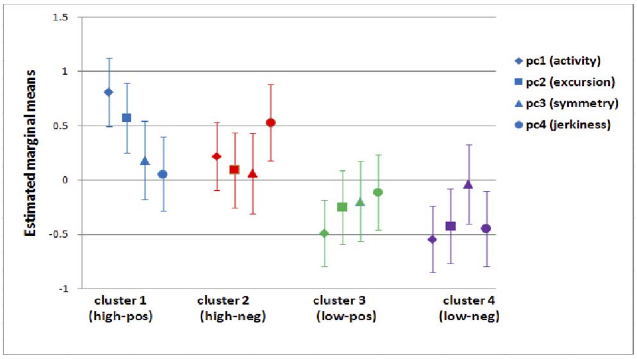


Fig. 5. The four clusters (blue, red, green, and purple lines) with their corresponding values of the four principal components: (first PC) motion activity (lozenge-shape), (second PC) temporal and spatial excursion of movement (square), (third PC) spatial extent and postural symmetry (triangle), (fourth PC) motion discontinuity and jerkiness (circle); the y axis corresponds to the estimated marginal means for z-score.

4.5 Cluster Profiles

Fig. 5 shows that cluster 1 (high positive) groups portrayals with high loadings on the first and second principal components (PC), respectively, referring to the motion activity and to the temporal and spatial excursion of movement. The movement pattern in this group is actually one of a very high amount of activity and high movement excursion compared to the other clusters. Values of symmetry and especially the motion discontinuity or jerkiness (respectively the third and fourth PC) are, on the contrary, relatively lower referring to asymmetric and slightly discontinuous type of movements. Cluster 2 (high negative) identifies portrayals whose movements are also characterized by relatively high loadings on the first and second principal components, yet are much less accentuated than in cluster one. This second cluster also clearly distinguished itself by a high score of jerkiness, referring to movements that unfold in a discontinuous way. In cluster 3 (low positive), groups portrayals are characterized by low motor activity and a reduced spatiotemporal excursion of movements with respect to the portrayals aggregated in cluster 1 and 2. Movements are also characterized by relatively asymmetry and moderate jerkiness (i.e., greater smoothness). Portrayals grouped in cluster 4 (low negative) are characterized by the low activity, spatiotemporal excursion, and jerkiness in movement execution, yet displaying a symmetry score similar to the one observed in clusters 1 and 2.

The current results show that the 12 emotions we addressed in this study can be distinguished on the basis of four components of head and hand movement features. The emotion categories grouped according to their position in the arousal/valence dimensions, presented a high intern homogeneity (low intraclass variability) with respect to the

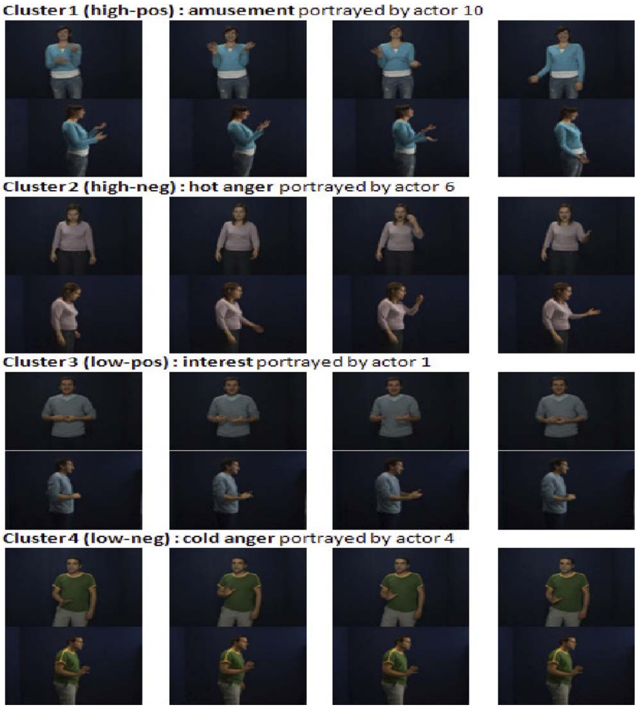

Cluster1 (high-pos) : amusement portrayed by actor 10
Cluster2 (high-neg) : hot anger portrayed by actor 6
Cluster3 (low-pos) : interest portrayed by actor 1
Cluster4 (low-neg) : cold anger portrayed by actor 4

Fig. 6. The selected four typical portrayals of each cluster, identified as the closest to their cluster centroid. From top to bottom, amusement, hot anger, interest, and cold anger emotions, respectively, portrayed by actors 10, 6, 1, and 4. From left to right, the four key-frames of each portrayal.

four cluster types. Cluster overlap on symmetry (see Fig. 5, estimated means of the four clusters converge around 0) confirms that information on the dynamics of movement performance is required to disambiguate the results obtained on the sole basis of postural information [7], [19].

Following [12], the most typical portrayals of each cluster were identified as the closest, based upon euclidean distance, to the cluster centroid in the principal components coordinate system. Snapshots of the frontal and side views of the four selected portrayals are displayed in Fig. 6).

It is worth pointing out that cluster profiles conform to the literature in nonverbal behavior research, developmental psychologies, clinical, and dance movement studies. According to [51], [74], a high amount of movement activity and variability clearly characterize high arousal emotions such as the one contained in clusters 1 and 2. High negative emotions grouped in cluster 2 are further related to motion discontinuity ("jerkiness"), confirming previous results by [59]. This finding is in line with a rating study showing that angry expressions were considered jerkier than happy expressions [52]. Finally, low positive and negative emotions of clusters 3 and 4 are often portrayed in the literature by continuous, temporally and spatially invariable movements. They are also low on movement activity compared to emotions from clusters 1 and 2. The combination of low activity and smoothness is typically ascribed to such emotions [59], [74].

In order to discard possible effects of individual movement characteristics (idiosyncratic styles) on cluster membership [6], we further computed the contingency between each of the 10 individual performers (who have 12 values each, corresponding to each of the 12 emotions) and cluster membership. Idiosyncratic style effects could arise in the case where some performers are characterized by specific clusters. Given the assumption of Chi-square owing that no more than 20 percent of the expected counts are less than five, the Fisher's Exact Test was used alternatively. Results showed that there is no association between an individual performer and a specific cluster (Fisher's Exact $\text{Test value} = 2.574$, p. value > 0.05). We may therefore conclude that the presented results are not an artifact of individual idiosyncratic styles, but may be ascribed to emotion expression alone.

5 Conclusion

This paper introduces an experimental framework and algorithms for the automated extraction in real time of a reduced set of features, candidate for characterizing affective behavior. Results can be used for developing techniques for lossy compression of emotion portrayals from a reduced amount of visual information related to human upper-body movements and gestures.

In order to evaluate the effectiveness of the experimental framework and algorithms, they were applied to extract a minimal representation from the GEMEP corpus and used as input to unsupervised machine-learning techniques. The analysis suggested that meaningful groups of emotions, related to the four quadrants of the valence/arousal space, could be distinguished. Further, these emotions were characterized by homogeneous movement patterns, as emerged by the cluster analysis. This is a remarkable result given the low quality and limited amount of information.

In addition, the identification of multiple clusters within the bodily modality extend findings from [63], while adding spatiotemporal features of movement into the equation. In particular, the 4D scheme proposed in the system for characterizing expressive behavior confirmed that dynamic aspects of motion features are complementary to postural and gesture shape-related information. These results also corroborate the recent view that bodily expressions of emotion constitute a relevant source of nonverbal emotion information [55], [71], [27], [74], [2].

In particular, the current results suggest that many of the distinctive cue parameters that distinguish the bodily expression of different emotions may be dynamic rather than categorical in nature—concerning movement quality rather than specific types of gestures, as they have been studied in linguistically oriented gesture research or work on nonverbal communication [10], [32], [36], [37], [49]. In consequence, a paradigm shift toward greater emphasis on the dynamic quality and unfolding of expressive behavior may help to provide a breakthrough in our ability to analyze and synthesize the mapping of emotion specificity into gestural, facial, and vocal behavior—with obvious benefits for leading edge affective computing applications. Of course, if on the one hand, this work represents a step toward the definition of a minimal representation characterizing affective behavior, on the other hand, some limitations have to be addressed in future work. For example, 1) our minimal representation can distinguish only among a small number of emotion expressions (four clusters), 2) it only includes the visual modality, and 3) it is limited to a categorical approach to emotion, whereas subtler emotion expressions and nuances should be taken into account. Future work includes the extension of the minimal representation to take into account other modalities

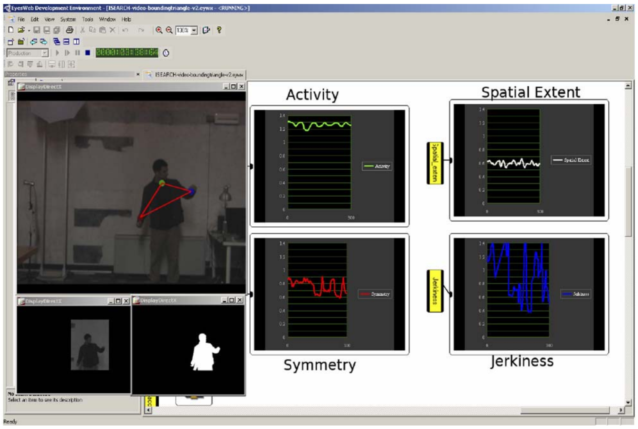


Fig. 7. The EU-ICT I-SEARCH project: An EyesWeb XMI experimental application implementing the minimal representation of emotion described in this paper. The snapshot in the figure shows the real-time extraction of activity, spatial extent, symmetry, and jerkiness from a single camera video input.

(e.g., audio features). We expect this to be a line of research worth pursuing, given the strong connection between body posture and gesture with speech and facial expression [46], [49], [68], [71], [55]. In this direction, we are working toward a bimodal approach consisting of the integration of the movement features with a number of (synchronized) audio features, using cross-modal techniques for audio analysis, i.e., inspired by the ones presented here on movement analysis [14]. Other extensions are in the direction of improving the feature extraction techniques, with particular reference to the dynamic features, in order to increase the number of emotion expression that can be distinguished. A more detailed analysis to be carried out on a larger number of higher-resolution data, aiming at investigating and describing the relative importance of each channel and of each feature is also a topic for our future work, as well as moving from acted emotion to natural expressions of emotion.

This research aims at contributing to concrete applications in affective computing, multimodal interfaces, and user-centric media applications, exploiting nonverbal affective behavior captured by a limited amount of visual information. The behavioral data were acquired by means of two video cameras: The same data could be obtained by a single video camera, e.g., the emerging 3D webcams (time-of-flight camera, MS Kinect, or stereo webcam). We are applying the proposed approach to 1) real-time networked communication and interaction applications to enhance the expressive interaction between remote users in networked music performance, and to 2) multimodal affective search of multimedia content, in the framework of the EU-ICT I-SEARCH Project. The EU-ICT I-SEARCH Project (EU 7th Framework, ICT, DGINFSO, Networked Media Unit) aims at developing novel multimodal search engines by exploiting users nonverbal affective behavior (Fig. 7). The cross-modal mapping between multimodal descriptors of multimedia content and the users nonverbal emotional expressions become a key-element of the search process.

Acknowledgments

The research presented in this paper has been partly funded by the EU 7th Framework Programme ICT Project I-SEARCH (a unified framework for multimodal content search), DG-INFSO, “Networked Media” Unit. The authors thank Carlo Chiorri, Tania Bänziger, Maurizio Mancini, Paolo Coletta, Alberto Massari, and the anonymous reviewers for their helpful comments and suggestions.

References

[1] http://www.affective-sciences.org/gemep., 2011.

[2] A.P. Atkinson, W.H. Dittrich, A.J. Gemmell, and A.W. Young, “Emotion Perception from Dynamic and Static Body Expressions in Point-Light and Full-Light Displays,” *Perception*, vol. 33, pp. 717-746, 2004.

[3] T. Balomenos, A. Raouzaiou, S. Ioannou, A. Drosopoulos, K. Karpouzis, and S. Kollias, “Emotion Analysis in Man-Machine Interaction Systems,” *Proc. Workshop Machine Learning for Multimodal Interaction*, pp. 318-328, 2004.

[4] T. Banziger and K.R. Scherer, “Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The GEMEP Corpus,” *Proc. Second Int'l Conf. Affective Computing and Intelligent Interaction*, pp. 467-487, 2007.

[5] T. Banziger and K.R. Scherer, “Chapter Blueprint for Affective Computing: A Sourcebook,” *Introducing the Geneva Multimodal Emotion Portrayal Corpus*, pp. 271-294, Oxford Univ. Press, 2010.

[6] D. Bernhardt and P. Robinson, “Detecting Affect from Non-Stylised Body Motions,” *Proc. Second Int'l Conf. Affective Computing and Intelligent Interaction*, pp. 59-70, 2007.

[7] N. Bianchi-Berthouze, P. Cairns, A. Cox, C. Jennett, and W.W. Kim, “On Posture as a Modality for Expressing and Recognizing Emotions,” *Proc. Workshop the Role of Emotion in HCI*, 2006.

[8] N. Bianchi-Berthouze and A. Kleinsmith, “A Categorical Approach to Affective Gesture Recognition,” *Connection Science*, vol. 15, no. 4, pp. 259-269, 2003.

[9] R.L. Birdwhistell, *Kinesics and Context: Essays on Body Motion Communication.* Univ. of Pennsylvania Press, 1970.

[10] A.F. Bobick, “Movement, Activity and Action: The Role of Knowledge in the Perception of Motion,” *Philosophical Trans. Royal Soc. B: Biological Sciences*, vol. 352, no. 1358, pp. 1257-1265, 1997.

[11] J.C. Borod, C.S. Haywood, and E. Koff, “Neuropsychological Aspects of Facial Asymmetry During Emotional Expression: A Review of the Normal Adult Literature,” *Neuropsychology Rev.*, vol. 7, no. 1, pp. 41-60, 1997.

[12] E. Braun, B. Geurten, and M. Egelhaaf, “Identifying Prototypical Components in Behaviour Using Clustering Algorithms,” *PLoS ONE*, vol. 5, no. 2, pp. e9361, 2010.

[13] P.E. Bull, “Body Movement Scoring System,” SSRC End-of-Grant Report, HR 6404/2, The Social Functions of Speech-Related Body Movement, pp. 1-16, 1981.

[14] A. Camurri, P. Coletta, C. Drioli, A. Massari, and G. Volpe, “Audio Processing in a Multimodal Framework,” *Proc. Int'l Conf. Audio Eng. Soc. Convention*, May 2005.

[15] A. Camurri, I. Lagerlöf, and G. Volpe, “Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques,” *Int'l J. Human-Computer Studies, Elsevier Science*, vol. 59, pp. 213-225, July 2003.

[16] A. Camurri, B. Mazzarino, and G. Volpe, “Expressive Interfaces,” *Cognition, Technology, and Work*, vol. 6, no. 1, pp. 15-22, 2004.

[17] A. Camurri, G. Volpe, G. De Poli, and M. Leman, “Communicating Expressiveness and Affect in Multimodal Interactive Systems,” *IEEE Multimedia*, vol. 12, no. 1, pp. 43-53, Jan.-Mar. 2005.

[18] G. Castellano and M. Mancini, “Analysis of Emotional Gestures for the Generation of Expressive Copying Behaviour in an Embodied Agent,” *Gesture-Based Human-Computer Interaction and Simulation*, Springer, pp. 193-198, 2009.

[19] G. Castellano, M. Mortillaro, A. Camurri, G. Volpe, and K. Scherer, “Automated Analysis of Body Movement in Emotionally Expressive Piano Performances,” *J. Music Perception*, vol. 26, no. 2, pp. 103-119, 2008.

[20] G. Castellano, S.D. Villalba, and A. Camurri, “Recognising Human Emotions from Body Movement and Gesture Dynamics,” *Proc. Second Int'l Conf. Affective Computing and Intelligent Interaction*, 2007.

[21] P. Cordier, M. Mendes France, J. Pailhous, and P. Bolon, “Entropy as a Global Variable of the Learning Process,” *Human Movement Science*, vol. 13, no. 6, pp. 745-763, 1994.

[22] M. Costa, W. Dinsbach, A.S.R. Manstead, and P.E.R. Bitti, "Social Presence, Embarrassment, and Nonverbal Behavior," *J. Nonverbal Behavior,* vol. 25, no. 4, pp. 225-240, 2001.

[23] S. Dahl and A. Friberg, "Visual Perception of Expressiveness in Musicians' Body Movements," *Music Perception,* vol. 24, no. 5, pp. 433-454, 2007.

[24] A.C. Davison and D.V. Hinkley, *Bootstrap Methods and Their Application.* Cambridge Univ. Press, 1999.

[25] B. De Gelder, "Towards the Neurobiology of Emotional Body Language," *Nature Rev. Neuroscience (Print),* vol. 7, no. 3, pp. 242-249, 2006.

[26] W.T. Dempster and G.R.L. Gaughran, "Properties of Body Segments Based on Size and Weight," *Am. J. Anatomy,* vol. 120, no. 1, pp. 33-54, 1967.

[27] W.H. Dittrich, T. Troscianko, S.E.G. Lea, and D. Morgan, "Perception of Emotion from Dynamic Point-Light Displays Represented in Dance," *Perception-London,* vol. 25, pp. 727-738, 1996.

[28] E. Douglas-Cowie, N. Campbell, R. Cowie, and P. Roach, "Emotional Speech: Towards a New Generation of Databases," *Speech Comm.,* vol. 40, nos. 1/2, pp. 33-60, 2003.

[29] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap.* Chapman & Hall, 1993.

[30] P. Ekman, "Differential Communication of Affect by Head and Body Cues," *J. Personality and Social Psychology,* vol. 2, no. 5, pp. 726-735, 1965.

[31] P. Ekman and W.V. Friesen, "Head and Body Cues in the Judgment of Emotion: A Reformulation," *Perceptual and Motor Skills,* vol. 24, no. (3 PT 1), pp. 711-724, 1967.

[32] P. Ekman and W.V. Friesen, *The Repertoire of Nonverbal Behavior.* Mouton de Gruyter, 1969.

[33] P. Ekman and W.V. Friesen, "Constants across Cultures in the Face and Emotion," *J. Personality and Social Psychology,* vol. 17, pp. 124-129, 1971.

[34] R. el Kaliouby and P. Robinson, "Generalization of a Vision-Based Computational Model of Mind-Reading," *Proc. First Int'l Conf. Affective Computing and Intelligent Interaction,* pp. 582-589, 2005.

[35] T. Flash and N. Hogan, "The Coordination of Arm Movements: An Experimentally Confirmed Mathematical Model," *J. Neuroscience,* vol. 5, no. 7, pp. 1688-1703, 1985.

[36] N. Freedman, "Hands, Words and Mind: On the Structuralization of Body Movements during Discourse and the Capacity for Verbal Representation," *Communicative Structures and Psychic Structures: A Psychoanalytic Interpretation of Comm.,* pp. 109-132, 1977.

[37] S. Frey and J. Pool, "A New Approach to the Analysis of Visible Behavior," Forschungsberichte aus dem Psychologischen Institut, Universität Bern, 1976.

[38] D. Glowinski, A. Camurri, G. Volpe, N. Dael, and K. Scherer, "Technique for Automatic Emotion Recognition by Body Gesture Analysis," *Proc. IEEE CS Workshops Computer Vision and Pattern Recognition,* pp. 1-6, 2008.

[39] H. Gunes and M. Piccardi, "Bi-Modal Emotion Recognition from Expressive Face and Body Gestures," *J. Network and Computer Applications,* vol. 30, no. 4, pp. 1334-1345, 2007.

[40] H. Gunes and M. Piccardi, "Automatic Temporal Segment Detection and Affect Recognition from Face and Body Display," *IEEE Trans. Systems, Man, and Cybernetics, Part B,* vol. 39, no. 1, pp. 64-84, Feb. 2009.

[41] S. Hashimoto, "KANSEI as the Third Target of Information Processing and Related Topics in Japan," *Proc. AIMI Int'l Workshop KANSEI,* pp. 101-104, 1997.

[42] J.L. Horn, "A Rationale and Test for the Number of Factors in Factor Analysis," *Psychometrika,* vol. 30, no. 2, pp. 179-185, 1965.

[43] A. Hreljac, "The Relationship between Smoothness and Performance during the Practice of a Lower Limb Obstacle Avoidance Task," *Biological Cybernetics,* vol. 68, no. 4, pp. 375-379, 1993.

[44] A. Kapoor, W. Burleson, and R.W. Picard, "Automatic Prediction of Frustration," *Int'l J. Human-Computer Studies,* vol. 65, no. 8, pp. 724-736, 2007.

[45] A. Kapur, A. Kapur, N. Virji-Babul, G. Tzanetakis, and P.F. Driessen, "Gesture-Based Affective Computing on Motion Capture Data," *Proc. First Int'l Conf. Affective Computing and Intelligent Interaction,* pp. 1-7, 2005.

[46] S. Kita, I. Van Gijn, and H. Van der Hulst, "Movement Phases in Signs and Co-Speech Gestures, and Their Transcription by Human Coders," *Proc. Int'l Gesture Workshop Gesture and Sign Language in Human-Computer Interaction,* pp. 23-36, 1998.

[47] R. Laban and L. Ullmann, *The Mastery of Movement.* Plays, 1971.

[48] I. Laso-Ballesteros and P. Daras, "User Centric Future Media Internet," technical report, EU Commission, Sept. 2008.

[49] D. McNeill, *Hand and Mind: What Gesture Reveals about Thought.* Univ. of Chicago Press, 1992.

[50] A. Mehrabian, *Nonverbal Comm.* Aldine, 2007.

[51] M. Meijer, "The Contribution of General Features of Body Movement to the Attribution of Emotions," *J. Nonverbal Behavior,* vol. 13, no. 4, pp. 247-268, 1989.

[52] J. Montepare, E. Koff, D. Zaitchik, and M. Albert, "The Use of Body Movements and Gestures as Cues to Emotions in Younger and Older Adults," *J. Nonverbal Behavior,* vol. 23, no. 2, pp. 133-152, 1999.

[53] S. Mota and R.W. Picard, "Automated Posture Analysis for Detecting Learner's Interest Level," *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshop,* vol. 5, p. 49, 2003.

[54] B.P.O Connor, "SPSS and SAS Programs for Determining the Number of Components Using Parallel Analysis and Velicer's MAP Test," *Behavior Research Methods Instruments and Computers,* vol. 32, no. 3, pp. 396-402, 2000.

[55] M. Pantic, A. Pentland, A. Nijholt, and T.S. Huang, "Human Computing and Machine Understanding of Human Behavior: A Survey," *Proc. Int'l Conf. Artificial Intelligence for Human Computing,* pp. 47-71, 2007.

[56] P.R. Peres-Neto, D.A. Jackson, and K.M. Somers, "How Many Principal Components? Stopping Rules for Determining the Number of Non-Trivial Axes Revisited," *Computational Statistics and Data Analysis,* vol. 49, no. 4, pp. 974-997, 2005.

[57] R.W. Picard, *Affective Computing.* MIT Press, 1997.

[58] J.R. Pijpers, R.R.D. Oudejans, F. Holsheimer, and F.C. Bakker, "Anxiety-Performance Relationships in Climbing: A Process-Oriented Approach," *Psychology of Sport and Exercise,* vol. 4, no. 3, pp. 283-304, 2003.

[59] F.E. Pollick, H.M. Paterson, A. Bruderlin, and A.J. Sanford, "Perceiving Affect from Arm Movement," *Cognition,* vol. 82, no. 2, pp. 51-61, 2001.

[60] C.L. Roether, L. Omlor, and M.A. Giese, "Lateral Asymmetry of Bodily Emotion Expression," *Current Biology,* vol. 18, no. 8, pp. R329-R330, 2008.

[61] J.A. Russell, "A Circumplex Model of Affect," *J. Personality and Social Psychology,* vol. 39, no. 6, pp. 1161-1178, 1980.

[62] J.A. Russell, J.A. Bachorowski, and J.M. Fernandez-Dols, "Facial and Vocal Expressions of Emotion," *Ann. Rev. Psychology,* vol. 54, no. 1, pp. 329-349, 2003.

[63] K.R. Scherer and H. Ellgring, "Multimodal Expression of Emotion: Affect Programs or Componential Appraisal Patterns," *Emotion,* vol. 7, no. 1, pp. 158-171, 2007.

[64] P. Siple, *Understanding Language through Sign Language Research.* Academic Press Inc., 1978.

[65] R. Timmers, M. Marolt, A. Camurri, and G. Volpe, "Listeners' Emotional Engagement with Performances of a Scriabin Etude: An Explorative Case Study," *Psychology of Music,* vol. 34, no. 4, pp. 481-510, 2006.

[66] E. Todorov and M.I. Jordan, "Smoothness Maximization along a Predefined Path Accurately Predicts the Speed Profiles of Complex Arm Movements," *J. Neurophysiology,* vol. 80, no. 2, pp. 696-714, 1998.

[67] J.L. Tracy and R.W. Robins, "The Prototypical Pride Expression: Development of a Nonverbal Behavior Coding System," *Emotion,* vol. 7, no. 4, pp. 789-801, 2007.

[68] J. Van den Stock, R. Righart, and B. de Gelder, "Whole Body Expressions Influence Recognition of Facial Expressions and Emotional Prosody," *Emotion,* vol. 7, pp. 487-494, 2007.

[69] W.F. Velicer, "Determining the Number of Components from the Matrix of Partial Correlations," *Psychometrika,* vol. 41, no. 3, pp. 321-327, 1976.

[70] S. Vieillard and M. Guidetti, "Children's Perception and Understanding of (Dis)similarities among Dynamic Bodily/Facial Expressions of Happiness, Pleasure, Anger, and Irritation," *J. Experimental Child Psychology,* vol. 102, no. 1, pp. 78-95, 2009.

[71] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social Signals, Their Function, and Automatic Analysis: A Survey," *Proc. 10th Int'l Conf. Multimodal Interfaces,* pp. 61-68, 2008.

[72] P. Viviani and C. Terzuolo, "Trajectory Determines Movement Dynamics," *Neuroscience,* vol. 7, no. 2, pp. 431-437, 1982.

[73] H.G. Wallbott, "Movement Quality Changes in Psychopathological Disorders," *Normalities and Abnormalities in Human Movement, Medicine and Sport Science,* vol. 29, pp. 128-146, 1989.

[74] H.G. Wallbott, "Bodily Expression of Emotion," *European J. Social Psychology,* vol. 28, pp. 879-896, 1998.

[75] H.G. Wallbott and K.R. Scherer, "Cues and Channels in Emotion Recognition," *J. Personality and Social Psychology,* vol. 51, no. 4, pp. 690-699, 1986.

[76] A.C.C. Williams, "Facial Expression of Pain: An Evolutionary Account," *Behavioral and Brain Sciences,* vol. 25, no. 4, pp. 439-455, 2003.

[77] J.H. Yan, "Effects of Aging on Linear and Curvilinear Aiming Arm Movements," *Experimental Aging Research,* vol. 26, no. 4, pp. 393-407, 2000.

Donald Glowinski received the MSc degree in cognitive science from the Ecole des Hautes Etudes en Sciences Sociales (EHESS), the MSc degree in music and acoustics from the Conservatoire National Suprieur de Musique et de Danse de Paris (CNSMDP), the MSc degree in philosophy from the Sorbonne-Paris IV, and the PhD degree in computing engineering from the InfoMus Lab—Casa Paganini, in Genoa, Italy, under the direction of Professor Antonio Camurri. His background covers scientific and humanistic academic studies, as well as high-level musical training. He was the chairman of the Club NIME 2008 (New Interfaces for Musical Expression), Genoa, 2008. His research interests include user-centric, multimodal, and social aware computing. He works in particular on the modeling of automatic gesture-based recognition of emotions in real-word scenarios. He is a member of the IEEE.

Nele Dael received the MSc degree in psychology and works as a doctoral student at the Swiss Center for Affective Sciences at the University of Geneva, Switzerland, under the direction of Professor K. Scherer. Her current research project covers the expression and perception of emotion through body posture, movement, and gesture. Her work includes the development of new methodologies for the description and analysis of human affective behavior. Her research interests include the behavioral study of affective phenomena in an interdisciplinary setting, including engineering and ethology in particular.

Antonio Camurri received the master's degree in electric engineering and the PhD degree in computer engineering in 1984 and 1991, respectively. He is an associate professor at DIST, University of Genova (Faculty of Engineering, Computer Engineering Curriculum), where he teaches "Human Computer Interaction" and "Multimedia Systems." His research interests include multimodal intelligent interfaces, interactive systems, sound and music computing, kansei information processing, computational models of emotions, with a special focus on empathy and entrainment, and interactive multimodal systems for theatre, music, dance, museums, interactive multimodal systems for therapy, rehabilitation, independent living. He is a founder and scientific director of the InfoMus Lab at DIST-University of Genova (www.infomus.org). He was the president of AIMI (Italian Association for Musical Informatics), is a member of the Executive Committee (ExCom) of the IEEE CS Technical Committee on Computer Generated Music, is an associate editor of the international *Journal of New Music Research*, and is a member of the Board of the European Institute of Enactive Systems. Since 1994, he has coordinated and has been local project manager of EU (IST 5 and 6 FP, ICT 7FP, CRAFT, and Culture 2007) Projects. He is the author of more than 100 international scientific publications. In 2005, he founded—and is the director of—he Casa Paganini—InfoMus International Research Centre of University of Genoa (www.casapaganini.org).

Gualtiero Volpe received the PhD degree in computer engineering,in 2003. He is an assistant professor at DIST-InfoMus Lab. His research interests include intelligent and affective human-machine interaction, modeling and real-time analysis and synthesis of expressive content in music and dance, and multimodal interactive systems. He is a member of the Board of Directors of AIMI (Italian Association for Musical Informatics). In 2005, he was the chairman of the Fifth International Gesture Workshop and guest editor of a special issue of the *Journal of New Music Research on Expressive Gesture in Performing Arts and New Media*. He was a cochair of NIME 2008 (New Interfaces for Musical Expression), Genova, held in June 2008, and of the eNTERFACE09 Summer Workshop on Multimodal Interfaces, Genova, held in July 2009.

Marcello Mortillaro received the PhD degree in the psychology of communication and linguistic processes with a thesis entitled "Multicomponential Analysis of Emotional Experience" from the Catholic University of Milan, Italy, in 2007. He works as a postdoctoral researcher at the Swiss Center for Affective Sciences from the University of Geneva, Switzerland. His research interests include vocal, facial, and bodily expression of emotions, as well as appraisal constituents of emotions. His work has a specific focus on the integration of multimodal nonverbal expressions, both in terms of production and recognition, and affective computing.

Klaus Scherer received the PhD degree in psychology from Harvard University, in 1970. He is the director of the Swiss Center for Affective Sciences and founder of GERG in the Department of Psychology at the University of Geneva. He has conducted many research programs, financed by granting agencies and private foundations in the USA, Germany, and Switzerland, directed at the study of cognitive evaluations of emotion-eliciting events and on facial and vocal emotion expression. He has reported this work in numerous publications in the form of monographs, contributed chapters, and papers in international journals. He has edited several collected volumes and handbooks and coedits the Affective Science Series for Oxford University Press. He was a founding coeditor of the journal *Emotion*. He is a member of several international scientific societies and a fellow of the American Psychological Association and the Acoustical Society of America. He has been elected a member of Academia Europea and an honorary foreign member of the American Academy of Arts and Sciences. He also pursues activities directed at the practical application of scientific research findings in industry, business, and public administration. He directs several long-term applied research programs in the area of organizational behavior and human resources, on psychological assessment, and on speech technology.