Vibrational Artificial Subtle Expressions: Conveying System's Confidence Level to Users by Means of Smartphone Vibration

Takanori Komatsu
Meiji University
Tokyo, Japan
tkomat@meiji.ac.jp

Kazuki Kobayashi
Shinshu University
Nagano, Japan
kby@cs.shinshu-u.ac.jp

Seiji Yamada
National Institute of Informatics and SOKENDAI, Tokyo, Japan
seiji@nii.ac.jp

Kotaro Funakoshi and Mikio Nakano
Honda Research Institute Japan Co., Ltd.
Wako, Japan
{funakoshi, nakano}@jp.honda-ri.com

ABSTRACT
Artificial subtle expressions (ASEs) are machine-like expressions used to convey a system's confidence level to users intuitively. So far, auditory ASEs using beep sounds, visual ASEs using LEDs, and motion ASEs using robot movements have been implemented and shown to be effective. In this paper, we propose a novel type of ASE that uses vibration (vibrational ASEs). We implemented the vibrational ASEs on a smartphone and conducted experiments to confirm whether they can convey a system's confidence level to users in the same way as the other types of ASEs. The results clearly showed that vibrational ASEs were able to accurately and intuitively convey the designed confidence level to participants, demonstrating that ASEs can be applied in a variety of applications in real environments.

Author Keywords
Artificial subtle expressions (ASEs); Vibration information; Smartphone.

ACM Classification Keywords
H.5.2. User Interfaces: Evaluation/methodology; J.4. Social and behavioral sciences: Psychology.

INTRODUCTION
Although various kinds of smart devices, such as smartphones, smartwatches, or tablet PCs, are rapidly becoming popular and widespread in our daily lives, there is still little possibility that these devices will ever be 100% reliable due to speech recognition errors, ambiguity in user input, and so on [2,18,20,22]. This means that there is still some possibility for these devices to present incorrect information to users.

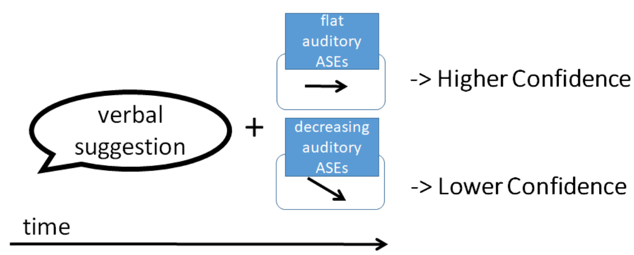

Figure 1. Auditory artificial subtle expressions (ASEs)

To manage such inevitable situations, some studies have been focusing on displaying a system's confidence level to users, and this was confirmed to be effective in various aspects of interaction between humans and systems [1,3,5,7,10,11]. For example, Komatsu et al. [14] proposed using artificial subtle expressions (ASEs) as machine-like expressions used to convey a system's confidence level to users intuitively in a complementary manner. Although an ASE itself does not have any specific meaning, the combination of an ASE with the main protocol has a specific meaning. Specifically, they proposed two simple beeping sounds used as "auditory ASEs": a flat sound (flat auditory ASE) and a sound with a decreasing pitch (decreasing auditory ASE). These auditory ASEs were added after the system's verbal suggestions. They then showed that suggestions followed by decreasing auditory ASEs intuitively conveyed a low system confidence level to users (Figure 1). Up to now, several studies on auditory ASEs have reported on the advantages of auditory ASEs, that is, being suitable for imperfect systems [15], language-independent

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ACM ISBN 978-1-4503-5620-6/18/04…$15.00
https://doi.org/10.1145/3173574.3174052

interpretations being possible [16], and faster response time compared with speech sounds [17][1].

ASEs can be implemented not only as auditory ASEs but also as visual or motion ASEs [8,13,31]. For example, Funakoshi et al. [8] proposed visual ASEs for human-robot dialogue. An LED was implemented on the chest of an interface robot and started blinking at 1/30-second even-intervals when a speech signal was detected and stopped blinking when the system started replying. They found that visual ASEs reduced the number of speech collisions between the robot and users, giving the users a positive impression of the robot. Yamada et al. [31] proposed motion ASEs, in which a robot slowly hesitates by turning to a human before giving advice with low confidence.

In this study, we focus on a situation where a user is using a smartphone, and this smartphone represents certain information with ASEs to the user. We propose using vibrations as ASEs (vibrational ASEs), which has the following two benefits in this situation.

- Most smartphones already have vibration motors, so it is easy to implement vibrational ASEs on them.
- Auditory ASEs are implemented as a time variation of sound pitch (frequency of a sound), visual ASEs as a time variation of the blinking of LEDs (frequency of blinking), and motion ASEs as a time variation of the velocity of a robot's rotation (angular velocity of robot). Vibrational ASEs would be implemented as a time variation of the frequency of vibrations, so it is expected that vibrational ASEs can be interpreted similar to other ASEs due to their physical characteristics, i.e., time variations of frequency or velocity that are essentially the same physical dimensions.

Many studies in human-computer interaction have tried to use vibrational information as a communication channel with users based on the perceptual features of haptic sense [9,27]. Specifically, vibrational information was utilized to notify the user of the progress of the system's functions [6], urgent information for pedestrians [28], timing awareness for oral presenters [30], CO_2 emission alerts for car drivers [26], or tactile icons for blind people [23,24]. This vibrational information was also utilized to represent the system's emotional states to users [19,25,29]. However, so far, there have been no studies that use vibrational information for conveying the system's confidence level to users in a complementary manner.

In this paper, we first describe the design of vibrational ASEs and then describe experiments done to confirm whether vibrational ASEs can convey a system's confidence level to users in the same way as auditory ASEs. Finally, we discuss the pros and cons of vibrational ASEs on the basis of the results of our experiments.

DESIGN OF VIBRATIONAL ARTIFICIAL SUBTLE EXPRESSIONS

We think that vibrational ASEs should be added after a system's verbal suggestions to convey a higher or lower confidence level to users like auditory ASEs. In the case of auditory ASEs, a flat sound (a flat auditory ASE) conveys a higher confidence level, and a sound with a decreasing pitch (decreasing auditory ASE) conveys a lower confidence level to users. We thus prepared two vibration patterns used as vibrational ASEs: one to convey a system's higher confidence level, and the other, a lower confidence level. We assumed that vibrations with a fixed frequency (flat vibrational ASE) would convey a higher confidence level and that vibrations with a continuously or gradually decreasing frequency (decreasing vibrational ASE) would convey a lower confidence level; that is, the time variation pattern of the frequency of vibrations was almost the same as that of auditory ASEs (Figure 2).

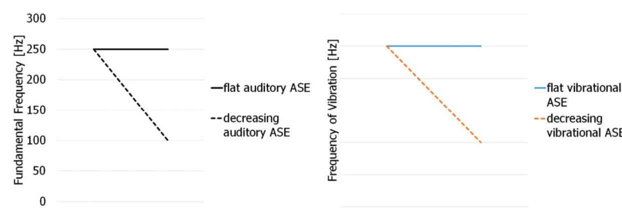

Figure 2. (left) Auditory ASEs and (right) vibrational ASEs at beginning

However, we cannot control the degree of frequency of the vibration motors on a smartphone because the vibration API of smartphones does not provide a method for controlling the frequency of the motors. It only provides a method for controlling the on/off statuses on the basis of vibration patterns. Therefore, we prepared the following two patterns for flat and decreasing vibrational ASEs (Figure 3).

- Flat vibrational ASE: eight 0.1-second long vibrations with 0.1-second long intervals [Figure 3(a)].
- Decreasing vibrational ASE: five 0.1-second long vibrations. Lengths of intervals are 0.1, 0.2, 0.3, and 0.4 seconds [Figure 3(b)].

We preliminarily confirmed that gradually lengthening the intervals between vibrations made users feel the sensation of deceleration or decreasing velocity, while a fixed length of vibration intervals made users feel the sensation of a constant velocity. The total durations of flat and decreasing vibrational ASEs were both 1.5 seconds.

[1] Although auditory ASEs look similar to Earcons [4], there are significant differences [14]; that is, an Earcon is a brief sound that has an arbitrary mapping to a certain meaning, while an auditory ASE has inevitable mapping to a meaning based on psychological findings. Moreover, Earcons play a main role in communication protocols, while ASEs play a complementary one.

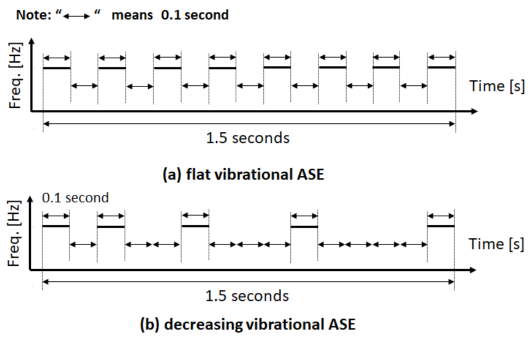

Figure 3. Vibrational ASEs. (a) Flat vibrational ASE and (b) decreasing vibrational ASE.

EXPERIMENT 1

Overview

We conducted an experiment to investigate whether the proposed vibrational ASEs were able to convey a system's confidence level to users accurately and intuitively; that is, whether the flat vibrational ASE conveys a higher confidence level and the decreasing vibrational ASE a lower confidence level.

As an experimental setting, we used a "driving treasure hunting" video game (Figure 4). In this game, the game image scrolls forward on a straight road as if the participant is driving a car with a navigation system and with small three mounds of dirt appearing along the way. A coin is inside one of the three mounds, while the other two mounds contain nothing. The game ends after the participant encounters 24 sets of mounds (24 trials).

The purpose for each participant is to get as many coins as possible. Which of the three mounds has the coin is randomly assigned. In each trial, the navigation system to the left of the driver seat (circled in the left image of Figure 4) told them in which mound it expected the coin to be by using a verbal suggestion followed by a vibrational ASE. The participant could freely accept or reject the navigation system's suggestions. In each trial, even after the participant selected one mound among the three, he/she was not told whether the selected mound had the coin or not (only a question mark appeared from the opened treasure box, as shown in the right image of Figure 4). Here, if the participant received feedback on whether his/her selection was correct or not, he/she usually started to solve a three-armed bandit problem [12] by considering strategies like "after estimating the probability distribution on which mound has a coin in the initial trials, select the mound with the highest probability." We thus avoided such strategic behavior by providing no feedback. The participants were then informed of their total numbers of coins only after they finished all 24 trials. It took about three minutes to complete this game. This game setting was commonly used in most former studies on ASEs [14-17], and the design in this study is fairly standard in human factor literature; it is a variation of a lane change task, which is a ISO-standardized driving task (ISO 17387:2008).

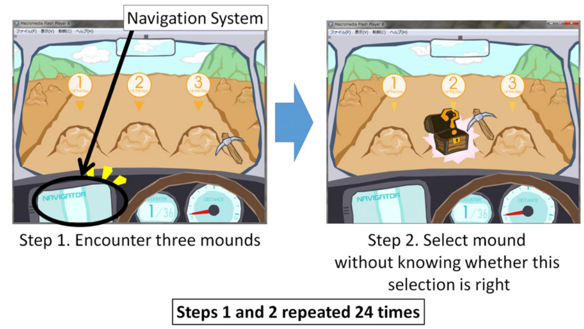

Figure 4. Driving treasure hunting video game

Stimuli

In this experiment, the navigation system used Japanese speech to suggest to the participants the expected location of the coin, that is, "ichi-ban (no. 1)," "ni-ban (no. 2)," or "san-ban (no. 3)." These speech sounds were created by adding robotic-voice effects to the recorded speech of one of the authors. The duration of these three sounds was 0.5 seconds.

One of two vibrational ASEs was played immediately after the verbal suggestion (Figure 5). These two ASEs were the flat vibrational ASE or decreasing vibrational ASE. The suggestions followed by decreasing vibrational ASEs were designed to inform users of the system's lower level of confidence in its suggestions, while those with flat vibrational ASEs were to inform them of a higher level of confidence.

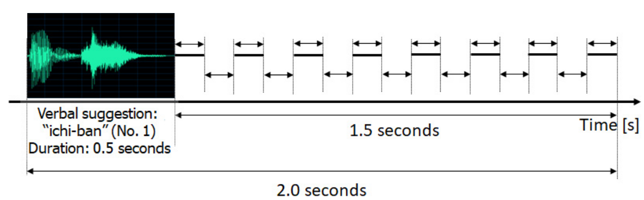

Figure 5. Speech waveform of suggestion "ichi-ban" with flat vibrational ASE

Here, the total length of a suggestion and a vibrational ASE was exactly 2.0 seconds (suggestion: 0.5 seconds, vibrational ASE: 1.5 seconds, and no interval between them). There were 6 variations of stimulus (3 suggestions × 2 ASEs). Among 24 trials of mound selection, each of these 6 variations was presented to the participants 4 times in a random order.

While the participants' played the game, we measured the response time, which was defined as the duration between the onset of the verbal suggestions and the participants' mound selection, and this was automatically measured by the experimental system implemented in the game environment. After playing the game, the participants were asked to answer the following two questions.

Q1: "Did you feel the vibrations during this game?"

Q2: "How many patterns of vibration did you notice?"

Participants

20 Japanese undergrads and graduate students participated (15 males and 5 females; 19 – 24 years old). These participants voluntarily responded to a call for participants from the authors. The participants were asked to be seated at a table, and an experimenter passed out a consent form with instructions on the experiment. These instructions and the experimenter never mentioned or explained the vibrational ASEs given after suggestions made to the participants. Afterward, a smartphone (Arrows M03, Fujitsu Limited, OS: Android 6.0.1) was passed to the participants, and they were asked to start the video game while holding the smartphone in their hands (Figure 6). The game was implemented in a web browser (Google Chrome: 60.0.3112.116) with JavaScript and the PixiJS library[2].

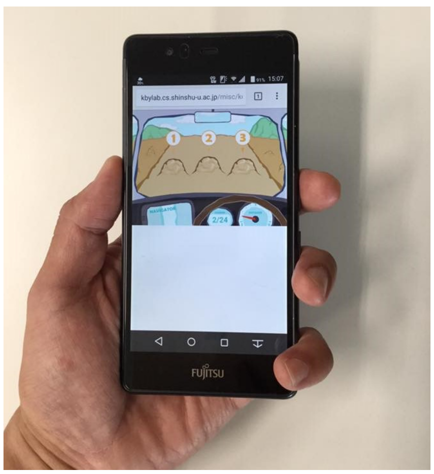

Figure 6. Smartphone and driving treasure hunting video game

Results

To investigate whether the vibrational ASEs could convey the designed confidence level to the participants or not, we calculated the rejection count, which indicates how many system suggestions were rejected by the participants (maximum rejection count: 12 times for each confidence level). For all 20 participants, the average rejection count of the 12 flat vibrational ASEs was 1.55 (SD = 1.99), while that of the 12 decreasing vibrational ASEs was 4.60 (SD = 4.13, Table 1).

Table 1. Average rejection counts for flat and decreasing vibrational ASEs

Type of vibrational ASE	Average rejection count
Flat	1.55 (SD = 1.99)
Decreasing	4.60 (SD = 4.13)

The rejection counts were analyzed with a dependent t-test (independent variable: flat/decreasing vibrational ASEs, dependent variable: rejection counts). The results showed significant differences between these two ASEs [$t(19) = 2.62$, $p = .017$, Glass' delta = 1.53], so we observed that suggestions with decreasing vibrational ASEs showed higher rejection counts compared with those with flat vibrational ASEs. Consequently, we confirmed that vibrational ASEs can convey the designed confidence level to the participants in the same way as auditory ASEs.

From the participants' answers to the two questions, Q1 and Q2, presented after the experiment, we found that all 20 participants felt vibrations during the game, but only 8 participants answered that there were two vibration patterns. We then compared the average rejection counts of the participants who were unaware that there were two vibrations with those of the other participants who were aware of this fact (Table 2). A 2 × 2 mixed ANOVA [within independent variable: vibration patterns (flat/dec.), between independent variable: awareness of two vibrations (Yes/No), dependent variable: rejection counts] showed that there were no significant differences in the interaction effects [$F(1,39) = 0.23$, n.s., effect size $f = 0.11$] and in the main effect of the between independent variable [$F(1,39) = 2.12$, n.s., effect size $f = 0.34$], but there was a significant difference in the main effect of the within independent variable [$F(1,39) = 6.82$, $p < .05$, effect size $f = 0.62$]. These results showed that all participants responded to the given stimuli in a similar way regardless of their awareness of the vibration patterns. This result clearly indicates that the participants responded to these vibrations intuitively without deeply considering what these patterns meant.

Table 2. Average rejection counts for two vibrational ASEs in terms of the participants' awareness of vibration patterns

Participants who noticed two types of vibrations	Average Rejection count for Flat ASE	Average Rejection count for Dec. ASE
Yes: 8 participants	2.00 (SD = 2.50)	5.75 (SD = 3.93)
No: 12 participants	1.25 (SD = 1.48)	3.83 (SD = 4.07)

Discussion

Definition of ASEs

The former study [14] stated that ASEs should satisfy two design requirements and two functional requirements simultaneously. Specifically, the two design requirements are as follows.

- Simple: ASEs should be implemented on a single modality.
- Complementary: ASEs should only have a complementary role in communication and should not interfere with communication's main protocol.

The two functional requirements are as follows.

- Intuitive: ASEs should be understandable by users who do not know about the ASEs beforehand.

[2] http://www.pixijs.com/

- Accurate: ASEs should convey the designer's intended meanings accurately.

In terms of design requirements, while the system's verbal suggestion was implemented as sounds, the vibrational ASE itself was implemented as only vibration. Therefore, the design requirement of "simple" was satisfied. The main protocol of this system was to tell participants the expected location of a coin as a verbal suggestion, while the ASE protocol was to indicate the system's confidence level in a complementary manner, so the design requirement of "complementary" was satisfied.

In terms of functional requirements, although the participants did not receive any instructions on vibrational ASEs beforehand, they interpreted the vibrational ASEs in accordance with the authors' intention. Moreover, the results of experiment 1 showed that the participants responded to these vibrations without deeply considering what their patterns meant. Thus, the functional requirement of "intuitive" was satisfied. At a glance, the average rejection count for decreasing ASEs was 4.6 (about 38%), which is not very high, so we cannot immediately conclude that the requirement of "accurate" was satisfied. However, even if the participants correctly understood the meanings of ASEs, it is not always true that they behave in accordance with their understandings. Therefore, we think that a significant difference between the rejection counts between flat and decreasing ASEs would strongly suggest that the requirement of "accurate" was satisfied, and actually, there was such a difference. Thus, the other functional requirement of "accurate" can be considered as satisfied, though we will explore a better way to directly measure the accuracy based on participants' behaviors. Consequently, the proposed vibrational ASEs satisfy all four of the ASE requirements.

Participants' reactions to vibrational ASEs

The average response time of all 20 participants was 1.76 seconds (SD = 0.95), but some participants responded to the system's suggestions very quickly; in fact, eight participants responded within less than 0.9 seconds. Figure 7 shows that the verbal suggestions with flat and decreasing vibrational ASEs were completely the same until 0.9 seconds; thus, these eight participants selected a mound without utilizing the vibrational ASEs (the average reaction time for these eight participants was 0.77 seconds). The average rejection count for the flat vibrational ASEs for these eight participants was 2.38 (SD = 1.84), while that for the decreasing vibrational ASEs was 1.88 (SD = 1.62), and there were no significant differences between the two vibrational ASEs [$t(7) = 1.32$, $p = .22$, Glass' delta = −.27].

The remaining 12 participants showed that the average rejection count for flat vibrational ASEs was 1.0 (SD = 2.04), while that for decreasing vibrational ASEs was 6.41 (SD = 4.50), and there were significant differences between them [$t(11) = 3.38$, $p = .006$, Glass' delta = 2.65]. However, the average reaction time for these 12 participants was 2.42 seconds (SD = 0.67).

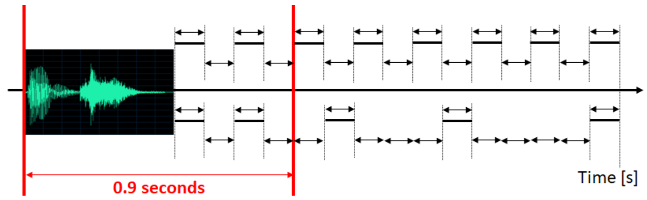

Figure 7. Comparison of flat vibrational ASEs and decreasing ones in terms of time variations

Therefore, although vibrational ASEs were found to be effective, it can be said that one problem with them is that the participants' response times for the given stimuli were longer. To reduce the response time for a given stimuli, we considered presenting the verbal suggestions and vibrational ASEs simultaneously because the main protocol, i.e., verbal suggestion, was auditory information, while the complementary protocol, i.e., vibrational ASEs, was vibrational information, so these two different types of information would not interfere with each other even if presented simultaneously. This simultaneous presentation could make users utilize the vibration for their selections and also shorten the length of a given stimuli itself.

EXPERIMENT 2

Overview

We thus conducted a consecutive experiment to investigate whether simultaneously presenting verbal suggestions and the proposed vibrational ASEs could convey a system's confidence level to users accurately and intuitively. The experimental setting and procedure was completely the same as that of experiment 1 except for the timing at which the verbal suggestions and vibrational ASEs were presented.

Stimuli

In this experiment, the vibrational ASEs were presented simultaneously with the verbal suggestions (Figure 8). The length of the suggestions with vibrational ASEs was exactly 1.5 seconds (suggestions: 0.5 seconds, vibrational ASEs: 1.5 seconds).

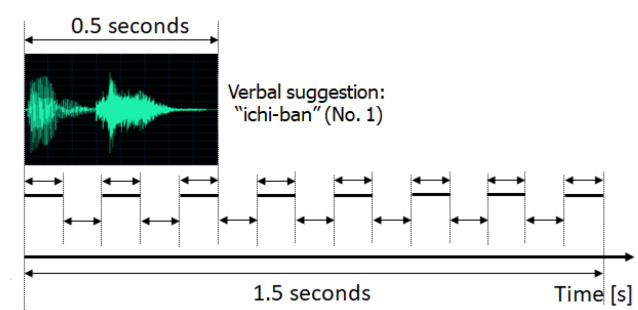

Figure 8. Speech waveform of suggestion "ichi-ban" and flat vibrational ASE in experiment 2

Here, the vibrational ASEs were implemented as time variations of vibration patterns in the same way as experiment 1, and the main protocol of this system was to tell participants of the expected location of the coin with a verbal suggestion, while the ASE protocol was to indicate the system's confidence level in a complementary manner. These two do not interfere with each other even though they are presented simultaneously, so the vibrational ASEs in this experiment also satisfied the two design requirements of "simple" and "complementary."

Participants

20 Japanese undergrads and graduate students participated (14 males and 6 females; 20 – 23 years old). These participants voluntarily responded to a call for participants, and they did not participate in experiment 1.

Results

To investigate whether vibrational ASEs presented simultaneously with verbal suggestions can convey the designed confidence level to the participants or not, we calculated the rejection counts. For all 20 participants, the average rejection count of the 12 flat vibrational ASEs was 1.65 (SD = 2.01), while that of the 12 decreasing vibrational ASEs was 5.30 (SD = 3.58, Table 3).

Table 3. Average rejection counts for flat and decreasing vibrational ASEs in experiment 2

Type of vibrational ASE	Average rejection count
flat	1.65 (SD = 2.01)
decreasing	5.30 (SD = 3.58)

The rejection counts were analyzed with a dependent t-test (independent variable: flat/decreasing vibrational ASEs, dependent variable: rejection counts). The results showed significant differences between the two ASEs [$t(19) = 3.92$, $p = .001$, Glass' delta = 1.82], so we confirmed that suggestions with decreasing vibrational ASEs showed a higher rejection count compared with those with flat vibrational ASEs, the same as in experiment 1. Therefore, we confirmed that vibrational ASEs presented simultaneously with verbal suggestions also could convey the designed confidence level to the participants, so the functional requirement "accurate" was satisfied. The participants were not given any instructions on vibrational ASEs, but they interpreted them appropriately, so the functional requirement "intuitive" was also satisfied. Thus, we confirmed that simultaneously presenting verbal suggestions and the proposed vibrational ASEs also succeeded in satisfying all four requirements of ASEs.

Table 4. Average rejection counts for two vibrational ASEs in terms of the participants' awareness of vibration patterns

Participants who noticed two types of vibrations	Average Rejection count for Flat ASE	Average Rejection count for Dec. ASE
Yes: 11 participants	1.45 (SD = 2.06)	5.18 (SD = 3.90)
No: 9 participants	1.89 (SD = 1.91)	5.44 (SD = 3.13)

From the participants' answers for the two questions, Q1 and Q2, presented after the experiment, we found that all 20 participants had noticed that they felt vibrations during the game, and 11 participants answered that there were two vibration patterns. We then compared the average rejection counts of the participants who were unaware or aware of the two vibrations (Table 4). A 2 × 2 mixed ANOVA [within independent variable: vibration patterns (flat/dec.), between independent variable: awareness of two vibrations (Yes/No), dependent variable: rejection counts] showed that there were no significant differences in the interaction effects [$F(1,39) = 0.01$, n.s., effect size $f = 0.02$] and in the main effect of the between independent variable [$F(1,39) = 0.13$, n.s., effect size $f = 0.08$], but there was a significant difference in the main effect of the within independent variable [$F(1,39) = 14.39$, $p < .01$, effect size $f = 0.89$]. These results also clearly indicate that the participants still responded to these vibrations intuitively, the same as in experiment 1.

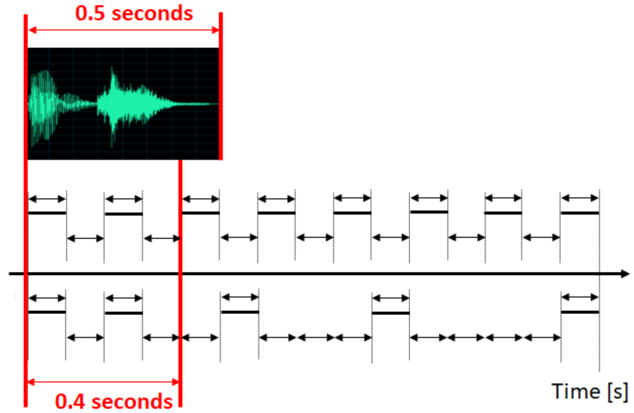

Figure 9. Comparison of flat vibrational ASEs and decreasing ones in terms of time variations in experiment 2

The average response time for all 20 participants was 1.52 seconds (SD. = 0.55), so this was significantly faster compared with that of the 12 participants who utilized vibrational ASEs when making selections in experiment 1 (average reaction time was 2.42 seconds) [independent t-test: $t(30) = 4.1$, $p = .0003$, $d = 1.51$].

Among all 20 participants in experiment 2, 3 participants responded in less than 0.9 seconds (the shortest average response time was 0.707 seconds). Figure 9 shows that the verbal suggestions with flat and decreasing vibrational ASEs were completely the same until 0.4 seconds, so it can be said that all 20 participants selected a mound after listening to the verbal suggestions and that they could utilize the differences in the two vibrational ASEs when selecting mounds.

DISCUSSION AND CONCLUSION

Pros of vibrational ASEs

One of the noteworthy achievements of this study was that it was found that vibration information can be utilized as ASEs in the same way as auditory, visual (LED-based), and motion ASEs. As already mentioned, most smart devices are equipped with vibration motors, so the results of this study will drastically increase the applicability of the ASEs into various kinds of devices. Furthermore, vibration is haptic information, so only users who hold such devices can feel the vibration. This means that vibrational ASEs would convey certain information that users do not want anyone to know. Clarifying the pros and cons of each type of ASE would be helpful for considering how to actually apply these ASEs.

The other noteworthy achievement was that simultaneous usage of vibrational ASEs and verbal suggestions was shown to be possible. If the modalities of the main protocol and complementary protocol were the same, simultaneous usage would be difficult; for example, if verbal suggestions and auditory ASEs were presented simultaneously, the verbal suggestions would mask or overlap the auditory ASEs, and vice versa, so eventually, the main protocol and complementary protocol would interfere with each other. However, the results of this study clearly showed that it is possible to present the main protocol and complementary protocol simultaneously in the case that the modalities of these protocols are different. This finding would contribute to expanding the variety of stimuli for users and also to shortening the length of the stimuli.

Cons of vibrational ASEs

Although flat and decreasing vibrational ASEs succeeded in conveying the designed confidence level to the participants, the duration of these ASEs was longer (1.5 seconds) compared with the length of verbal suggestions (0.5 seconds). This length could interfere with users' ability to comfortably interact with these systems. As already noted, we at first considered that the frequency of the decreasing vibrational ASE should continuously or gradually decrease (Figure 2), but we could not implement this because the vibration API of smartphones does not provide a method for controlling the frequency of the vibration motors. Therefore, we are now considering how to overcome this limitation. If we can implement a continuously or gradually decreasing frequency vibration on smart devices, we expect to find that the duration of the vibrational ASEs should be much shorter.

Although only users who hold a device with vibration motors can feel the vibration, most smartphones usually make vibration sounds when their motors are activated, so the users can know whether the motors are active or not via vibration sounds without feeling the vibrations. In our experiments as well, the smartphone made vibration sounds, so we could not confirm whether the participants interpreted the meanings of vibrational ASEs from their vibrations or from the sounds. In fact, all 40 participants in both experiments reported that they noticed the vibrations during the experiment, so it is reasonable to believe that these participants interpreted the meaning of the vibrational ASEs from the vibrations. However, to precisely confirm that this is correct, an additional experiment must be conducted in which the vibration sounds made during vibration are somehow masked, e.g., asking the participants to wear noise-canceling headphone or conducting an experiment in a noisy environment. However, in an opposite way, it would be interesting to focus on the effects of vibration sounds without feeling the vibrations as a novel type of auditory ASE. This approach is similar to that of Moore et al.'s study [21], which focused on the aural noise of the servomotors of robots.

Conclusion and Future work

In this paper, we proposed a novel type of ASE that uses vibration, called vibrational ASEs. We implemented the vibrational ASEs on a smartphone and conducted experiments to confirm whether they can convey a system's confidence level to users in the same way as the other types of ASEs. The results clearly showed that vibrational ASEs were able to accurately and intuitively convey the designed confidence level to the participants.

On the basis of the results of this study, we found a wide variety of future research directions. First, we would like to investigate the effectiveness of combining various kinds of main protocols and different ASEs because this study was focused on only the combination of verbal suggestions and vibrational ASEs. We can currently use various kinds of ASEs, e.g., auditory, visual, motion, and vibrational ones, so considering appropriate combinations of various kinds of main protocols and different ASEs would expand the applicability of the ASEs.

Second, we are also considering expanding the variety of expressions of vibrational ASEs. Specifically, we are considering using "increasing ASEs" as a new variation of ASE and different transition patterns for increasing or decreasing ASEs in linear or logarithmic ways. Based on former studies of vibration information in human-computer interaction [26,27], various on/off patterns or different frequencies of vibrations would be possible candidates. These new expressions could convey a system's confidence levels in much more detail or other kinds of information to users.

In parallel with the above rather fundamental investigation, we are also curious about concrete applications of vibrational ASEs. For example, the vibrations could be implemented not only in smart devices but in the handles or levers that users are holding. This means that vibrational ASEs can be applied for most driving or operating consoles, e.g., cars, trains, planes, heavy machines, and cranes. Considering how to apply vibrational ASEs into these realistic applications would be our next target, e.g., a car navigation system expressing decreasing vibrational ASEs together with speech like "This route A would be faster than route B if there was no snow on it" in a cold morning in winter.

In this study, we focused on a situation in which a user is holding a smartphone. We are now curious about much more ubiquitous situations, like when a user is wearing smart glasses and a smart watch and holding a smartphone. In this situation, there would be a suitable information modality for each device, e.g., visual ASEs are good for glasses, so it is very exciting to consider how to apply various kinds of ASEs in such situations.

REFERENCES

1. Stavros Antifakos, Nicky Kern, Bernt Schiele, and Adrian Schwaninger. 2005. Towards Improving Trust in Context Aware Systems by Displaying System Confidence, In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI2005)*, ACM Press (2005), 9-14.
2. Victoria Bellotti, and Keith Edwards. 2001. Intelligibility and Accountability: Human Considerations in Context-aware Systems, *Human-Computer Interaction 16*, 2 (2001), 193-212.
3. Mohamed Benzeghiba, Renato De Mori, Olivier Deroo, Stephane Dupont, Teodora Erbes, Denis Jouvet, Luciano Fissore, Pietro Laface, Alfred Mertins, Christophe Ris, Richard Rose, Vivek Tyagi, and Christian Wellekens. 2007. Automatic Speech Recognition and Speech Variability: A Review, *Speech Communication 49*, 10-11 (2007), 763-786.
4. Meera M. Blattner, Denise A. Sumikawa, and Robert M. Greenberg. 1989. Earcons and Icons: Their Structure and Common Design Principles, *SIGCHI Bulletin. 21*, 1 (1989), 123-124.
5. Hua Cai and Yingzi Lin. 2010. Tuning Trust Using Cognitive Cues for Better Human-machine Collaboration, In *Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society (HFES2010)*, 2437-2441 (5).
6. Jessica R. Cauchard, Janette L. Cheng, Thomas Pietrzak, and James A. Landay. 2016. ActiVibe: Design and Evaluation of Vibrations for Progress Monitoring, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI2016)*, ACM Press (2016), 3261-3271.
7. Jinjuan Feng and Andrew Sears. 2004. Using Confidence Scores to Improve Hands-free Speech Based Navigation in Continuous Dictation Systems, *ACM Transactions on Computer-Human Interaction 11*, 4, ACM Press (2004), 329-356.
8. Kotaro Funakoshi, Kazuki Kobayashi, Miko Nakano, Takanori Komatsu, and Seiji Yamada. 2010. Non-humanlike Spoken Dialogue: a Design Perspective, In *Proceedings of the 11th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL2010)*, 176-184.
9. Jihong Hwang and Wonil Hwang. 2011. Vibration perception and excitatory direction for haptic devices, *Journal of Intelligent Manufacturing 22* (2011), 17-27.
10. Eric Horvitz and Matthew Barry. 1995. Display of Information for Time-critical Decision Making, In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann (1995), 296-305.
11. Eric Horvitz. 1999. Principles of Mixed-Initiative User Interfaces, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI1999)*, ACM Press (1999), 159-166.
12. Michael N. Katehakis and Arthur F. Veinott, Jr. 1987. The Multi-Armed Bandit Problem: Decomposition and Computation, *Mathematics of Operation Research 12*, 2 (1987), 262-268.
13. Kazuki Kobayashi, Kotaro Funakoshi, Seiji Yamada, Mikio Nakano, Takanori Komatsu, and Yasunori Saito. 2012. Impressions Made by Blinking Light Used to Create Artificial Subtle Expressions and by Robot Appearance in Human-robot Speech Interaction, In *Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN2012)*, 215-220.
14. Takanori Komatsu, Kazuki Kobayashi, Seiji Yamada, Kotaro Funakoshi, and Mikio Nakano. 2010. Artificial Subtle Expressions: Intuitive Notification Methodology for Artifacts, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI2010)*, ACM Press (2010), 1941-1944.
15. Takanori Komatsu, Kazuki Kobayashi, Seiji Yamada, Kotaro Funakoshi, and Mikio Nakano. 2012. How Can We Live with Overconfident or Unconfident Systems?: A Comparison of Artificial Subtle Expressions with Human-like Expression, In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society (CogSci2012)* (2012), 1816-1821.
16. Takanori Komatsu, Kazuki Kobayashi, Seiji Yamada, Kotaro Funakoshi, and Mikio Nakano. 2015. Investigating Ways of Interpretations of Artificial Subtle Expressions Among Different Languages: A Case of Comparison Among Japanese, German, Portuguese and Mandarin Chinese, In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society (CogSci2015)* (2015), 1159-1164.
17. Takanori Komatsu, Kazuki Kobayashi, Seiji Yamada, Kotaro Funakoshi, and Mikio Nakano. 2017. Response Times When Interpreting Artificial Subtle Expressions are Shorter than with Human-like Speech Sounds, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI2017)*, ACM Press (2017), 3501-3505.
18. Bing Liu and Ian Lane. 2016. Joint Online Spoken Language Understanding and Language Modeling with Recurrent Neural Networks, In *Proceedings of the 17th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL2016)* (2016), 22-30.
19. Deepa Mathew. 2005. vSmileys: Imaging Emotions through Vibration Patterns, *In Proceedings of Alternative Access: Feelings & Games 2005*, 75-80.
20. Grégoire Mesnil, Yann Dauphin, Kaisheng Yao, Yoshua Bengio, Li Deng, Dilek Hakkani-Tur, Xiaodong He,

Larry Heck, Gokhan Tur, Dong Yu, and Geoffrey Zweig. 2015. Using Recurrent Neural Networks for Slot Filling in Spoken Language Understanding, *IEEE Transactions on Audio, Speech, and Language Processing 23* (3), (2015), 530-539.

21. Dylan Moore, Hamish Tennent, Nikolas Martelaro, and Wendy Ju. 2017. Making Noise Intentional: A Study of Servo Sound Perception, In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI2017)*, ACM Press (2017), 12-21.
22. Atsunori Ogawa and Atsushi Nakamura. 2012. Joint Estimation of Confidence and Error Causes in Speech Recognition, *Speech Communication 54*, 9 (2012), 1014-1028.
23. Huimin Qian, Ravi Kuber, and Andrew Sears. 2010. Towards developing perceivable tactile feedback for mobile devices, *International Journal of Human-Computer Studies 69*, 11 (2011), 705-719.
24. Huimin Qian, Ravi Kuber, and Andrew Sears. 2013. Developing tactile icons to support mobile users with situationally-induced impairments and disabilities, In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS2013)*, ACM Press (2013), Article No. 47.
25. Shafiq Ur Réhman and Li Liu. 2008. Vibrotactile Emotions on a Mobile Phone, In *Proceedings of IEEE International Conference on Signal Image Technology and Internet Based Systems (SITIS2008)* (2008), 239 - 243.
26. A. Riener, A. Ferscha, P. Frech, M. Hackl, and M. Kaltenberger. 2010. Subliminal vibro-tactile based notification of CO_2 economy while driving, In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI2010)*, ACM Press (2010), 92-101.
27. Jonghyun Ryu, Jaehoon Jung, and Seungmoon Choi. 2008. Perceived Magnitudes of Vibrations Transmitted Through Mobile Device, In *Proceedings of Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems 2008*, 139-140.
28. Bahador Saket, Chrisnawan Prasojo, Yongfeng Huang, Shengdong Zhao. 2013. Designing an Effective Vibration-Based Notification Interface for Mobile Phones, In *Proceedings of the 2013 conference on Computer supported cooperative work companion (CSCW2013)*, ACM Press (2013), 1499-1504.
29. Sichao Song and Seiji Yamada. 2017. Expressing Emotions through Color, Sound, and Vibration with an Appearance-Constrained Social Robot, In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI2017)*, ACM Press (2017), 2-11.
30. Diane Tam, Karon E. MacLean, Joanna McGrenere, and Katherine J. Kuchenbecke. 2013. The design and field observation of a haptic notification system for timing awareness during oral presentations, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI2013)*, ACM Press (2013), 1689-1698.
31. Seiji Yamada, Kazunori Terada, Kazuki Kobayashi, Takanori Komatsu, Kotaro Funakoshi, and Mikio Nakano. 2013. Expressing a Robot's Confidence with Motion-based Artificial Subtle Expressions, In *Extended Abstract of the SIGCHI Conference on Human Factors in Computing Systems (CHI2013)*, 1023-1028.