UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Can Monaural Auditory Displays Convey Directional Information to Users?

Permalink
https://escholarship.org/uc/item/19r6m57h

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Authors
Komatsu, Takanori
Yamada, Seiji

Publication Date
2016

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Can Monaural Auditory Displays Convey Directional Information to Users?

Takanori Komatsu (tkomat@meiji.ac.jp)
Department of Frontier Media Science (FMS), Meiji University
4-21-4 Nakano, Tokyo 1648525 JAPAN

Seiji Yamada (seiji@nii.ac.jp)
National Institute of Informatics/SOKENDAI/Tokyo Institute of Technology
2-1-2 Hitotsubashi, Tokyo 1018340 JAPAN

Abstract

The purpose of this study is to build a monaural auditory display to convey four pieces of directional information (upward, downward, rightward, and leftward) to users effectively and intuitively without the need for wearing headphones or preparing more than one speaker. We prepared five types of monaural auditory displays consisting of triangle wave sounds and conducted an experiment to investigate which kinds of displays succeeded in conveying the four pieces of information to participants. As a result, we could confirm that one of the prepared monaural auditory displays, designed as a "progress bar" on the basis of the mental-number line and spatial-number association of the response code effect, succeeded in conveying the four pieces of information more effectively compared with the other candidate sets (its average correct rates were about 0.88). This result thus strongly shows that this monaural auditory display was quite useful for conveying primitive spatial information to users

Keywords: Monaural auditory display; Directional information; Mental-number line; Spatial-number association of response code effect.

Introduction

Most user interfaces convey various types of information to users as visual information appearing on visual displays, such as LCDs or LEDs, or as sound information emitted from speakers as speech or non-speech sounds. Currently, visual displays can be used to convey rich information to users, while sound information still has an indispensable role, especially for some users, such as visually impaired persons, those whose eyes are focused on a task at hand, or workers who must move around in their environment. Furthermore, the omnidirectional nature of auditory perception allows listeners to obtain information from sound information, regardless of their location or orientation (Walker & Kramer, 2006a).

In terms of speech and non-speech sounds, it is said that speech sounds are good for conveying precise statements to listeners, but there are also some flaws; that is, speech sounds are limited only to those who can understand the utilized language; paralinguistic information in speech sounds, e.g., gender, inflections, or tones of voices, can cause unexpected interpretations (Nass & Brave, 2005), and understanding such statements places a cognitive load on users, interfering with tasks they are engaged in (Rouben & Terveen, 2007). In comparison, non-speech sounds could contain an immediate and intrinsic relationship between the display dimensions and the information that is conveyed, e.g., in a Geiger counter, wherein radiation measurements are presented to users in the form of a clicking sound, with more clicks representing higher radiation levels (Walker & Kramer, 2006b), so understanding these sounds is quite easy and intuitive for users. Therefore, these sounds have been studied as "auditory displays" (Gregory, 1994; Walker & Kramer, 2006a) and actually used in various kinds of user interfaces for many years (Patterson, 1982; Pollack & Ficks, 1954), especially for conveying warnings or alarms (Edworthy & Stanton, 1999; Edworthy *et al.*, 1991) and for indicating events or objects on graphical user interfaces, such as by using an earcon (Blattner *et al.*, 1989; McGookin & Brewster, 2004), e.g., notifying that a USB device is connected or removed with a brief piece of music consisting of three or four notes, or an auditory icon (Gaver, 1986; Gaver, 1993), e.g., notifying that a file has been closed with the sound of a closing metal cabinet. The sounds have also been used for presenting spatial information to users (Loomis *et al.*, 1998; Walker & Lindsay, 2006).

Auditory displays used as warnings/alarms and earcons/auditory icons are simply emitted from a monaural speaker, while ones used for representing spatial information are emitted from headphones or more than one speaker. These studies use beacon-like sounds, e.g., a beacon sound played in a cyclic on-off pattern that increases in tempo when a user move closer to a target, or stereo sounds to represent left or right directions for differentiating the times sound reaches the two ears (interaural time difference) or the sound pressure level reaches the two ears (interaural level difference). Therefore, these users must pay attention to beacon-like sounds continuously or wear headphones, or keep the same distance and orientation from the speakers, so this would hinder users engaged in tasks or in information gathering to comprehend their surrounding environment. Furthermore, most studies on auditory displays require a training phase in which users get used to the prepared displays. We regard this training phase as a severe constraint for users.

The purpose of this study is then to build monaural auditory displays to convey directional information to users effectively and intuitively without such physical constraints and a training phase. As specific spatial information, we focus on four primitive directions: upward, downward, rightward, and leftward. These four pieces of directional information could consist of a 2D vertical plane in front of a user, so we regard these pieces of information as fundamental information for spatial recognition. Specifically, we prepared five types of monaural auditory displays consisting of triangle wave sounds and conducted an experiment to investigate which kinds of displays succeeded in conveying the four pieces of directional information to the participants.

We expect that successful auditory displays can convey directional information to users without the several constraints mentioned above, and various auditory displays (including earcon, alarms, and spatial information) can be emitted from monaural speakers. This would make users receive the fruitful benefits of auditory displays regardless of their locations or orientations.

Prepared Auditory Displays

In terms of designing the auditory displays, several studies argued that "data-to-sound mappings that seem intuitive to a sound designer may actually result in less effective performance (Walker & Kramer, 2005)," or "an effective and practical approach to sonification can only result from the determination of actual listener preference and performance (Walker & Lane, 2001)." Therefore, we decided not to propose the best candidate display set on the basis of deep deliberation but to prepare the several candidate sets on the basis of a broad range of possibilities and explore the best one of them.

First, we prepared two auditory displays that indicate upward and downward as continuously increasing and decreasing pitched sounds, respectively [Figure 1] because several studies (Pratt, 1930; Roffler & Butler, 1968) showed strong evidence that "high pitch is consistently mapped to high positions" and this mapping has been used in various applications successfully (Meijer, 1992). Concretely, a triangle wave sound 0.35 s in duration with increasing pitch [onset fundamental frequency (F0): 250 Hz and end F0: 400 Hz] indicates upward, and a triangle wave sound 0.35 s in duration with decreasing pitch indicates downward [onset F0: 250 Hz and end F0: 100 Hz]. Therefore, these upward and downward displays were commonly used among the following five candidate sets for rightward and leftward[1].

- Set 1 [Figure 2]: Pitched sounds increasing and decreasing in a stepwise manner indicate rightward and leftward, respectively. This was designed on the basis of the spatial-musical association of the response code (SMARC) effect (Rusconi *et al.*, 2006); that is, low-pitched tones are associated with the left side of a keyboard like a piano, while high-pitched tones are associated with the right side. Concretely, a triangle wave sound 0.7 s in duration decreasing 25 Hz from an onset F0 of 400 Hz every 0.1 s in a stepwise manner (end F0: 250 Hz) indicates leftward, and a triangle wave sound 0.7 s in duration increasing 25 Hz from an onset F0 of 250 Hz every 0.1 s in a stepwise manner (end F0: 400 Hz) indicates rightward.

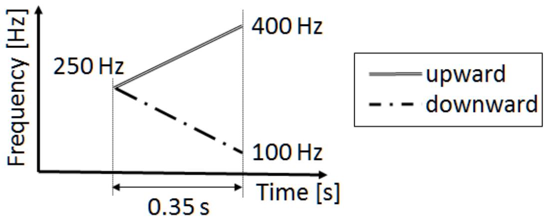

Figure 1: Commonly used upward and downward

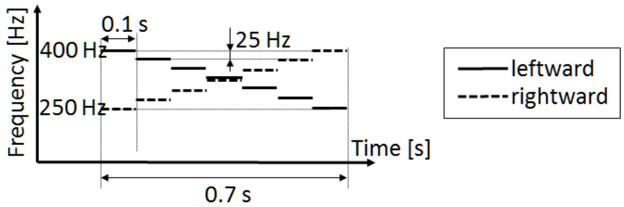

Figure 2: Leftward/rightward of Set 1

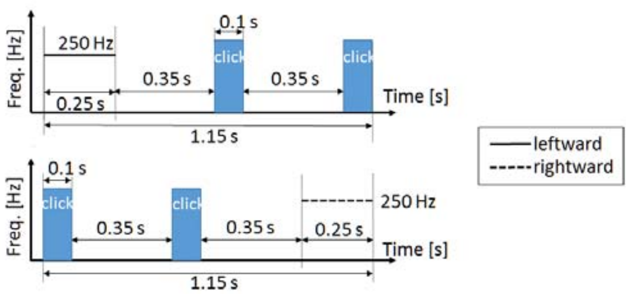

Figure 3: Leftward/rightward of Set 2

[1] URL will be listed here for listening to the prepared auditory displays if this paper is accepted.

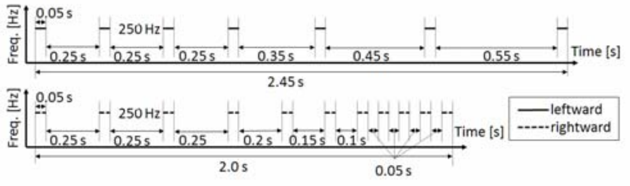

Figure 4: Leftward/rightward of Set 3

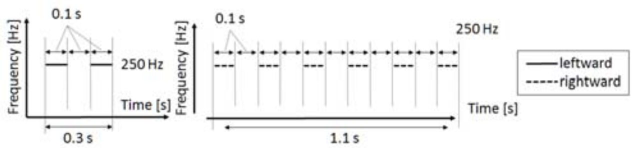

Figure 5: Leftward/rightward of Set 4

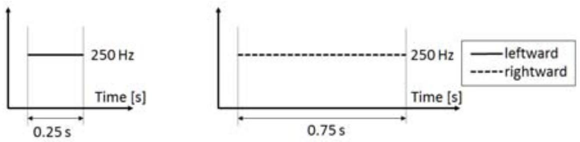

Figure 6: Leftward/rightward of Set 5

- Set 2 [Figure 3]: "Sound-click-click" indicates leftward, and "click-click-sound" indicates rightward. This was designed on the basis of the concept of eye-music (Abboud *et al.*, 2014); that is, the sounds appear to scan a picture from left to right, i.e., elements to the left will be heard first. Concretely, "sound" was a triangle wave sound 0.25 s in duration with a fixed F0 of 250 Hz, and "click" was a recorded sound of a metronome, and its duration was 0.1 s. The interval among these sounds was set at 0.35 s.
- Set 3 [Figure 4]: A sound played 7 times with intervals between sounds becoming gradually wider (after first four sounds; from 0.25 to 0.55 s) indicates leftward, while a sound played 11 times with intervals becoming gradually narrower (from fourth sounds to sixth sounds; from 0.25 to 0.05 s) indicates rightward. This was also designed on the basis of the concept of the above mentioned eye-music, but changing the sound intervals represents deceleration and acceleration, respectively. The sound was a triangle wave sound 0.05 s in duration with a fixed F0 of 250 Hz.
- Set 4 [Figure 5]: A sound played twice indicates leftward, and a sound played six times indicates rightward. This was designed on the basis of the concept of the mental-number line (MNL) (Dehaene *et al.*, 1993) and the spatial-number association of the response code (SNARC) effect (Fias *et al.*, 1996); that is, smaller numbers are associated with the left side, while larger numbers are associated with the right side. The sound was a triangle wave sound 0.1 s in duration with a fixed F0 of 250 Hz, and its interval was also 0.1 s.
- Set 5 [Figure 6]: A shorter sound (duration: 0.25 s) indicates leftward, and a longer one (duration: 0.75 s) indicates rightward. This was also designed on the basis of the concept of the MNL and the SNARC effect, but these concepts were represented as a "progress bar." The F0 of these sounds was fixed at 250 Hz.

Actually, the authors arbitrarily designed the parameters of the sounds, such as the durations (e.g., 0.25 s in Set 2, 0.05 s in Set 3, or 0.1 s in Set 4), numbers (e.g., 7 times for leftward and 11 times for rightward in Set 3 or twice for leftward and 6 times for rightward), or fundamental frequency (most of onset F0 was 250 Hz). The purpose of this experiment was not to find the best parameters of these sounds but to find the best pattern of these sounds as an auditory display.

Experiment

Participants

20 Japanese undergrads (14 men and 6 women; average age was 19.9 years old; all right-handed) participated in this experiment. The mother tongue of all of the participants was Japanese, in which reading and writing is done from left to right, so the SNARC effect was the same as that in Western countries.

Procedure

In an experiment room, each participant was asked to sit at a table in which a USB speaker (JVC Kenwood Cooperation, NC-SP1) was placed in front of her/him. This speaker was connected to an experimenter's laptop PC, and the experimenter, who was placed on the other side of a room partition, played the selected auditory display manually. The sound pressure at the participants' head level was set at about 50 dB (FAST, A). Participants were asked to answer with one of the four pieces of directional information (upward, downward, rightward, or leftward) after hearing one of the prepared auditory displays described in the former section.

The participants experienced all five candidate sets (Set 1 to Set 5) in random order, and in each set, the four pieces of information were presented four times, also in random order. This means that each participant answered 80 times (4 pieces of directional information × 4 times × 5 candidate sets), and this took about 10 minutes. Note that this experiment did not have any training phases in which the participants got used to the prepared auditory displays.

Result

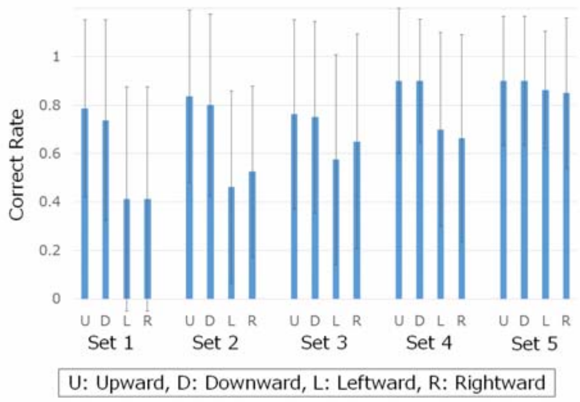

Figure 7: Correct rates in terms of directional information and candidate (error bars represent standard deviations)

Table 1: Average correct rates for each set and each piece of directional information.

	up	down	left	right
Set 1	0.7875 (0.3646)	0.7375 (0.4144)	0.4125 (0.4628)	0.4124 (0.4628)
Set 2	0.8375 (0.356)	0.8 (0.3758)	0.4625 (0.3975)	0.525 (0.3527)
Set 3	0.7625 (0.3911)	0.75 (0.3953)	0.575 (0.4337)	0.65 (0.443)
Set 4	0.9 (0.3)	0.9 (0.225)	0.7 (0.4)	0.6625 (0.4277)
Set 5	0.9 (0.2669)	0.9 (0.2669)	0.8625 (0.2433)	0.85 (0.3102)

To investigate which candidate set succeeded in conveying the directional information to the participants correctly, we calculated the correct rates, indicating how much the presented display was interpreted correctly. Table 1 and Figure 7 show the correct rates in terms of the four pieces of information and five candidate auditory display sets.

The correct rates were analyzed by using a two-way within-participants design ANOVA [independent variable 1: directional information (four levels: upward, downward, leftward, and rightward), independent variable 2: candidate set (five levels: Sets 1 to 5), dependent variable: correct rates]. The results of the ANOVA showed significant differences in the interaction effect [$F(12, 228) = 2.72$, $p < .01$, effect size: $\eta^2 = 0.15$] and in both of the main effects [candidate set: $F(4,76) = 4.2$, $p < .01$, effect size: $\eta^2 = 0.29$, directional information: $F(3,57) = 10.1$, $p < .01$, effect size: $\eta^2 = 0.14$]. The simple main effects of both of the independent variables were analyzed, and the results showed significant differences in correct rates for leftward and rightward among the candidate sets [leftward: $F(4,76) = 5.23$, $p < .01$, rightward: $F(4,76) = 3.72$, $p < .01$] and in the rates for Sets 1, 2, and 4 among the four pieces of information [Set 1: $F(3,57) = 10.92$, $p < .01$, Set 2: $F(3,57) = 10.64$, $p < .01$, Set 4: $F(3,57) = 4.17$, $p < .05$], and the results showed a marginal difference in the rates for Set 3 [$F(3,57) = 2.50$, $p < .10$].

A multiple comparison using an LSD test on the simple main effect of independent variable 2 (candidate set) showed that the rates for leftward for Sets 4 and 5 were significantly higher than those for Sets 1 and 2, the rates for Set 5 were significantly higher than those for Set 3 ($MSe = 2.0429$, $p < .05$), the rates for rightward for Sets 4 and 5 were significantly higher than those for Set 1, and the rates for Set 5 were significantly higher than those for Set 2 ($MSe = 2.3150$, $p < .05$). A multiple comparison using an LSD test on the simple main effect of independent variable 1 (directional information) showed that the rates for upward and downward were significantly higher than those for leftward and rightward in Set 1 ($MSe = 1.2088$, $p < .05$), those for Set 2 ($MSe = 1.0851$, $p < .05$), those for Set 4 ($MSe = 1.2406$, $p < .05$), and those for leftward in Set 3 ($MSe = 1.0037$, $p < .05$).

The results of this statistical analysis can be summed up as follows.

- Sets 4 and 5 showed significantly higher correct rates for leftward and rightward compared with Sets 1 and 2.
- Except for Set 5, the correct rates for upward and downward were higher than those for leftward and rightward in the remaining four candidate sets.

Therefore, we could confirm that the monaural auditory display prepared as Set 5 succeeded in conveying the four pieces of information more effectively compared with the other candidate sets. Moreover, the average correct rates among the four pieces of information for Set 5 were about 0.88 without any training, so these rates are undoubtedly high. This result thus strongly shows that this monaural auditory display was quite useful for conveying the primitive spatial information to users.

Discussion and Conclusions

In this study, we investigated which kinds of our prepared monaural auditory displays can convey four pieces of directional information to users. The information for upward and downward was correctly understood, while that for rightward and leftward in Set 5, designed as a progress bar on the basis of the MNL and the SNARC effect, was correctly understood. Conveying such information to users in an auditory manner is traditionally done as beacon-like or stereo sounds, so our approach (using monaural auditory displays to convey directional information) and the acquired results are completely original and novel. The other noteworthy point of these monaural auditory displays is that a training phase is unnecessary for users, while most auditory studies require these phases. This means that our proposed auditory display is not only effective but also quite intuitive for users to understand directional information.

As already mentioned, the parameters of the sounds utilized in this experiment, such as durations, numbers, or fundamental frequency, were arbitrarily designed by the authors. Now that we confirmed that the monaural auditory display prepared as Set 5 succeeded in conveying the four pieces of information more effectively, we are planning to find the best parameter sets of these auditory sets, e.g., the appropriate duration for indicating the four pieces of information or appropriate F0 values.

Currently, we have two ideas for concrete applications that use our proposed auditory display. One is to use it independently to indicate the relative direction to users such as in a huge warehouse to find a desired product or in a server machine placed in a huge server rack to convey alarms or warnings to users to indicate which server machine has a problem. The other is to use the proposed display together with the speech sounds used in car navigation to facilitate users' understanding of these sounds. The way of combining the speech sounds and auditory display is almost the same as those of artificial subtle expressions (Komatsu *et al.*, 2010a; Komatsu *et al.*, 2010b); for example, 0.2 s after the speech sound "turn left," a sound indicating a left pattern (0.25 s of sound) is played. We expect that this combination can help new learners of certain languages or people who have right-left confusion to understand directional information from navigation systems smoothly, e.g., an American driving in Japan with Japanese navigation systems. Therefore, this paper is regarded as fundamental research or a first step toward such realistic applications.

Acknowledgments

This study was partially supported by JSPS KAKENHI Grant Numbers 26118005 and 16K00379.

References

Abboud, S., Hanassy, S., Levy-Tzedek, S., Maidenbaum, S., & Amedi, A. (2014). EyeMusic: Introducing a "visual" colorful experience for the blind using auditory sensory substitution, *Restorative neurology and neuroscience, 32*, 247–257.

Blattner, M. M., Sumikawa, D, A., & Greenberg, R. M. (1989). Earcons and Icons: Their Structure and Common Design Principles, *Human-Computer Interaction, 4*, 11–44.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude, *Journal of Experimental Psychology, 122*, 371–396.

Edworthy, J. and Stanton, N. A. (Eds.) (1999). *Human Factors in Auditory Warning*, UK: Ashgate Publishing Limited.

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: relationship between warning sound parameters and perceived urgency, *Human Factors, 33 (2)*, 205–231.

Fias, W., Brysbaert, M., Geypens, F., & d'Ydewalle, G. (1996). The Importance of Magnitude Information in Numerical Processing: Evidence from the SNARC Effect, *Mathematical Cognition, 2*, 95–110.

Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces, *Human-Computer Interaction, 2*, 167–177.

Gaver, W. W. (1993). Synthesizing auditory icons, *Proceedings of INTERCHI'93,* 228–235.

Gregory, K. (Eds.) (1994). *Auditory Display: Sonification, Audification, And Auditory Interfaces*, Addison-Wesley.

Komatsu, T., Yamada, S., Kobayashi, K., Funakoshi, K., & Nakano, M. (2010a). Artificial Subtle Expressions: Intuitive Notification Methodology of Artifacts, *Proceedings of the 28th ACM Conference on Human Factors in Computing Systems*, 1941–1944.

Komatsu, T., Yamada, S., Kobayashi, K., Funakoshi, K., & Nakano, M. (2010b). Proposing Artificial Subtle Expressions as an Intuitive Notification Methodology for Artificial Agents' Internal States, *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*, 447–452.

Loomis, J. M., Golledge, R. G., & Klatzky, R. L. (1998). Navigation System for the Blind: Auditory Display Modes and Guidance, *Presence, 7 (2)*, 193–203.

McGookin, D. K. & Brewster, S. A. (2004). Understanding Concurrent Earcons: Applying Auditory Scene Analysis Principle to Concurrent Earcon Recognition, *ACM Transaction on Applied Perception, 1 (2)*, 130–155.

Meijer, P. B. L. (1992). An Experimental System for Auditory Image Representations, *IEEE Transactions on Biomedical Engineering, 39 (2)*, 112–121.

Nass, C. and Brave, S. (2005). *Wired for Speech*. Cambridge, MA: The MIT Press.

Patterson, R. D. (1982). *Guideline for auditory warning systems on civil aircraft* (Paper No. 82017.) London, UK: Civil Aviation Authority.

Pratt, C. C. (1930). The spatial character of high and low tones, *Journal of Experimental Psychology, 13 (3)*, 278–285.

Pollack, I. & Ficks, L. (1954). Information of elementary multidimensional auditory displays, *Journal of Acoustical Society of America, 26*, 155–158.

Roffler, S. K. & Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane, *Journal of Acoustic Society of America, 43 (6)*, 1255–1259.

Rouben, A. & Terveen, L. (2007). Speech and Non-Speech Audio: Navigational Information and Cognitive Load, *Proceedings of the 13th International Conference on Auditory Display*, 468–475.

Rusconi, E., Kwan, B., Giordano, B. L., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect, *Cognition, 99 (2)*, 113–129.

Walker, B. N. & Kramer, G. (2005). Mapping and Metaphors in Auditory Displays: An Experimental Assessment, *ACM Transaction on Applied Perception, 2(4)*, 407–412.

Walker, B. N. & Kramer, G. (2006a). Auditory Displays, Alarms, and Auditory Interfaces. In W. Karwowski (Eds.), *International Encyclopedia of Ergonomics and Human Factors (2nd ed.)*, New York, NY: CRC Press, 1021–1025.

Walker, B. N. & Kramer, G. (2006b). Sonification. In W. Karwowski (Eds.), *International Encyclopedia of Ergonomics and Human Factors (2nd ed.)*, New York, NY: CRC Press, 1254–1256.

Walker, B. N. & Lane, D. M. (2001). Psychophysical Scaling of Sonification Mappings: A comparison of visually impaired and sighted listeners, *Proceedings of the 7th International Conference on Auditory Display*, 90–94.

Walker, B. N. & Lindsay, J. (2006). Navigation Performance with a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice, *Human Factors, 48 (2)*, 265–278.