# Facial Expression Recognition as a Creative Interface

**Roberto Valenti[1], Alejandro Jaimes[2], Nicu Sebe[1],**
[1]University of Amsterdam, [2]IDIAP Research Institute
{rvalenti,nicu}@science.uva.nl; alex.jaimes@idiap.ch

## ABSTRACT
We present an audiovisual creativity tool that automatically recognizes facial expressions in real time, producing sounds in combination with images. The facial expression recognition component detects and tracks a face and outputs a feature vector of motions of specific locations in the face. The feature vector is used as input to a Bayesian network which classifies facial expressions into several categories (e.g., angry, disgusted, happy, etc.). The classification results are used along with the feature vector to generate a combination of sounds and images that change in real time depending on the person's facial expressions. We explain the basic components of our tool and several possible applications in the arts (performance, installation) and medical domains.

#### Author Keywords
Affective, multimodal, interface, sonification, facial therapy interface, gesture-based interaction.

#### ACM Classification Keywords
H5.2. User Interfaces: Auditory (non-speech) feedback.

#### INTRODUCTION
Computer vision can be used to facilitate unobtrusive, natural, and rich interaction in a variety of applications. One of the most exciting of these is new media art, in which vision can play a major role in performance and interactive installations. At the same time, in recent years the analysis of emotional signals has taken on great importance as researchers have realized that emotions form an important part of communication between humans and between humans and machines.

In this paper, we present a system that combines our interests in both, in an application that recognizes facial expressions in real time and generates sounds and image combinations. In our setup a person's face is captured by a camera. Our system uses a model based non-rigid face tracking algorithm to extract facial motion features (motion units) that serve as input to a Bayesian network classifier used for recognizing different facial expressions. The output of the feature detector and classifier are used to generate sounds and to combine images whose parameters vary depending on the facial expressions.

The system we have developed, which we describe below, can be used for creative activities in the arts or for medical applications (e.g., facial physiotherapy [1], among others).

**Related Work.** The closest work we are aware of was presented by Funk et. al. [1]. Our system differs on the following aspects: (1) we classify facial expressions rather than only detect changes in particular facial regions; (2) in contrast to using 7 face regions [1], we extract 12 motion units; and (3) we use a non-rigid face tracking algorithm instead of a face detector. In practice, these differences mean that our system could be used reliably with a wearable camera (e.g., in a performance) and we could create a richer set of outputs by combining expression classification with the motion unit values.

#### FACIAL EXPRESSION RECOGNITION
Ekman and Friesen [3] developed the Facial Action Coding System (FACS) to manually code, following prescribed rules, facial expressions where movements on the face are described by a set of action units (AUs) which roughly correspond to muscles. The inputs are still images of facial expressions, often at the peak of the expression.

Most automatic methods [4, 5] are based on Ekman's work and extract features from images or video and use them as inputs to a classifier. Although the classification is of facial expressions and not emotions (one can feel angry and smile), the output is one of a set of pre-selected "basic" emotion categories (happiness, surprise, fear, disgust, sadness, and anger). Most automatic approaches to recognize facial expressions differ mainly in the features extracted and in the classifiers used to distinguish between the different facial expressions.

Our system (described in detail in [2]) tracks 12 facial motion units in the following categories (fig. 1): vertical movement of the lips, horizontal movement of the mouth corners, vertical movement of the mouth corners, vertical movement of the eyebrows, lifting of the cheeks, and blinking of the eyes. The 7 facial expressions that the system classifies are: neutral, happy, angry, disgusted, afraid, sad, and surprised.



IUI-08, January 13–16, 2008, Canary Islands, Spain.

**Figure 1.** Our system's interface (top) and the 12 motion units extracted. Each facial expression is assigned a probability value [0,1]. Arrows show correlations between motion units, learned when facial expressions are classified.

## SONIFICATION

Several sonification approaches exist. For the sake of simplicity, we distinguish only three types frequently used: *direct sonification* (also referred to as audification), *parameter mapping*, and *model-based sonification*. In the first type, raw time-series data is mapped to amplitude and other attributes so the data itself becomes the waveform (after scaling and filtering if needed). In the second type, properties of the data are used only to set parameters in a sound waveform (e.g, pitch of a sine wave can be set dynamically based on features computed from the data, for example, the pitch can be changed only if the acceleration of a motion unit is greater than a threshold *t*), and in the third type specific models are built to produce sounds for a particular data set and interaction scenario (e.g., see [7]).

The tool we have developed lends itself for all three types and we are currently experimenting with different combinations. In particular, we have found the Pure Data [6] environment suitable for interactively testing different direct, and parameter mapping sonifications. For example, at setup time we interactively tweak various parameters that remain fixed for a particular "session" and simply use the data for each of the 12 motion units to generate sounds. Each motion unit is mapped to a waveform with specific parameters (e.g., pitch, etc.), and the waveform's amplitude is varied according to the motion of the unit as a person interacts with the system in real time. The probability measures [0,1] for each of the recognized facial expression categories are used to set the musical scale (e.g., high scale for expressions such as happy).

The images, which are shown in combination with the generated sounds, depend on pre-determined labels assigned to them that subjectively match the facial expression being recognized (e.g., "happy" or "sad" images).

The face is reliably tracked and motion units are extracted with sufficient accuracy. In spite of this, however, we find that most people cannot accurately control some areas of the face and learning to do so could be a challenge. Interestingly enough, as pointed out in [1], this can actually have great benefits in medical applications, or in particular types of performance art in which the performers have strong control of different facial muscles. Applications of our system include performance (for instance in Butoh dance), meditation (a person could use the system and hear sounds matching her "mood"), physical therapy feedback, artistic creation, and others.

## CONCLUSIONS AND FUTURE WORK

The current implementation allows anyone to effectively experiment with facial movements and expressions to create combinations of sounds and images. In the future we plan to perform user studies to better determine how to map the parameters extracted, understand how people interact with the system, explore further sonification approaches, and combine the outputs with automatic audio and image analysis algorithms (e.g., match "happy" expressions to "happy" sounds from large music collections).

## ACKNOWLEDGMENTS

The authors would like to thank Michael J. Lyons for his valuable comments. The work of R. Valenti and N. Sebe was supported by the MIAUCE European Project and the work of A. Jaimes by the Swiss National Science Foundation, through the National Center of Competence in Research (NCCR) on "Interactive Multimodal Information Management (IM2)", http://www.im2.ch and as well as by the European PASCAL Network of Excellence.

## REFERENCES

1. M. Funk, K. Kuwabara, and M. J. Lyons, "Sonification of Facial Actions for Musical Expression", in proc., *Intl. Conference on New Interfaces for Musical Expression (NIME05)*, Vancouver, BC, Canada, June, 2005.
2. R. Valenti, N. Sebe, T. Gevers, "Facial Expression Recognition: A Fully Integrated Approach," in proc. *ICIAP 2007*, Modena, Italy, September, 2007.
3. P. Ekman and W. Friesen. *Facial Action Coding System: Investigator's Guide*. Consulting Psychologists Press, 1978.
4. M. Pantic and L. Rothkrantz. "Automatic analysis of facial expressions: The state of the art," *IEEE Transactions on PAMI*, 22(12):1424–1445, 2000.
5. B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognition*, 36:259–275, 2000.
6. Puredata (http://www.puredata.org)
7. Y. Visell and J.R. Cooperstock, "Modeling and Continuous Sonification of Affordances for Gesture-Based Interfaces," *13th Intl. Conf. on Auditory Display*, Montreal, Canada, June 2007.