A DESIGN GUIDE-LINE OF AUDITORY DISPLAY FOR ELECTRIC APPLIANCE

Takanori Komatsu

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

Eiji Hayashi

Carnegie Mellon University,
5000 Forbes Ave,
Pittsburgh, PA 15213, United States
ehayashi@cs.cmu.edu

ABSTRACT

The auditory channel is important for communication between computers and users because of its properties, such eye-free communication and strong attention grabbing properties. However, interpreting the meanings of sounds is not a trivial task. Users have to learn and memorize the mapping between sounds and their meanings for each device. Therefore, as the number of devices increases, this becomes challenging for users. To mitigate the challenge, it is desirable to use sounds that users can understand intuitively. Thus, investigating the intuitiveness of sounds is of significant interest. In this work, we investigated 2,012 sounds consisting of 48 earcons, 80 auditory icons, and 1,884 beep sequences through a series of user studies using Amazon Mechanical Turk as well as a lab study that validated the results of the Mechanical Turk studies. The results provided a guideline for designing sounds that users might understand more intuitively.

1. INTRODUCTION

The auditory channel is important for computers to communicate information to users [20]. Although the development of graphical user interfaces allows computers to communicate more and more information to users through the visual channel with a higher bandwidth, the auditory channel still has its advantages over the visual channel, such as eye-free communication and its strong attention grabbing properties [6]. This is especially true for mobile phones. Because there are many cases where mobile phones initiate interaction with users, such as when push notifications are received from servers, if users are not looking at the displays, communicating information through the visual channel is infeasible. Therefore, devices have to rely on the auditory channel first to communicate. Simple devices, such as digital audio recorders or microwaves, offer other examples of devices that rely on the auditory channel in communicating with their users. These devices, in most cases, have very limited visual displays, such as single line displays or even just a few LEDs. Yet, these devices have multiple informational states that they have to communicate to users.

The environment is typically full of sounds played by multiple devices. However, interpreting the meaning of these sounds is not a trivial task. Most users would have experienced challenges like these: “I am sure I heard a short melody from the next room, but I could not figure out which appliance played that sound” or “My washing machine beeps during operation, but I do not understand what it means.” These problems are due to the fact that communication via sounds strongly relies on users’ knowledge [9]. Users have to learn the mapping between a sound and its meaning. However, as the number of devices around us increases, learning and memorizing the mapping for each device is becoming increasingly difficult for users. Therefore, it is of great importance to use intuitive sounds when communicating informational states to users so that users can interpret the meaning of these sounds without intense learning.

In this paper, we investigated the intuitiveness of sounds used by electric appliances. Specifically, we evaluated 2,012 auditory signals consisting of 48 earcons [1,3], 80 auditory icons [8,23], and 1,884 beep sequences in terms of their intuitiveness. We adopted crowd-sourcing to make evaluating the large number of sounds feasible, as opposed to prior pieces of work that investigated small sets of sounds [5,9,18]. We conducted a series of three user studies consisting of four tasks by using Amazon Mechanical Turk to evaluate the intuitiveness of the sounds as well as a lab study where we validated the results of the Mechanical Turk study to compensate for its low internal validity.

Through these three user studies, our paper makes four contributions. First, it provides a novel case study where we evaluated sounds through crowd-sourcing. Second, it provides empirical data about what sounds are used by electric appliances to communicate with users. Third, it also provides empirical results regarding the intuitiveness of sounds including beep sequences that have not been investigated intensively in existing works. Finally, we provide a guideline design on the basis of empirical data about what sounds should be used to communicate which informational states.

2. RELATED WORK

There have been several studies on the intuitiveness of simple auditory signals in conveying information to users. There are two types of sounds, earcons [1,3] and auditory icons [8,23], that have been investigated thoroughly in existing work.

Blattner et al. [1] defined earcons as “nonverbal audio messages used in the user-computer interface to provide information to the users about some computer object, operation or interaction,” and Brewster et al. [3] further stated that “earcons are abstract musical tones composed of short, rhythmic sequences of pitches with variable intensity, timbre and register.” Brewster et al. [4] also stated that, because of their flexibility, earcons could be easily designed to extend any object, operation, and interaction by means of their proposed guidelines. However, it could be difficult to

https://doi.org/10.21785/icad2019.002

design sounds appropriate for communicating certain informational states to users because of a lack of concrete guidelines that explain the relationships between informational states and earcons.

Gaver [8] introduced the concept of auditory icons. Gaver defined auditory icons as everyday sounds that conveyed information about computer events through analogy with everyday events. For example, the sound of shattering dishes could represent the drop of a virtual object into a virtual recycle bin. Gaver eventually argued that these auditory icons are an intuitively accessible way to use sounds to give information to users.

There have been many studies that have compared earcons with auditory icons in terms of their effectiveness [9], e.g., learnability or memorability [2,7]. While many studies have reported that auditory icons were generally perceived as easy to learn [2,7,18] and quicker in understanding [5], some have also reported that earcons were more pleasant and appropriate for actual applications than auditory icons [22]. Although the existing work has demonstrated that earcons and auditory icons effectively convey information to users, both have their limitations. With regard to earcons, one limitation is the arbitrary relationships between sounds and the information communicated by the sounds. Because of the arbitrary mappings, users in one study had to memorize the mappings to understand the meaning of the sounds correctly [23]. With regard to auditory icons, metaphoric mappings were not always easy to find [16]. Thus, it is difficult to design appropriate auditory icons for all informational states that computer systems have to communicate to users.

Currently, some electric appliances that can play rich sounds use earcons and/or auditory icons when interacting with users. However, most appliances still use rather simple auditory signals like beep sounds. One reason is that various international or domestic standards organizations have published standards for such simple auditory signals for the visually impaired or elderly. The American National Standard Institute (ANSI/INCITS 389-393), the International Organization for Standardization (ISO 11429), and the Japanese Industrial Standard Committee (JIS S0013) are representative standards organizations that deal with auditory signals. These organizations determine standards like "its pitch should be more than 250 Hz and less than 2,000 Hz" and one beep should indicate "start" and two beeps "finish." However, the relationship between signals and events is unintuitive [16]. This suggests that the design guidelines for such auditory signals are not clear either. Moreover, up to now, little study has been done to compare the effectiveness or intuitiveness of beep sounds with those of earcons or auditory icons.

German architect Ludwig Mies van der Rohe adopted the motto "less is more" to describe his aesthetic approach to arranging the numerous necessary components of a building to create an impression of extreme simplicity by enlisting every element and detail to serve multiple visual and functional purposes. Recently, a similar design concept has become popular in HCI studies [19]. Recent electric appliances can present rich information to users through their high-resolution displays or stereo sound systems. However, providing too much information could overwhelm users' cognitive resources [13,17]. Thus, more work has been started with a focus on simple ways of communicating information [10,11].

Harrison et al. [10] experimentally showed that the various blinking patterns of the small LEDs of mobile phone were interpreted differently by users, and these patterns succeeded in informing users of the informational states of mobile phones, such as low-battery and the presence of notifications. Similarly, Harrison et al. [11] also proposed Kinecticons, graphical icons with simple motions that can convey various informational states to users. In terms of simple auditory signals, Komatsu et al. [14] proposed artificial subtle expressions (ASEs) for intuitively notifying users of artifacts' internal states (specifically, their confidence level).

Therefore, it is worthwhile to investigate whether various patterns of simple auditory signals (sounds like beep sounds) could inform users of the informational states of appliances as well as blinking LEDs or Kinecticons.

3. METHOD

In this paper, we conducted three user studies by using Amazon Mechanical Turk to investigate the intuitiveness of sounds. In the first user study, we asked participants to report the names of electric appliances or devices that use sounds to convey information. In the second study, we asked them to list the informational states that these devices expressed by using sounds. Finally, in the third study, we investigated the mapping between the informational states extracted in the second study and 2,012 different sounds by using Amazon Mechanical Turk, and we validated the results in a lab study.

In terms of the adequacy of crowdsourcing experiments, Komarov et al. [15] already reported that "there were no significant differences between the two settings (lab experiment and MTurk experiment) in the raw task completion times, error rates," so we assumed that our experimental setting was a reasonable one.

4. USER STUDY #1: EXCRACTING DEVICES

In the first user study, we created a human intelligence task (HIT) that asked each Amazon Mechanical Turk worker to list 15 electric appliances that used sounds to communicate their states. We paid $0.05 for each completed HIT. Although most electric appliances use sounds to communicate some meaning, we intended to extract those sounds of which people recognized the usage. This allowed us to collect appliances that use sounds to express various states rather than limited states, such as power on and off.

Table 1. Seven representative appliances in rich and simple sound groups

Rich Sound Group	Simple Sound Group
Mobile phones	Microwaves
Laptops	Refrigerators
Desktop computers	Washing machines
Televisions	Cars
Alarm clocks	Ovens
DVD players	Coffee machines
Music players	Doors

We collected 690 electric appliances listed by 46 workers in 7 days. The workers listed 146 unique electric appliances in total. Then, the two authors individually categorized all of the appliances into two categories: appliances capable of playing complicated sounds, such as melodies (rich sound group) and those capable of playing only simple sounds, such as beep sounds (simple sound group). Other than one

disagreement, all of the categorizations by the two authors were agreed upon. We also solved the disagreement through discussion. We thought that there could be differences between appliances in these two categories. If devices in the former category used simple sounds to communicate states, designers intentionally chose the simple sounds rather than other possible rich sounds. In contrast, in the latter category, designers were forced to choose simple sounds because the devices in this category could not play complicated sounds. This difference potentially affected what sounds were used to convey states in these devices.

For each category, we extracted 7 devices that were listed by participants more than 10 times (Table 1). These 14 devices were used in user study #2.

5. USER STUDY #2: EXCRACTING INFORMATIONAL STATES

In this step, using Amazon Mechanical Turk, we extracted the informational states that the 14 devices chosen in the first user study communicate to users via the auditory channel. We asked workers to list up to 10 artificial sounds that a specified electric appliance played to express informational states. In the task, we explicitly defined artificial sounds to mean sounds or brief melodies played by electrical appliances as indicators. We also explained that the artificial sounds did not include mechanical noise, such as the seek noise from hard disk drives, nor recorded music for listening to, such as songs stored in music players. For each artificial sound, we asked the five questions shown in Table 2 to obtain the characteristics of the artificial sounds in detail. We paid $0.10 for each completed HIT.

We collected 700 responses in total (50 responses for each of the 14 devices chosen in the first study). In total, 384 unique Mechanical Turk workers completed this task in 14 days. Furthermore, 700 responses listed 1,785 descriptions of sounds. Of the 1,785 descriptions, 156 were descriptions for mobile phones, 97 for microwaves, 194 for laptops, 95 for refrigerators, 183 for desktop computers, 153 for washing machines, 122 for televisions, 139 for cars, 120 for alarm clocks, 116 for ovens, 118 for DVD players, 96 for coffee machines, 105 for MP3 players, and 91 for doors.

Table 2. We asked participants to list up to 10 artificial sounds played by a given device chosen in the first study and to answer these questions for each artificial sound.

#	Questions
1	What do you believe this sound is attempting to communicate?
2	Is the sound repeating?
3	If repeating, how long (in seconds) is one cycle?
4	If not repeating, how long (in seconds) is the sound?
5	Is the sound a sequence of beeps or a melody?

On the basis of the 1,758 answers for question 1, we consolidated the informational states that the workers thought the sounds were trying to convey. Consequently, we obtained eight informational states that the electric appliances listed in Table 1 communicate to their users via sounds. We did not find substantial differences between the appliances in the rich and simple sound groups in this process. The consolidated informational states were used in user study #3 in which the mapping between sounds and informational states was investigated (Table 3).

Table 3. Extracted 8 informational states by consolidating 700 responses from workers about the states that electric appliances are trying to communicate via sound.

#	Informational States
1	The device acknowledges your input.
2	The device is reporting that there is a message or a notification.
3	The device is reporting that there is a warning, an alert, or an error.
4	The device is turning on, booting up, or warming up.
5	The device is sleeping, suspended, or hibernating.
6	The device is thinking, computing, or processing.
7	The device is ready to execute a task, a process, or a command.
8	The device completed a task, a process, or a command.

We also investigated the characteristics of the artificial sounds on the basis of workers responses to other questions (Table 4). The results show that most sounds used to communicate informational states are beep sequences. For all the devices we tested, we found statistically significant differences between the ratios of responses that reported that the sounds were beep sequences and those that reported that the sounds were melodies ($p < 0.01$ in Fisher's exact test). In fact, workers reported that melodies are used only for specific cases, such as ringtones on mobile phones or computers booting up/shutting down. These results indicate that, although the number of devices capable of playing rich sounds has increased recently, many devices still use simple sounds to convey informational states to users.

Table 4. Workers' descriptions of artificial sounds used to convey informational states. They reported that 66.6% of the sounds were beep sequences.

Is the sound a sequence of beeps or a melody?		Is the sound repeating?	
Beeps	1,186	Yes	782 (43.8%)
Melodies	543 (30.4%)	No	983 (55.1%)
If not repeating, how long (in seconds) is the sound?		If repeating, how long (in second) is one cycle?	
Beeps	2.68 sec (SD=2.60)	Beeps	4.58 sec (SD=6.34)
Melodies	3.27 sec (SD=3.17)	Melodies	8.02 sec (SD=10.7)

The answers to the other questions also characterize what sounds are used to communicate informational states (Table 4). Roughly half the sounds repeat a sequence multiple times. The average lengths of the sequences are 4.58 seconds for beeps and 8.02 seconds for melodies. Similarly, the average lengths of non-repeating sounds are 2.68 seconds for beeps and 3.27 seconds for melodies. We used these data in designing the sounds used in user study #3.

6. USER STUDY #3: MAPPING BETWEEN STATES AND SOUNDS

Finally, we investigated the mapping between the states (Table 3) and sounds. We investigated 2,012 sounds consisting of 48 sounds composed as earcons, 80 sound effects as auditory icons, and 1,884 beep sequences. Essentially, for each sound, we asked multiple Amazon Mechanical Turk workers to choose one of the informational states that they thought a sound was trying to convey after listening to it. Then, we analyzed the distributions of workers' responses. If the distribution for a sound was skewed to one state, this indicated that workers were likely to interpret the sound as an intuitive indication of the state. In this regard, we compared composed sounds, sound effects, and beep sequences. Furthermore, we provide the results of a

qualitative analysis on the relationships between the compositions of sounds and the states chosen by workers.

We evaluated 2,012 sounds in 3 rounds. In the first round, for each beep sequence, we asked 10 Amazon Mechanical Turk workers to choose one of the informational states that a given sound was trying to communicate. Through this process, we excluded beep sequences that were not intuitive for workers to interpret. In the second round, we asked 50 workers to evaluate the 48 composed sounds, 80 sound effects, and 238 beep sequences that passed the first round. Finally, in the third round, we conducted a lab study to verify the results from the second round.

6.1. Sounds

We used 2,012 sounds consisting of 48 composed sounds, 80 sound effects, and 1,884 sequences of beeps.

For the composed sounds, we hired two music composers and asked them to design sounds that represented the eight informational states (Table 3) according to earcon design guidelines [4]. Both had at least 5 years of experience in composing music on computers, and each composed 24 earcons (3 composed sounds for each of the 8 states). We paid them $120 each. Hereinafter, we refer to the 48 sounds as composed sounds.

For the sound effects, we downloaded 200 sound effects from a web site where royalty-free sound effects are distributed. After downloading the effects, the 2 authors chose 10 sounds that were likely to be related to each of the states in Table 3. In total, they chose 80 sound effects. We added the sound effects to our investigation to mitigate one limitation regarding the composed sounds. Although we believed our composed sounds had reasonable quality, the quality relied on the music composers' skills. Thus, to compensate for this limitation, we added sound effects that were made by various people with different skill levels.

The third type of sounds was the beep sequences. In generating the sequences, we took an exhaustive approach. Essentially, we generated all possible beep sequences under four constraints: the number of beeps, length of a beep, pitch of a beep, and gaps between beeps. Under these constraints, we generated 1,884 beep sequences. We refer to this set of sounds as beep sequences. In the following, we further describe the constraints and how we generated the sequences.

6.1.1. Length of sequences

According to the results of user study #2, the average length of the indicators (non-repeated sounds) was about three seconds (Table 4). Although there were no explicit guidelines for the length of composed sounds and sound effects, the examples of such sounds were mostly within three seconds [12]. Thus, we decided to limit the length of sequences to three seconds to make the length comparable to other sounds.

6.1.2. Length of beeps

One set of guidelines for earcons [4] showed that the sound length should not be less than 0.0825 seconds. Furthermore, the length of a single beep is mostly shorter than one second. Thus, we decided to limit the length of a beep sound to either 0.1 or 0.5 seconds.

6.1.3. Number of beeps

Because we decided to limit the length of the sequences to 3 seconds and a beep sound could be 0.5 seconds, we limited the number of beep sounds in a sequence up to three (i.e., one, two, or thee) considering that there should be some pauses between the beep sounds in a sequence.

6.1.4. Pauses

Because we decided to use up to three beep sounds in a sequence, there could be two pauses between the beep sounds. Manipulating these gaps could have affected how people perceived the sequences. The guidelines for earcons [4] showed that a 0.1-second gap between sounds was recognized by users as where one sound finishes and another starts. Thus, we decided to use two different lengths for the pauses: 0.1 and 0.5 seconds.

6.1.5. Pitch

We decided to use three different pitches: high (1,200 Hz), medium (850 Hz), and low (500 Hz). Using the three different pitches allowed three beep sounds in a sequence to have different pitches. These frequencies were chosen on the basis of the results of Edworthy et al. [7] and Roy [21], who reported, "The warning indication shall be a steady alarm/horn with a frequency of 800 Hz," "the clear indication shall be a bell or simulated chime tone with a frequency of 1,200 Hz," and "the pitch should be no lower than 250 Hz."

6.1.6. Played once or repeated

Additionally, because the results of user study #2 showed that half the sounds were repeated sounds, we played the sequences either only once or three times.

In summary, there were six different beep sounds (three different pitches and two different lengths) and two different lengths of pauses between the beep sounds (short or long). The sequences consisted of one to three beep sounds and pauses between them. Additionally, the sequences were played either only once or three times. All of these combinations gave us 1,884 sequences of beeps. These 1,884 beep sequences consisted of 12 sequences with 1 beep: 6 possible beeps and 2 possible ways of playing the sequences (i.e., played once or repeated), 144 sequences with 2 beeps: 6 possible beeps, 2 possible pauses, 6 possible beeps, and 2 possible ways of playing the sequences, and 1,728 sequences of 3 beeps: 6 possible beeps, 2 possible pauses, 6 possible beeps, 2 possible pauses, 6 possible beeps, and 2 possible ways of playing the sequences.

6.2. Evaluation Method

As described in the previous section, we gathered 48 composed sounds, 80 sound effects, and 1,884 beep sequences. To investigate how people map these sounds to the eight states extracted in user study #2, we created a task using Amazon Mechanical Turk. In the task, workers could play a given sound by using a user interface (Figure 1).

In the first question, the HIT asked workers to transcribe a four-digit number read verbally in English. This question validated whether the workers could play sounds and paid

reasonable attention to the HIT. After that, there were questions about the mapping. In the questions, the workers were instructed to play a sound, and, then, to choose one of the eight informational states (Table 3) that they felt the sound tried to convey while imagining that their mobile phones played the sound. We chose mobile phones because the results of user study #1 indicated that most people said that mobile phones used auditory signals to communicate states. Alternatively, the workers could also choose "The device is reporting something not included in this list" if they felt the sounds represent a certain state that was not included in the list, or they could choose "The device made a random sound that does not have any meaning" if they felt the sound did not mean anything. The orders of the choices were randomized.

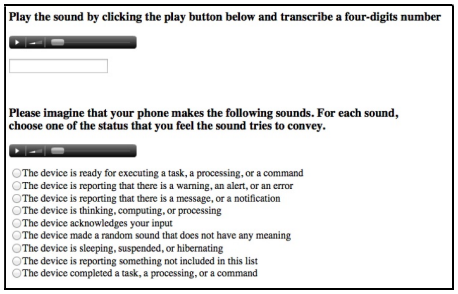

Figure 1. Screenshot of our investigation system

In one HIT, we asked the workers to evaluate five sounds one by one. We paid $0.10 for each completed HIT. The combinations of the five sounds were randomized. Because we needed a large amount of responses from the workers, we did not limit the workers to a specific region.

To evaluate the 2,012 sounds efficiently, we conducted two rounds of evaluation with Amazon Mechanical Turk and one validation in a lab study. In the first round, we asked the workers to answer questions regarding the 1,884 beep sequences. Because the beep sequences were generated by using an exhaustive approach, there were likely to be many sequences that were difficult to interpret. Therefore, we did the first round to eliminate such sequences. In the first round, we asked 10 workers to choose the state for each sound. Thus, each sound received 10 responses regarding the states that the workers thought the sounds tried to convey. Then, we eliminated sounds if the distributions of the responses were not skewed. For instance, if, for a beep sequence, 10 responses were evenly distributed among 10 choices, this indicated that the sound was not interpreted consistently. Thus, we eliminated such sequences in the first round to reduce the number of sequences. The sequences that passed the first round were further evaluated in the second round. All composed sounds and sound effects were also evaluated in the second round because we had a relatively small number of sounds for them. In the second round, we asked 40 workers to answer the same question as that in the first round; we asked them to choose a state that they thought a given sound tried to convey. Then, we analyzed the distribution of the workers' responses to investigate the intuitiveness of these sounds.

Finally, we validated the results obtained in the second round in a lab study where we asked 10 participants to evaluate the sounds that showed statistically significant results in the second round.

6.2.1. First round: eliminating beeps difficult to interpret

In the first round, we collected 18,840 responses (10 responses for each beep sequence). Seventy-four unique workers completed the task in 5 days. Out of 18,840 responses, we removed 475 responses because workers failed to transcribe the four-digit numbers correctly. Most sequences had 10 responses for each; however, because we removed 475 responses, some had 8 or 9 responses. Thus, we normalized the difference by dividing the number of responses in which a state was chosen by the number of total responses given to a sequence.

For each sequence of beeps, we focused on the states with the highest ratio of responses. Intuitively, if the ratio is high, it indicates that the workers agreed that a sequence meant a certain state and that it is easy to interpret, while, if the ratio is low, it indicates that the workers did not agree and that it is difficult to interpret.

There were three peaks with ratios around 0.2, 0.3, and 0.4 (Figure 2). Because we asked the workers to choose 1 of 10 choices, a few of the workers would have made the same choices by chance. This would have caused the peaks around 0.2 and 0.3. Thus, we decided to put a threshold at 0.4 and eliminated the sequences with a ratio smaller than 4.0. As a result, we extracted 238 beep sequences, which we further evaluated in the second round.

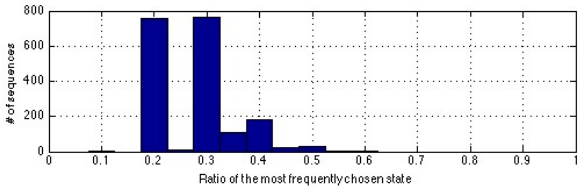

Figure 2. Graph shows distribution of mode responses divided by total number of responses for each sequence of beeps

6.2.2. Second round: comparison between sound types

In the second round, we evaluated 366 sounds that included 48 composed sounds, 80 sound effects, and 238 beep sequences. We collected 40 responses from workers for each sound in the same way as the first round. In total, we collected 14,640 responses. The task was completed by 114 unique workers in 3 days. Each worker evaluated 219 sounds on average. No worker evaluated the same sound more than once. The workers took 88 seconds to complete one HIT (i.e., transcribing a four-digit number and tagging five sounds) on average. We excluded 735 responses because workers did not transcribe the four-digit numbers correctly. We also removed 25% of the responses from the HIT that took less than 45 seconds to complete because this duration was too short to complete this HIT. In the following, we analyze the rest of the 10,406 responses.

Table 5 shows the distribution of the workers' responses. To calculate the distribution, first, we normalized the responses for each sound because each sound had a slightly different number of responses due to the removal of the low quality responses. More specifically, for each sound, we calculated the ratios by dividing the number of workers who chose a specific state by the total number of workers who evaluated the sounds. Then, we calculated the averages of the ratios to obtain the ratios shown in Table 5. We will use the distribution as a baseline for further analysis.

We then conducted 2 × 2 chi-square tests for each sound and each state to evaluate whether there was a statistically significant difference between the responses given to the sounds and the expected numbers of responses on the basis of a baseline. If at least one number in the four cells was smaller than five, we used a chi-square test with Yate's correction to evaluate the sound-state pair instead of the standard chi-square test. We regarded the differences as statistically significant when $p < 0.01$. We had to be careful when interpreting the results. Because we had 2,928 (366 sounds multiplied by 8 states) tests, we expected to have 29.3 combinations become statistically significant by chance. However, still, we were able to analyze overall trends because the statistical significances observed by chance were randomly distributed over all combinations.

Table 5. Distribution of workers' responses aggregated for all of 366 sounds

#	States	Rates
1	The device acknowledges your input.	0.098
2	The device is reporting that there is a message or a notification.	0.108
3	The device is reporting that there is a warning, an alert, or an error.	0.130
4	The device is turning on, booting up, or warming up.	0.091
5	The device is sleeping, suspended, or hibernating.	0.075
6	The device is thinking, computing, or processing.	0.127
7	The device is ready to execute a task, a process, or a command.	0.084
8	The device completed a task, a process, or a command.	0.099

Table 6. Numbers of sounds that had statistically significant ($p < 0.01$) differences between observed responses and baseline

State	Composed	Sound Effects	Beep Sequences
1	6 (12.5%)	12 (15.0%)	1 (0.4%)
2	1 (2.0%)	1 (1.3%)	3 (1.2%)
3	0 (0.0%)	0 (0.0%)	16 (6.7%)
4	8 (16.7%)	1 (1.3%)	2 (0.8%)
5	1 (2.0%)	0 (0.0%)	3 (1.2%)
6	0 (0.0%)	0 (0.0%)	15 (6.3%)
7	0 (0.0%)	0 (0.0%)	5 (2.1%)
8	0 (0.0%)	0 (0.0%)	6 (2.5%)
Total	16 (33.3%)	14 (17.5%)	51 (21.4%)

Table 6 shows the number of sounds that had statistically significant differences in the 2 × 2 chi-square tests for each state. The table clearly indicates that the workers interpreted the three sound types with different trends. The workers interpreted the composed sounds mostly as the acknowledgement of user inputs (state 1) or indications of system boot-up (state 4). Similarly, the workers mostly interpreted the sound effects as the acknowledgement of user inputs (state 1). In contrast, the workers interpreted the beep sequences as warnings (state 3) or indications of processing (state 6). Additionally, some beep sequences were interpreted as indications of executing a task (state 7) and of completing a task (state 8).

6.2.3. Intended states and interpretations

As mentioned, the composed sounds used in this study were composed by music composers to convey one of eight states (states 1 to 8). Similarly, the sound effects were selected by the researchers to represent one of the eight states. We investigated the relationships between the states that the sounds were supposed to convey and the states that the workers chose as interpretations of these sounds. The results indicate the ease or difficulty of composing/choosing sounds that convey intended states to users.

Table 7. Confusion matrix between informational states that sounds were composed/chosen to convey and informational states that workers interpreted

		Interpreted States							
		1	2	3	4	5	6	7	8
Intended States	1	**5**	0	0	0	0	0	0	0
	2	0	**1**	0	0	0	0	0	0
	3	0	0	**0**	1	0	0	0	0
	4	0	0	0	**3**	0	0	0	0
	5	0	0	0	2	**0**	0	0	0
	6	0	0	0	2	0	**0**	0	0
	7	0	0	0	0	1	0	**0**	0
	8	1	0	0	0	0	0	0	**0**

(a) Confusion Matrix of Composed Sound

		Interpreted States							
		1	2	3	4	5	6	7	8
Intended States	1	**5**	1	0	0	0	0	1	0
	2	0	**0**	0	0	0	0	0	0
	3	2	0	**0**	0	0	0	0	0
	4	1	0	0	**1**	0	0	0	0
	5	1	0	0	0	**0**	0	0	0
	6	0	0	0	0	0	**0**	0	0
	7	2	0	0	0	0	0	**0**	0
	8	1	1	0	0	0	0	0	**0**

(b) Confusion Matrix of Sound Effects

Table 7 shows the confusion matrixes for the sound-state pairs with statistically significant differences in the 2 × 2 chi-square tests. The rows and the columns denote the states that the sounds were intended to convey and the states that workers interpreted, respectively. The numbers in the cells denote the number of sounds. For instance, the bottom-left cell in Table 7 (a) shows that one composed sound that was intended to convey state 8 was interpreted to mean state 1. The numbers in the diagonal cells denote the sounds for which the workers' interpretations were the same as the intended informational states.

The composed sounds that were intended to convey state 1 (acknowledgement of user inputs) were mostly interpreted correctly. However, the other composed sounds were interpreted as state 4 (indications of turning on, booting up, or warming up) regardless of the intended states. Similarly, most sound effects were interpreted as state 1 (acknowledgement of user inputs) regardless of intended states.

These results gave us important design implications. Although rich sounds, such as composed sounds and sound effects, are expressive, they may not be as intuitive enough as we think for users to interpret their meanings. In contrast, users could more intuitively interpret simple beep sequences that conveyed some states. Therefore, when designing sounds, designers should choose appropriate sound types on the basis of the information that they intend to convey by using the sounds.

6.2.4. Third round: validation

As we already mentioned, we expected to have 29.3 combinations become statistically significant in the second round because we tested 2,930 combinations by using $p < 0.01$ as a threshold. We believe that the combinations that became statistically significant by chance were distributed

randomly across all combinations and that they would not have affected the analyses of general trends. However, to investigate the relationships between the sounds and users' interpretations of these sounds, we further validated the combinations (Table 6) that were statistically significant in the second round in a lab study, which would have higher internal validity than studies using Amazon Mechanical Turk.

We recruited 10 university students (8 males and 2 females). Their ages ranged from 21 to 25 with a mean age of 22.7. We paid $5 each for their participation. In the study, we asked participants to listen to the 81 sounds listed in Table 6 one by one and to rate the sounds. The participants were asked to rate a given sound for each state by using a 5-point Likert scale in terms of how strongly they agreed or disagreed that a sound conveyed a state (five denoted strongly agree and one denoted strongly disagree). Consequently, we obtained 648 ratings (8 ratings for each sound) from one participant. The orders of the sounds were randomized. The study took about one hour to complete.

We analyzed the data by using a one-way ANOVA to investigate whether the participants were likely to interpret the sounds as shown in the second round (within-participant design, treating eight informational states as independent variables and the ratings as dependent variables). Table 9 shows the sounds that had one specific informational state with a significantly higher average rating than all of the other seven states. These states were the same as those that were statistically significant in the second round. This ensured that these sounds were likely to be interpreted as indications of specific information with a high confidence.

7. DESIGN GUIDELINE FOR AUDITORY SIGNALS

Table 9 shows the relationships between sounds and their interpretations. On the basis of these results, we extracted a design guideline for auditory signals that communicate four informational states for which we found sounds with statistically significant differences.

- **State 1: Acknowledgement**: Two sound effects with quite short durations of less than 0.08 sec were interpreted as state 1. This indicates that users are likely to interpret sounds with very short durations as indications of the acknowledgement of user inputs. This also could explain why no beep sequences were interpreted as acknowledgement because the shortest beep sounds were 0.1 sec in our design.
- **State 3: Warning/Alert**: Eleven out of the 12 beep sequences extracted in the second round with ".H," "_H," ".h," or "(h" elements were interpreted as state 3. Thus, the beep sounds with the higher frequency in the middle of beep sequences were interpreted as indications of warning/alert.
- **State 4: Turing On/Booting Up**: All four composed sounds in melodies 5.0 sec long were interpreted as state 4. Operating systems, such as Microsoft Windows or macOS, or smartphones use rather long melodies to indicate turning on/booting up. Prior exposure to these devices would have led users to interpret the melodies as indications of state 4.
- **State 6: Processing**: All four beep sequence sounds that include at least two elements among "_," "M," or "L" were interpreted as state 6. Thus, utilizing beep sounds with medium or lower frequencies with a longer duration and longer interval in the beep sequences would be interpreted as state 6. It seems that the combination of longer sounds without high-pitched sounds and longer intervals are interpreted as relaxing situations (not like "warning/alert").

Table 8. Notations representing beep sequences

Notation	Meanings
H, M, L	Long beep sounds (0.5 sec) with high (1200 Hz), medium (850 Hz), and low (500 Hz) frequencies
h, m, l	Short beep sounds (0.1 sec) with high (1200 Hz), medium (850 Hz), and low (500 Hz) frequencies
.	Short pause (0.1 sec) between beep sounds
	Long pause (0.5 sec) between beep sounds
[ ]	Sequence played only once
( )	Sequence repeated three times

Table 9. Twenty-two sounds that succeeded in indicating specific informational states (with hyperlinks)

State	Sounds
1	2 sound effects: 0.06 sec, average F0: 880 Hz, 0.08 sec, 2,000 Hz, sounds like high-pitched ringing
2	None
3	12 beep sequences: [M_M.h], [H.H], [H.H.H], (l.H.h), (L.H_H), (L.H.m), (m.H.M), (M.m.M), (M.H.H), (h_H.l), (h.M), (H_H.m)
4	4 manually composed sounds: these sounds were melodies of about 5.0 sec
5	None
6	4 sequences: (L_M.l), (M_m_m), (h.L.L), (h.M_h)
7	None
8	None

Thus, to convey the above four informational states, we recommend utilizing the above guideline for preparing a specific melody or beeps for various electric appliances.

For the other four informational states, no sounds showed statistically significant differences in the average ratings compared with the other seven states in the validation round. However, many devices use auditory signals to convey these informational states to users. For instance, it is common to notify users that they have received e-mails by using sounds. Our results indicate that these auditory signals are less likely to be intuitive for users to interpret. Thus, devices have to communicate more information via other channels, such as text shown on a display, to compensate for the lack of intuitiveness in the auditory signals.

8. CONCLUSION

In this paper, we evaluated the intuitiveness of sounds through crowd-sourcing. By using crowd-sourcing, we explored a much larger design space, including beep sequences, than in the existing work. In the first user study, we extracted 14 devices that used the auditory channel to communicate informational states to users on the basis of 690 responses. In the second study, we collected 1,785 descriptions of sounds used to communicate informational states in the 14 devices. We then consolidated the descriptions into eight informational states that were frequently communicated via the sounds. Afterwards, in the third study, we investigated the intuitiveness of 2,012 sounds consisting of 48 composed sounds, 80 sound effects, and 1,884 beep sequences. More specifically, we asked Amazon

Mechanical Turk workers to listen to the sounds and choose one of the states that they felt the sounds represented. On the basis of an evaluation of 33,480 responses that we collected in a series of two Amazon Mechanical Turk studies, we found that the beep sequences were good at communicating notifications of warnings and status updates indicating that systems are processing commands, whereas the sound effects were mostly interpreted as indications of systems booting up, and very brief sounds were mostly interpreted as indications of acknowledgement. Finally, through a lab study, we validated the results from the user studies conducted with Amazon Mechanical Turk to provide a guideline for designing sounds used in electric appliances to communicate four informational states.

We designed our studies carefully; however, there are some limitations. In our studies, we asked workers and participants to imagine that a mobile phone made a sound. However, in practice, interpretations of sounds could depend on prior contexts. For instance, if a user started a task and heard a sound, s/he would interpret the sound as an indication of completion. We still believe that there would be many cases where users have to interpret sounds with little context, especially for mobile phones and computers because there are many background processes or push notifications on these devices. Nevertheless, the effects of context need to be further investigated. Finally, although we generated 1,884 beep sequences by using an exhaustive approach, the search space was still limited by the constraints that we set in generating beep sequences. There is a potential to improve the intuitiveness of beep sequences by modifying other properties.

Although this work still leaves unanswered questions, such as how we can design intuitive sounds for the other four states (states 2, 5, 7, and 8 in Table 6), this study presents an interesting methodology for evaluating sounds as well as a novel exploration in a large design space of beep sequences.

9. REFERENCES

[1] Blattner, M. M., Sumikawa, D. A., and Greenberg, R. M. Earcon and Icons: Their Structure and Common Design Principles. In Proc. of SIGCHI Bull 21, 1, ACM Press (1989), 123-124.

[2] Bonebright, T. L. and Nees, M. A. Memory for Auditory Icons and Earcons with Localization Cues. In Proc. of ICAD 2007 (2007), 419-422.

[3] Brewster, S. A. Using Non-Speech Sounds to Provide Navigation Cues. ACM Transactions on Computer-Human Interaction 5, 2, ACM Press (1998), 224-259.

[4] Brewster, S. A., Wright, P. C., and Edwards, A. D. N. Experimentally Derived Guidelines for the Creation of Earcons. In Adjunct Proc of HCI 95, Huddersfield, UK (1995).

[5] Bussemakers, M. P. and De Hann, A. When It Sounds Like a Duck and It Looks Like a Dog: Auditory Icons vs. Earcons in Multimedia Environments. In Proc. of ICAD 2000 (2000), 184-189.

[6] Cohen, M. H., Giangola, J. P., and Balogh, J. Voice User Interface Design, Addison-Wesley, MA, USA (2004).

[7] 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.

[8] Gaver, W. W. The SonicFinder: An Interface That Uses Auditory Icons. Human-Computer Interaction 4, 1 (1989), 67-94.

[9] Garzonis, S., Jones, S., Jay, T. and O'Neill, E. Auditory Icon and Earcon Service Notifications: Intuitiveness, Learnability, Memorability and Preference. In Proc. of SIGCHI 2009, ACM Press (2009), 1513-1522.

[10] 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.

[11] Harrison, C., Horstman, J., Hsieh, G., and Hudson, S. E. Unlocking the Expressivity of Point Lights. In Proc. of SIGCHI 2012, ACM Press (2012), 1683-1692.

[12] Hermann, T., Hunt, A., and Neuhoff, J. G (eds). The Sonification Handbook, Logos Publishing House, 2011.

[13] Keller, J. M. Development and Use of the ARCS Model of Instructional Design. Journal of Instructional Development 10, 3 (1987), 2-10.

[14] Komatsu, T., Yamada, S., Kobayashi, K., Funakoshi, K. and Nakano, M. Artificial Subtle Expressions: Intuitive Notification Methodology for Artifacts. In Proc. of SIGCHI 2010, ACM Press (2010), 1941-1944.

[15] Komarov, S., Reinecke, K., and Gajos, K. Z. Crowdsourcing Performance Evaluations of User Interfaces. In Proc. CHI'13 (2013), 207-216.

[16] Kramer, G (eds). Auditory Display - Sonification, Audification, and Auditory Interfaces, Addison-Wesley (1994).

[17] Krug, S. Don't Make Me Think!, New Riders, CA, USA (2005).

[18] Leung, Y. L., Smith, S., Parker, S., and Martin, R. Learning and Retention of Auditory Warnings. In Proc. of ICAD'97 (1997), 288-299.

[19] Maeda, J. The Laws of Simplicity, The MIT Press, MA, USA (2006).

[20] Nass, C. and Brave, S. Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship, The MIT Press, MA, USA (2005).

[21] Roy D. Patterson. Guidelines for the Design of Auditory Warning Sounds. Institute of Acoustics, 11(5):17-25

[22] Sikora, C. A., Roberts, L., and Murray, L. Musical vs. Real World Feedback Signals. In Proc. of SIGCHI 1995, ACM Press (1995), 220-221.

[23] 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.