Towards Real-Time Affect Detection Based on Sample Entropy Analysis of Expressive Gesture

Donald Glowinski and Maurizio Mancini

InfoMus Lab, DIST, University of Genova, Italy
{donald.glowinski,maurizio.mancini}@dist.unige.it

Abstract. Aiming at providing a solid foundation to the creation of future affect detection applications in HCI, we propose to analyze human expressive gesture by computing movement Sample Entropy (SampEn). This method provides two main advantages: (i) it is adapted to the non-linearity and non-stationarity of human movement; (ii) it allows a fine-grain analysis of the information encoded in the movement features dynamics. A realtime application is presented, implementing the SampEn method. Preliminary results obtained by computing SampEn on two expressive features, smoothness and symmetry, are provided in a video available on the web.

Keywords: Expressive gesture, Sample Entropy, EyesWeb XMI.

1 Introduction

When do bodily channels convey actual information about a person's emotional state?

Research on nonverbal communication corroborate the view that bodily expressions constitute a relevant source of affective information [22,33,35]. In the last 10 years, an increasing number of affect detection systems have been developed and a great number of features characterizing an affective content have been proposed: movement direction and kinematics, arm extension and so on. As observed by Calvo [2], an assumption in Affective Computing is that emotions occur occasionally during usually affective-free interaction whereas contemporary theories maintain that affect is constantly influencing behavior. The challenge is to "model these perennially present, but somewhat subtle, manifestations of emotion" (p.32, [2]).

In this paper, we address this challenge by considering dynamic entropy of expressive features. Each movement potentially embodies an affect-related information content but this affect-related information content is subtly encoded in the temporal evolution of the movement features. For example, an upward movement may convey several affective meanings: when it follows a still posture, it may convey surprise; when it is a portion of many upward movements expressing anger, it may confirm the subject's angry state.

We develop a real-time affect detection system that is built on the *Sample Entropy* (*SampEn*) method. Tests have been conducted on a reduced amount of visual information related to human upper-body movements [13].

S. D´Mello et al. (Eds.): ACII 2011, Part I, LNCS 6974, pp. 527–537, 2011.

The paper is organized as follows: Section 2 introduces the key concepts of our approach; in Section 3 we detail a realtime application for computing dynamic entropy of movement expressive features; we conclude the paper in Section 4.

2 Background

2.1 Bounding Triangle

Our framework is based on a minimal and efficient representation of human upper body movement. We consider a *bounding triangle* related to the three blobs' centroids of the user's hands and head (see Figure 1). Recent studies [10,13] showed that this minimal representation of human upper body movement provide sufficient information to automatically distinguish between meaningful groups of emotions, related to the four quadrants of the Russel's valence/arousal space [27]. Previous evaluation of emotion recognition performance based on this minimal representation further assessed that human observers could discriminate between high and low arousal emotions [10].

By basing on this minimal user representation we ensure: (i) robust expressivity/emotional analysis; (ii) simplified identification of dynamics factors contributing to the communication of an expressive (e.g., emotional) content; (iii) real-time implementation of our framework.

2.2 Dynamic Expressive Behavior Analysis

Starting from the Kurtenbach and Hulteen's definition of gesture as "a movement of the body that contains information", a gesture is considered *expressive* if the information it carries has an expressive content, i.e., an "implicit message" [6].

Expressive content of a gesture can provide information on the emotional state, mood and personality of the person [34]. Researchers [16][9][34] have investigated human expressive motion and determined qualifiers such as slow/fast, small/large, weak/energetic, unpleasant/pleasant. Behavior expressivity has been correlated to energy in communication, to the relation between temporal/spatial characteristics of gestures, and/or to personality/emotion.

According to Camurri et al.'s framework [4], expressive gesture analysis is accomplished by three subsequent layers of processing: low-level physical measures (e.g., position, speed, acceleration of body parts); overall gesture features (e.g., motion fluency, impulsiveness); high-level information describing semantic properties of gestures (affect, emotion, attitudes).

A growing body of research in affecting computing and in psychology argue that temporal dynamics of human behavior (i.e., timing and duration of behavioral features) can be decisive in distinguishing between observed behavioral expressions [22][18]. [11] extended pilot studies by Castellano et al. [5] and defined a set of dynamics features, derived from the temporal profiles of expressive variations (e.g., the ratio between the gesture movement main peak duration and the total gesture duration, or the number of gesture movement local maxima in a given time span).

The system developed in [13] for characterizing expressive behavior showed that dynamic aspects of motion features are complementary to postural and gesture shape-related information.

2.3 Entropy Measure

The majority of the computational tools used for analyzing behavior dynamics [5] are based on traditional time and frequency domain measures but they fail to account for some properties of human movement: (i) non-linearity (small perturbations can cause large effects) and (ii) non-stationarity (the statistical properties change with time).

The concept of entropy has first been coined in thermodynamics referring to the amount of energy that is inaccessible for work (Shannon [30] applied the concept of information entropy to the development of information theory of communication. First entropy interpretation of human movement variability in terms of information theory constructs were proposed by Fitts to evaluate speed-accuracy task [8]. This classic investigation led to numerous studies examining motor behavior from an information processing perspective ([15] for a review). However, these studies have been often restricted to observing, for example, what happens in single points of a trajectory without considering dynamics, i.e., that consecutive points are part of the same trajectory.

In time series analysis, entropy has been newly defined as a quantity measuring the mean rate of new information production [25]. The Kolmogorov-Sinai (KS) entropy measures the decrease of uncertainty by knowing the current state of the system given its past history. Methods to estimate the K-S entropy were first developed in the field of nonlinear dynamic analysis and chaos [Lake, 16,20] by Grassberger and Procaccia [14], Eckmann and Ruelle [7].

The ApEn statistic and its last, most used modification, *SampEn* (*Sample Entropy*) method, was developed within this conceptual framework, respectively by Pincus [23] and Richman and Moorman [25], to compute the K-S entropy for real-world, noisy time series of finite length. High values of SampEn indicate disorder, smaller values indicate greater regularity. SampEn has been applied to a variety of physiological (heart rate, EMG, see [29] for a review).

Most recent application deal with behavioral data (e.g., investigating postural control mechanisms [24]) and some specifically address affective and social dynamics [12,17].

3 Realtime Implementation

Figure 1 shows the application we developed to perform realtime estimation of entropy in dynamic expressive movement features. The current implementation of our application does not perform any affect analysis: that is, according to the framework described by Camurri et al. [4] our application performs low-level and gesture-level measurements. The application has been implemented in the EyesWeb XMI platform (`http://www.eyesweb.org`), including the EyesWeb Gesture Processing Library for motion features extraction.

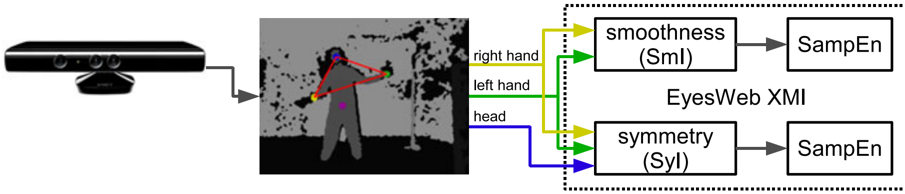

Fig. 1. The user's silhouette is captured through a Kinect acquisition device and body parts are extracted. Head and hands positions are sent to the EyesWeb XMI platform that computes the smoothness and symmetry SampEn.

3.1 Extracting Bounding Triangle Form Human Silhouette

We track the user's body using a Kinect acquisition device (`http://www.xbox.com/en-US/kinect`). Figure 2 shows and example of such tracking: on the left image the user's silhouette is extracted from the background; on the right image the user silhouette is segmented into body parts (e.g., head, shoulders, trunk and so on) by the functions provided by the Kinect open driver OpenNI [20].

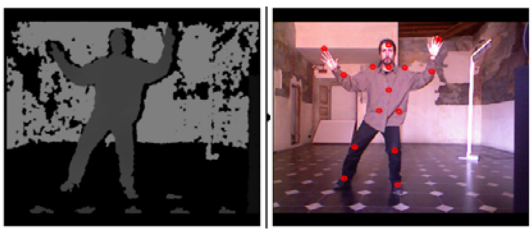

Fig. 2. Realtime tracking of user body configuration performed using Kinect

As the user enters the application space for the first time an initialization phase is required: the user must stand for about 3 seconds in the PSI pose (hands up, legs slightly opened).

Then the Kinect tracking is started and 3 user's body parts are considered: (1) head, (2) left hand and (3) right hand. Their 2D coordinates are extracted realtime and provided as input to the computation of dynamic features symmetry and smoothness, see Section 3.2. Finally, the dynamic features values as provided as input to the entropy estimation modules, described in Section 3.3.

3.2 Extracting Dynamic Features: Smoothness and Symmetry Index

As explained in Section 3, we focus movement analysis on a simplification of the human body: the bounding triangle, determined by the user's head as the top vertex and the user's hands as the triangle basis. An example of bounding triangle is shown in Figure 1.

Smoothness Index (SmI). Research in [32] demonstrates a correspondence between (i) smooth trajectories performed by human arms, (ii) minimization of the third-order derivative of the hand position (called *jerk* in physics) and (iii) correlation between hand trajectory curvature and velocity. In our work we use an approach similar to (iii) to determine if a trajectory is smooth or not starting from the trajectory curvature and velocity.

The left and right hand positions (x_l, y_l) and (x_r, y_r) are buffered in two time series consisting of 30 elements. These structures are managed as FIFO buffers, that is, as a new element is pushed in the structure, the oldest element pops out.

The time series are then provided as input to the smoothness computation algorithm: for every sample (x_h, y_h) (where h coiuld be l or r) in the buffer we compute curvature k and velocity v as:

$$k_h(x_h, y_h) = \left| \frac{{x_h}'{y_h}'' - {y_h}'{x_h}''}{({x_h}'^2 + {y_h}'^2)^{\frac{3}{2}}} \right| \qquad v_h(x_h, y_h) = \sqrt{{x_h}'^2 + {y_h}'^2} \tag{1}$$

where ${x_h}'$, ${y_h}'$, ${x_h}''$ and ${y_h}''$ are the first and second order derivatives of x_h and y_h for hand h. To compute derivatives values we apply a Savitzky-Golay filter [28] that provides as output both the filtered signal and an approximation of the $n-th$ order smoothed derivatives. As mentioned above, we define our algorithm for computing smoothness by taking inspiration from [32], that is, we compute correlation between trajectory curvature and velocity. We consider the Pearson correlation coefficient for two variables, that is, in our algorithm, $log(k_h)$ and $log(v_h)$:

$$\rho_h(k_h, v_h) = \frac{\sigma_{log(k_h),log(v_h)}}{\sigma_{log(k_h)}\sigma_{log(v_h)}} \tag{2}$$

However, k_h and v_h are computed over a relatively "short" time window, so we could approximate the covariance $\sigma_{log(k_h),log(v_h)}$ with 1, as the k_h and v_h variate (or not) approximately at the same time:

$$\rho'_h(k_h, v_h) = \frac{1}{\sigma_{log(k_h)}\sigma_{log(v_h)}} \tag{3}$$

That is, the Smoothness Index SmI_h for hand h is equal to $\rho'_h(k_h, v_h)$. As we compute these value for both the left and right hand, we finally compute the mean value of both hands SmI:

$$SmI = (SmI_l + SmI_r)/2 \tag{4}$$

Symmetry Index (SyI). Symmetry/asymmetry of emotion expression has been first studied in face expressions. Results revealed general hemisphere dominance in the control of emotional expression. A seminal work by [1] using static pictures of emotional expressions with one side of the face replaced by the mirror image of the other (*chimeric face stimuli*) showed that left hemiface is further related to expressivity. Roether et al. recently showed that human gait display lateral asymmetries also in human emotional full-body movement [26]. Motion capture data of twenty four actors recorded during neutral walking and emotionally expressive walking (anger, happiness, sadness) showed that the left body side moved 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 symmetric puppets.

A few studies accounted for the relationship between upper-body movements symmetry and 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 [19].

Spatial hands symmetry is computed with respect to the vertical axis and with respect to the horizontal axis. Horizontal Symmetry Index ($SyI_{horizontal}$) 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 (Eq 5).

$$SyI_{horizontal} = \frac{||x_B - x_L| - |x_B - x_R||}{|x_R - x_L|} \tag{5}$$

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. Similarly, vertical Symmetry Index ($SyI_{vertical}$) is computed by the difference between the y coordinates of hands. A first measure related to spatial symmetry (SyI) results from the ratio of the measures of horizontal and vertical symmetries (Eq 6).

$$SyI = \frac{SyI_{horizontal}}{SyI_{vertical}} \tag{6}$$

Dynamic update of features. Movement features such as symmetry and smoothness are dynamical features, as explained in [3]: their updated values do not only depend on the last frame of the user's movement data (i.e., the last position of the user's head and hands) but it should also consider the recent user movement *history*. We include such dynamic properties by performing an incremental update of movement features, as shown in Figure 3.

To do that, we store the features values at previous times $t-1$ and $t-2$: $SyI(t-1), SyI(t-2), SmI(t-1), SmI(t-2)$. At time t, we compute the *detected* movement features values $SyI_{det}(t)$ and $SmI_{det}(t)$, as explained above. Finally we update movement features values by *weighting* the detected values by the difference between the current and previous feature values:

$$SyI(t) = SyI(t-1) + ((SyI_{det}(t) - SyI(t-1)) * |SyI(t-1) - SyI(t-2)|) \tag{7}$$

$$SmI(t) = SmI(t-1) + ((SmI_{det}(t) - SmI(t-1)) * |SmI(t-1) - SmI(t-2)|) \tag{8}$$

In our application, the dynamically updated values of SyI(t) and SmI(t) are used to compute Sample Entropy as described in the following Section.

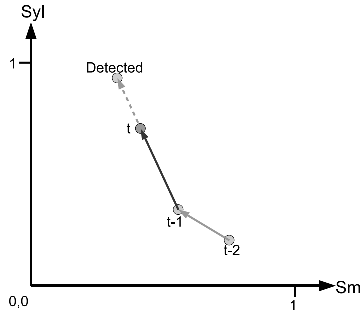

Fig. 3. Dynamic update of movement features

3.3 Computing SampEn

The following algorithm for computing SampEn is introduced in [25].

Given a standardized one-dimensional discrete time series of length N, $X = \{x_1, ..., x_i..., x_N\}$:

1. construct vectors of length m (similarly to the time delay embedding procedure) [31,21],

$$u_i(m) = \{x_i, ..., x_{i+m-1}\}, 1 \leq i \leq N - m \tag{9}$$

2. compute the correlation sum $U_i^m(r)$ to estimate similar subsequences (or *template vectors*) of length m within the time series:

$$U_i^m(r) = \frac{1}{(N-m-1)} \sum_{i=1, i \neq j}^{N-m} \Theta(r - \parallel u_i(m) - u_j(m) \parallel_\infty) \tag{10}$$

where $u_i(m)$ and $u_j(m)$ are the template vectors of length m formed from the standardized time series, at time i and j respectively, N is the number of samples in the time series, r is the tolerance (or *radius*), Θ is the Heaviside function, and $\|\|_\infty$ is the maximum norm defined by $\parallel u_i(m) - u_j(m) \parallel_\infty) = max_{0 \leq k \leq m-1} \mid x_{j+k} - x_{i+k} \mid$.

3. calculate the average of U_i^m, i.e., the probability that two vectors will match in the m-dimensional reconstructed state space

$$U^m(r) = \frac{1}{(N-m)} \sum_{i=1}^{N-m} U_i^m(r) \tag{11}$$

4. set $m = m + 1$ and repeat steps 1-4
5. calculate the sample entropy of X_n

$$SampEn(X_n, m, r) = -ln \frac{U^{m+1}(r)}{U^m(r)} \tag{12}$$

Sample Entropy computes the negative average natural logarithm of the conditional probability that subsequences similar for m points in the time series remain similar (as defined by Eq. 3) when one more point $(m + 1)$ is added to those sequences. Small values of SampEn indicate regularity.

In the proposed implementation, we compute SampEn on two time series containing the values of the two dynamic features SmI and SyI, that is, given:

$$SmI_{ts} = \{SmI(t - N), ..., SmI(t)\}, SyI_{ts} = \{SyI(t - N), ..., SyI(t)\} \quad (13)$$

we compute:

$$SampEn(SmI_{ts}, m, r), SampEn(SyI_{ts}, m, r) \quad (14)$$

3.4 Output

An example of the output provided by our application is provided in Figure 4. A demo video is available at:
`http://www.mauriziomancini.org/downloads/acii2011.m4v`

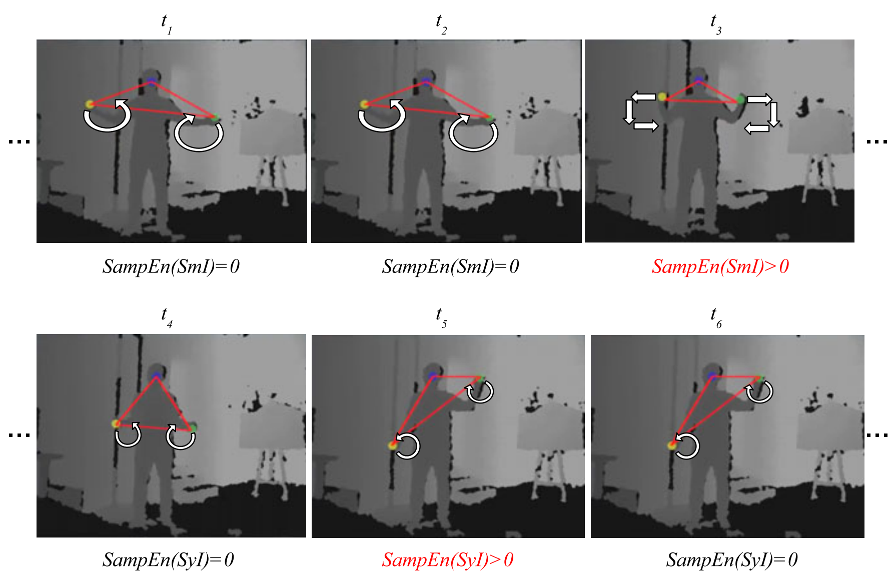

Fig. 4. An example of SampEn computation: user's smoothness is constant (high) between t_1 and t_2, so $SampEn(SmI)$ is zero; then, between t_2 and t_3 smoothness decreases, so $SampEn(SmI)$ increases; user's symmetry decreases between t_4 and t_5, so $SampEn(SyI)$ increases; finally, symmetry is constant (low) between t_5 and t_6, so $SampEn(SyI)$ is zero

4 Conclusion

Our research work aims to contribute to develop affective detection applications. To provide solid foundations for such applications, we suggest that expressive gesture analysis should consider the information content conveyed by movement features such as smoothness and symmetry. We also propose to compute Sample Entropy, a measure adapted to the non-linear dynamics of human movement.

Future work includes: (a) applying the proposed approach to realtime classification of emotion portrayals to refine the evaluation in terms of recognition performance, error analysis and real time aspects; (b) extension to 3D analysis (using the Kinect device) to include other significant information, such as distance and forward or backward movement with respect to camera or person.

Acknowledgements. This work is partly supported by the EU FP7 project MIROR (the project is co-funded by the European Community under the Information and Communication Technologies (ICT) theme of the Seventh Framework Programme, grant agreement 258338).

References

1. Borod, J., Haywood, C., Koff, E.: Neuropsychological aspects of facial asymmetry during emotional expression: A review of the normal adult literature. Neuropsychology Review 7(1), 41–60 (1997)
2. Calvo, R., D'Mello, S.: Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Transactions on Affective Computing 1(1) (2010)
3. Camurri, A., Ferrentino, P.: Interactive environments for music and multimedia. Multimedia Systems 7(1), 32–47 (1999)
4. Camurri, A., Volpe, G., De Poli, G., Leman, M.: Communicating Expressiveness and Affect in Multimodal Interactive Systems. IEEE Multimedia, 43–53 (2005)
5. Castellano, G., Mortillaro, M., Camurri, A., Volpe, G., Scherer, K.: Automated Analysis of Body Movement in Emotionally Expressive Piano Performances. Music Perception 26(2), 103–119 (2008)
6. Douglas-Cowie, E., Campbell, N., Cowie, R., Roach, P.: Emotional speech: Towards a new generation of databases. Speech Communication 40(1-2), 33–60 (2003)
7. Eckmann, J., Ruelle, D.: Ergodic theory of chaos and strange attractors. Reviews of Modern Physics 57(3), 617–656 (1985)
8. Fitts, P.: The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology 47(6), 381–391 (1954)
9. Gallaher, P.E.: Individual differences in nonverbal behavior: Dimensions of style. Journal of Personality and Social Psychology 63(1), 133–145 (1992)
10. Glowinski, D., Bracco, F., Chiorri, C., Atkinson, A., Coletta, P., Camurri, A.: An investigation of the minimal visual cues required to recognize emotions from human upper-body movements. In: Proceedings of ACM International Conference on Multimodal Interfaces (ICMI), Workshop on Affective Interaction in Natural Environments (AFFINE). ACM, New York (2008)
11. Glowinski, D., Camurri, A., Volpe, G., Dael, N., Scherer, K.: Technique for automatic emotion recognition by body gesture analysis. In: Computer Vision and Pattern Recognition 2008. CVPR Workshops. IEEE Computer Society, Los Alamitos (2008)
12. Glowinski, D., Coletta, P., Volpe, G., Camurri, A., Chiorri, C., Schenone, A.: Multiscale entropy analysis of dominance in social creative activities. In: Proceedings of the International Conference on Multimedia, pp. 1035–1038. ACM, New York (2010)

13. Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., Scherer, K.: Towards a minimal representation of affective gestures. IEEE Transactions on Affective Computing (99), 1–13 (2011)
14. Grassberger, P., Procaccia, I.: Measuring the strangeness of strange attractors. Physica D: Nonlinear Phenomena 9(1-2), 189–208 (1983)
15. Hong, S., Newell, K.: Entropy conservation in the control of human action. Nonlinear Dynamics, Psychology, and Life Sciences 12(2), 163 (2008)
16. Johansson, G.: Visual perception of biological motion and a model for its analysis. Perception and Psychophysics 14, 201–211 (1973)
17. Kim, J., Andre, E.: Four-Channel Biosignal Analysis and Feature Extraction for Automatic Emotion Recognition. In: Biomedical Engineering Systems and Technologies, pp. 265–277 (2009)
18. Kleinsmith, A., Bianchi-Berthouze, N., Steed, A.: Automatic recognition of non-acted affective postures. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics (99), 1–12 (2011)
19. Mehrabian, A.: Nonverbal Communication. Aldine (2007)
20. OpenNI, `http://www.openni.org`
21. Packard, N., Crutchfield, J., Farmer, J., Shaw, R.: Geometry from a time series. Physical Review Letters 45(9), 712–716 (1980)
22. Pantic, M., Pentland, A., Nijholt, A., Huang, T.: Human Computing and Machine Understanding of Human Behavior: A Survey. In: Huang, T.S., Nijholt, A., Pantic, M., Pentland, A. (eds.) ICMI/IJCAI Workshops 2007. LNCS (LNAI), vol. 4451, pp. 47–71. Springer, Heidelberg (2007)
23. Pincus, S.: Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences of the United States of America 88(6), 2297 (1991)
24. Ramdani, S., Seigle, B., Lagarde, J., Bouchara, F., Bernard, P.: On the use of sample entropy to analyze human postural sway data. Medical Engineering & Physics 31(8), 1023–1031 (2009)
25. Richman, J., Moorman, J.: Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology- Heart and Circulatory Physiology 278(6), H2039 (2000)
26. Roether, C., Omlor, L., Giese, M.: Lateral asymmetry of bodily emotion expression. Current Biology 18(8), 329–330 (2008)
27. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology 39(6), 1161–1178 (1980)
28. Savitzky, A., Golay, M.J.E.: Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 36(8), 1627–1639 (1964)
29. Seely, A., Macklem, P.: Complex systems and the technology of variability analysis. Crit Care 8(6), R367–R384 (2004)
30. Shannon, C., Weaver, W.: The mathematical theory of information (1949)
31. Takens, F.: Detecting strange attractors in turbulence. In: Dynamical Systems and Turbulence, Warwick 1980 pp. 366–381 (1980)

32. Todorov, E., Jordan, M.I.: Smoothness maximization along a predefined path accurately predicts the speed profiles of complex arm movements. Journal of Neurophysiology 80(2), 696–714 (1998)
33. Vinciarelli, A., Pantic, M., Bourlard, H.: Social signals, their function, and automatic analysis: a survey. In: Proceedings of the 10th International Conference on Multimodal Interfaces, pp. 61–68. ACM, New York (2008)
34. Wallbott, H.G., Scherer, K.R.: Cues and channels in emotion recognition. Journal of Personality and Social Psychology 51(4), 690–699 (1986)
35. Wallbott, H.: Bodily expression of emotion. Eur. J. Soc. Psychol. 28, 879–896 (1998)