REAL-TIME SMILE SONIFICATION USING SURFACE EMG SIGNAL AND THE EVALUATION OF ITS USABILITY

Yuki Nakayama[1] *Yuji Takano*[2] *Masaki Matsubara*[3] *Kenji Suzuki*[4] *Hiroko Terasawa*[3,5]

[1]Graduate School of Library, Information and Media Studies
[2]Graduate School of Systems and Information Engineering
[3]Faculty of Library, Information and Media Science
[4]Faculty of Engineering, Information and Systems
[5]JST PRESTO
[1,2,3,4]University of Tsukuba
[1,3]1-2, Kasuga, Tsukuba, Ibaraki 305-8550 Japan
yuki@slis.tsukuba.ac.jp

ABSTRACT

We propose a real-time, interactive system for smile recognition and sonification using surface electromyography (EMG) signals. When a user smiles, a sound is played. The surface EMG signal is mapped to pitch using a conventional scale. The timbre of the sound is a synthetic sound mimicking bubbles.

In the experiment, eight participants evaluated the effects of smile-based sonification feedback. Participants expressed smiles in a condition that there was feedback or no feedback. We investigated what type of effects the feedback had on smiling by analyzing surface EMG signals and interviewing the subjects.

The results suggest that the sonified feedback could facilitate the expression of spontaneous smiles. In addition, results suggested that both the attack and release of a smile were similarly perceived using visual or auditory feedback.

1. INTRODUCTION

Smiles are one of the most basic facial expressions exhibited by humans and are generally associated with positive emotions, such as pleasure, amusement, and enjoyment. People smile in order to convey a kind attitude toward others. A smile can be spontaneous or intentional, with a wide range of underlying emotions and intentions, from pure joy or trust to sarcasm. A smile is not only a facial expression but also a powerful social tool.

Physically, a smile is a facial gesture that involves lifting the corners of the mouth upward. However, the manner in which one delivers this gesture is highly dynamic, personal, and situational. For example, a person can smile in a slow and subtle manner, or in a fast and fluttering manner. Such a variation of muscular motion results in a wide range of smile-based expressions.

In this study, we propose a real-time system for smile detection and sonification using surface electromyography (EMG). This system has the following characteristics: (1) the variety of muscular motions in smiles becomes audible, and (2) the sound is pleasant and entertaining, so that users are motivated to produce more smiles.

The potential applications of such a system are vast, monitoring one's own and other's smiles is useful to represent emotions. In this study, we sonify one's own smile with EMG and provides it back for self-monitoring in order to enhance or augment smile production. However, in this paper, we focus primarily on the development of the system, as well as user evaluation of the system.

This work is licensed under Creative Commons Attribution - Non Commercial 4.0 International License. The full terms of the License are available at http://creativecommons.org/licenses/by-nc/4.0

2. BACKGROUND

2.1. Emotions

Facial expressions, including smiles, convey our emotions. Sonification of facial expression could function as medium for the emotional communication. It can also augment emotional contagion. We are particularly interested in smile sonification because we value positive emotions represented by smiles.

There are two notable theories for the cognition of emotions, namely the "category theory" and the "dimension theory".

The category theory suggests that people around the world express the six basic emotions of enjoyment, anger, sadness, fear, disgust, and surprise [1]. Furthermore, facial expressions associated with the six basic emotions are recognized regardless of nationality and cultural background. According to the category theory, smiles are generally considered as expressions of enjoyment.

The dimension theory considers emotions as a continuous change on a coordinate axis. Russell proposed a circumplex model of affect, which shows a position of feelings using adjectives on the coordinate axes, where the horizontal dimension is "pleasure - displeasure" and the vertical dimension is "arousal - sleep" [2]. Fig. 1 illustrates this circumplex model of affect. The highlighted region in Fig. 1 shows the range of affects that can result in smiling. This wide variety of feelings underlying smiling can potentially explain the diverse nuances in this facial expression.

In this study, we employ both of these two theories. As described in Sec. 3, we first detect smiles by classification, and within

the smiling time-frame, we use parameter-mapping sonification. In other words, the detection stage is analogous to the category theory, and the parameter-mapping sonification stage is analogous to the dimension theory.

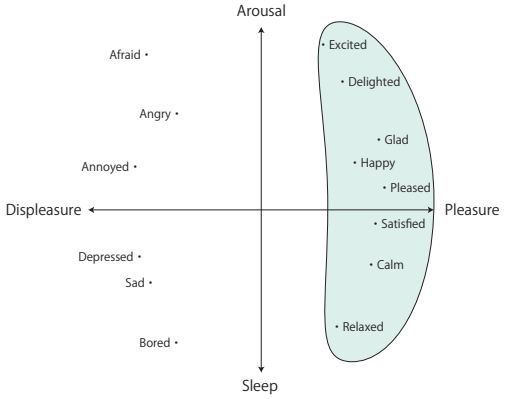


Figure 1: A circumplex model of affect (Colored area is area of feelings which can be related to smile)

2.2. EMG

We employed surface EMG (sEMG) signals to retrieve facial muscular motion. For our purposes, the sEMG signal has two advantages. First, there is little spatial limitation. Second, real-time capacity of the sEMG signal is high.

Previously, Patil et al. [3] and Funk et al. [4] used image processing for sonification of facial expressions. Their systems used optical flows for different parts of the face. Such image processing schemes require a camera; moreover, the user has to stay in front of the camera at an appropriate, fixed angle and distance, resulting in spatial limitations imposed by the relative positions of the camera and the face. The users are required to fix their faces and cannot turn their heads freely, resulting in limitations of both posture and view.

In contrast, the electrodes for measuring sEMG signals are both small and light. They can be attached on the face without interrupting the view. Therefore, there is more freedom for the spatial arrangement, posture, and view. Moreover, the EMG signals can capture the earliest stages of facial expression that are not yet visible. The subtle musculature motions that eventually grow in magnitude to alter the facial expression can be observed with the EMG signals. Therefore, EMG-based smile recognition is potentially faster than camera-based smile recognition. In our study, we employed a smile recognition algorithm using sEMG signals [5] to trigger a sound synthesis system at the starting and ending points of a smile.

3. SYSTEM

We implemented a real-time smile sonification system. The system uses the sEMG signals measured on the forefront and sides of the face (four channels). The system synthesizes a sound in real time. Figure 2 shows a schematic of the system.

The system is constructed in three modules. The first is the signal processing module, which performs filtering for noise reduction and calculates RMS value of sEMG signals. The second is the facial expression classification module, which consists of the support vector machine (SVM) learning model and classifier. The third is the sonification module, which synthesizes the sound using the RMS of sEMG signals triggered by the SVM result. The sonification sounds are played only when the SVM detects a smile.

The signal processing and facial expression classification modules are implemented using C#. The sonification module is implemented using SuperCollider. The system uses Open Sound Control (OSC) [6] to send the result of facial expression classification, and the features of sEMG signals from C# runtime environment to the SuperCollider.

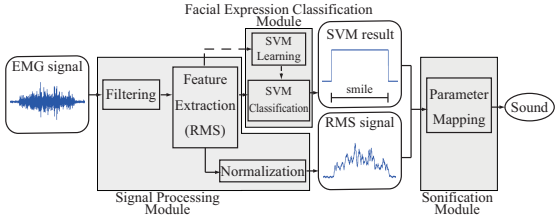


Figure 2: Schematic of smile sonification system

3.1. Signal Processing Module

In the signal processing module, the system first performs noise reduction, and then calculates the features for machine learning from the sEMG signals. The classification system uses Root Mean Square (RMS) as its feature [7]. RMS is considered conveys all necessary information about a signal's amplitude.

For noise reduction of sEMG signals, the system uses comb and bandpass filters. First, the system removes the power source noise by comb filtering. Next, the system limits the frequency spectrum to 30 - 450 (Hz) using a bandpass filter. The system performs two kinds of RMS calculations; one is for facial expression classification and the other is for sonification.

For the facial expression classification (described in Sec. 3.2), the RMS is calculated with a window of 150 ms. We use this RMS for learning and classifying the facial expression, which is essentially a binary decision of smile or no-smile. For sonification, the RMS is calculated with a time window of 50 ms. This small time window is indicative of the extremely fast changes during facial muscle movements, and these dynamic fluctuations are reflected as sound in our proposed system. The frame shift is 1 ms for the learning phase and 25 ms for the classification and sonification phase.

When a smile is expressed, sEMG signals on the sides of the face tend to change; however, no changes are experienced on the forefront of face. Therefore, we analyzed the sEMG signal changes at the side of face for the sonification of smiles.

After calculating the RMS for sonification, the data were normalized. $R_m(n)$, which is the normalized RMS, is calculated in Equation (1), where m is measurement part, n is the number of samples, $r_m(n)$ is the RMS for sonification before normalization, R_{0_m} is the average RMS for sonification at the time when a neutral face is learned (Sec. 3.2), and R_{max_m} is the maximum value

of RMS for sonification at the time when a big smile is learned (Sec. 3.2).

$$R_m(n) = \frac{r_m(n) - R_{0_m}}{R_{max_m} - R_{0_m}} \quad (1)$$

$$R_m(n) = \begin{cases} 1 & (R_m(n) > 1) \\ R_m(n) & (otherwise) \end{cases}$$

3.2. Facial Expression Classification Module

In the facial expression classification module, a smile is classified using a SVM, which is a two-class classifier. We use LIBSVM as the classification and learning algorithm [8]. SVM has a strong generalization capability against unclassified patterns and is computationally inexpensive. These features of SVM make it possible to classify a smile in real time.

sEMG signals are highly individual and not generalizable. These parameters vary depending on individual differences and electrode position. Therefore, the system first conducts calibration for each user by recording sEMG of five known facial expressions (neutral, frown, bite, smile and big smile) for four seconds each, and learns the user's signal pattern and intensity. Afterwards, the system classifies smiles based on the learned patterns and displays the results using sound in real time.

3.3. Sonification Module

In the sonification module, the system receives RMS signals for sonification of one side of face (two channels) and facial expression classification results. The synthesized sound is produced in real time, but only while one is smiling. We used parameter mapping sonification (PMSon) for sonification.

We had three desiderata for sonification: understandability, enjoyability, and pleasantness. Understandability means that users can easily understand the movements associated with facial expressions using sound. Enjoyment means that the sound can encourage and facilitate the expression of spontaneous smiles. Pleasantness means that the sound does not make users uncomfortable and does not disturb their spontaneous smiles.

We hypothesized that mapping to pitch using a scale would be suitable to satisfy the criteria of understandability and enjoyment. Value ranges of $R_m(n)$ (Equation (1), 0 - 1) were divided by the number of scale elements, with equal spacing. The C-major pentatonic scale was used. The timbre of the sound was a pure tone sine wave.

We thought that this mapping would satisfy the understandability criterion because a change of pitch reflects the movement of facial expressions. We also thought it could satisfy the enjoyment criterion because the sound reflects the movement of facial expressions and changes frequently. However, the timbre of pure-tone sine waves resulted in an artificial impression. Therefore, the timbre of pure tones did not satisfy the requirement of pleasantness.

In order to satisfy the pleasantness criterion, we therefore used a bubble sound model. A synthetic algorithm for bubble sounds is written in "Designing Sound" [9]. In implementing the bubble sound using SuperCollider, we referred to "Bubbles", implemented by Dan Stowell [10] in "Designing Sound in SuperCollider" [11]. As a result, we discovered that the timbre of bubble sounds could satisfy the pleasantness criterion. With our bubble sound implementation, when the RMS signal maintains the same value, the sound is not played. Only when the RMS signal value changes does it produce a bubble grain sound, thereby creating a kind of rhythm according to the muscle movements. In addition, as the pitch becomes higher, the amplitude of the sound becomes smaller—like the sound of natural bubbles.

4. EXPERIMENT

4.1. Conditions

Eight normal listeners (5 males, 3 female; aged 22 - 33) participated in the experiment. First, they received information about the experiment and instructions on the experimental procedures. Next, they were outfitted with the sEMG electrodes. They were requested to sit in a chair and listen to sounds played from stereo speakers, which were placed approximately 1.5 meters away from them (Fig. 3). The experiment was approved by the IRB at the Faculty of Library, Information and Media Science at the University of Tsukuba.

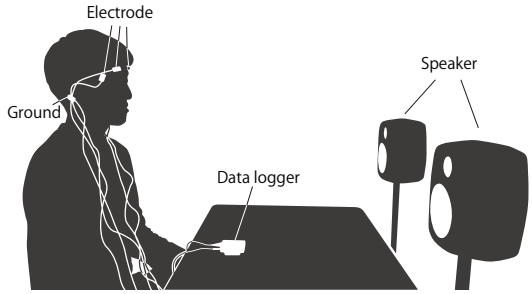


Figure 3: Experiment setup

4.2. Methods

After wearing the electrodes and practicing the system for a while, participants were instructed to smile under three feedback conditions. The first condition was no-feedback, in which participants expressed smiles without any feedback. The second condition was "mirror", in which participants expressed smiles with visual feedback using a mirror but without sound. The third condition was "sound," in which participants expressed smiles with sonification-based auditory feedback, but without a mirror. We recorded sEMG signals in each condition.

We asked participants to hold their smiles for approximately two to three seconds. However, we did not specify what kind of smile they needed to express. After the feedback experiment, we had a questionnaire and an interview with participants.

5. ANALYSIS & DISCUSSION

5.1. sEMG analysis

5.1.1. Symmetry of a Smile

We analyzed the sEMG signals during the smile in terms of the left and right balance. According to Rinn et al., spontaneous facial

expressions are symmetric; however, deliberate facial expressions are asymmetric [12]. Considering this, we hypothesized that if auditory feedback facilitated the expression of spontaneous smiles, then the difference between right and left muscles involved in forming smiles with auditory feedback should become small.

Taking this into consideration, we calculated RMS energy using a window size of 500 ms for the sEMG signals from the two channels on the sides of the face. Then, using this RMS signal, we integrated the difference between right and left signals for each smile. The integral of the sEMG difference between right and left is associated with the difference of the muscle movements for the right and left sides of the face during the smile. The smaller the integral, the more symmetric the smile. As shown in Fig. 4, the integral of the sEMG difference between right and left is smallest in the sound condition.

In the sound condition, the integral of the difference was smaller than in the other conditions without auditory feedback. Therefore, auditory feedback may have facilitated the expression of spontaneous smiles. In the sound condition, five of the eight participant's integral of the difference was the smallest of all conditions. However, this tendency just missed being statistically significant with the Wilcoxon signed rank test (p-value = 0.078).

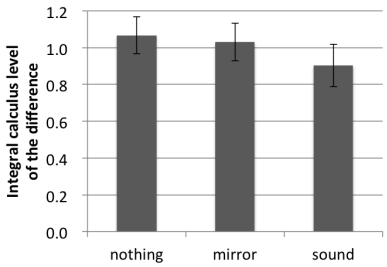


Figure 4: Integrated difference of right and left signals (normalized). We used the average of all integrated levels to normalize. The integral is smallest in the sound condition.

5.1.2. *Time Envelope of a Smile*

We investigated the time envelope of the sEMG signals as we were interested in the temporal aspects of smiles. Sometimes, a smile starts very quickly and ends quickly. Alternatively, a smile can gradually emerge, or a smile can sustain with fluctuations. Such variations in the time envelope of smiles are analogous to that of musical sounds, which are often modeled with the attack, decay, sustain, release (ADSR) model (Fig. 5). We therefore analyzed the time envelope of smiles using the ADSR model. Fig. 6 shows an example RMS of EMG signals and its envelope. The RMS signal shows the gradual rising and slow decay, which resembles ADSR model.

We divided the period of a smile into ten equally spaced time intervals. Within each interval, the maximum value of RMS signal of the sEMG were computed, and the time-envelope was defined as a connection of these ten maximum values. We defined the section from the start to the first local maximum as "attack," and the section from the last local maximum to the end as "release." We then used the time between start and attack as the "attack time," and the time between release and the end as the "release time." Finally, we analyzed the ratio of the attack and the release time compared to the duration of a smile. Fig. 7 shows the average ratio of "attack time" and "release time" in each experimental condition.

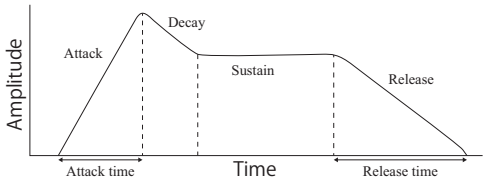


Figure 5: Time envelope of ADSR model

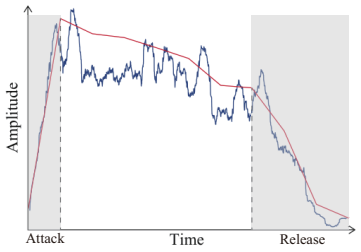
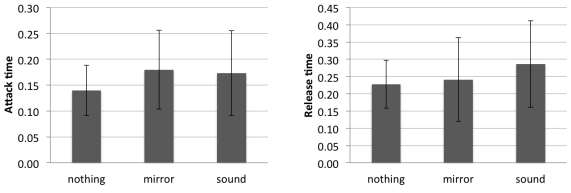


Figure 6: Example RMS of EMG signals (blue line) and its envelope (red line). Left gray area represents the attack and the right gray area represents the release.

According to Fig. 7, both the attack and release time ratios became higher in the feedback condition (i.e., sound and mirror) than in the no feedback condition (i.e., nothing). The higher attack and release time ratios suggest that smiles tend to appear and disappear gradually. The visual and auditory feedback had a similar effect on the ratio of attack and release times. For the majority of subjects, we observed increases in attack or release ratios due to feedback for all the experimental conditions; however, the amount of increase only trended toward being statistically significant according to the Wilcoxon signed rank test (p-value > 0.1 for attack time and p-value = 0.055 for release time).



(a) attack (b) release

Figure 7: Result of time envelope analysis in attack and release

5.2. Interview

During the interview, we received the following comments:

> "It was fun to have a feedback of my own moving."
>
> "It was interesting that I can understand how I smiled with the pitch of sound."
>
> "I could smile easier when there were sounds."

Overall, the participants indicated that the sonification reflected their smiles appropriately. Moreover, most of the participants commented that the system was "fun" or "interesting". To summarize, our interactive sonification system seems to deliver an enjoyable and intuitive feedback experience to all participants.

6. CONCLUSION & FUTURE WORK

In our study, we implemented a real time smile sonification system using sEMG signals in order for people to recognize their own smiling using auditory information, and conducted a user evaluation test.

Using parameter mapping sonification, our system sonifies a smile from sEMG signals that was measured from the sides of the face. The sound synthesis was designed to satisfy our target criteria (understandability, enjoyability, pleasure). Considering these criteria, we employed a mapping in which the strength of a smile was associated with pitch using a musical scale with the timbre of bubble sounds.

In the user evaluation test, we evaluated the effects of feedback on smiling. Furthermore, we analyzed the effects of feedback by using the integral of the left-right difference and the time envelope according to the ADSR model. The results indicated that auditory feedback facilitated the expression of spontaneous smiles. In addition, the effects of visual and auditory feedback on the rate of attack and release times were similar. Participants' comments on the enjoyment of using the system were mostly positive. However, there was no statistically significant outcome in our analysis due to the small number of participants.

In future work, we intend to test other types of sounds for the system. We also plan to develop a wearable device so that people can use the system more easily. We are interested if this system could help or enhance the emotional communication among multiple people.

7. ACKNOWLEDGEMENT

This work was funded by JST PRESTO. We thank Y. Morimoto for help in designing sound. We would like to thank everyone who kindly participated in our experiment.

8. REFERENCES

[1] P. Ekman, "An argument for basic emotions," *Cogn. & Emot.*, vol. 6, no. 3-4, pp. 169–200, 1992.

[2] J. A. Russell, "A circumplex model of affect." *J. Per. Soc Psychol*, vol. 39, no. 6, p. 1161, 1980.

[3] V. Patil *et al.*, "Sonification of Facial Expression Using Dense Optical Flow on Segmented Facial Plane," *International Conference on Computing and Control Engineering (ICCCE)*, 2012.

[4] M. Funk *et al.*, "Sonification of facial actions for musical expression," in *Proceedings of the 2005 Conf. NIME.* National University of Singapore, 2005, pp. 127–131.

[5] Y. Takano and K. Suzuki, "Affective communication aid using wearable devices based on biosignals," in *Proc. of the 2014 Conf. Interact. Design and Children (IDC2014).* New York, NY, USA: ACM, 2014, pp. 213–216.

[6] "http://opensoundcontrol.org/".

[7] A. Phinyomark *et al.*, "Feature extraction and reduction of wavelet transform coefficients for emg pattern classification," *Elektronika ir Elektrotechnika*, vol. 122, no. 6, pp. 27–32, 2012.

[8] "http://www.csie.ntu.edu.tw/˜cjlin/libsvm/".

[9] A. Farnell, *Designing Sound*. Cambridge, Massachusetts: The MIT Press, 2010.

[10] "http://www.mcld.co.uk/research/#phd".

[11] "http://en.wikibooks.org/wiki/Designing_Sound_in_SuperCollider".

[12] W. E. Rinn, "The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions." *Psychol. Bull*, vol. 95, no. 1, p. 52, 1984.