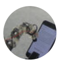
Providing Ambient Information as Comfortable Sound for Reducing Cognitive Overload

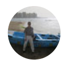
Conference Paper · November 2015

DOI: 10.1145/2832932.2832985

3 authors, including:



Mohammed Al-Sada
Qatar University



Tatsuo Nakajima
Waseda University

Providing Ambient Information as Comfortable Sound for Reducing Cognitive Overload

Kohki Ikeuchi
Department of Computer Science and Engineering
Waseda University
k.ikeuchi@dcl.cs.waseda.ac.jp

Mohammed AlSada
Department of Computer Science and Engineering
Qatar University
Department of Computer Science and Engineering
Waseda University
alsada@dcl.cs.waseda.ac.jp

Tatsuo Nakajima
Department of Computer Science and Engineering
Waseda University
tatsuo@dcl.cs.waseda.ac.jp

ABSTRACT

Everyday, we receive large amounts of information through our various senses. Since the majority of information is perceived through eye sight, other senses are often left unutilized. As a result, in order to utilize disengaged senses in our daily interactions, especially auditory, we propose a method to extract and transform real-world's visual information to auditory information. Our method is based on converting a real-world's objects' location and distance attributes to sound attributes, which are later be combined to form music. To verify our approach, we developed a prototype that we called *Music Sonar*, with which we carried out a preliminary user study followed by a questionnaire. The objective of *Music Sonar* is to reduce the cognitive overload due to visual information by converting visual feedback to ambient auditory feedback, allowing us to use our visual attention for more essential tasks. The user study results confirmed general validity of our approach, despite some shortcomings. Participants also provided a number of usability and interactivity related insights regarding the music based feedback. Finally, we provided our future direction of this research project.

Author Keywords

Ambient information; Music; Cognitive Load;

ACM Classification Keywords

H.5.2. User Interfaces, Interaction styles

ACE '15, November 16-19, 2015, Iskandar, Malaysia
ACM 978-1-4503-3852-3/15/11.
http://dx.doi.org/10.1145/2832932.2832985

INTRODUCTION

Everyday, we receive a large amount of information through our various senses. A sight of neon billboards or smells of flowers are daily encounters of different mediums from which we are able to perceive various information. As digital devices play larger roles in our lives, we are becoming more dependent on visual feedback for daily interactions. As a consequence, other senses, such as olfactory or tactile, are often disengaged from any activity and are not being effectively utilized.

With the latest advances in wearable and mobile devices, it is possible to use these devices to convert and distribute real-world visual information to other senses [8]. Nevertheless, current non-visual feedback methods, in modern devices, are very primitive in comparison with visual feedback. For instance, during a phone call, the amount of information that is conveyed through a ringtone is not comparable to that of the device's screen that may contain caller's picture, name and other related information.

As such, non-visual feedback, especially aural feedback, is mainly designed to supplement other types of feedback [9]. [1] Moreover, previous research have collectively indicated the potential of aural based interaction, whether in supporting complicated user tasks [2] or in non-visual interactive systems [6]. Sonification, which is defined as the use of non-speech audio feedback to convey information [1, 4], has also been well studied as an affective aural feedback approach. The majority sonification researches mainly focus on pitch, loudness, spatialization and conversion-duration from data to sound, without losing the meaning of the data [1].

Although, there have already been a significant progress in previous aural feedback researches, the overall results have not been well studied from the perspective of human

[1] Except for systems intended for the visually impaired.

attention management. Human attention management is an essential research topic in ubiquitous computing research areas that aims at overcoming the human cognitive overload problem [7]. Accordingly, the overall objective of our research project is to investigate an approach to deliver information as an ambient sound that is comparable to visual information in terms of fidelity and quantity. The domain of our mentioned target is to verify our approaches effectiveness in the area of human attention management.

In this paper, our contribution is to present the results of a case study that shows the potential of our sonification approach in reducing users' cognitive overload. Therefore, we have developed a prototype system, called *Music Sonar*. To verify our approach, we carried out a preliminary user study to investigate the feasibility and effectiveness of our method, essentially from the perspective of overcoming the cognitive overload problem that was absent in most previous research literature. The user study was also followed by a questionnaire and an interview that measured different characteristics of our prototype. Overall, the user study yielded promising result by showing the potential of our sonification-based techniques in reducing cognitive overload. Finally, we provided a number of additional results and insights regarding our approach, which would later be used for the project's future direction.

AURAL NOTIFICATION APPROACHES

Traditional auditory notifications can roughly be divided into two categories:

A) **Non-Verbal Notifications:** The alarming tones of a car's parking sensor are a typical example of non-verbal notifications. Such tones are usually very short in length and deliver small amounts of information. For instance, a car's parking sensor would deliver high pitch and continuous tones when detecting a near object, and vice versa. Moreover, this type of notifications usually consists of dull and short sounds; which are mainly intended to draw attention, not to be enjoyable to listen to. Therefore, since these notifications are mainly meant to announce certain events, information capacity is low and sounds are usually unpleasant to listen to.

B) **Verbal Notification:** A typical example of this category is flight notifications in airports. In comparison to non-verbal notifications, speech conveys a higher amount of information. Moreover, to emphasis conveying the intended information, it is typical to immediately repeat the message, either partially or fully. Consequently, verbal messages demand a longer listening time and could significantly increase the listeners' cognitive load; since listeners need to pay attention to most of the verbal notification to get the intended message.

AMBIENT SONIC NOTIFICATION

In contrary to the previously mentioned approaches, our approach intends to emphasize the following three main characteristics: 1) Large information capacity; 2) Short and comfortable listening experience; 3) Low cognitive load.

We believe that the three mentioned qualities could be integrated by dynamically generating short and continuous bursts of information-conveying music. Music is very potent in comprising the intended qualities due to the following factors:

- **Easily distinguishable**: A continuous stream of sounds like music is usually much easier to notice.
- **Comfortable**: In contrary to music [10], typical notifications, like warning tones, are not pleasant to listen to, especially for extended periods of time. Thus, these notifications consume more human cognitive resources.
- **Information Capacity**: Scale, volume, timbre and localization are the main components of sound. By introducing harmony and rhythm to arrange different sounds, music can be formed. Since each component of the music's structure can reflect certain information, music could be used to convey complex and high capacity auditory notifications [5].

MUSIC SONAR

The main objective of *Music Sonar* is to convert visually detected objects to music based notifications to reduce human cognitive overload. We extend Gaver et al [3] approach with *Music Sonar* by converting an object's distance and location attributes to sound attributes. Generated sounds are later combined with melodies to compose music.

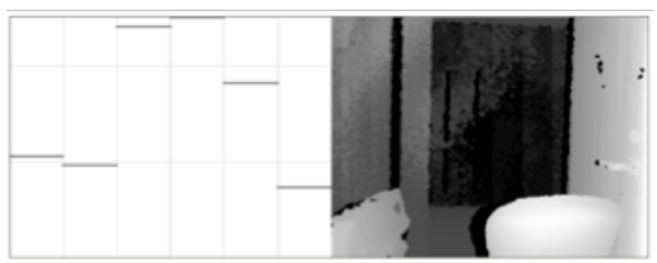


Figure 1: Some Screenshots in Music Sonar

To detect objects, we utilize an infrared (IR) camera (Microsoft Kinect[2]) which can extract depth information of detected objects. The selected number of detectable objects is directly proportional to the selected value of IR image's vertical segmentation (Blocks). For example, if the value of detectable objects is set to 6, the segmented IR image would have 6 blocks. Figure 1 illustrates *Music Sonar*. In the figure, the left image shows a visualized screen representing a top down view of positional relationship among detected objects and a user, where the user is shown

2 https://www.microsoft.com/en-us/kinectforwindows/

in the center of the bottom line. The right image shows a screenshot from the IR camera.

Upon IR image segmentation, the basic processing pipeline is illustrated in Figure 2. The processing pipeline in *Music Sonar* can be summarized in the following steps:

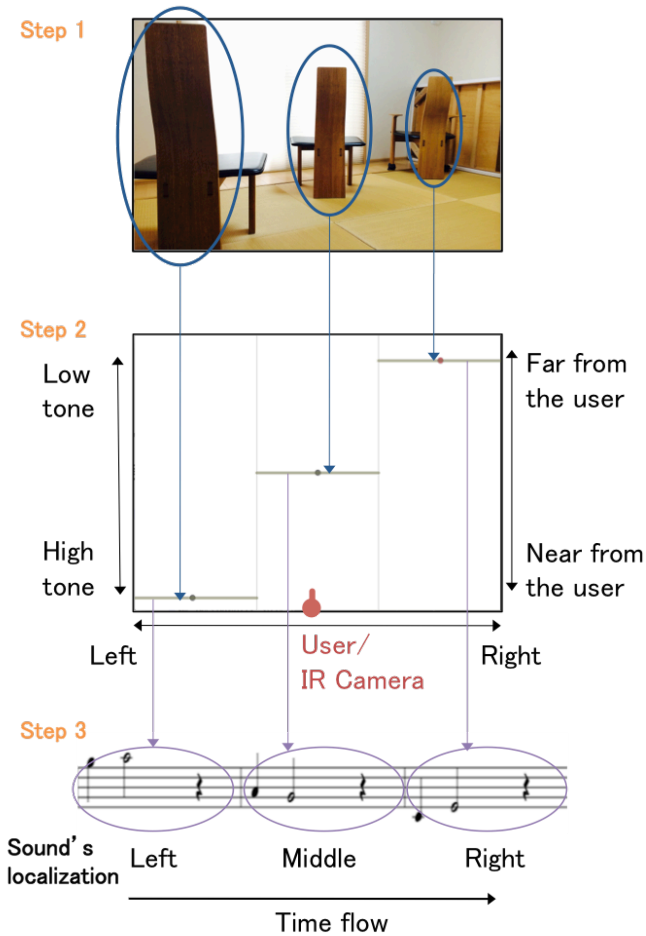


Figure 2: Processing Steps in Melody Mode

(1) Distance and Location Extraction: Depth information is extracted for each of the detected objects based on the set value of blocks. This information is used to estimate the direction and distance between the user and each object.

(2) Sound Generation: Detected objects' distance and location attributes are respectively converted to pitch and location attributes of sound. The pitch of the generated sound is directly determined by the distance of the detected object; as the object gets the closer, the pitch becomes higher, and vice versa. Likewise, the generated sound's localization is matched with the location of the detected object.

(3) Music Generation: The generated sounds, from step 2, are combined together to form a melody. Depending on the music mode, sounds can be formed as either tones (Melody Mode) or chords and tones (Harmony Mode).

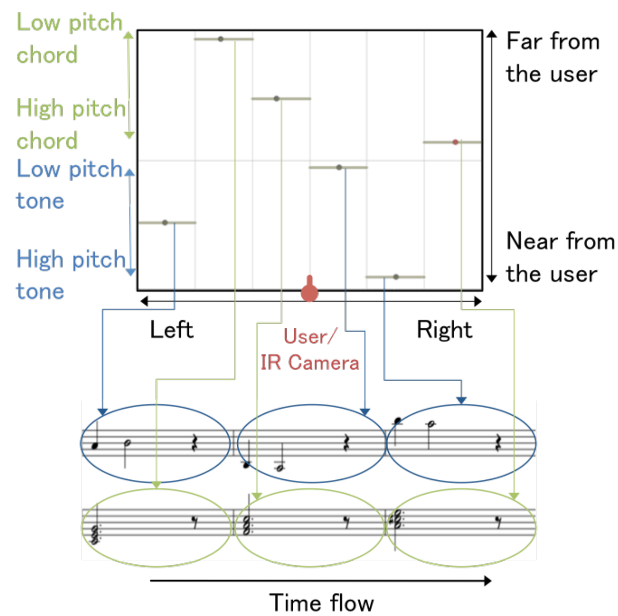


Figure 3: Processing Steps in Harmony Mode

The details of each mode are displayed in Figure 2 and Figure 3, respectively. Music is directly played to the user after step 3. When a music burst is being played to the user (9 second for 3 blocks, and 18 seconds for 6 blocks), the system repeats executing the mentioned 3 steps again to generate the next music burst. For generating music, *Music Sonar* uses synthesized piano sound; it also uses the synthesized sound of bass drum and hi-hat sound in the drum set for putting a rhythm.

EVALUATION

Preliminary Study

The preliminary study is aimed to measure the effects of two independent variables, which are the number of blocks (3 and 6 blocks) and sound mode (melody/harmony). The dependent variable was the accuracy of identifying an object's location and distance attributes. Each participant was blindfolded so he/she could not see any objects during the study. The study included 11 participants (8 males, 3 females, mean age is 22) all of whom where locally recruited students.

Prior to the study, each participant was briefed about our approach; we taught them the meaning of each sound, after which they had a short familiarization session with *Music Sonar*. During the familiarization session, participants understood the mapping between each sound and it's corresponding object. After that, each participant had four trials to correspond to the four experiment-conditions. During each trial, participants were asked to indicate where an object is located based on the generated auditory feedback.

Results

The results indicated a direct correlation between the number of blocks and accuracy. As the number of blocks increased, the accuracy decreased. In 3-blocks melody and harmony modes, the average accuracy was 97.0% (SD=0.29) and 48.5% (SD=1.17) respectively. In contrary, 6-blocks melody and harmony modes yielded an accuracy of 43.9% (SD=1.53) and 18.2% (SD=1.0) respectively.

We believe that the mentioned results were mainly due to the added complexity of sound structure in Harmony Mode with 6-blocks. However, we believe that performance would significantly improve with more training, since participant thought that with adequate training and familiarization time they could improve their performance.

Thus, after the study, we asked the following questions to the participants, and their answers are shown in Table 1.

Q1: Do you think the initial practice period was sufficient?

Q2: Do you think your accuracy would increase if you are given more time to practice?

The results show that our approach offers ambient sound that does not heavily consume human attentions, but may requires more learning time than traditional approaches.

	3 block, Melody Mode	3 block, Harmony Mode	6 block, Melody Mode	6 block, Harmony Mode
Q1	64%	36%	45%	9%
Q2	73%	82%	91%	63%

Table 1: Questionnaire of Practice's Effects

The post user study interviews and questionnaires also indicate a number of the participants' preferences. We asked the following questions to the participants, and Table 2 shows their answers of the questions.

Q1: Which is the best mode for conveying information?

Q2: Which mode is most enjoyable to listen to as music?

Q3: Which mode is most enjoyable to listen to for longer periods of time?

Q4: Which mode do you mostly prefer?

Participants favored listening to notifications that are based on music over plain sounds. Moreover, the majority of participants preferred the 3-blocks harmony mode over the rest of the modes as auditory feedback. Participants justified their choice by saying that it provides the best balance of simplicity and comfort. One of interesting results in the study is that the most comfortable sound is not the most enjoyable. Based on the participants' interview session, we believe that the mention comfort preferences are due to the fact that participants considered the ease of perceiving information from the aural feedback as one of the critical factors in determining the comfort of a sound.

	3 block, Melody Mode	3 block, Harmony Mode	6 block, Melody Mode	6 block, Harmony Mode	Not applicable
Q1	91%	9%	0%	0%	0%
Q2	0%	27%	0%	46%	27%
Q3	18%	27%	0%	37%	18%
Q4	37%	63%	0%	0%	0%

Table 2. Listening Preferences of Participants

CONCLUSION AND FUTURE DIRECTION

The user study results are very encouraging to further investigate our approach; since results confirmed that music based notifications are feasible in convening large amounts of information in complex sound structures without significantly increasing human cognitive load. We intend to expand our research to investigate daily usage of music as an information conveying medium in realistic scenarios. For instance, our approach could be utilized for people's environmental awareness while commuting, especially when most people's attention is fragmented among digital devices. Such application would minimize the risk of accidents caused by the increasing demand of visual interaction with digital devices while walking, without overloading users' cognitive resources. In addition, we also intend to carry out further user studies to understand more characteristics of our approach. In particular, we will investigate how our approach would reduce visual information overload and overcome the current issues in human attention management.

REFERENCES

1. Gael Dubus, Roberto Bresin. 2013. A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities. *PLoS ONE,* Vol. 8, No. 12, e82491. doi:10.1371/ journal.pone.0082491.
2. William W. Gaver, Randall B. Smith, and Tim O'Shea. 1991. Effective sounds in complex systems: the ARKOLA simulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '91).
3. William W. Gaver. 1993. Synthesizing auditory icons. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (CHI '93).
4. Thomas Herman. 2008. Taxonomy and definitions for sonification and auditory display*,* In P*roceedings of the 14th International Conference on Auditory Display*.

5. Robert H. Lorenz, Alex Berndt, Rainer Groh. 2013. Designing auditory pointers. In *Proceedings of the 8th Audio Mostly Conference.*

6. Matthew Marx and Chris Schmandt. 1996. MailCall: message presentation and navigation in a nonvisual environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '96).

7. Tadashi Okoshi, Jin Nakazawa, Hideyuki Tokuda. (2014) Attelia: Sensing User's Attention Status on Smart Phones. In *the Adjunct Proceedings of ubicomp '14*.

8. Yolanda Vazquez-Alvarez and Stephen A. Brewster. 2011. Eyes-free multitasking: the effect of cognitive load on mobile spatial audio interfaces, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11).

9. Jeff Wilson, Bruce N. Walker, Jeffrey Lindsay, Craig Cambias, Frank Dellaert. 2007. SWAN: System for Wearable Audio Navigation. In *Proceeding 11th IEEE International Symposium on Wearable Computers.*

10. Singo Yamano, Takamitsu Hamajo, Shunsuke Takahashi, Keita Higuchi. 2011. Single-Modal Mobile Navigation Using Directionally Annotated Music. In *Proceedings of the 19th Workshop on Interactive Systems and Software.*