PUBLICATION P4

Ilmonen, Tommi and Takala, Tapio. Detecting Emotional Content from the Motion of an Orchestra Conductor. Gesture in Human-Computer Interaction and Simulation, 6th International Gesture Workshop, Lecture Notes in Artificial Intelligence 3881, pages 292–295, Springer Berlin / Heidelberg, 2006

Detecting Emotional Content from the Motion of an Orchestra Conductor

Tommi Ilmonen and Tapio Takala

Helsinki University of Technology,
Telecommunications Software and Multimedia Laboratory
Firstname.Lastname@tml.hut.fi

Abstract. In this paper we present methods for analysis of the emotional content of human movement. We have studied orchestra conductor's movements that portrayed different emotional states. Using signal processing tools and artificial neural networks we were able to determine the emotional state intended by the conductor. The test set included various musical contexts with different tempos, dynamics and nuances in the data set. Some context changes do not disturb the system while other changes cause severe performance losses. The system demonstrates that for some conductors the intended emotional content of these movements can be detected with our methods.

1 Introduction and Background

Emotionally aware computer systems are a new and interesting research field. Emotion detection is a crucial part of making a system that can take emotions into account. Especially facial expression and voice have been tackled by researchers. These modalities have also been combined for more robust gesture recognition [1]. Even mice have been made emotion-aware. Physiological signals can also be used to track the emotional state of a person [2]. Kang has published a system that tracks the emotional state of a video stream [3].

In contrast little research has been published that would use hand or body motion as the starting point. Drosopoulos has published a system that aims to detect emotions based on the gestures that a person makes [4].

Movements usually convey more than emotions, there is a context for them. A unique feature of our research is that context changes were included in the tests. In these tests the context changes were presented by change of musical parameters; dynamics, tempo and character.

2 Collecting and Processing Data

Emotions cannot be measured directly. In these tests we asked a person to manifest a feeling. Conducting brings another layer of expression parallel with the emotional content — musical content and context. Nuances (staccato, legato)

S. Gibet, N. Courty, and J.-F. Kamp (Eds.): GW 2005, LNAI 3881, pp. 292–295, 2006.

and dynamics (piano, forte) also affect the motion. The musical environment is the context of the motion.

The conductors wore a data suit that contained magnetic motion tracking sensors. The conductor was asked to conduct short a passage of music without a band with given emotional content. The emotional and musical content changed during the passage. Musical parameters were also varied and expressed simultaneously, resulting in superposition of parameters. Variables were dynamics (piano / forte) and character (staccato / no character / legato). Passages were performed in four tempos (about 56, 80, 120 and 160 beats per minute). Since the emotions also varied the number of permutations of all parameters gets easily very large. The data set used in analysis contains about 20 takes per conductor with each take containing four combinations of parameters. Since we want a separate test set a few extra takes were also recorded — duplicating some of the parameter combinations of the first data set.

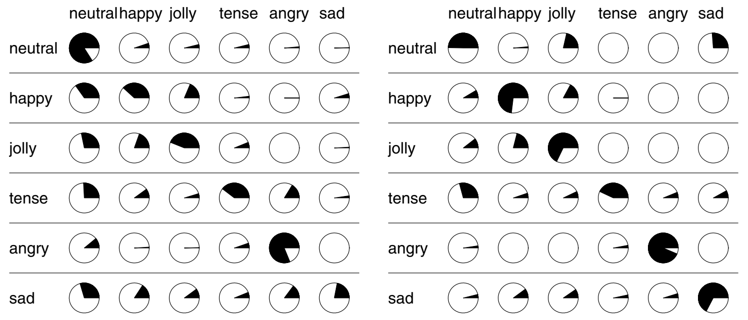
The process by which humans detect the emotional content of motion is not known. At any rate the form of the motion must play a role in the recognition process. To somehow mimic this process we calculate parameters (features) from the motion. This calculation acts as preprocessing for ANNs. The preprocessing must transform the absolute motion curves to more general and abstract parameters. These parameters should not be affected by tempo, nuances or dynamics. We tested various preprocessing methods. We used the Cartesian coordinate position and rotation metrics as the basic information. These parameters were then transformed to velocity, acceleration and curvature spectrograms, histograms and filterbank outputs. We used artificial neural networks (ANNs) as analysis tool. To produce the figures in this paper, only self-organizeng maps (SOM) were used.

3 Results

In general, we found that the system can detect emotions implied by hand movements. The performance of the system depends heavily on the conductor. In these tests only two conductors participated.Results in this paper were obtained with the conductor that was easier for the system to analyze.

It was found that the characters legato and staccato confused the ANNs greatly. Since we found no way to fix the problem these characteristics were left totally out of the analysis. As a result the parameters that were varied in the test set were dynamics, tempo and emotion. We were only interested in emotion — the other parameters were varied to represent changing context. The effect of tempo range used was briefly tested. When the system only needs to analyze tempo 80 beats per minute it performs much better than when given the full range of tempos (see figure 1).

A widely used analysis approach is to classify the emotions to a fixed number of emotions e.g. categorization of emotions to N mutually exclusive classes. In this case we used neural networks that had N outputs — one output per emotion. The output with largest activation is then assumed to represent the emotion. Based on this one can calculate the confusion matrix that indicates how well



(a) The confusion matrix when a wide range of tempos is used

(b) The confusion matrix when the tempo is fixed to 80 beats per minute

Fig. 1. Confusion matrices in various situations

the analysis worked (figure 1(a)). In the matrix the title of each row indicates the intended emotion and the pie charts on the row indicate how the system interpreted the motion samples corresponding to that emotion. In the ideal case the diagonal elements of the matrix would be one and all others zero. By using random choice, one would get 17 percent (one out of six) of choices correct. In figure 1(a) the ratio of correct choices is between 20 and 80%.

By considering the emotional space as a low-dimensional continuum one can drop the number of dimensions of the emotional space. We used two-dimensional space illustrated in figure 2. The same figure shows how emotions we used were mapped to the space. We chose locations that appeared feasible. The ANN was trained to create an output vector that is correctly located in this emotional space. Figure 3 displays results obtained with one ANN. In the optimal case the ANN output of each emotion would be clustered to match the positions given in figure 2.

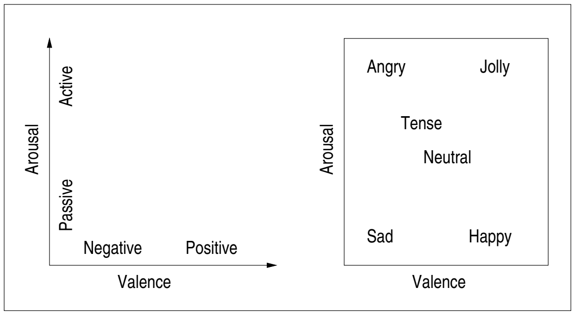


Fig. 2. Definition of arousal/valence space and positions of emotions used in the space

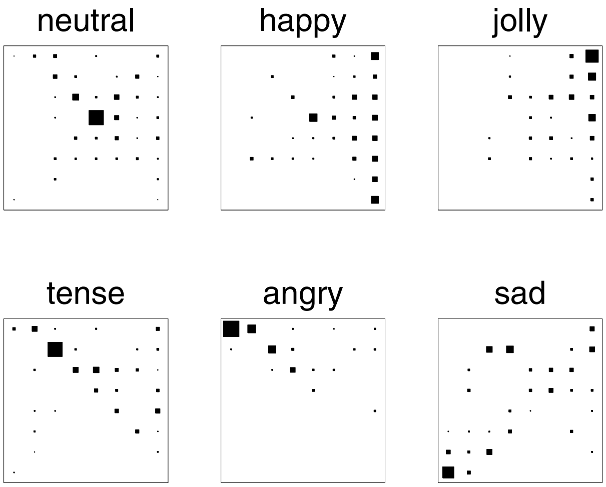


Fig. 3. Distribution of ANN-estimations in arousal/valence space

4 Conclusions and Future Work

These are first results on a new field of study — how to determine the emotional content of human motion. As a qualitative result we can state that the gestural manifestation of emotions can be detected with computers. With only two test persons little can be said about how the system might work in the general case.

References

1. Chen, L., Tao, H., Huang, T., Miyasato, T., Nakatsu, R.: Emotion recognition from audiovisual information. In: IEEE Second Workshop on Multimedia Signal Processing. (1998) 83–88
2. Healey, J., Picard, R.: Digital processing of affective signals. In: Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing. Volume 6. (1998) 3749–3752
3. Kang, H.B.: Affective content detection using hmms. In: MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia, ACM Press (2003) 259–262
4. Drosopoulos, A., Balomenos, T., Ioannou, S., Karpouzis, K., Kollias, S.: Emotionally-rich man-machine interaction based on ges-ture analysis. In: Universal Access in HCI: Inclusive Design in the Information Society, Crete, Greece (2003) 1372–1376