Displays 47 (2017) 32–39

Contents lists available at ScienceDirect

Displays

journal homepage: www.elsevier.com/locate/displa

The sound of smile: Auditory biofeedback of facial EMG activity ☆

Yuki Nakayama [a,*], Yuji Takano [b], Masaki Matsubara [c], Kenji Suzuki [d], Hiroko Terasawa [c,e,*]

[a] Graduate School of Library, Information and Media Studies, University of Tsukuba, 1-2, Kasuga, Tsukuba, Ibaraki 305-8550, Japan
[b] Graduate School of Systems and Information Engineering, University of Tsukuba, Japan
[c] Faculty of Library, Information and Media Science, University of Tsukuba, 1-2, Kasuga, Tsukuba, Ibaraki 305-8550, Japan
[d] Faculty of Engineering, Information and Systems, University of Tsukuba, Japan
[e] JST PRESTO, Japan

ARTICLE INFO

Article history:
Received 1 October 2015
Received in revised form 15 June 2016
Accepted 19 September 2016
Available online 21 September 2016

Keywords:
Sonification
Parameter mapping
EMG signals
Smile detection

ABSTRACT

In this paper, a real-time interactive system for smile detection and sonification using surface Electromyography (sEMG) signals is proposed. 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 that mimics bubbles.

In a user testing of smiling tasks, 14 participants underwent the system and are required to produce smiles under three conditions, i.e., auditory feedback with sonification, visual feedback with mirror, and no feedback. The impression of the system is evaluated through questionnaires and interviews with the participants. In addition, we analyzed the total amount of muscular activity and temporal envelope patterns of the sEMG during smiling.

The questionnaire and interview showed that users felt that (1) the sonification system well reflects their facial expressions, and (2) the sonification system was enjoyable. The users also expressed that the auditory feedback condition is easier to smile with, as compared to the visual feedback or no feedback conditions. However, the analysis of sEMG did not provide a quantitative difference among the three conditions, which is most likely due to the experiment design, which lacks socially engaging settings.

1. Introduction

Smiling is one of the most basic human facial expressions generally associated with positive emotions such as pleasure, amusement, and enjoyment. People smile to convey a kind expression towards 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 is 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 such a 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 proposed a real-time system for smile detection and sonification using surface electromyography (sEMG). This system has the following characteristics: (1) the variation of muscular motions in smiles becomes audible, and (2) the sound is pleasant and entertaining, to encourage users to smile more.

The potential applications of such a system are vast. Our work provides a fundamental technology to aurally represent the nonverbal elements of emotional communication. People who are visually challenged or who need to train their own facial expressions will benefit from this technology. For example, visually impaired people have difficulty in recognizing subtle smiles without vocalization nor teeth. Some autistic children have difficulty in understanding people's faces. These people may benefit from hearing the various degrees of facial expression using our smile sonification system. Some people who need to train their own facial expressions, such as people who have to communicate with clients in person or patients of facial paralysis, can practice their smile using this system. Compared to the camera-based smile sonification system, our EMG-based sonification system has no limitations on the smiling person's posture, angle, and position against camera, thus allowing more degree of freedom to the users. Here, we did not design or evaluate our system for a particular application; rather, we tried to design the fundamental mechanism, and evaluated its usability with ordinary people, so that we construct the core system for a broad range of useful applications.

☆ This paper was recommended for publication by Richard H.Y. So.
* Corresponding authors.
E-mail addresses: yuki@slis.tsukuba.ac.jp (Y. Nakayama), terasawa@slis.tsukuba.ac.jp (H. Terasawa).

http://dx.doi.org/10.1016/j.displa.2016.09.002
0141-9382/© 2016 The Authors. Published by Elsevier B.V.

In this study, we sonify a user's smile with sEMG and provide it to the user for self-monitoring in order to enhance or augment smile production. In addition, we focus primarily on the system development, as well as the user evaluation of the system.

2. Background

2.1. Emotions and facial expressions

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

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

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

On the other hand, the dimension theory considers emotions as a continuous change on a coordinate axis. Russell proposed a circumplex model of affect, which shows the position of feelings using adjectives on the coordinate axes, where the horizontal dimension represents "pleasure - displeasure" and the vertical dimension represents "arousal - sleep" [2]. Fig. 1 illustrates this circumplex model of affect. The highlighted region in Fig. 1 shows a range of affective states that can results in smiling. This wide variety of feelings underlying our smiles can potentially explain the diverse nuances of facial expressions. Thus, this theory can explain facial expressions continuously.

As a matter of fact, we experience emotions both categorically and continuously. Fig. 2 shows two examples of facial expressions. Both facial expressions are categorized as a smile; however, both smiles are somehow different. As suggested by the category and dimension theories, we recognize a facial expression not only as a category but we also perceive a continuous difference in facial expressions.

In this study, we employed both these theories. As described in Section 3, we first detect smiles by classification, and within the

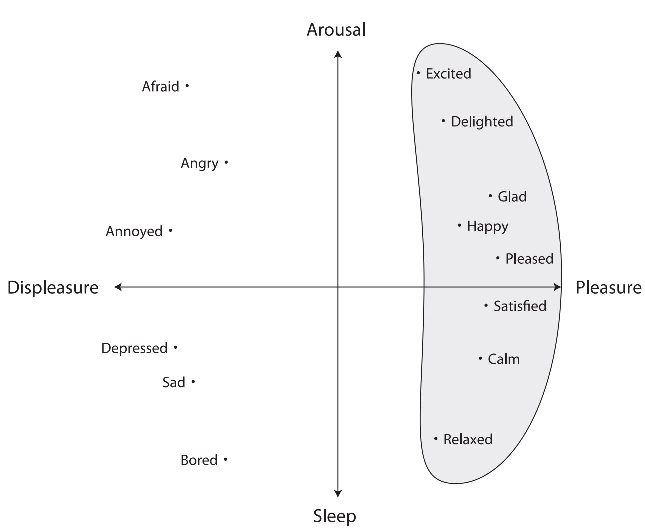

Fig. 1. Circumplex model of affect (colored area represents the area of feelings that can be related to smiles). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

smiling time frame, we use parameter-mapping sonification to represent the temporal change and strength of the smile. In other words, the detection stage is similar to the category theory, and the parameter-mapping sonification stage is similar to the dimension theory. In short, our system recognizes smiles both categorically and continuously. Furthermore, we think that this process is in agreement with the symbolic and analogic representation as suggested by Kramer in the book, "Auditory Display: Sonification, Audification, and Auditory Interfaces" [3]. The detection of a smile is similar to symbolic representation while the representation of temporal change and strength of the smile is similar to analogic representation. Thus, this system is a hybrid of symbolic and analogic representations.

2.2. Surface Electromyography (sEMG)

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

There are two major exiting methods for facial recognition. The first is image processing and the second is sEMG. Image processing is used generally for facial expression analysis [4,5]. It can recognize various facial expressions and measure a movement of a face without using the electrodes for measuring a sEMG signal. Previously, Patil et al. [6] and Funk et al. [7] used image processing for the 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 that are imposed by the relative positions of the camera and the face. The users are also required to maintain their faces in a fixed position and are unable to turn their heads freely, resulting in both posture and view limitations.

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 spatial arrangement, posture, and view. Moreover, the sEMG signals can capture the earliest stages of facial expression that are not yet visible. The subtle muscular motions that eventually grow in magnitude to alter the facial expression can be observed with the sEMG signals. Therefore, sEMG-based smile detection is potentially faster than the camera-based smile detection. We have already reported the potential use of facial EMG signals for the development of a wearable device for reading smiles [8]. In this study, we have used a wearable device and a smile detection algorithm using sEMG signals [9] to trigger a sound synthesis system at the starting and ending points of the smile.

3. System

We implemented a real-time smile sonification system. The system uses sEMG signals measured on the forefront and sides of the face (four channels, sampling frequency = 1000 Hz). The system synthesizes a sound in real time. Fig. 3 shows a schematic of the system.

The system consists of three modules. The first is the signal-processing module, which reduces noise and calculates the root-mean-square (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 SVM detects a smile.

The signal processing and facial expression classification modules are implemented using C#. The sonification module is

Fig. 2. Example of two different facial expressions that are categorized as a smile.

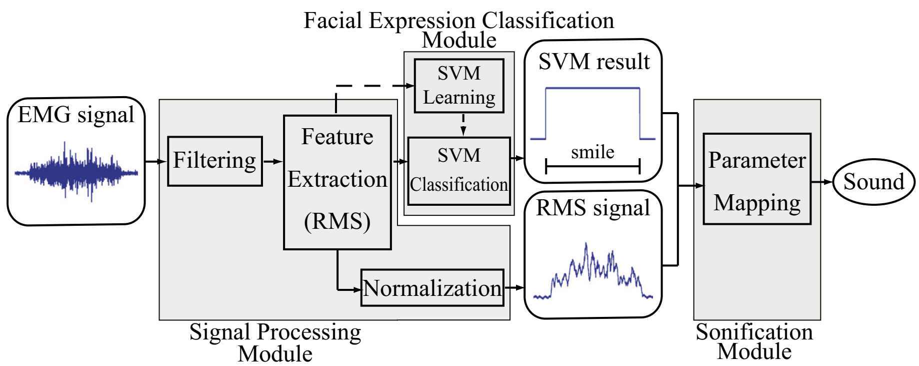

Fig. 3. Schematic of smile sonification system.

implemented using SuperCollider. The system uses open sound control (OSC) [10] to send the results of the facial expression classification and the features of the sEMG signals from the C# runtime environment to the SuperCollider.

We used a head-mounted interface, proposed by our collaborators, to capture facial expressions (Fig. 4) [9]. sEMG signals are measured by the interface and sent to the signal processing module through Bluetooth wireless communication. The interface has sockets including dry electrodes. The position of the sockets and the length of the headband are adjustable. Hence, the interface can be adjusted accordingly to different head sizes and shapes.

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 RMS as its feature [11]. RMS is considered to convey all the necessary information about a signal's amplitude. The signal's amplitude is important information of the EMG signals [12].

For the 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 between 30 and 450 Hz using a bandpass filter. The system performs two kinds of RMS calculations; one for facial expression classification and the other for sonification.

For the facial expression classification (described in Section 3.2), the RMS is calculated with a time window of 150 ms. We use this RMS for learning and classifying the facial expression, which is essentially a binary decision of smile or non-smile. For sonification, the RMS is calculated with a time window of 50 ms. The small time window indicates the extremely fast changes during facial muscle movements, and these dynamic fluctuations are reflected as sound in our proposed system. 50 ms is recommendation value of time window [13]. The frame shift is 1 ms for the learning phase and 25 ms for the classification and sonification phases.

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

After calculating the RMS for sonification, the data are normalized. $R_m(n)$, which is the normalized RMS, is calculated using Eq. (1), where m is the 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 (Section 3.2), and R_{max_m} is the maximum value of RMS for sonification at the time when a big smile is learned (Section 3.2).

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

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

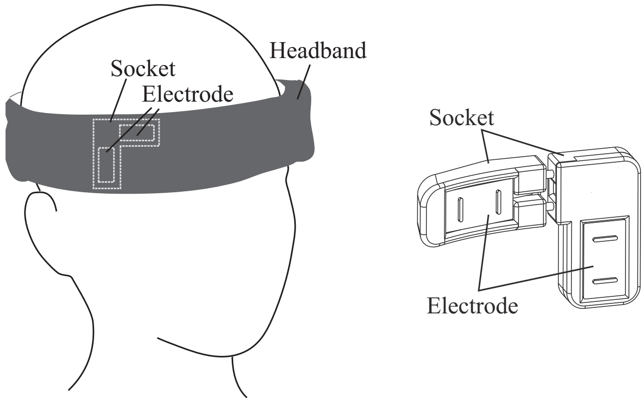

Fig. 4. Interface overview.

3.2. Facial expression classification module

In the facial expression classification module, we employed SVM to classify sEMG signals into smile and non-smile patterns, using the LIBSVM environment [14]. SVM is a two-class classifier that has a strong generalization capability against unclassified patterns and is computationally inexpensive. SVM defines optimal hyperplane as the linear decision function with maximal margin between the learning data of the two classes [15]. These advantages of SVM enable the classification of smile in real time possible.

sEMG signals are highly individual and not generalizable. These parameters vary depending on the individual differences and electrode position. In fact, the same person shows different sEMG strength and patterns from day to day. Therefore, the system first conducts a calibration for each user on a specific day, by recording the sEMG of four known facial expressions (neutral, bite, smile, and big smile) for 2 s each, and then learns the user's signal pattern and intensity. Next, the system classifies the smiles based on the learned patterns and displays the results using sound in real time. The input of the classification is the RMS of four channels, and SVM is employed as the classification method in which smile and big smile expressions are categorized as the "smile" class, while neutral and bite facial expressions are categorized as the "non-smile" class.

3.3. Sonification module

In the sonification module, the system receives RMS signals for the sonification of sides of the face (two channels) and facial expression classification results. The synthesized sound is produced in real time, given only when the user is smiling. We used parameter mapping sonification (PMSon) for sonification [16, Chapter.15].

We had three desirable qualities for sonification: understandability, enjoyability, and pleasantness. Understandability means that users can easily understand the movements associated with the facial expressions using sound. Enjoyability 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 affect their spontaneous smiles.

In this system, we employed bubble-like sounds for sonification. Each grain of sound (i.e., a bubble) is a very short "pop" sound with a slightly rising pitch. When a smile occurs on a face, several bubble sounds are synthesized, where starting pitch of each bubble becomes high according to the strength (i.e., the value of RMS signals) of the smile, and the density of the bubbles is associated with the dynamic change of the smile (i.e., deviation in the RMS signals).

The design of our system was finalized after preparing several prototypes. First, we tested mappings from RMS signals to pure tone with the sliding pitch (i.e., voltage-frequency oscillator), amplitude of pure tone (i.e. loudness change), and harmony (i.e. more complex harmony when the RMS signal is higher). Among them, the sliding pitch mapping was most favored during informal listening. The listeners could easily recognize the changes in the RMS signals, and enjoyed the changes. However, the timbre of pure tone sine waves resulted in an artificial impression that did not satisfy the requirement of pleasantness. Therefore, we employed bubble-like pop timbre in addition to the sliding pitch mapping. We believed that natural sounds would be more pleasant and would not prevent spontaneous smiles. This bubble-like sound helped in achieving a positive response during informal listening, and so we decided to use this mapping technique.

We synthesized the bubble sound so as to be able to control the sound parametrically. Our synthetic algorithm for bubble sounds is based on the description in "Designing Sound" [17]. In implementing the bubble sound using SuperCollider, we referred to "Bubbles" implemented by Dan Stowell [18] in "Designing Sound in SuperCollider" [19]. As a result, we discovered that the timbre of bubble sounds could satisfy the pleasantness criterion.

The initial pitch of a pop sound is controlled in the following manner: Value ranges of the RMS signal $R_m(n)$ (Eq. (1), 0–1) were divided by the number of scale elements, with equal spacing. The C-major pentatonic scale was used (C4–C5). Each pop sound starts from this pitch and its pitch is slightly bent upward. A decaying envelope was added onto the pop sound. With our bubble sound implementation, when the initial pitch maintains the same value, the sound will not be played. Only when the initial pitch value changes, a bubble grain sound is produced, thereby creating a rhythm according to the muscular movements. In addition, as the pitch becomes higher, the amplitude of the sound becomes smaller, similar to the sound of natural bubbles.

4. Experiment

4.1. Conditions

Fourteen subjects with normal hearing (9 males and 5 females; aged between 21 and 26) 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 on a chair and listen to the sounds played from the stereo speakers, which were placed approximately 1.5 m away from them (Fig. 5). The experiment was approved by the IRB at the Faculty of Library, Information and Media Science at the University of Tsukuba.

4.2. Methods

Fig. 6 shows the flow of this experiment. After wearing the head-mounted interface and experiencing the system for a period of time, participants answered questionnaire-A about their impressions on using the system. Next, participants were instructed to smile under three feedback conditions. The first condition was "nothing", 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 the sEMG signals for each condition. The order of the conditions is randomized for counterbalancing considerations. We asked the participants to hold their smiles for approximately 2–3 s. However, we did not specify what kind of smile they needed to express. After the feedback experiment, we conducted a questionnaire-B on the ease of smiling under each condition as well as an interview with the participants. All questionnaires includes 5-point rating scale of evaluations with free description columns.

5. Results on subjective evaluation

Participants answered questionnaire about their impression of using the system (Fig. 6, Questionnaire-A). Figs. 7–9 show the results of the questionnaires. Majority of the participants answered that the sonification could reflect their own smiles and that they enjoyed the system. In addition, most of them did not mind using the head-mounted interface.

After smiling under the three conditions, participants gave their feedback on the ease of smiling under each condition. Most of them answered they could easily smile under the "sound" condition as compared to the others (Fig. 10, Questionnaire-B). There is a

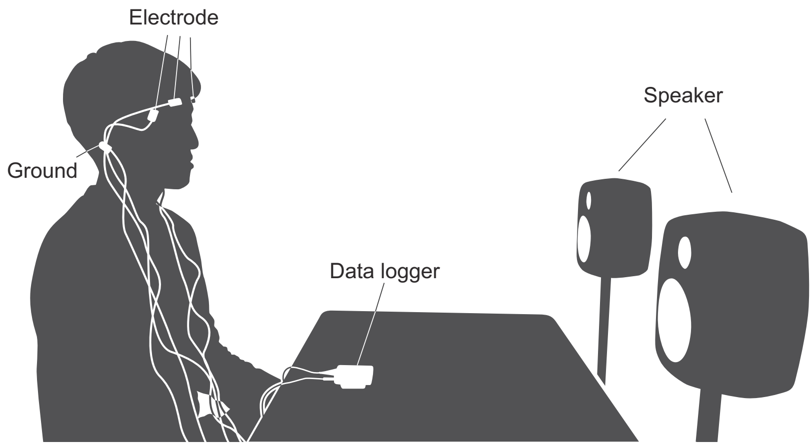

Fig. 5. Experiment setup.

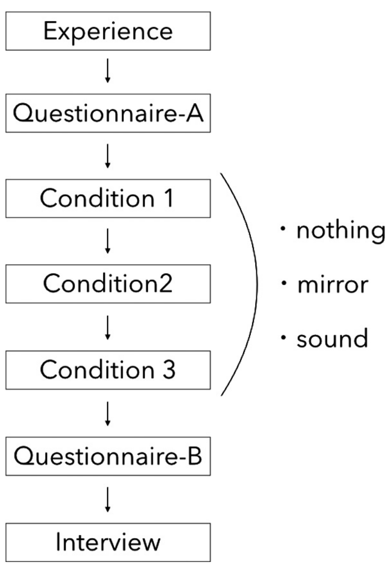

Fig. 6. Experiment flow.

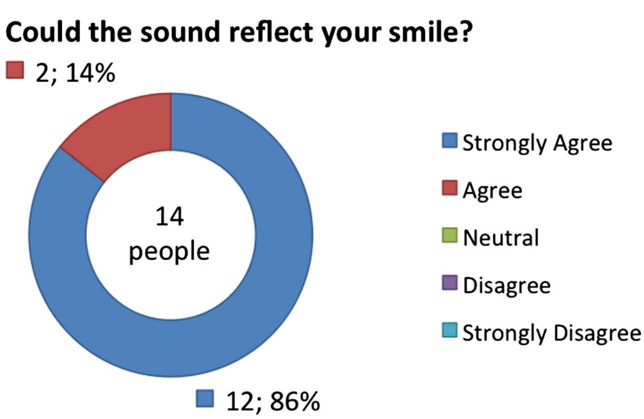

Fig. 7. Participants' feedback on whether the sound could reflect their own smiles. (Majority of them agreed that the sound could reflect their own smiles.)

significant difference of the Wilcoxon signed-rank test between the "sound" condition and the other conditions ($p < 0.001$).

We received the following comments from the interview with the participants and from the free description columns in the questionnaires:

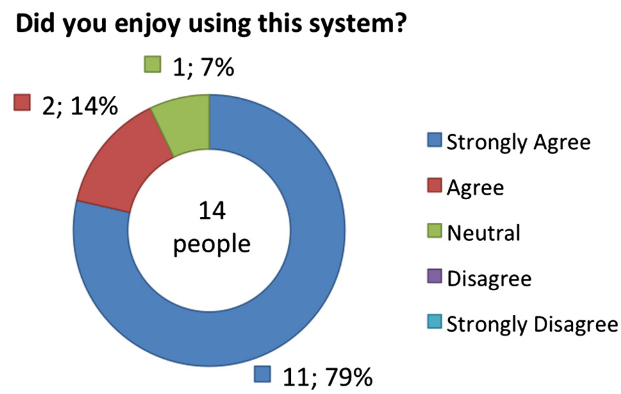

Fig. 8. Participants' feedback on whether they enjoyed the system. (79% of the participants enjoyed the system.)

"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."

"I could smile more by the sound."

Overall, the participants indicated that the sonification reflected their smiles appropriately. In addition, most of the participants described the system as "fun" and "interesting." Therefore, our interactive sonification system can deliver an enjoyable and intuitive feedback experience to all participants.

6. Results on objective evaluation

We analyzed the sEMG signals for each condition on the muscle activity and temporal change. We calculated the RMS using a time window size of 150 ms for analysis.

6.1. Total amount of muscle activity

We calculated the integrated value of RMS of both sides of the face, and added them together. The higher the integrated value, the bigger is the smile. Most of the participants reviewed that the sonification is fun, and thus, we hypothesized that the integrated value is higher under the "sound" condition as compared to the other conditions. Fig. 11 shows the average integral calculus level. There was no significant difference between the condition factor and the order factor of analysis of variance (ANOVA).

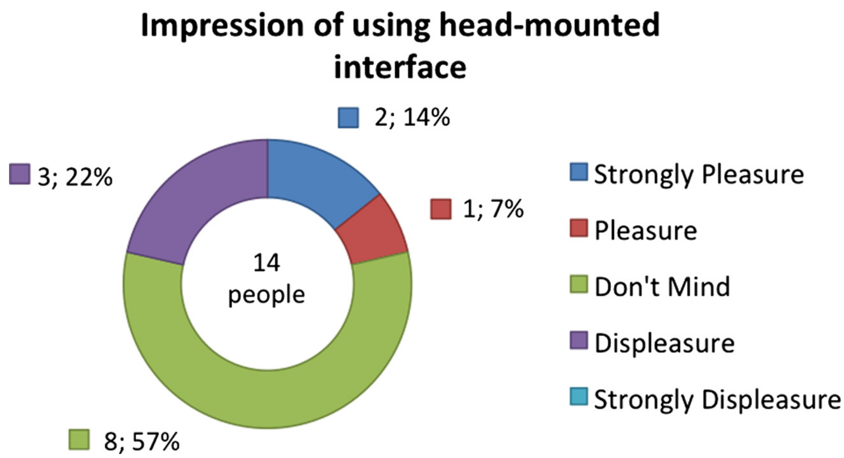

Fig. 9. Participants' feedback on their impression of using the head-mounted interface.

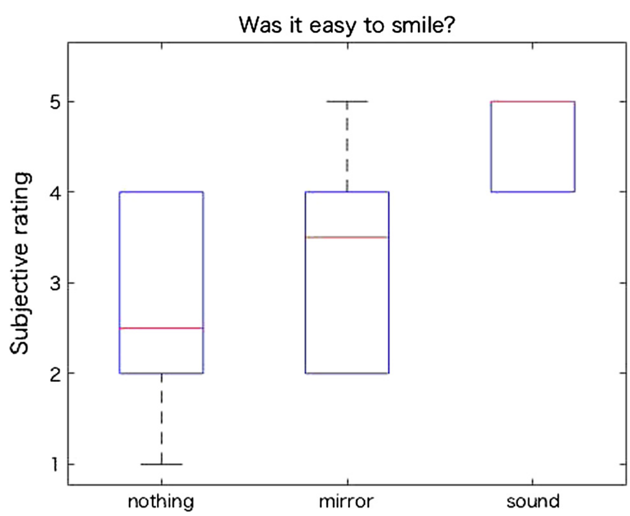

Fig. 10. Boxplot of participants' feedback on the ease of smiling under each condition with 5-point rating scale (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree and 5 = strongly agree).

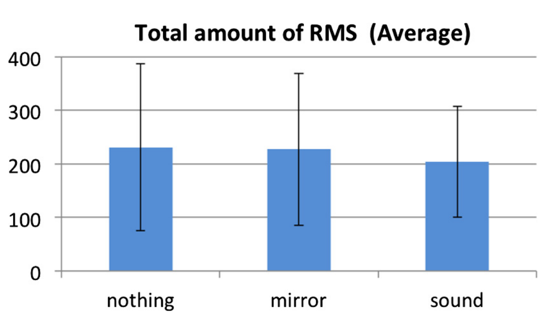

Fig. 11. Total amount of RMS.

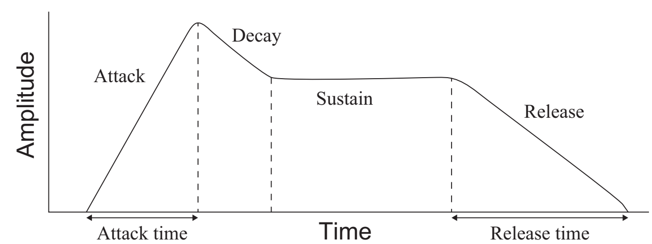

Fig. 12. Time envelope of ADSR model.

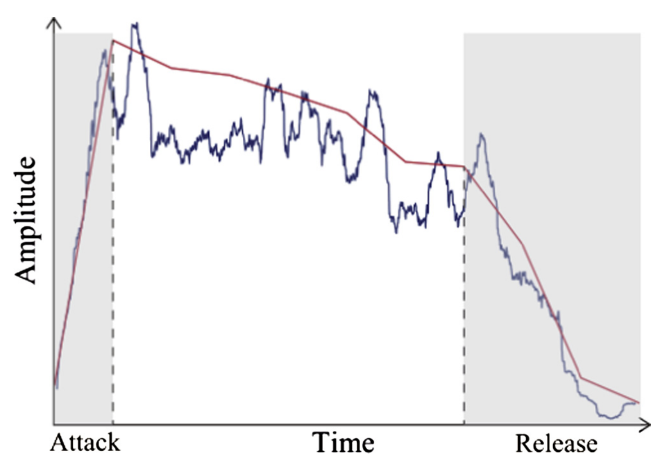

Fig. 13. Example of the RMS of sEMG signals (blue line) and its envelope (red line). Left gray area represents the attack while the right gray area represents the release. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6.2. Time envelope of a smile

We investigated the time envelope of sEMG signals, as we are interested in the temporal aspects of smiles. Often, a smile starts and ends quickly. Alternatively, a smile gradually emerges or it 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. 12). We therefore analyzed the time envelope of smiles using the ADSR model. Fig. 13 shows an example of the RMS of sEMG signals and its envelope. The RMS signal shows a gradual rising followed by a slow decay, which resembles the 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 sEMG was computed, and the time-envelope was defined as a connection to these ten maximum values. We defined the section between the start and the first local maximum as “attack” and the section between the last local maximum and the end as “release.” We then defined 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

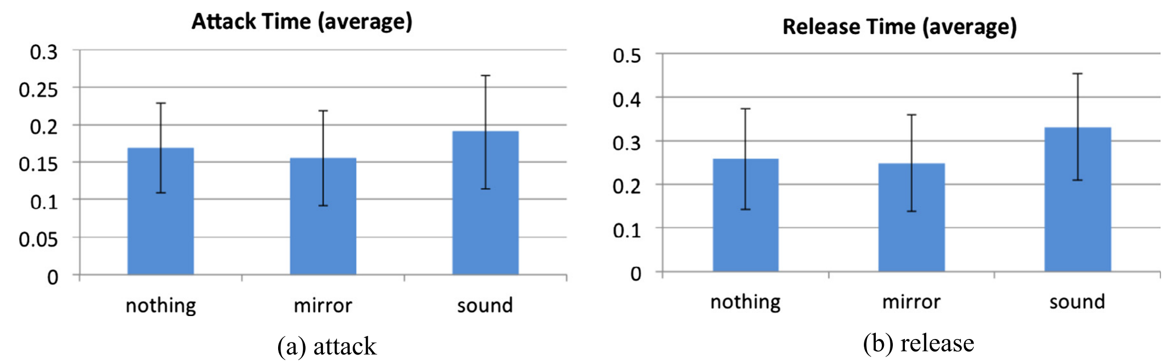

Fig. 14. Result of time envelope analysis in attack and release.

and the release time compared to the duration of a smile. Fig. 14 shows the average ratio of "attack time" and "release time" under each experimental condition.

According to Fig. 14, both attack and release time ratios are higher in the "sound" condition as compared to the other conditions. The higher attack and release time ratios suggest that the smiles tend to appear and disappear gradually. However, there was no significant difference.

7. Discussion

7.1. Feedback from users

According to the questionnaires, the participants answered that the sonification reflects the user's own smile, and the overall system experience was entertaining. Many of them judged that sonification system is the best, as compared to the other conditions (i.e., mirror and nothing.) Between the "nothing" and "mirror" conditions, participants showed contrasting judgments. Some highly preferred the "nothing" condition, while the others favored the "mirror" condition. Those who had chosen the "mirror" condition to be better, commented that the visual feedback helped them to smile easier compared to without any feedback. The participants who deemed the "nothing" condition to be better described that they feel awkward and timid about looking into their own smile in the mirror. Compared to the "nothing" and "mirror" conditions, the bubble sound feedback condition was easily accepted by all users. As the users' impression is highly dependent on the sound design, the difference of emotional acceptance cannot be generalized. Nevertheless, our experiment showed that a carefully designed sound system can provide a smiling bio-feedback in a pleasant manner for both people who are happy and unhappy with the direct visual feedback.

7.2. Evaluation tasks

From the analysis of sEMG signals, we did not find any significant statistical difference between the three conditions, although the users showed higher preference towards the sound feedback condition. We speculate that this contrasting results are attributed to the laboratory experiment setting. The experimenter reported that such unusual smiling task by a single participant without any conversation with others seemed to be a little awkward, even though it is not stressful. One of the participants also commented that the smiling task felt like an "operation." Since the evaluation task is designed in such a way that the participants are asked to "produce" a smile, we assumed that the difference between the feedback conditions were not well observed. Given that we received very positive subjective evaluations from the questionnaires, we would like to conduct future studies to investigate the effect of sonification in a social setting, where participants can smile more naturally during a conversation with one another.

7.3. System calibration

Participants who are not familiar in "producing" smiles, had difficulty in smiling during the calibration stage. They can smile naturally during conversations, but when they are asked to smile during the calibration, they failed to do so and thus the sEMG signals at the sides of their faces were very weak. In the informal follow-up sessions after the experiment, we tried to conduct a natural conversation with the participants and use the signal generated from their smiles during the conversation to calibrate the system. This approach worked better than the formal experimental procedure. Therefore, we believed that this system can be further improved through calibration with naturally induced smiles. However, it will be essential to maintain the controlled experiment settings at the same time.

8. 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 smiles using auditory information. We also conducted a user evaluation test with both objective and subjective measures.

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

In the user evaluation test, we evaluated the impression of the system through subjective evaluation by questionnaire and the effects of feedback on smiling from objective evaluation by sEMG signals. Based on the subjective evaluation, we determined that the system can reflect the user's smile and provide an enjoyable feedback. In addition, a user could smile more easily with the system auditory feedback as compared to the visual feedback and no feedback. In objective evaluation, we analyzed the time envelope according to the ADSR model and muscle activity under each condition. From this evaluation, we did not find any statistically significant difference or trends among different conditions. We observed that the facial EMG activity has significant information about various different patterns of smiles.

In future work, we intend to test other types of sounds for the system and study the effect of sonification under a social setting. We are interested to investigate if such a system could help enhance emotional communication among multiple individuals. Although the potential application of this system is vast, we are most interested in supporting visually impaired people. We conducted a survey on the visually impaired people, and found that they have difficulty in recognizing other's smile without laughter or teeth. We now think this system can support them to recognize various types of subtle smiles. For example, it would be great if this system helped the visually impaired children to recognize their parent's smiles, which would thus improve the parent-child relationship. Moreover, visually impaired children could use this system to train their own smile expression. Furthermore, blind people often have difficulty in making voluntary facial expressions [20]. Training with this system since childhood could facilitate them to develop voluntary facial expressions more confidently.

Acknowledgement

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

References

[1] P. Ekman, An argument for basic emotions, Cognit. Emot. 6 (3–4) (1992) 169–200.
[2] J.A. Russell, A circumplex model of affect, J. Pers. Soc. Psychol. 39 (6) (1980) 1161–1178.
[3] G. Kramer, Auditory Display: Sonification, Audification, and Auditory Interfaces, Perseus Publishing, 1993.
[4] Y. Yacoob, L. Davis, Computing spatio-temporal representations of human faces, in: Proceedings of the Computer Vision and Pattern Recognition Conference, IEEE Computer Society, 1994, pp. 70–75.
[5] D. Terzopoulos, K. Waters, Analysis and synthesis of facial image sequences using physical and anatomical models, IEEE Trans. Pattern Anal. Mach. Intell. 15 (6) (1993) 569–579.
[6] V. Patil, M.Q. Akhtar, A. Parab, A. Fernandes, Sonification of facial expression using dense optical flow on segmented facial plane, in: International Conference on Computing and Control Engineering (ICCCE), 2012.
[7] M. Funk, K. Kuwabara, M.J. Lyons, Sonification of facial actions for musical expression, in: Proceedings of the International Conference on NIME, 2005, pp. 127–131.
[8] A. Gruebler, K. Suzuki, Design of a wearable device for reading positive expressions from facial EMG signals, IEEE Trans. Affect. Comput. 5 (3) (2014) 227–237.
[9] Y. Takano, K. Suzuki, Affective communication aid using wearable devices based on biosignals, in: Proceedings of the 13th International Conference on Interact. Design and Children, New York, 2014, pp. 213–216.
[10] opensoundcontrol.org — An Enabling Encoding for Media Applications. <http://opensoundcontrol.org/> (accessed: 2015/09/24).
[11] A. Phinyomark, A. Nuidod, P. Phukpattaranont, C. Limsakul, Feature extraction and reduction of wavelet transform coefficients for EMG pattern classification, Elektron. Elektrotech. 122 (6) (2012) 27–32.
[12] K. Tomohiro, M. Tadasi, K. Tohru, S. Tsugutake, Practical usage of surface electromyogram, Tokyo Denki University Press, Japan, 2006.
[13] P. Konrad, The ABC of EMG: A Practical Introduction to Kinesiological Electromyography, Noraxon, Inc., 2005.
[14] LIBSVM – a Library for Support Vector Machines. <http://www.csie.ntu.edu.tw/cjlin/libsvm/>.
[15] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[16] T. Hermann, A. Hunt, J.G. Neuhoff (Eds.), The Sonification Handbook, Logos Publishing House, Berlin, G, 2011.
[17] A. Farnell, Designing Sound, The MIT Press, Cambridge, Massachusetts, 2010.
[18] Dan Stowell EPSRC Research Fellow. <http://www.mcld.co.uk/research/#phd> (accessed: 2015/09/24).
[19] Designing Sound in SuperCollider. <http://en.wikibooks.org/wiki/Designing_Sound_in_SuperCollider> (accessed: 2015/09/24).
[20] J.S. Fulcher, Voluntary facial expression in blind and seeing children, Arch. Psychol. 272 (1942).