Sonify Your Face: Facial Expressions for Sound Generation

Roberto Valenti[1], Alejandro Jaimes[2] and Nicu Sebe[3]

[1]Intelligent Systems Lab Amsterdam, University of Amsterdam, The Netherlands.
[2]Yahoo! Research, Barcelona, Spain
[3]Department of Information Engineering and Computer Science, University of Trento, Italy.

rvalenti@science.uva.nl; ajaimes@yahoo-inc.com; sebe@disi.unitn.it.

ABSTRACT

We present a novel visual creativity tool that automatically recognizes facial expressions and tracks facial muscle movements in real time to produce sounds. The facial expression recognition module 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 that change in real time depending on the person's facial expressions. We explain the artistic motivation behind the work, the basic components of our tool, and possible applications in the arts (performance, installation) and in the medical domain. Finally, we report on the experience of approximately 25 users of our system at a conference demonstration session, of 9 participants in a pilot study to assess the system's usability, and discuss our experience installing the work at an important digital arts festival (RE-NEW 2009).

Categories and Subject Descriptors

J.5 [Computer Applications]: Arts and Humanities - fine arts
I.2 [Artificial Intelligence]: Vision and Scene Understanding - Modeling and recovery of physical attributes

General Terms

Algorithms, Measurement, Design, Experimentation, Human Factors.

Keywords

Affective computing, multimodal interface, sonification, facial therapy interface, gesture-based interaction, facial expressions.

1. INTRODUCTION

Facial expressions are an important channel of nonverbal communication. Many animal species display facial expressions, but expressions are highly developed particularly in the primates, and perhaps most of all, in humans. Even though the human species has acquired the powerful capabilities of a verbal language, the role of facial expressions in person-to-person interactions remains substantial. Messages of the face that provide commentary and illustration about verbal communications are significant in themselves[1].

Facial expressions play a major role in the arts, from sculpture and painting to performance, traditional and avant-garde. In some cultures, particularly in the East, facial expressions are intricately linked with performance themes and styles. In Kathakali, a traditional dance from the Indian state of Kerala, facial expressions are emphasized with colorful makeup. In contrast, in Japanese Butoh dance, white make up is used to emphasize facial and corporal expressions. Both dances, although very different in style place particular importance on facial expressions through small, subtle, often slow movements. Performers train to severely limit their eye blinking and to control their facial muscles.

Figure 1. Kathakali makeup (top left), Butoh (top right), sculpture by Ron Mueck, and Gorgon sculpture (ancient Greece).

In sculpture, since ancient times, facial expressions have played a major role and they continue to do so. Expressions

[1]http://www.face-and-emotion.com/dataface/expression/expression.jsp

are tightly linked to feelings, thus they help us communicate emotions (some examples in **Figure 1**).

In spite of the importance of facial expressions, most people are unaware of many of their facial muscles. Professional actors and performers train to develop the skills to control important facial muscles and to emphasize non-verbal aspects of communication.

These ideas form the motivation of our work. We wanted to build a system that could be used in different settings, and that brought attention to the importance of facial expressions by emphasizing their communicative power in non-explicit ways and in creative settings. The motivation can be further described as follows:

- Performance: inspired by Kathakali and Butoh, we wanted to create a tool that could be used in a performance that had a strong focus on facial expressions.
- Installation: we wanted a tool that could be used to create an interactive art installation that encouraged participants to think about facial expressions, their facial muscles, and more importantly, to have a different experience communicating with their face (**Figure 2**).
- Play: we were not interested in creating a musical instrument, but rather a playful interface that while bringing attention to the issues described, would be fun to play with.

Computer vision can play a major role in performance and interactive installations and 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 machines.

The tool we present in this paper combines our interests in Computer Vision and art. The tool recognizes facial expressions in real time and generates sounds combining the recognized facial expressions and the motion of particular parts of the face. In our setup a person's face is captured by a camera (**Figure 2**, bottom). 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 whose parameters vary depending on the facial expressions.

An interactive art installation using the tool was recently presented at RE-NEW 2009[2]. In the setup depicted in **Figure 2**, the idea is to have 4 people facing each other and generating sounds by modifying their facial expressions. As stated above, the idea behind this is to create a different communicative experience that brings attention to facial muscles.

Figure 2. Installation at RE-NEW 2009. In the top image four people interact with the system. The installation included a high quality professional sound mixer and professional speakers.

In addition to applications in the arts, the tool we have developed can be used for medical applications (e.g., facial physiotherapy [4], among others).

1.1. Related Work

Research related to the work presented can be grouped in two areas: facial expression recognition and sonification.

Research in facial expression recognition spans several decades [11], but has gained practical interest recently for affective computing applications. In virtual reality environments and animation films, for example, synthesized facial expressions modeled on automatically recognized human facial expressions can be more effective. Facial expression recognition can also be useful in communicating with computers: information kiosk systems, for instance, could use such techniques to communicate more effectively (e.g., notice an angry customer and connect to a human operator).

[2] http://re-new.org/

Research on sonification is comparatively recent, but the field is growing quickly, with many application areas. For example, sonification can be very effective in discovering patterns in interactively analyzing very large volumes of data. It has also been used to understand brain signals and to train users of brain-machine interfaces. In addition, there are many applications for performance art, interactive installations, and in creating new musical instruments—our main area of interest.

Many researchers have studied approaches to sonify gestures (hand, body, etc.) and, in particular, several works have been presented for creating music using the face. The authors of [6][7], for instance, use eye movements for music generation, while the authors of [13][8] use the mouth.

The closest body of work we are aware of was presented by Funk et. al. [4]. 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 [4], 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. In [15] we presented a brief description of our system. In this paper we expand the technical details and report on the experience of users during the presentation of the demo described in [14] at the International Conference of Intelligent User Interfaces, 2008.

The main contributions of this paper can be summarized as follows:

- it is the first work, to the best of our knowledge, to perform sonification of facial expression recognition and
- it is a first attempt to report on the experience of non-expert users of a facial expression sonification system.

The rest of the paper is organized as follows. In section 2 we describe our facial expression recognition system. Section 3 gives an overview of sonification techniques and explains the methods we use. In section 4 we discuss user feedback, discuss applications in section 5 and conclude in section 6.

2. FACIAL EXPRESSION RECOGNITION

Ekman and Friesen [1] 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 [11][3] 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 of emotions (one can feel angry and smile), the output is typically 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 [1]) 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. Here we only summarize the technical details and we direct the interested reader to the original contribution.

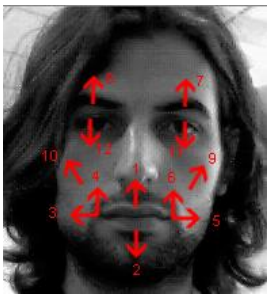


Figure 3. The 12 extracted motion units. Each facial expression is assigned a probability value [0,1].

The face tracking use in our system is based on a system developed by Tao and Huang [18] called the piecewise Bezier volume deformation (PBVD) tracker. The face tracker uses a model-based approach where an explicit 3D wireframe model of the face is constructed (see **Figure 3**).

A generic face model is warped to fit the detected facial features. The face model consists of 16 surface patches embedded in Bezier volumes. The surface patches defined this way are guaranteed to be continuous and smooth.

Once the model is constructed and fitted, head motion and local deformations of the facial features such as the eyebrows, eyelids, and mouth can be tracked. First the 2D image motions are measured using template matching between frames at different resolutions. Image templates from the previous frame and from the very first frame are both used for more robust tracking. The measured 2D image motions are modeled as projections of the true 3D motions onto the image plane. From the 2D motions of many points on the mesh, the 3D motion can be estimated by solving an overdetermined system of equations of the projective motions in the least squared sense.

The recovered motions are represented in terms of magnitudes of predefined motion of various facial features. Each feature motion corresponds to a simple deformation on the face, defined in terms of the Bezier volume control

parameters. We refer to these motions vectors as motion-units (MU's). Note that they are similar but not equivalent to Ekman's AU's, and are numeric in nature, representing not only the activation of a facial region, but also the direction and intensity of the motion. The MU's used in the face tracker are shown in **Figure 4**.

The MU's are used as the basic features for the classification scheme.

For the classification, we notice that the facial motion features are correlated and consequently learning the dependencies among the facial motion units could potentially improve classification performance, and could provide insights as to the "structure" of the face, in terms of strong or weak dependencies between the different regions of the face, when subjects display facial expressions.

For the sonification system we decided to learn a Tree-augmented Bayesian Network (TAN) classifier. In the TAN classifier structure the class node has no parents and each feature has as parents the class node and at most one other feature, such that the result is a tree structure for the features. Two examples of the learned correlations between the MUs are shown in Figure 2.

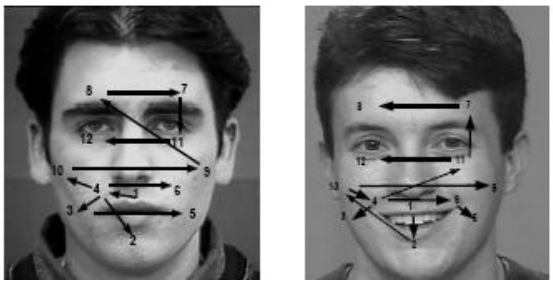


Figure 4. The learned correlations between motion units when facial expressions are classified.

The 7 facial expressions that the system classifies are: neutral, happy, angry, disgusted, afraid, sad, and surprised.

3. 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 [1]).

The tool we have developed lends itself for all three types and we are experimenting with different combinations. In particular, we have found the Pure Data (Pd) [12] environment suitable for interactively testing different direct, and parameter mapping sonifications.

As explained in section 4, there isn't a single sonification setup that is appropriate for every type of usage scenario. Our goal, therefore, is to experiment with different setups and customize them according to the particular application (section 5). Next we describe the setups we have tested.

3.1. Features

Our facial expression recognition module (FERM) is written in C++ and connects to the Pd environment via a network socket. By "listening" to a port, the system is able to either continuously read data from the FERM or sample it at a desired rate. In the current FERM implementation we obtain a 19 dimension feature vector every 1/25 of a second. The feature vector, as described in the previous section, contains the following values (see **Figure 4**):

- f_1: vertical movement of *upper* lips
- f_2: vertical movement of *lower* lips
- f_3: *horizontal* movement of left mouth corner
- f_4: *vertical* movement of left mouth corner
- f_5: *horizontal* movement of right mouth corner
- f_6: *vertical* movement of right mouth corner
- f_7: vertical movement of right eyebrow
- f_7: vertical movement of left eyebrow
- f_8: vertical movement of right cheek
- f_9: vertical movement of left cheek
- f_{10}: vertical movement of left cheek
- f_{11}: vertical movement of right eyelid
- f_{12}: vertical movement of left eyelid
- f_{13}: neutral
- f_{14}: happy
- f_{15}: angry
- f_{16}: disgusted
- f_{17}: afraid
- f_{18}: sad
- f_{19}: surprised

Features f_1 to f_{12} produce a real number between [-1.0, 1.0] and features f_{13} to f_{19} produce a number between [0.0, 1.0]. Note that all displacements are relative to the face (not absolute in image coordinates).

3.2. Graphical User Interface

The current system consists of the FERM and the sonification module (in Pd) and each has its own user interface (**Figure 5**).

The FERM can receive input from a camera, a video file, or a still image and can also record video of the interaction. Once the user has selected the input mode, the system

attempts to locate the face. In the center of the picture captured from the camera the interface shows four lines depicting a square: the user places his face inside the square and once the face is detected the system indicates it by overlaying a light blue mask on it and showing a "face found" label. The user then presses the emotion button, the mask disappears and is replaced by a mask of points overlaid on the face (**Figure 5**). The facial expression category labels on the right show the probability of each expression (see appendix).

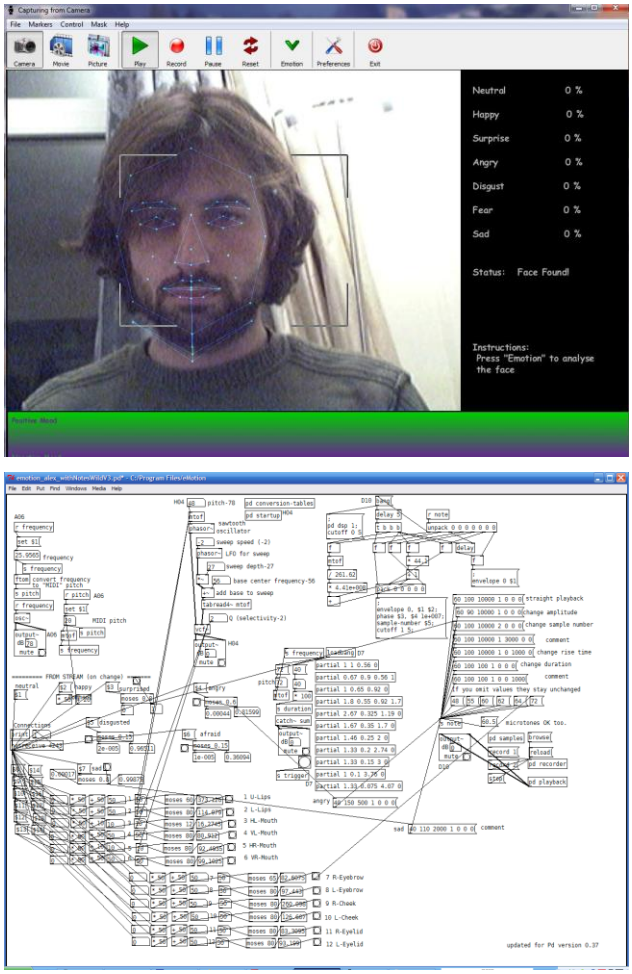


Figure 5. The FERM graphical user interface (top) and the Pd patch used in our system (bottom).

Once the recognition process starts, the points overlaid on the face move as the face moves and start blinking if the face is "lost" by the face tracker. As shown by the images in the appendix, the system is robust enough to allow the user to step back from the original position by about 60 cm.

The FERM GUI provides enough visual cues for the user to know what the system is recognizing at any given time and whether it is properly tracking the face.

The Pd patch (and corresponding interface, bottom of **Figure 5**) has 4 main components and for each the user can manually control the volume and other parameters. In the demonstration sessions described in the next section, at setup time we interactively tweak various parameters to show the user what sounds are generated for the different features. The Pd patch has several bang and message patches that show when certain sounds are activated and the values used, and each of the four components below has its own volume control. These are useful in explaining the functionality to users and in the experimentation in generating different sounds.

3.3. Feature Mapping

The Pd patch, in its basic configuration has four main sound generation components:

- A cosine wave oscillator.
- A sweeping filter.
- A sampler that allows interactive loading of sound files.
- Additive synthesis.

The basic mapping configuration is performed as follows:

- Oscillator: pitch is modified by the happiness feature (f_{14})
- A sweeping filter: selectivity is modified by the horizontal movement of the mouth (f_3, f_5), and pitch by the vertical movement of the lips (f_1, f_2).
- A sampler: our patch allows us to interactively load a sound file and modify various parameters such as pitch, duration, rise time, etc. We selected three parameter settings and associate them with the features sad (f_{18}), and angry (f_{15})
- Additive synthesis: generates a bell-type of sound that is triggered by the feature surprise (f_{19})

The rest of the features are mapped in similar ways and all values are properly scaled according to the parameters they modify.

As described earlier in this section, our system uses direct and parameter mapping. The pitch of the oscillator, for example, is modified constantly (every 1/25 of a second) as the probability of "happiness" changes (more happiness, higher pitch). This can be considered a form of direct mapping since the values of the happiness probability directly change the sound waveform. For some of the other features, (e.g., surprise), changes in their probabilities do not have a direct affect on the sounds heard. Instead, different thresholds of surprise generate different tones (so if surprise is below a certain threshold, no sound is generated; for happiness there is constant sound unless the value of the happiness probability is zero). Although direct and parameter mapping could be considered equivalent in this case we find the distinction important in designing the interaction. For some features it makes more sense if

sounds are heard when they reach certain levels, whereas for others hearing constant changes seems more suitable.

With the basic mapping just described we are experimenting with some ideas on how to sonify facial expressions. Tying high happiness probability with high pitch, for example, seems intuitive, as is generating bell-like sounds when there is surprise. In spite of this, it is difficult to determine what might be the "best" mapping, particularly for features that have no particular meaning outside the context of a performance or culture: if the mappings just described for happiness and surprise seem intuitive, what might be intuitive mappings for cheek movements?

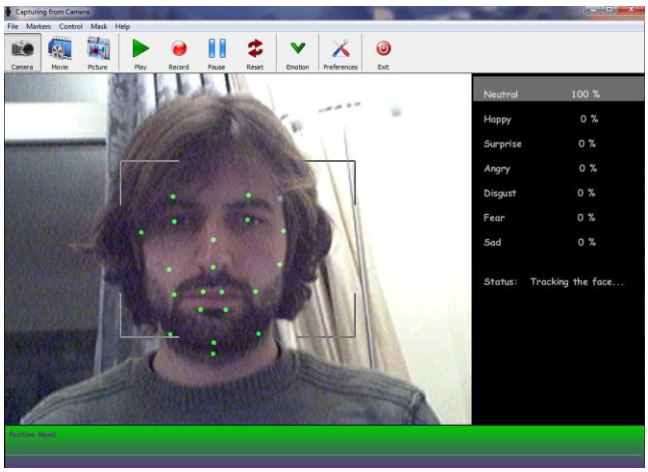


Figure 6. FERM interface during a session. Note the probabilities given to different facial expression categories in the appendix.

4. USER FEEDBACK AND DISCUSSION

We have performed two pilot studies to collect feedback on the design of the system. The first one consisted of informal interviews, and in the second one we asked the set of questions proposed by Nielsen as attributes of usability [16]. In both cases the participants were individual users (i.e., the setup consisted of a single computer).

4.1. User Feedback

At the IUI conference demonstration session [14] around 25 persons tried the system (between 5-10 minutes per person). The participants, who were attendees to the conference, ranged in age from early 20s to late 50s (with approximately 50-50 men to women ratio, around 70% graduate students or young professors and the rest senior persons). There are no significant differences in the performance of the FERM for users with glasses or facial hair (details on person-dependent and person-independent tests of the FERM only are given in [1]). Below we summarize some of the verbal feedback that we collected at the IUI conference:

- Most people liked the system, but a few did not like the sounds that it produced. Some suggestions were made to have several sound libraries so the same functionality could be tried with sound sets chosen by the particular user to make the system more fun (e.g., generate animal sounds, bells, string sounds, etc.). This made it clear that ideally, the particular sounds produced should be customized according to personal taste.
- Younger persons in the group were keener to use the system and found it "cool." They also adapted themselves more easily and saw the potential of the application.
- Several people found it difficult to understand what sounds were being generated by each facial expression and/or facial movement. As a result, during the demonstrations we "switched off" all of the sounds, and switched them on one at a time to make it easier to understand which movements mapped to which sounds.
- Some people wanted to be able to play particular tunes (e.g., how do I play happy birthday?).
- Some participants felt uncomfortable with the idea of a computer recognizing their "emotions"[3] and did not really understand why anyone would be interested in generating sounds this way as opposed to playing an instrument. One particular user (female, 72 years old) stated that a traditional instrument (she mentioned a piano) was more expressive because it involved touch, but agreed on the benefits of the technology for physical therapy and applications for persons with disabilities.
- Some participants commented that they would have liked to see a mapping of sounds that corresponded to the emotions (i.e., generating happy sounds when happy or grave sounds when sad). Although these ideas were considered and are part of the system (e.g., high happiness probability produces high pitch), these mappings are not as obvious as some participants would have liked.
- It was also pointed out that it would be desirable for the system to be less sensitive to small expression changes and for the sounds to "hold" for a period of time.

[3] Many people failed to make a distinction between emotions and facial expressions. In other words, we found that most people are unlikely to think of the difference and do not realize the system does not recognize emotions.

In general, the system raised a lot of interest, and most people had a lot of fun with it. However, it is important to mention that in some cases the interest shifted from the sounds being generated to the technical details of how the facial expression recognition component works.

In a second evaluation, done in 2010, 9 people used the system. For this second evaluation we took into account the feedback received in previous sessions and modified the sound outputs by including sound samples (e.g., of water, bells, etc.) and making the sound transitions more subtle.

We asked the participants the following questions, letting them answer on a scale of 1 (bad) to 7 (good) based on Nielsen's attributes of usability [16]:

1. It was easy to learn to use this system
2. I believe I became creative quickly using this system
3. I feel comfortable using this system
4. It was easy to understand which actions generated sounds
5. I enjoyed using the system

We outline the results of this second evaluation (Table 1):

Table 1. Results of the questionnaire used in the second evaluation. Nine subjects participated and rated the system on a scale from 1(bad) to 7(good): Ea: Ease, Cr: Creativity, C: Comfort, U: Understanding, and E: Enjoyment.

Subject	Ea	Cr	C	U	E	AVG
1	5	4	5	5	6	5
2	6	5	5	6	**7**	5.8
3	6	3	6	3	5	4.6
4	6	6	5	5	**7**	5.8
5	4	4	**7**	4	6	5
6	5	4	5	5	6	5
7	4	5	5	4	5	4.6
8	6	5	6	4	**7**	5.6
9	4	4	**7**	4	**7**	5.2
AVG	5.1	4.4	5.7	4.4	6.2	
STDV	0.9	0.9	0.9	0.9	0.8	

In addition to the questionnaire, participants provided verbal feedback. The general feeling is that people enjoyed (as in "funny") making sounds with their faces. None of the participants seemed to really have control of the interface.

Results are also shown in Figure 7. Enjoyment had the largest average score, while creativity and understanding had the lowest. It is "easy" to explain these results: in the version of sound mapping used in this second study, the sound chnages are more sublte than in the first version. This means participants had a harder time understanding the mapping between specific facial movements and sounds. At the same time, this was likely to have caused a bit of frustration as participants felt they could not really create much since they did not feel they were really in control.

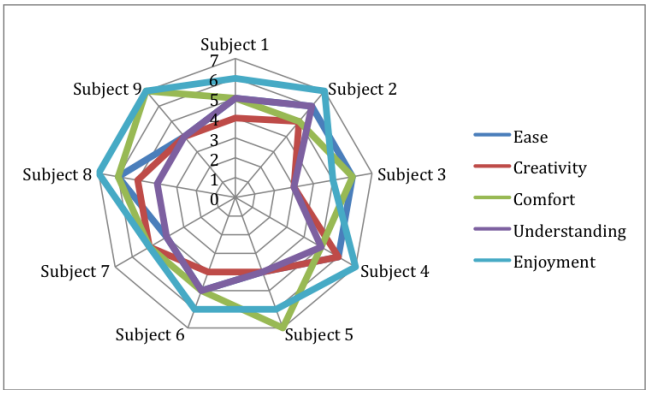


Figure 7. Results of the questionnaire used in the second evaluation.

Comments at the festival where the second sound mapping was also used were pretty much in line with the general comments received.

4.1. Observations

Based on the user comments we can make the following observations:

- In developing a musical interface of this type, the usage scenario and type of user is important and should determine how the sonification is done and the types of sounds produced. Building a generic system for non-experts would require significant work on designing a natural and effective interface that is easy to understand and allows sufficient changes to the sound generation (the authors of [4] created an interface that could be easier for general users).
- The system is easy to use in the sense that anyone can "play" it to generate sounds. As with a real instrument, however, training is required, not only to understand the sound mappings, but also to learn to control different facial muscles (e.g., most people can only blink with one eye, or have little or no control of certain facial areas).

The feedback received raised several important questions and closely relates to observations made in the literature [1][5][10].

5. APPLICATIONS

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 [4], 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 and Khatakali dance), meditation (a person could use the system and hear sounds matching her "mood"), physical therapy feedback, artistic creation, and others.

6. CONCLUSIONS AND FUTURE WORK

We have presented a novel visual creativity tool that automatically recognizes facial expressions and tracks facial muscle movements in real time to produce sounds. The system allows anyone to experiment with facial movements and facial expressions for musical expression. In the future we plan to address many of the comments obtained from the sessions in which non-experts have used our system. Future work includes further investigation into how to map the parameters extracted, 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). We also plan to use our system in installation and performance settings.

ACKNOWLEDGMENTS

The authors would like to thank Michael Lyons for his valuable comments in the initial version of the paper.

[1] C. Dobrian, and D. Koppelman, "The 'E' in NIME: Musical Expression with New Computer Interfaces," NIME 2006, Paris, France.

[2] P. Ekman and W. Friesen. Facial Action Coding System: Investigator's Guide. Consulting Psychologists Press, 1978.

[3] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," Pattern Recognition, 36:259–275, 2000.

[4] 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.

[5] M. Gurevich, J. Trevino, "Expression and Its Discontents: Toward and Ecology of Musical Creation," NIME 2007.

[6] A.J. Hornof, T. Rogers, and T. Halverson, "EyeMusic: Performing live music and multimedia compositions with eye movements," NIME 2007: Conference on New Interfaces for Musical Expression. In proceedings on pp. 299-300.

[7] J. Kim, G. Schiemer, and T. Narushima, "Coulog: Playing with Eye Movements", NIME 2007, New York City, USA.

[8] M. J. Lyons, M. Haehnel, "Designing, Playing, and Performing with a Vision-based Mouth Interface," NIME 03, Montreal, Canada.

[9] N. Moody, N. Fells, and N. Bailey, "Ashitaka: an audiovisual Instrument," NIME 2007.

[10] S. Oore, "Learning Advanced Skills on New Instruments," NIME 2005.

[11] M. Pantic and L. Rothkrantz. "Automatic analysis of facial expressions: The state of the art," IEEE Transactions on PAMI, 22(12):1424–1445, 2000.

[12] Puredata (http://www.puredata.org)

[13] G. C. de Silva, T. Smyth, M. J. Lyons, "A Novel Face-tracking Mouth Controller and its Application to Interacting with Bioacoustic Models," NIME 2004.

[14] R. Valenti, A. Jaimes and N. Sebe, "Facial Expression Recognition As a Creative Interface". In ACM International Conference on Intelligent User Interfaces, 2008.

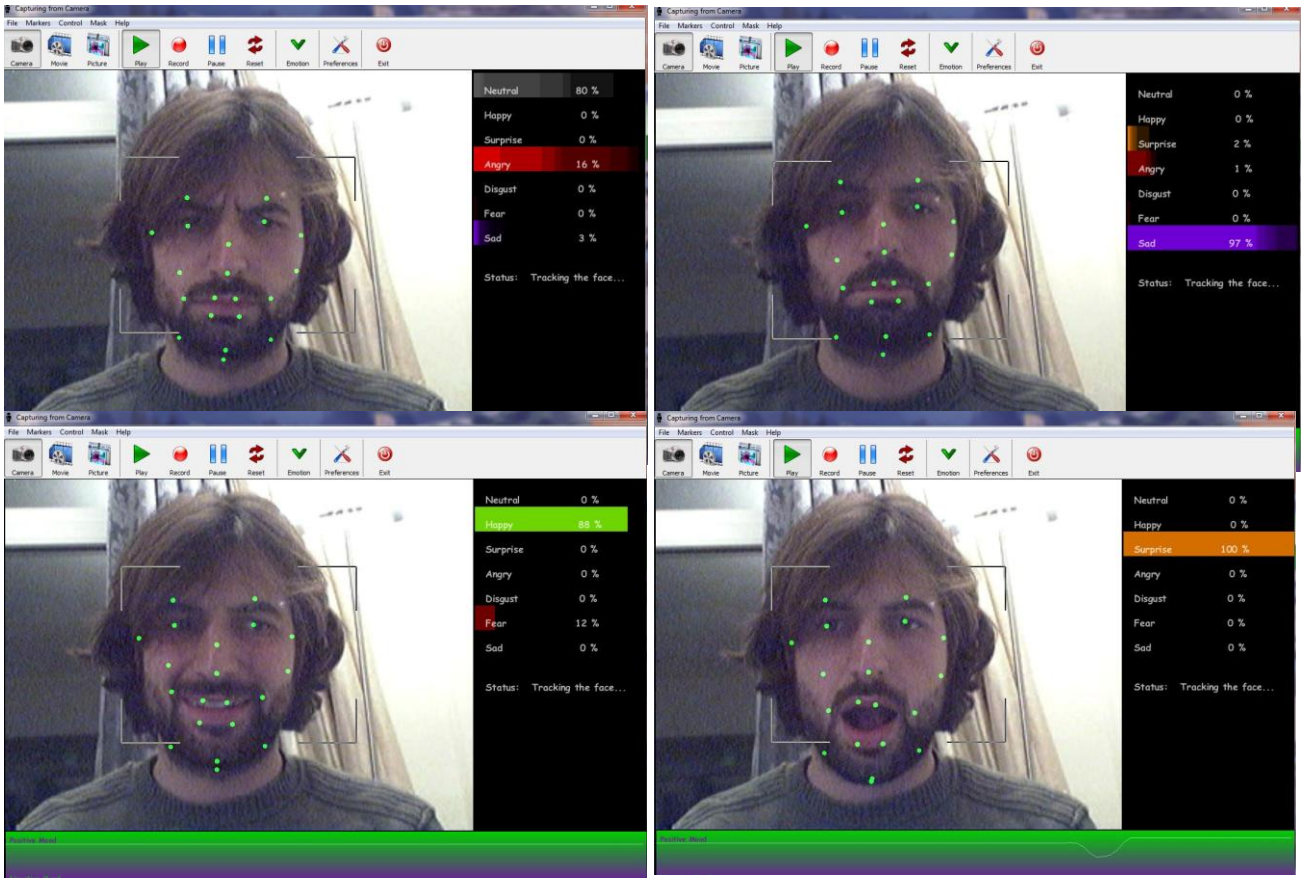
[15] R. Valenti, A. Jaimes and N. Sebe, "Facial Expression Recognition As a Creative Interface". In ACM International Conference on Intelligent User Interfaces, 2008.

[16] J. Nielsen, Usability Engineering. Academic Press, 1999.

[17] 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.

[18] H. Tao and T.S. Huang, "Connected Vibrations: A Modal Analysis Approach to Non-rigid Motion Tracking," IEEE Conf. on Computer Vision and Pattern Recognition, 1998.

APPENDIX



The system is better at recognizing some expressions such as happy and surprise and less accurate at others such as disgust.