Augmenting Expressivity of Artificial Subtle Expressions (ASEs): Preliminary Design Guideline for ASEs

Takanori Komatsu
Meiji University
4-21-1 Nakano,
Tokyo 1648525, Japan
tkomat@meiji.ac.jp

Kazuki Kobayashi
Shinshu University
4-17-1 Wakasato,
Nagano 3808553, Japan
kby@shinshu-u.ac.jp

Seiji Yamada
National Institute of Informatics
2-1-2 Hitotsubashi,
Tokyo 1018430, Japan
seiji@nii.ac.jp

Kotaro Funakoshi
Honda Research Institute Japan, Co., Ltd.
8-1 Honcho, Wako 3510188, Japan
funakoshi@jp.honda-ri.com

Mikio Nakano
Honda Research Institute Japan, Co., Ltd.,
8-1 Honcho, Wako 3510188, Japan
nakano@hp.honda-ri.com

ABSTRACT
Unfortunately, there is little hope that information-providing systems will ever be perfectly reliable. The results of some studies have indicated that imperfect systems can reduce the users' cognitive load in interacting with them by expressing their level of confidence to users. Artificial subtle expressions (ASEs), which are machine-like artificial sounds to express the confidence information to users added just after the system's suggestions, were keenly focused on because of their simplicity and efficiency. The purpose of the work reported here was to develop a preliminary design guideline for ASEs in order to determine the expandability of ASEs. We believe that augmenting the expressivity of ASEs would lead reducing the users' cognitive load for processing the information provided from the systems, and this would also lead augmenting users' various cognitive capacities. Our experimental results showed that ASEs with decreasing pitch conveyed a low confidence level to users. This result were used to formulate a concrete design guideline for ASEs.

Author Keywords
Artificial subtle expressions (ASEs); Confidence; Design guideline; Inflection pattern.

ACM Classification Keywords
H.5.2. [User Interfaces]: Evaluation/methodology; J.4 [Social behavioral science]: Psychology.

General Terms
Experimentation; Human Factors; Design.

AH '14, March 07 - 09 2014, Kobe, Japan

http://dx.doi.org/10.1145/2582051.2582091

INTRODUCTION

Speech interface systems that, like Siri or Google speech recognition, can understand and express speech sounds are becoming common [9,25] because they enable users to obtain information while engaging in their own primary tasks without facing or manually operating the information-providing systems. A user can, for example, drive a car while hearing commands from a navigation system. For various reasons, however, such as noise in the sensors or the incompleteness of data, the reliability of such systems is often limited [3]. Some recent studies showed that displaying system confidence information increased a user's positive impressions of the systems [1,7]. For example, Antifakos et al. [1] showed that users easily adapt to systems if system confidence is displayed on a computer display. Expressing a system's level of confidence to the user is therefore becoming a crucial issue for systems communicating with humans.

Most people intending to express a system's level of confidence think of using human-like verbal expressions such as "probably," "definitely," or "95% confident." Komatsu et al. [21], however, showed that when the system's suggestions were wrong, expressing confidence levels by using human-like verbal expressions gave users a poorer impression of the system than did expressing these levels with machine-like artificial sound expressions. This phenomenon can be explained by the Uncanny Valley hypothesis. [1] That is, an artifact's human-like appearance or behavior makes users overestimate the artifact's abilities. Some investigators have therefore argued the dangerousness of the human-like verbal expressions and the effectiveness of the machine-like sound expressions. These

[1] http://www.popularmechanics.com/technology/engineering/robots/4343054

machine-like expressions are called "artificial subtle expressions" (ASEs) [23].

ASEs have usually been used after the system's verbal suggestions. Two types of ASEs (one is a flat sound; the other is a sound with a decreasing pitch) have been shown to be effective for expressing higher or lower confidence levels (Figure 1). There are, however, no concrete design guidelines for these ASEs, such as how steeply the pitch should decrease.

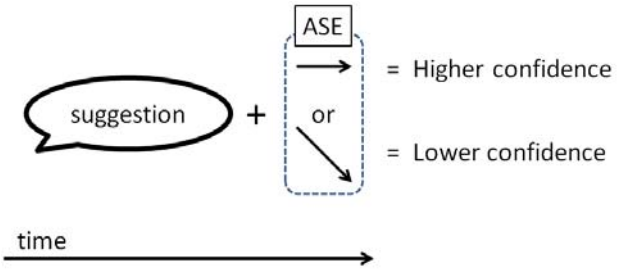

Figure 1. Artificial subtle expressions (ASEs).

The purpose of this study was to identify elements relevant to a design guideline for ASEs. We believe that augmenting the expressivity of ASEs would lead reducing the users' cognitive load for processing the information provided from the systems, and this would also lead augmenting users' various cognitive capacities. With regard to specific design elements, we focused on the following four factors (Figure 2).

1. Timing of ASEs (before or after the system's verbal suggestions).
2. Interval between suggestions and ASEs.
3. Inflection pattern of ASEs (flat, increasing pitch, or decreasing pitch).
4. Range of pitch variation (deeper or shallower change).

We used two experiments to investigate the effects of these four factors in determining how users interpret the ASEs. The results were used to formulate a design guideline for ASEs. The guideline's limitations and directions for future research on ASEs are clarified later in this paper.

RELATED WORK

Expressing confidence levels

Although there is little hope that information-providing systems will ever be perfectly reliable [2,16], people's interactions with imperfect ones have been analyzed only sparsely [27]. Recently, however, some studies have been focusing on displaying system confidence levels to users, and these studies have shown that it is actually effective for various aspects of interaction between humans and systems [11,17,18,26]. For example, Cai et al. [7] showed how expressing to users the levels of confidence that the presented information is accurate plays an important role in improving the users' performance as well as their impressions. Furthermore, Antifakos et al. [1] showed that users adapt to systems easily if system confidence is displayed. Horvitz and Barry [18] proposed a context-aware system that can estimate the expected value of revealed information to enhance computer displays for time-critical applications. Expressing the system's confidence to users is therefore an important requirement for user interfaces.

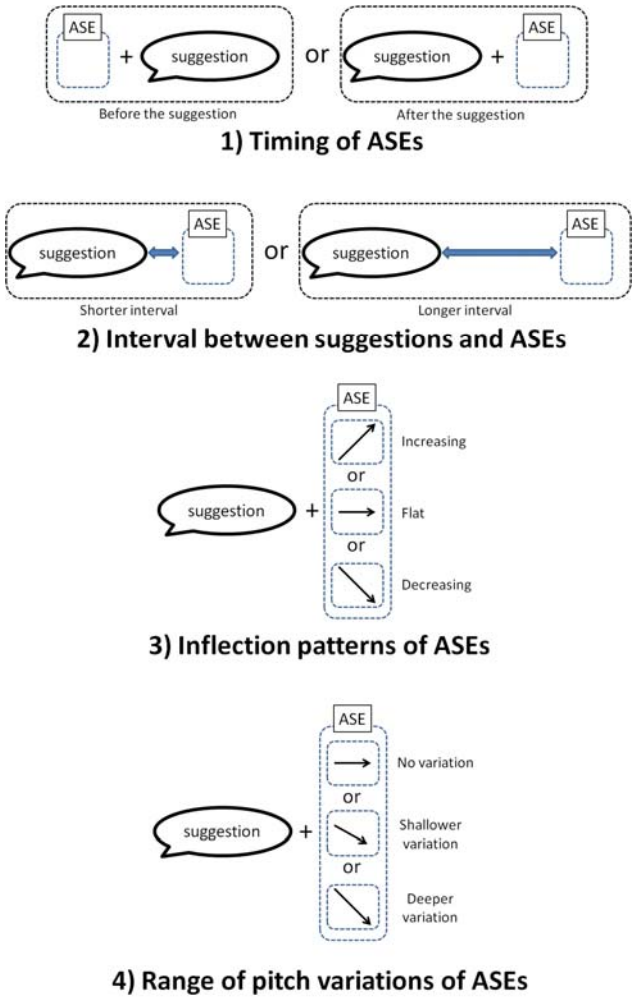

Figure 2. Four design elements for ASEs.

In the studies described above, the confidence level was expressed by using human-like expressions conveyed by speech sounds from speakers or by using words on computer displays. Komatsu et al. [20,21,23], in contrast, proposed that the system's confidence level be conveyed to users by using machine-like artificial sounds they called artificial subtle expressions (ASEs). They prepared two simple artificial wave sounds (sounding like beep sounds) they used as ASEs—one was a flat sound (flat ASE), and the other was a sound with a decreasing pitch (decreasing ASE)—and these ASEs were added just after the end of system's verbal suggestions. They then showed that suggestions with decreasing ASEs conveyed a low system confidence level to users intuitively [23]. They also showed

that expressing the levels of confidence by using human-like expressions gave users a poorer impression of the system than did expressing these levels by using machine-like expressions when the system's suggestions were inconsistent [21].

Although these studies showed the effectiveness ASEs, they investigated only two types of ASEs (flat and decreasing ASEs), and currently there are no concrete design guidelines for ASEs. That is, neither what range of decreasing pitch can work for decreased ASEs nor how long the silence between a suggestion and an ASE should be is clear.

Simple auditory signals

There have been several studies on the effectiveness of using simple auditory signals to convey information to users. These auditory signals can be classified into two categories: earcons [4–6] and auditory icons [12,13,30]. Blattner et al. [4] defined earcons as "nonverbal audio messages used in user-computer interfaces to provide information to users about some computer objects, operations, or interactions." Brewster et al. [5,6] said that "because of their flexibility, earcons can be easily designed to extend any object, operation, and interaction by means of their proposed guidelines." Gaver [12,13] introduced the concept of using auditory icons of everyday sounds to convey information about computer events through analogy with everyday events. For example, the sound of shattering dishes can be represented by the drop of a virtual object into a virtual recycle bin [12]. Gaver argued that these auditory icons are an intuitively accessible way to use sound to organize information for users.

Although earcons and auditory icons have been shown to be effective in conveying information to users, both of them have their limitations. With earcons, because of the arbitrary mappings between sounds and communicated information, users have to memorize the mappings to correctly understand the meaning of the sounds [6]. With auditory icons, metaphoric mappings are not always easy to find its actual meaning on user's operations [12], so it is difficult to design appropriate auditory icons for all the information a computer system has to communicate to its users.

Actually, these earcons and auditory icons were basically not designed for conveying the system's confidence level as well as ASEs.[2] Moreover, earcons and auditory icons consist of a lot of elements (e.g., earcons are "musical tones composed of short, rhythmic sequences of pitches with variable intensity, timbre, and register" [9,25]), so their design guideline tends to be quite ambiguous and abstract. ASEs, however, consist of just a few elements, like "timing," "interval," "inflection pattern," and "range of pitch variations," so the investigation of this study that explores the effects of these elements can clarify a design guideline for ASEs.

Expressivity of simple information

Recent electric appliances can show rich information to users through their high-resolution displays or stereo sound systems. Showing too much information, however, may overwhelm the users' cognitive resources [19,24]. Simpler ways of communicating information are therefore getting attention. For example, Harrison et al. [14] showed that different blinking patterns of mobile phones' small LEDs were interpreted differently by users and that they succeeded in informing users of the states of the mobile phones, such as low battery and notifications. And Harrison et al. [15] proposed Kinecticons, which are a collection of graphical icons with simple motions. Kinecticons successfully conveyed various kinds of information to users.

It is then worthwhile to explore which kinds of simple auditory signals can inform users of the states of appliances as well as blinking LEDs and Kinecticon can. Exploring possible variations of ASEs should therefore be also worthwhile. It is then expected that the result of this investigation will contribute to proposing a specific design guideline for ASEs.

EXPERIMENT 1: TIMING AND INFLECTION

To explore a design guideline for ASEs, we focused on four factors as design elements 1) the timing of ASEs, 2) the interval between suggestions and ASEs, 3) the inflection patterns of ASEs, and 4) the range of pitch variations. We first investigated how two of them, the timing of ASEs and the inflection pattern of ASEs, affected participants' interpretations of the ASEs. The complexity of the statistical analysis was reduced by investigating the effects of the other two factors (the interval between the suggestions and ASEs and the range of pitch variations) in another experiment (Experiment 2) based on the results of Experiment 1.

In both experiments we evaluated two types of participants' behaviors that we think reflect their way of interpreting ASEs: one was how long it took the participants to react to the presented stimuli (reaction time), and the other was how often the participants rejected the system's suggestions (rejection count).

Environment

We used a "driving treasure hunting game" video game as the experimental environment (Figure 3). This game was the same one used by Komatsu et al. [21]. In this game, the game image scrolls forward on a straight road, as if the participant were driving a car with a navigation system, with three small 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

[2] Substantial differences between earcons, auditory icons, and ASEs were already discussed in Komatsu et al. [22].

participants encounter 36 sets of mounds (36 trials). The goal of the participants is to get as many coins as possible. The coin was randomly among the three mounds. In each trial, the navigation system to the left of the driver's seat (circled in the top image in Figure 3) told them which mound it expected the coin to be in by using speech. The participants could freely accept or reject the navigation system's suggestions. The experimenter never mentioned or explained the ASEs to the participants. In each trial, the participants selected one of the three mounds but were not told whether or not it had the coin (the selected mound just showed a question mark and a closed treasure box, as depicted in the middle image in Figure 3). The participants were informed of their total number of coins only after they finished all 36 trials.

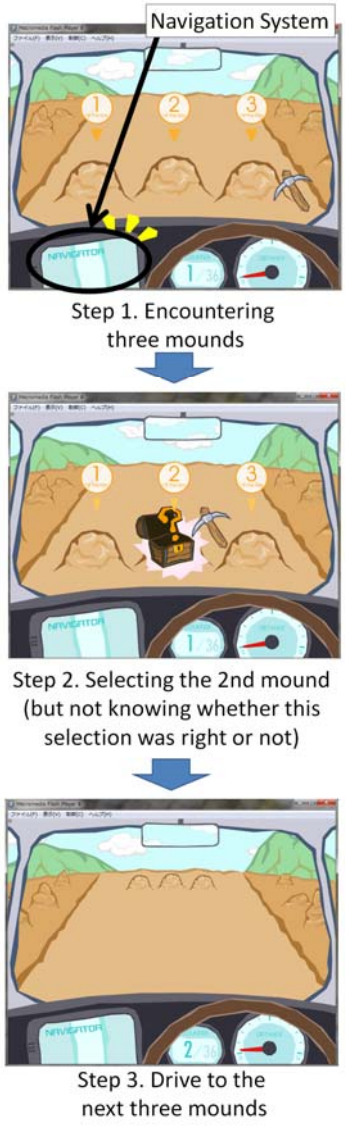

Figure 3. Driving treasure hunting video game.

At a glance, this game environment seems to be too simple and too abstract be used to investigate the effectiveness of ASEs and their relevance to concrete applications, but this game succeeded in realizing a primal situation of user interfaces: does the user accept the system's suggestions or not? So the results of this experiment should be relevant to various kinds of concrete applications.

Stimuli

The navigation system used Japanese speech sounds to suggest to the participants the expected location of the coin: "ichi-ban (no. 1)," "ni-ban (no. 2)," and "san-ban (no. 3)." These speech sounds were created by adding robotic voice effects to the recorded speech sounds of one of the authors in order to eliminate human-like impressions from the speech sounds.

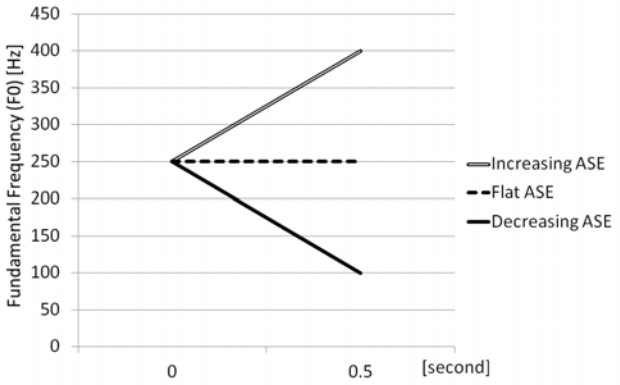

Figure 4. Flat, increasing, and decreasing ASEs.

We then prepared three different inflection types of the ASEs, which were triangle wave sounds 0.5 seconds in duration with different inflection patterns, to express the confidence level of the system's suggestions (Figure 4). Yuan et al [31] reported that the average speaking rates in English conversation are about 150 words per minutes (i.e., about 0.4 seconds per word), so we determined the duration of ASEs is 0.5 seconds with excluding the effects of articles, such as "a" or "the."

- **Flat ASEs**: Onset F0 (fundamental frequency) was 250 Hz, and end F0 was also 250 Hz.
- **Decreasing ASEs**: Onset F0 was 250 Hz, and end F0 was 100 Hz.
- **Increasing ASEs**: Onset F0 was 250 Hz, and end F0 was 400 Hz.

The F0 ranges of human's speech sounds are usually centered around values ranging from 500 Hz (male voice) to 1,000 Hz (female voice), and the center of the F0 range of the ASEs was set to 250 Hz so that the ASEs F0 range would not overlap with the human's F0 range and make a human-like impression.

We then combined the system's suggestions with these ASEs with different timings: either the suggestions were

first or the ASEs were first. Therefore there were 18 stimuli (3 suggestions × 3 different inflections of ASEs × 2 different timings) to express to the participants. Regardless of the order, the interval between the suggestions and the ASEs was 0.2 seconds. Therefore the experimental design in this experiment was a 2 × 3 within-participants design; that is, within-factor #1 was the timing of ASEs (before or after the suggestions), and within-factor #2 was the inflection pattern of the ASEs (flat, decreasing, or increasing), as shown in Figure 5.

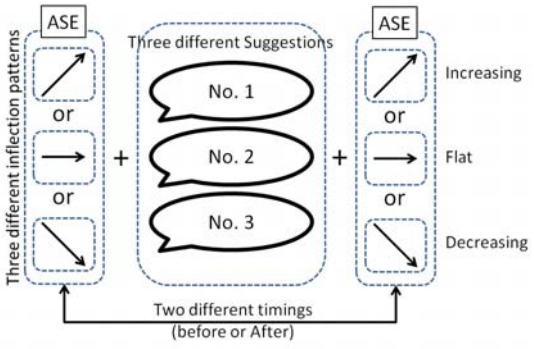

Figure 5. Experimental design of Experiment 1.

Procedure

Twenty-one Japanese university students (15 men and 6 women, 19–24 years old) participated. In the 36 trials, all participants experienced the 18 stimuli two times in random order.

We used a web-based experiment system to record the participants' behaviors in regard to selecting which mound contained the coin according to the given suggestions (rejection count) and how long it took from the beginnings of presenting the stimuli until the participants' selected the mounds (reaction time). First, the system displayed a consent form and the instructions to the experiments. Before starting the video game, the participants were asked to listen to the test sounds via a speaker or headphones and to adjust the sound volume to a comfortable level. [3] Afterwards, they played the driving treasure hunting video game.

Results

Rejection count

First, to investigate the effects of the timing and inflection patterns of the ASEs on the participants' behavior in terms of how often they rejected the system's suggestions, we counted how many of the system's suggestions the participants rejected. For all participants, the average rejection counts are summarized in Table 1.

	Before	After
Decreasing ASEs	2.10 (SD = 2.29)	2.33 (SD = 2.42)
Flat ASEs	1.48 (SD = 2.06)	1.05 (SD = 1.56)
Increasing ASEs	1.10 (SD = 1.69)	0.95 (SD = 1.59)

Table 1. Rejection counts for combinations of two factors (timing of ASEs and inflection patterns of ASEs). Note that for each cell the maximum rejection count was 6.

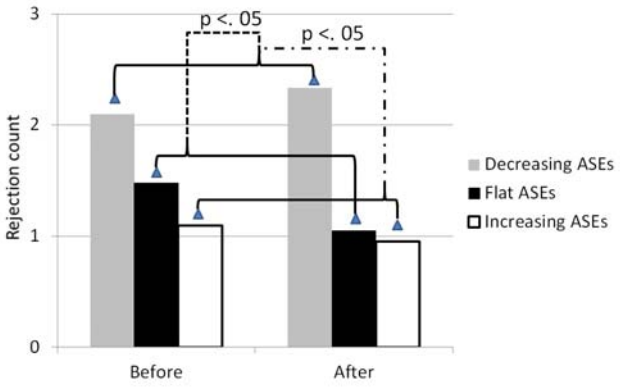

Figure 6. Rejection counts for combinations of two factors (timing of ASEs and inflection patterns of ASEs) and the significant differences between them.

These rejection counts were then analyzed by using a 2 × 3 within-participant plan ANOVA [within independent variable #1: timing of ASEs (before or after the suggestions), within independent variable #2: inflection patterns of ASEs (flat, increasing, or decreasing), dependent variable: rejection count]. The results of the ANOVA showed that there were no significant differences in the interaction effect [$F(2,40) = 1.76$, n.s.] or in the main effect of within independent variable #1 (timing of ASEs) [$F(1,20) = 0.41$, n.s.], but there was a significant difference in the main effect of within independent variable #2 (inflection patterns) [$F(2,40) = 5.51$, $p < .01$]. The simple main effect of the within independent variable #2 was then analyzed by using the LSD test [MSe = 3.03], and the results showed that there were significant differences between the decreasing ASEs and the flat ASEs [$p < .05$] and between the decreasing ASEs and the increasing ASEs [$p < .05$], but there was no significant difference between the flat ASEs and increasing ASEs (Figure 6).

Thus we observed that the timing of the ASEs did not affect the participants' rejection counts but the inflection patterns did. Specifically, the suggestions with decreasing ASEs showed higher rejection counts than did suggestions with flat or increasing ASEs.

[3] Komatsu and Nagasaki [22] already showed that the power information (sound volume) of ASEs does not affect the participants' ways of interpreting the ASEs. Therefore we did not control the sound volume of the ASEs.

Reaction Time

Next, to investigate the effects of these two factors of ASEs on the participants' behaviors in terms of how long it took from the beginnings of the system presenting stimuli until the participants' to select a mound, we focused on the reaction time. For all participants, the average reaction times are summarized in Table 2.

	Before	After
Decreasing ASEs	2.60 (SD = 0.76)	2.43 (SD = 1.00)
Flat ASEs	2.73 (SD = 0.65)	2.41 (SD = 1.10)
Increasing ASEs	2.55 (SD = 0.75)	2.40 (SD = 1.37)

Table 2. Reaction times [seconds] for combinations of two factors (timing of ASEs and inflection patterns of ASEs).

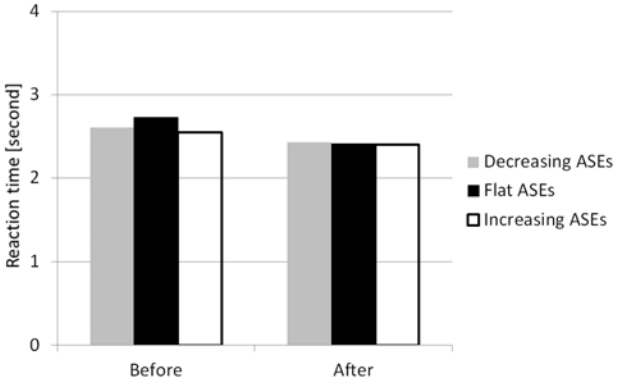

Figure 7. Reaction times for combinations of two factors (timing of ASEs and inflection patterns of ASEs). None differed significantly from any of the others.

These reaction times were then analyzed by using a 2 × 3 within-participant plan ANOVA [within independent variable #1: timing of ASEs (before or after the suggestions), within independent variable #2: inflection patterns of ASEs (flat, increasing, or decreasing), dependent variable: reaction time]. The results of the ANOVA showed that there were no significant differences in the interaction effect [$F(2,40) = 0.19$, n.s.], in the main effect of within independent variable #1 (timing of ASEs) [$F(1,40) = 2.81$, n.s.] and #2 (inflection patterns) [$F(2,40) = 0.19$, n.s.]. Thus we observed that neither the timing nor inflection patterns of ASEs affected the participants' reaction time (Figure 7).

Summary of Experiment 1

In this experiment, we focused on the effects of two factors (timing of ASEs and inflection patterns of ASEs) on users' rejection counts and reaction time. The results can be summarized as follows.

- **Timing of ASEs**: No effect on reaction time or rejection count.
- **Inflection pattern**: No effect on reaction time, but significant effect on rejection count. The suggestions with decreasing ASEs showed higher rejection counts than did those with flat or increasing ASEs.

EXPERIMENT 2: INTERVAL AND RANGE OF PITCH VARIATION

In Experiment 1 we focused on two of the four factors: namely, the timing of ASEs and the inflection pattern of ASEs. In Experiment 2 we focused on the other two: the interval between the suggestions and ASEs and the range of pitch variations. We investigated how these two factors affected the participants' interpretation of the ASEs. The experimental environment was the same driving treasure hunting video game used in Experiment 1.

Stimuli

The navigation system used the same speech sounds used in Experiment 1. We then prepared three different types of ASE, which were triangle wave sounds 0.5 seconds in duration with different ranges of pitch variation. As the result of Experiment 1showed that the increasing ASEs had the same effects that the flat ASEs did, In Experiment 2 we used flat ASEs and decreasing ASEs with two different ranges of pitch variation (Figure 8).

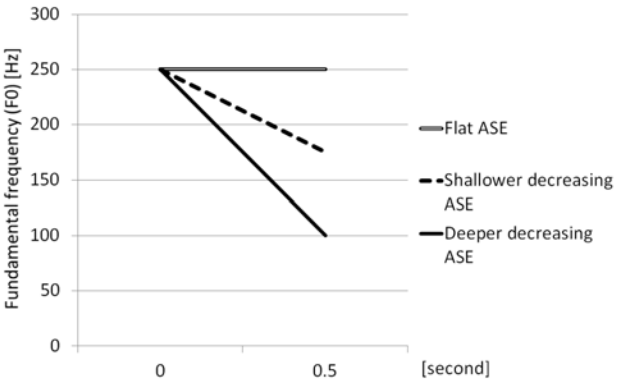

Figure 8. Flat, deeper decreasing, or shallower decreasing ASEs.

- **Flat ASEs**: Onset F0 (fundamental frequency) was 250 Hz, and end F0 was 250 Hz (as in Experiment 1).
- **Deeper decreasing ASEs**: Onset F0 was 250 Hz, and end F0 was 100 Hz (same as the decreasing ASEs used in Experiment 1).
- **Shallower decreasing ASEs**: Onset F0 was 250 Hz, and end F0 was 175 Hz. This was a newly prepared one.

We then combined the system's suggestions with these ASEs with different intervals between the end of the suggestions and the beginnings of ASEs: 0.2 or 1.0 seconds. Campione and Véronis [8] reported that the most of silent pause durations are from 0.2 to 1.0 second based on the

analysis of about 6,000 pauses in 5.5 hours speeches in five languages, so we determined the interval between the suggestions and ASEs were 0.2 and 1.0 seconds.

Therefore there were 18 stimuli (3 suggestions × 3 ranges of pitch variations × 2 different intervals) to express to the participants. The timing of the ASEs was set after the system's suggestions because the timing of the ASEs did not have any effect on the interpretation of the ASEs (Figure 9). Therefore, the experimental design in Experiment 2 was a 2 × 3 within-participant design; that is, within-factor #1 was the interval between the end of the suggestions and the beginning of the ASEs (0.2 or 1.0 seconds), and within-factor #2 was the range of pitch variation (flat, deeper decreasing, or shallower decreasing).

Twenty Japanese university students (17 men and 3 women; 21 - 27 years old) participated. Among 36 trials, all participants experienced the 18 stimuli two times in random order. These participants did not participate in the former Experiment 1.

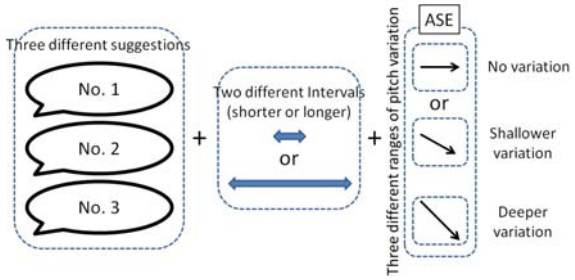

Figure 9. Experimental design of Experiment 2. .

Results

Rejection count

For all participants, the average rejection counts are summarized in Table 3. These rejection counts were analyzed by using a 2 × 3 within-participant plan ANOVA [within independent variable #1: interval between suggestions and ASEs (0.2 or 1.0 seconds), within independent variable #2: range of pitch variation (flat, deeper decreasing, or shallower decreasing), dependent variable: rejection counts]. The results of the ANOVA showed that there were no significant differences in the interaction effect [$F(2,38) = 0.00$, n.s.] and in the main effect of within independent variable #1 (interval between suggestions and ASEs) [$F(1,19) = 1.71$, n.s.], but there was a significant difference in the main effect of within independent variable #2 (range of pitch variation) [$F(2,38) = 32.3$, $p < .01$]. The simple main effect of the within independent variable #2 was then analyzed by using the LSD test [MSe = 4.17], and the results showed that there were significant differences between the flat ASEs and deeper decreasing ASEs [$p < .05$] and between the flat ASEs and shallower decreasing ASEs [$p < .05$], but there were no significant differences between the deeper and shallower decreasing ASEs (Figure 10).

Thus we observed that the interval between the suggestions and ASEs did not affect the participants' rejection counts, while the range of pitch variations did affect their counts. Specifically, the suggestions with both deeper and shallower decreasing ASEs showed higher rejection counts the flat ASEs did.

	Shorter interval	**Longer interval**
Deeper decreasing ASEs	3.85 (SD = 2.26)	4.00 (SD = 2.00)
Flat ASEs	0.60 (SD = 0.97)	0.75 (SD = 0.99)
Shallower decreasing ASEs	3.70 (SD = 2.03)	3.85 (SD = 2.08)

Table 3. Rejection counts for combinations of two factors (interval of ASEs and range of pitch variation of ASEs).

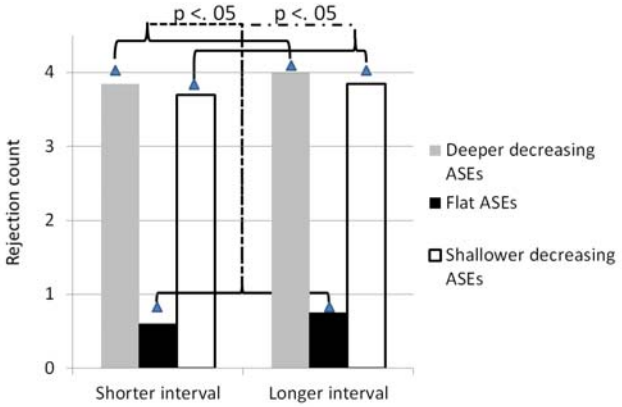

Figure 10. Rejection counts for combinations of two factors (interval of ASEs and the range of ASE pitch variation) and the significant differences between them.

Reaction Time

	Shorter interval	**Longer interval**
Deeper decreasing ASEs	2.70 (SD = 1.00)	3.04 (SD = 1.05)
Flat ASEs	2.45 (SD = 0.87)	3.32 (SD = 1.74)
Shallower decreasing ASEs	2.50 (SD = 0.60)	3.32 (SD = 1.74)

Table 4. Reaction times [second] for combinations of two factors (interval of ASEs and range of ASE pitch variation).

For all participants, the average reaction time is summarized in Table 4. These reaction times were then analyzed by using a 2 × 3 within-participant plan ANOVA [within independent variable #1: interval between suggestions and ASEs (0.2 or 1.0 seconds), within independent variable #2: range of pitch variations (flat, deeper decreasing, or shallower decreasing), dependent variable: reaction time]. The results of the ANOVA showed

that there were no significant differences in the interaction effect [F(2,38) = 1.40, n.s.] or in the main effect of within independent variable #2 (range of pitch variations) [F(2,38) = 0.05, n.s.], but there was a significant difference in the main effect of within independent variable #1 (interval of ASEs) [F(1,19) = 18.59, p < .01.] (Figure 11).

Thus we observed that the range of pitch variations did not affect the participants' reaction times but the interval between suggestions and ASEs did. Specifically, the longer interval between suggestions and ASEs showed longer reaction times than the shorter interval did.

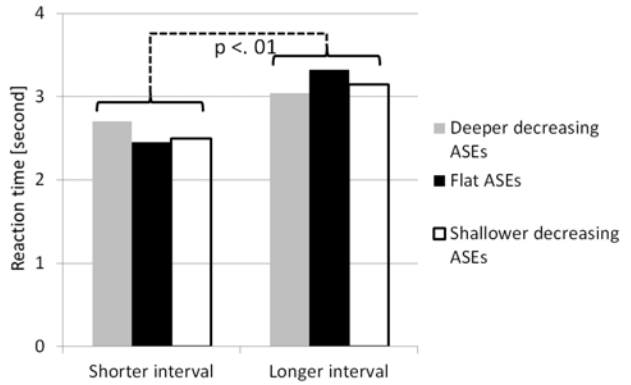

Figure 11. Reaction times for combinations of two factors (interval of ASEs and range of ASE pitch variation) and the significant differences between them.

Summary of Experiment 2

In Experiment 2 we focused on the effects of two factors (the interval between the end of suggestions and the beginnings of ASEs, and range of ASE pitch variation) on users' rejection counts and reaction time. The results can be summarized as follows.

- **Interval between suggestion and ASEs**: Significant effects on reaction time but not on rejection counts. Specifically, the longer interval showed longer reaction times.
- **Range of pitch variations**: No effects on reaction time but significant effects on rejection count. Specifically, the suggestions with both deeper and shallower decreasing ASEs showed higher rejection counts than did those flat ASEs.

DISCUSSION

Proposal of a design guideline

First we summarize the results of Experiments 1 and 2 in terms of the two dependent variables.

Reaction time

Although three of the four independent variables (timing of ASEs, inflection patterns of ASEs, and range of pitch variation) did not affect the participants' reaction time, the other independent variable (the interval between the suggestion and the ASE) did. Specifically, the longer interval showed a longer reaction time than the shorter interval did. However, the reaction time was counted from the beginning of the presentation of the verbal suggestion, so it is obvious and trivial that the reaction time for the stimuli with a longer interval were longer than those for the stimuli with a shorter interval. The only thing we could infer from this phenomena is that the participants unintentionally listened to the whole stimuli (verbal suggestions with ASEs) and used the given ASEs to select the mound, even though we did not or explain the ASEs to the participants before the experiments or even mention them.

Rejection count

Although two of the four independent variables (the timing of the ASEs and the interval between the suggestion and the ASEs) did not affect the participants' rejection count, the other two (inflection pattern and range of pitch variation) did. Specifically, ASEs that decreased by any amount (deeper decreasing or shallower decreasing) showed a higher rejection count than did increasing or flat ASEs. This result is in accordance with a conclusion of Edworthy and Hards [10]: that "small changes in F0 could produce the same change in perceived urgency as could a large one because the salient feature of a pitch change is its direction (up or down) rather its magnitude." So this result clearly showed that the interpretation of the ASEs was quite robust even though different timings or intervals were implemented in the system.

On the basis of the above summary of the experiments, we propose the following design guideline for ASEs.

1 To express high confidence in a system's suggestions to users, either before or after the suggestions simply add a triangle wave sound (it sounds like a "beep sound") 0.5 seconds in duration with a constant or increasing pitch. To express low confidence, either before or after the suggestions simply add a triangle wave sound 0.5 seconds in duration with a decreasing pitch.

   1.1 Whether the ASEs should be added after or before the suggestions depends on your choice or task. If you waver between "before the suggestion" and "after the suggestion," we recommend adding the ASEs after the suggestions. This is because, as shown in Figure 6, the experimental results presented a recognizably larger gap in rejection count between decreasing and flat ASEs "after" compared with "before," although ANOVA did not detect a statistically significant difference between the gaps in the two levels for "before" and "after."

2 You do not need to care about the length of the interval between the suggestion and ASEs. We however recommend adding a shorter interval between

the suggestion and ASEs because a longer interval will cause a user to have a longer reaction time, and this might make users feel frustrated.

3 You do not need to care about the pitch range of decreasing ASEs. The only thing that you need to care about is whether users can recognize that the pitch is decreasing when they them.

This guideline is so simple and flexible that it is quite easy to implement the ASEs in various kinds of interface systems that are required to give suggestions to users.

Limitations and future directions

In this study, we designed a guideline for ASEs by using a gaming environment in which participants simply needed to accept or reject the system's suggestions. The ASEs needed to convey only high or low confidence to the users, so the design guideline became rather simple. Although conveying high/low confidence to users is an abstract gaming task, it is quite important and effective for systems that need to tell users what they should do next, such as car navigation systems giving route guidance like "turn left" or "enter highway #I-8."

Currently, we are wondering whether this simple guideline is also effective for systems that are required to give much more complex information to users, such as those that need to express a degree of confidence level not simply expressed with "higher" or "lower," like information retrieval systems [29] or recommendation systems [28]. In such more complex systems, we speculate that not only decreasing ASEs but other inflection patterns of ASEs or ranges of pitch variation of ASEs will have specific meanings (although the different inflection patterns or ranges of pitch variation were interpreted as having the same meaning in this simple environment in this study). We are now planning to use other kinds of gaming environments to handle much more complex and flexible interaction with users. The results of such experiments in these more complex systems should expand the application range of ASEs.

The other concern we have to focus on is whether or not the participants' mother tongue affects their interpretation of ASEs. Currently, we believe that interpretation of the ASEs is not affected by the participants' mother tongue because the ASEs consist of quite simple artificial sounds and thus exclude most linguistic information. However, it can be said that the ASEs still include paralinguistic information, so clarifying this issue would be worthwhile. Currently, we are planning to conduct the same experiment with participants who are in several parts of Europe. The results of this experiment should also contribute to comprehending the application range of ASEs.

CONCLUSIONS

To reduce users' cognitive load when interacting with imperfect systems, some investigators have argued that these systems should express their level of confidence to users. Artificial subtle expressions (ASEs) were keenly focused on because of their simplicity and efficiency for expressing confidence information to users. Up to now, however, only two types of ASEs have been investigated (one is a flat sound, and the other is a sound with a decreasing pitch). The purpose of this study was then to develop a preliminary design guideline for ASEs in order to determine the expandability and application range of ASEs. We believe that augmenting the expressivity of ASEs would lead reducing the users' cognitive load for processing the information provided from the systems, and this would also lead augmenting users' various cognitive capacities. The design elements we focused on were the following four factors: the timing of ASEs, the interval between suggestions and ASEs, the inflection patterns of ASEs, and the range of pitch variation in ASEs.

Experiments showed that ASEs with decreasing pitch, regardless of the range of the variation (deeper decreasing or shallower decreasing) conveyed a lower confidence level to users than did ASEs with increasing or constant pitch. We therefore proposed an ASE design guideline that states that a triangle wave sounds 0.5 seconds in duration with a decreasing pitch after or before a suggestion can convey a system's low confidence level to users and that neither the length of the interval between suggestions and ASEs nor the range of the decreasing pitch of the ASEs interferes with the interpretation of ASEs. Although we have to carefully investigate whether this guideline can be used in much more complex systems and whether the ways ASEs are interpreted are affected by one's mother tongue, we believe this guideline is useful for many kinds of systems as it is quite simple and has a high flexibility.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number 25330319, and by joint research with Honda Research Institute Japan, Shinshu University, Meiji University and National Institute of Informatics, Japan.

REFERENCES

1. Antifakos, S., Kern, N., Shiele, B., and Schwaninger, A. Towards Improving Trust in Context Aware Systems by Displaying System Confidence, In *Proc. MobileHCI'05*, ACM Press (2005), 9-14.
2. Bellotti, V., and Edwards, K. Intelligibility and accountability: Human considerations in context-aware systems. *Human-Computer Interaction 16*, 2 (2001), 193–212.
3. Benzeghibaa, M., De Moria, R., Derooa, O., Dupont S., Erbesa, T., Jouveta, D., Fissorea, F., Lafacea, P., Mertinsa, A., Risa, C., Rosea, R., Tyagia, V., and Wellekensa, C. Automatic speech recognition and speech variability: A review, *Speech Communication 49*, 10-11 (2007), 763-786.

4. Blattner, M. M., Sumikawa, D. A. and Greenberg, R. M. Earcon and Icons: Their Structure and Common Design Principles. *SIGCHI Bull 21*, 1, ACM Press (1989), 123-124.
5. Brewster, S. A. Using Non-speech Sounds to Provide Navigation Cues. *ACM Transactions on Computer-Human Interaction 5*, 2, ACM Press (1998), 224-259.
6. Brewster, S.A., Wright, P.C. and Edwards, A.D.N. Experimentally derived guidelines for the creation of earcons. In *Adjunct Proc of HCI95* (1995).
7. Cai, H. and Lin, Y. Tuning Trust Using Cognitive Cues for Better Human-Machine Collaboration, In *Proc. HFES2010* (2010), 2437-2441(5).
8. Campione, E., and Véronis, J. A Large-Scale Multilingual Study of Silent Pause Duration, In *Proc. Speech Prosody 2002* (2002), 199 – 202.
9. Cohen, M. H., Giangola, J. P., and Balogh, J. *Voice User Interface Design*, Addison-Wesley, MA, USA, 2004.
10. Edworthy, J. and Hards, R. Learning Auditory Warnings: The Effects of Sound Type, Verbal Labeling and Imagery on the Identification of Alarm Sounds. *International Journal of Industrial Ergonomics 24*, 5 (1999), 603-618.
11. Feng, J., and Sears, A. 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.
12. Gaver, W. W. Auditory Icons: Using Sound in Computer Interfaces. *Human-Computer Interaction 2*, 2 (1986), 167-177.
13. Gaver, W. W. The SonicFinder: An Interface That Uses Auditory Icons. *Human-Computer Interaction 4*, 1 (1989), 67-94.
14. Harrison, C., Horstman, J., Hsieh, G. and Hudson, S. E. Unlocking the Expressivity of Point Lights, In *Proc. CHI'12*, ACM Press (2012), 1683-1692.
15. Harrison, C., Hsieh, G., Willis, K. D. D., Forlizzi, J. and Hudson, S. E. Kinecticons: Using Iconicgraphic Motion in Graphical User Interface Design, In *Proc. CHI'11*, ACM Press (2011), 1999-2008.
16. Higashinaka, R., Sudoh, L., and Nakano, M.: Incorporating Discourse Features into Confidence Scoring of Intention Recognition Results in Spoken Dialogue Systems, *Speech Communication 48*, 3-4 (2006), 417–436.
17. Horvitz, E. Principles of mixed-initiative user interfaces, In *Proc. CHI'99*, ACM Press (1999), 159–166.
18. Horvitz, E., and Barry, M. Display of information for time–critical decision making, In *Proc. 11th Conf. on Uncertainty in Artificial Intelligence*, Morgan Kaufmann (1995), 296–305.
19. Keller, J. M. Development and use of the ARCS model of instructional design, *Journal of Instructional Development 10*, 3 (1987), 2-10.
20. Komatsu, T., Kobayashi, K., Yamada, S., Funakoshi, K. and Nakano, M. Effects of Different Types of Artifacts on Interpretations of Artificial Subtle Expressions (ASEs), In *CHI'11 Ext. Abs* (2011), 1249-1254.
21. Komatsu, T., Kobayashi, K., Yamada, S., Funakoshi, K. and Nakano, M. How Can We Live with Overconfident or Unconfident Systems?: A Comparison of Artificial Subtle Expressions with Human-like Expression, In *Proc. CogSci2012* (2012), 1816-1821.
22. Komatsu, T., and Nagasaki, Y. Can we estimate the speaker's emotional state from her/his prosodic features? : Effects of F0 contour's slope and duration on perceiving disagreement, hesitation, agreement and attention, In *Proc. ICA2004* (2004), 2227-2230.
23. Komatsu, T., Yamada, S., Kobayashi, K., Funakoshi, K. and Nakano, M. Artificial Subtle Expressions: Intuitive Notification Methodology for Artifacts, In *Proc. CHI'10*, ACM Press (2010), 1941-1944.
24. Krug, S. *Don't make me think!*, New Riders, CA, USA, 2005.
25. Nass, C. and Brave, S. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*, The MIT Press, MA, USA, 2005.
26. Ogawa, A., and Nakamura, A. Joint estimation of confidence and error causes in speech recognition, *Speech Communication 54*, 9 (2012), 1014-1028.
27. Parasuraman, R. Human use and abuse of Automation. In M. Mouloua and J. Koonce, editors, *Human-Automation Interaction: Research and Practice*. Erlbaum Associates, 1997.
28. Re Roure, D. C. and Shadbolt, N. R. Capturing knowledge of user preferences: Ontologies in recommender systems, In *Proc. K-CAP'01*, ACM Press (2001), 100-107.
29. Sanderson, M. and Zobel, J. Information retrieval system evaluation: effort, sensitivity, and reliability, In *Proc. SIGIR'05*, ACM Press (2005), 162 - 169.
30. Walker, B. N. and Kramer, G. Mappings and Metaphors in Auditory Displays: An Experimental Assessment. *ACM Transaction on Applied Perception 2*, 4, ACM Press (2005), 407-412.
31. Yuan, J., Liberman, M., and Cieri, C. Towards an integrated understanding of speaking rate in conversation, In *Proc. Interspeech 2006* (2006), 541-544.