

feature. Using a 'window' of N values, we extracted the minimum, the maximum, the mean, the standard deviation, the skewness (measure of the asymmetry of the data around the sample mean) and the kurtosis (measure of how outlier-prone a distribution is) of each feature.

B. Heart Beat

The ECG signal is the manifestation of contractile activity of the heart. It can be used to measure heart rate (HR), inter-beat intervals (IBI) and to determine the heart rate variability (HRV). In our project, we just used the heart rate that was extracted from the ECG signal as a component for emotion recognition.

C. Live Images of Face

Live face image analysis has also been considered as a complementary means for emotion recognition. There are two main approaches for feature extraction from images: holistic methods and geometric methods. We are using a geometric method proposed by Cootes [3]. The method is mostly known under acronym ASM (Active Shape Model). Its basic idea is as follows: ASM is a statistical model which is able to deform to fit to a similar shape as the ones who were presented during training period (see Figure 2).

The main problem in ASM is the localization of the points of the model. There were several enhancement since the ASM was introduced. We would like to implement the LBP (Local binary patterns) to improve the fitting of the model.

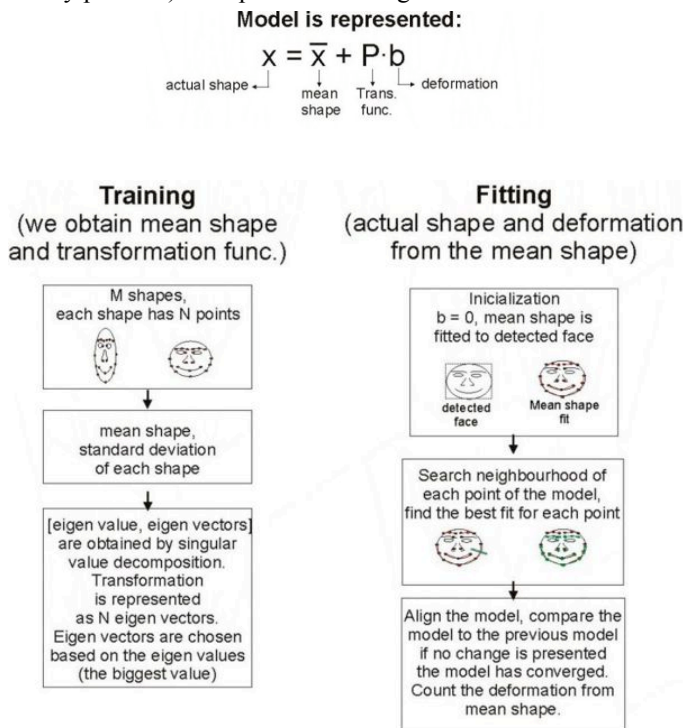


Figure 2 Live face emotion analysis

III. REAL-TIME SYNCHRONIZATION OF THE FEATURES

For the synchronization and sensor management, a kind of server will receive the information from the all sensors. This server will able to calibrate and normalize the values in different feature vectors and record them. Afterwards, it resends the normalized and also the raw value, if requested to each feature extraction process. It is better to send as many values as possible, to put more information into emotion recognition unit (see Figure 3).

In order to synchronize the process, a time mark is sent each t milliseconds. When this time mark is received within each extracted feature, two actions can be proposed:

- if the process works in real time, only the current response value is sent
- if the process is short enough, first the process is completed then the response value is sent

The server will wait for each process to respond (The mark time is sent back to check the synchronization), then send the data to the neural network in order to do the emotions classification.

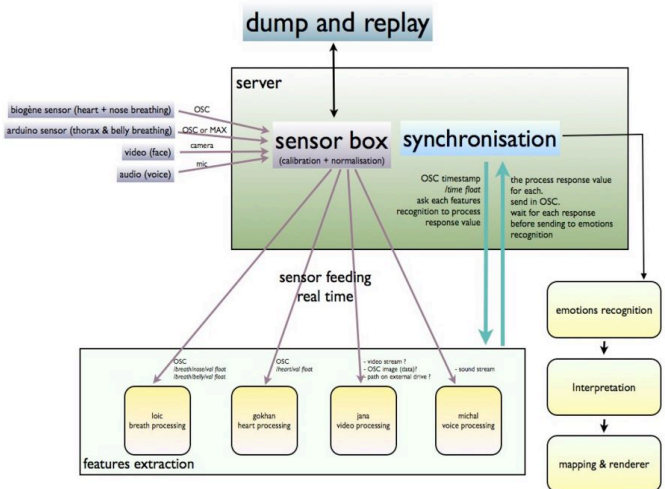
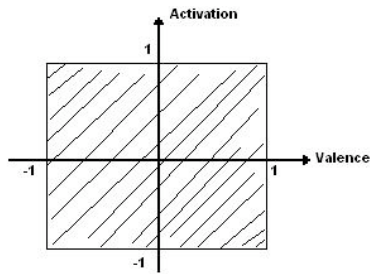


Figure 3 Feature extraction

IV. CLASSIFICATION OF THE FEATURES INTO EMOTIONS

After the extraction and the synchronization of the feature vectors from heart beat, breath and the facial images, classification among the emotions is made. It is not like labeling the feature vectors with an emotion but it is more like mapping a feature vector to the following 2 dimensional emotion space:



In order to map our N-dimensional feature vector to the two dimensional emotion space, Stuttgart Neural Network Simulator modified for Java (JavaNNS) is used. JavaNNS is a simulator for artificial neural networks, i.e. computational models inspired by biological neural networks. It enables the user to use predefined networks or create your own, to train and to analyze them.

The designed network that takes physiological features as input and gives the corresponding emotion point (A,V), is first trained with a labeled training set. After the training is completed, the model is exported out as a C++ function that can easily be mounted into the system.

Unfortunately, there are many difficulties in the process. One of the challenging points here is to extract right features that contain the distinguishing data on emotions and to label these feature vectors correctly. In addition to that, the right structure for the network has to be built. Back-propagation method for training is proven to be fast and accurate enough in most cases and thus it can be used in real time processing. For activation and output functions of the input and output nodes and the hidden nodes, many comprehensive tests and trials have to be conducted with the real data.

V.INTERPRETATION OF SENSOR DATA

Emotion is a complex phenomenon that no conceptual theory describes at a sufficient level at the moment. It involves percepts as well as cognition (concepts), conation (impulse) & affects (feelings, physiological systems). Everyday language provides a manageable simplification of concepts & experiences.

Emotional life is a large and complex space that we can't represent completely, because of technical, conceptual and experimental limitations.

- Technical: we need to be aware of what the multimodal inputs we have are capable of describing. We realize we have few parameters and they are superficial.
- Conceptual: are we interested in basic emotions that statistically occur more often (scientific approach), or are we going to fit subjective needs (artistic approach) ?
- Experimental: there is no obvious link, and we don't have knowledge to link measurements to the inner emotion of the human being sensed.

A subject found in a measured state equivalent to a state in which he was before does not mean that he is going through

the same particular emotion. Inversely, being in a particular emotional state could result in a comparable measure, even though it is not necessarily the case. Anyway, it would require one step more of analysis to link the representation to an inner emotion.

So, what is represented and what is not? Are we recognizing emotions, are we recognizing concepts, or are we recognizing face expressions + sensor features?

With an artistic approach, it is possible to rely on the feedback of the multimedia result on the performer's perception to create a link between capitation and actual inner emotion. If expressivity is placed at a sufficient level to allow this loop to exist, the performer addresses directly to the multimedia content that includes things we wouldn't normally accede, because the performer isn't even conscious about. Because it is explicit, this is the only space where we know where we are.

Thus, we need a system of representation with emphasis on experiential description.

Intuitively, we proposed to use the Activity / Valence representation, which we assume to be a broad enough concept to fit the view of several artists, or at least to be of interest for them. This representation is not describing the context (environment, cognition, affect...) but taking it into account. Emotional life is divided into 4 poles, 4 poetic worlds that the artists describe with general sentences. An example of a poetic world mapped to sound generators is given in the following figure 4.

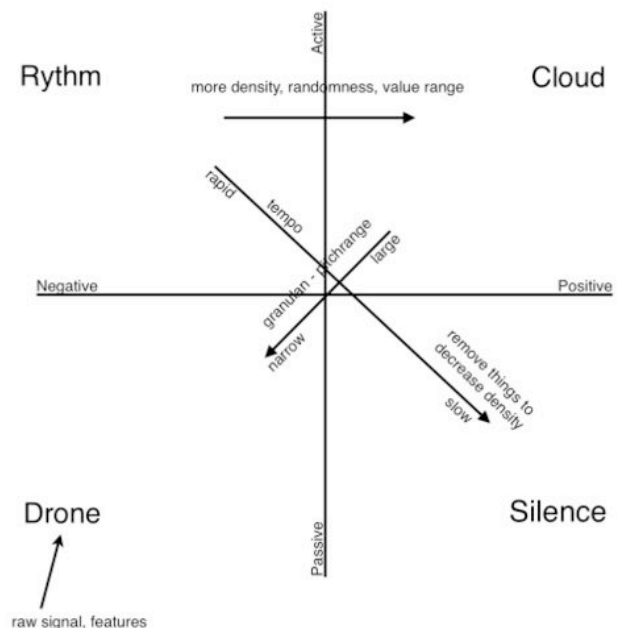


Figure 4 Poles of emotional life

VI.ARTISTIC WORK

We now present a few artistic applications that we made during the course of the workshop.

A. Sound Generation

Breathing and Thoracic Volume: For breathing we used the sensor signal to modulate a filtered noise. We used a white noise and a band-pass filter with a moderate coefficient of resonance. The signal from the breath temperature sensor was connected to both the amplitude of the output signal and the center frequency of the filter. We apply a flanging effect at the output of the filter in order to give coloration to the resulting sound. For the thoracic volume we used a similar method but with a voiced source as signal input to feed the filter.

A toolbox for Heart beat sonification: We designed a toolbox for Heartbeat sonification using two different methods. These methods have different levels of flexibility, in particular, concerning spectral transformation. The first system uses complex resonances. Using parallel resonance filter bank based on the CNMAT tools, we made a module which allows us to synthesize an arbitrary resonance model and to control and modify several high-level parameters. The sonic interaction designer can control and modify the number of resonances and the lower resonance frequency. This model can be modified by changing other parameters at a high-level control as the global resonance and the spectral amplitude slope. A second module was built to use sampling techniques. This module use a polyphonic player that allows playing recorded sounds by triggering it and avoids clicks due to a re-triggering when it's already playing a sound, by allocating a new playing voice to it. Modifiable parameters are speed (pitch modified), time of beginning for reading into the sound file and parameters of a linear amplitude envelope defined by fade in time and fade out time.

B. Live performances based on input signals

The point is to create a sound-space worked by the variation of the emotions felt during the act of performing. Michal Osowski wishes to build an « emotional curve » from analyzing the video of the performance and then to program automatic sounds on the computer that happen in relation to the variations worked by the curve.

The performance is shot by the camera of Ivan Chabanaud. Fabienne Gotusso creates the performance choosing as a point of departure to the improvisation to play with the sound of the German word « Holzwege » (meaning forest path) and the English word « to die ». The aim is to create some estrangement from the meaning of both words and then to liberate the games and associations of meaning produced by the pronunciation of the words relating to improvised acts. She chooses to improvise in a sandy path with a tree standing by. Then Michal and Fabienne manage to extract some range of emotions by watching the video of the performance to finally order them as a map.

Fabienne dresses in a black suit on which several captors are fixed. She begins to move after several necessary preliminary measures in order to identify her skeleton. Skeleton that will appear at the time on the computer screen as she is acting the performance. The skeleton reproduces in real time the motions composed by Fabienne. The inscription and architecture of the

motions are plainly faithful to the quality involved in the dancing. It is impossible not to think about E J Marey's works.

VII. CONCLUSION

Even though this workshop has not allowed the participants to design a full working system from signal capture to emotion recognition and reactive digital art, it has however had numerous positive achievements:

- through regular seminars, participants have discovered new methods and new techniques in signal analysis, machine learning, computer graphics and sound, live performance, and affective computing,
- through joint developments, the researchers have gained a better insight on the requirements of physiological signal analysis and interpretation,
- through human experience and joint collaboration, the participants have discovered shared scientific interests and defined new perspectives for future collaborative works.

Acknowledgement:

Part of this research was supported by the Grant Agency of Academy of Sciences of the Czech Republic, project No. 1ET101470416 .

REFERENCES

- [1] T. Jehan, A. Freed, and R. Dudas. "Musical applications of new iter extensions to Max/MSP", Proceedings of the International Conference on Computer Music, pages 504-507, 1999.
- [2] R. Cowie, "Emotional life: Terminological and conceptual clarifications", HUMAINE - D3i, 2006 <http://emotion-research.net/projects/humaine/deliverables/D3i%20final.pdf>
- [3] Cootes, T. F., Edwards, G. J., and Taylor, C. J. 2001. Active Appearance Models. IEEE Trans. Pattern Anal. Mach. Intell. 23, 6 (Jun. 2001), 681-685. DOI= <http://dx.doi.org/10.1109/34.927467>