An Experiment on the Feasibility of Spatial Acquisition Using a Moving Auditory Cue for Pedestrian Navigation

Yeseul Park
Seoul National University
1 Gwanak-ro, Gwanak-Gu,
Seoul 151-744, Korea
+82-2-880-7044
yspark@hcil.snu.ac.kr

Kyle Koh
Seoul National University
1 Gwanak-ro, Gwanak-Gu,
Seoul 151-744, Korea
+82-2-880-7044
kyle@hcil.snu.ac.kr

Heonjin Park
Seoul National University
1 Gwanak-ro, Gwanak-Gu,
Seoul 151-744, Korea
+82-2-880-7044
hounjini@hcil.snu.ac.kr

Jinwook Seo
Seoul National University
1 Gwanak-ro, Gwanak-Gu,
Seoul 151-744, Korea
+82-2-880-7044
jseo@snu.ac.kr

ABSTRACT
We conducted a feasibility study on the use of a moving auditory cue for spatial acquisition for pedestrian navigation by comparing its performance with a static auditory cue, the use of which has been investigated in previous studies. To investigate the performance of human sound azimuthal localization, we designed and conducted a controlled experiment with 15 participants and found that performance was statistically significantly more accurate with an auditory source moving from the opposite direction over users' heads to the target direction than with a static sound. Based on this finding, we designed a bimodal pedestrian navigation system using both visual and auditory feedback. We evaluated the system by conducting a field study with four users and received overall positive feedback.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation]: User Interfaces—auditory feedback, evaluation/methodology

General Terms
Measurement, Design, Experimentation, Human Factors.

Keywords
sound localization; spatial audio

1. INTRODUCTION
There have been several efforts to develop efficient navigation systems for pedestrians, such as route planners showing the shortest path to a destination. Spatial knowledge acquisition based on landmarks [10] such as traffic signs and wayfinding based on turn-by-turn instructions commonly used in car navigation cannot be directly applied because of fundamental differences in modalities and goals between drivers and pedestrians [7]. For example, while cars move on designated roads, pedestrians can more freely walk through parks and grass fields and can take shortcuts not accessible by car.

While visual stimuli can deliver the most accurate data with more information bandwidth than is possible with other sensory stimuli, they are not without their limitations. Exploring strange surroundings using a traditional pictorial map could prove a cognitive overload for users. People navigating to a destination need to recognize their orientation and position within the area as well as to predict changes in spatial relationships between themselves and environments upon a change in the direction of movement. Safety is another important issue to consider. A pedestrian concentrating on a map and walking on roadways always carries a risk of accident due to a lack of attention to his or her surroundings. To alleviate these issues, many pedestrian navigation systems suggest alternative solutions to lighten the cognitive burden through the use of additional sensory channels like auditory or haptic signals: [8], [9].

This study suggests a spatial acquisition method with moving auditory cues. The feasibility of the method was evaluated in a psychoacoustic experiment to compare human sound azimuthal localization performance between stationary and moving auditory cues. After confirming the feasibility, we designed a bimodal navigational system involving both visual and auditory feedback for pedestrians. Through a field study using the system, we have confirmed that moving auditory cues can help pedestrians navigate in a safe and engaging way.

2. DESIGN RATIONALE
There have been many attempts to use auditory cues for pedestrian navigation. Liljedahl et al. used a sound source moving from center to left or right of the pedestrian to indicate upcoming turns [3], while Fujimoto and Turk used a sound that swept from the center to the target direction [2]. Unlike [2] and [3], in which the sound moved from the center to a target position, in our prototype method, it moved from the 180° opposite direction towards the target position (Figure 1). We speculated that such movement would result in more accurate detection of direction because of 1) greater changes in intensity of the sound in each channel than in [2] and [3] and 2) the additional sound from one more (exact opposite) direction, which gave more azimuthal information than would have been possible with just one sound.

Also, we constantly updated the position of the sound relative to users' head movements using a compass. Although the front–back confusion is common in simulated surround sound using a stereo channel, even a small change in the directional movement of the sound could be detected [6] and help compensate for the error [12]. It was reasonable to assume that users do not hold their heads perfectly straight all the time while navigating through the street or even sitting in an office environment [5]. By introducing the

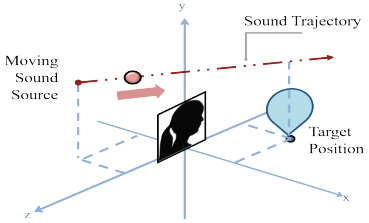

Figure 1. The sound moves from the opposite side to a target position, passing above the user's head with HRTF.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
ICMI '15, November 09 - 13, 2015, Seattle, WA, USA
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3912-4/15/11…$15.00
DOI: http://dx.doi.org/10.1145/2818346.2820779

Figure 2. A headphone with a smartphone mounted on top (Left) and the user interface for responding (Right).

constant directional update to the sound, we were able to achieve 360° of spatial acquisition.

However, we did not choose music as an auditory stimulus, as in [2], since the test results would have heavily depended on the melody, tempo, or type of music, making it difficult to compare performance. Also, human speech might draw too much attention from users, introducing safety hazards [13]. Instead, we used a repeated chime. The stimulus was generated as a non-individualized head-related transfer function (HRTF) using OpenAL (for Open Audio Library) on an Android application. The chime was repeated at a pace of 75 BPM while it took 20 seconds for it to move from the opposite side to the destination.

Although the primary goal of navigation systems is to direct users to their target destinations, they do not always provide the shortest or most efficient route. Brown et al.'s ethnographic study on designing a new navigation system for city tourists found that they usually use a map to locate or orient themselves as heading in a "roughly correct" direction and enjoy exploring an unfamiliar city [1]. Based on this idea, we did not design our navigation system solely focusing on achieving maximum accuracy in spatial acquisition. Instead of providing turn-by-turn navigation, our method was designed to help "wanderers" explore the city and reach their destinations within an acceptable timeframe. For our field study, which involved using a navigation system with both visual and auditory cues, we tested users' ability to reach roughly guided locations as wanderers without getting lost.

3. EXPERIMENT WITH A MOVING CUE

To evaluate the feasibility of the moving auditory cue, we designed a controlled user study.

3.1 Procedure

Fifteen adults (*age = 19 to 33, M = 25.88*) participated in the experiment as paid volunteers. All had normal hearing with no history of hearing problems of any kind. No subject had any previous experience with psychoacoustic experiments.

The experiment was carried out in a quiet classroom. A participant sat in a turning chair and wore headphones with a smartphone attached on top (Figure 2). The experiment consisted of two sessions, one for the stationary sound and one for the moving sound. Each session comprised 20 tests. The test was designed as a within-subject comparison, and the order of the sessions was counterbalanced to prevent learning effects. The participants were allowed to review instructions and conduct trial tests until they felt comfortable with the procedure. The orientation sensor was calibrated before every test.

Each participant was asked to try to figure out the azimuthal location of the sound. While participants' heads were locked to a chinrest in most previous psychoacoustic experiments, we allowed head movement as part of natural behavior, although answers had to be given in relation to the straightforward direction. There was no noticeable latency for head tracking. Once the direction was determined, participants recorded their answers by selecting one of the twelve buttons on the computer screen. The reaction time and judged location were recorded. The azimuthal direction of the sound source was distributed from 0° to 330° at intervals of 30° and was randomly picked.

In this experiment, the relationship between intended and responded azimuthal locations was measured only on the horizontal plane, rather than in 3D, to keep things simple.

3.2 Results

Based on the azimuthal localization responses of all participants shown in Figure 3, the accuracy of sound localization using the moving sound was on par with or significantly better than with the static sound (except at 120°). The moving sound performed slightly worse than the stationary sound at 150°, as well, although the difference was not significant. We wonder if this could be generalized to argue that the sound starting at the front and fading toward the back does not provide a more accurate azimuthal cue than does the stationary sound, although a more elaborate study is required in the future. Also, we currently do not know the cause of the especially high error rates at 60° and 150° for the stationary sound, although we speculate that it is because they are close to "nice" angles like 90° and 180°.

The Shapiro-Wilk normality test shows that the data for each type do not come from a normally distributed population ($p < .0001$). The Wilcoxon signed rank test shows that the difference in the absolute error angles between the static sound (*avg. rank of 85.09, $\mu = 14.13°$ $\sigma = 23.31$*) and the moving sound (*avg. rank of 83.44, $\mu = 21.7°$ $\sigma = 29.75$*) was significant ($Z = -3.63$, $p < .001$). It should be noted that the result could vary under different experimental setup factors, such as the use of headphones or a loudspeaker, whether head movement is allowed, the interval angle (30° for us), and the type of answers requested (discrete and forced choice vs. free answers). The RMS errors for the moving and stationary sounds were 27.28° ($\sigma = 11.35$) and 37.75° ($\sigma = 11.54$), respectively.

Overall, the front–back confusion error rates for the static sound and the moving sound among all participants were 7% and 2.67%, respectively. In particular, front–back confusion was most prominent for the two participants (P8 and P10) responsible for 76.2% of the entire confusion for the static sound. Their confusion rate dropped by 75% when they used the moving sound, implying that the moving sound not only helped improve the determination of azimuthal localization but also decreased front–back confusion, especially for those who suffered from significant front–back confusion.

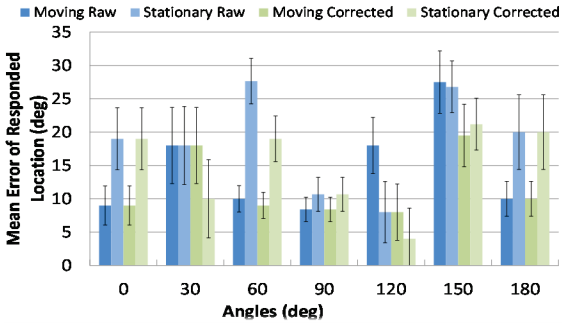

Figure 3. Mean error (in degree) for each azimuthal location. The results from the same angle from both sides were merged (i.e., ±30°, ±60°, ±90°, ±120°, and ±150°). The blue bars show the original errors, while the green bars show responses after front–back confusion is compensated.

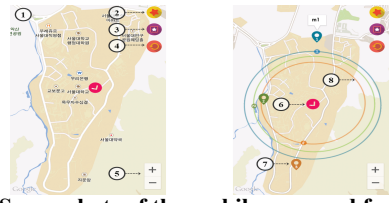

Figure 4. Screenshots of the mobile app used for the study.

The Wilcoxon signed rank test shows that the difference in the absolute error angles between the static sound (*avg. rank of 147.67, μ = 16.21s σ = 8.21*) and the moving sound (*avg. rank of 169.51, μ = 18.30s σ = 8.89*) was significant ($Z = -3.74$, $p < .001$). The result conforms to previous research findings (i.e. participants tend to delay their judgment for the moving sound until the sound has passed above their heads.) It remains to be seen how long the cycle must be in order to provide such an improvement, although it was certainly the case for 20 seconds.

In conclusion, we were able to confirm the feasibility of a moving auditory cue for sound localization on the horizontal plane as a more accurate method than a stationary sound.

4. FIELD STUDY

Encouraged by the results above, we built a bimodal navigation system for pedestrians using visual and moving auditory feedback. The system comprised five components: a map, a user state marker, a destination marker, an orientation guide as a visual cue of orientation, and an auditory cue using spatial sound with HRTF (Figure 4). The visual/auditory design would allow users to refer to the display as needed, which would be beneficial for "wanderers."

Users tend to zoom in on (Fig. 4–6, 8) a map when trying to orient and locate themselves. Because the action results in the destination markers (Fig. 4–7) being pushed outside the screen, the macroscopic orientation of a destination was indicated in this study via consistently sized circles (Fig. 4–8) around a user state marker (Fig. 4–6), making it harder to find the "rough" route without the help of a route planner. The orientation guide (Fig. 4–3) always showed the azimuthal direction of the final destination regardless of map scale, allowing users to see a big picture of the path to be taken.

An auditory cue informed users of the orientation of the destination. When a user tapped a target marker, the tinkling sound moved from the opposite side to an origin, then to the destination.

4.1 Implementation

The prototype was built on a Samsung Galaxy S III smartphone with SHURE SHR440 stereo headphones. The phone measured current location and orientation using its GPS, an orientation sensor. Spatial information was provided to users via both visual and auditory feedback. In the field study, to test a more realistic scenario, we did not use headphones with a smartphone attached on top, as we had in the experiment in the previous section, because readily available headphones do not usually come with a compass built in. Instead, we relied on the orientation of the smartphone. Heller et al.'s study, which reports that the head-yaw stays within a 10° angle to both sides, except for some delays between the head-turn and the body/device-turn [3], inspired us to make this decision. We also told participants to keep the device as straightforward from the head as possible.

4.2 Evaluation

Since there is no consensus on criteria for the quantitative evaluation of multimodal navigation systems, it is difficult to quantitatively analyze the performance of the system. Also, users' walking speeds and innate senses of spatial knowledge acquisition differ. Thus, we took a qualitative approach. The test was conducted to shed light on questions such as the following:

- Is this system really helpful in finding destinations?
- How can users control this system to find a way to the destination?

4.3 Design of the User Test

The test took place on a university campus. Four subjects (three of whom were female) aged 25 to 33 without any problems with sight, hearing, or mobility participated in the test. All participants had previous experience using smartphone navigation and had never been to the campus before.

All participants took part in adequate training sessions before the main test. In the main test, they were asked to navigate their way to three predesignated destinations marked on the map. While they were performing their tasks, they were asked to "think aloud" about their feelings and experience. The experimenter accompanied participants for safety and to record data. Post-test interviews were conducted after the main test collected users' subjective impressions of this system.

4.4 Observations

All participants successfully reached all three destinations without any significant disorientation. The mean time to complete the task was 25 min. and 30 sec. However, as mentioned above, completion time is not very meaningful because of individual differences and preferences among participants. We also did not observe any functional anomalies, like fatigue, among the participants within that time. The traces and the behavioral patterns gave us more valuable insights.

There were two main types of usage patterns for the moving auditory cue. Some participants showed a usage pattern of following the auditory cues throughout the whole test to find the destinations. They played the sound as they started the test and then kept the cue playing until they reached their destinations. A noticeable observation of this type of user is that they relied on the sound especially when they felt lost.

"I can get to a destination by following only the sound. It was easier to follow the movement of the sound than I thought" (P1).

"I lost my orientation when I got confused in the forked road, like just now. I will just follow the sound" (P3).

One participant also showed a usage pattern of not relying on the auditory cues so much. This participant used only the visual map to get closer to the target and used the auditory cues to find an accurate target position when nearing the destination.

"I prefer to wander to find the target when I am lost rather than to stop and look at the map standing up. It makes me more confused to find my position and orientation in the map. However, the sound cue is less confusing than the visual map" (P2).

4.5 Insights from the Interviews

In post-test interviews, the participants gave positive scores to the auditory cues and the orientation guide in a satisfactory survey with nine-point Likert-scale questions.

4.5.1 The Auditory Cue as a Moving Sound

Users gave positive feedback overall on the auditory cues (avg. 7.5 out of 9). No participants had previously used navigation with auditory cues. They felt a bit confused at first, but shortly got used to this system.

"I think that a novice needs to get used to the sound cue. I was confused at first, because the sound was moving. But now I feel I could get to my destination by simply following it" (P4).

"I like the moving sound because I can recognize the direction of moving easily. Even when I had some difficulties, I was able to figure it out as soon as it moved to the direction towards the destination from the opposite side" (P1).

Some users reported weak auditory cues. Three users had difficulty figuring out whether the sound was moving forward or backward. However, the constant update on the movement of sound relative to the device helped them reorient themselves.

"I could easily figure out whether the sound was moving to the left or right. But it was quite hard to distinguish front and back. So I had to turn my head around to find the direction" (P3).

Although OpenAL provided a binaural effect using HRTF, the difference was subtle to make substantial changes to the results.

4.5.2 The Orientation Guide

This feature also earned positive feedback but was less favored by users than the auditory cues (avg. 6.25 out of 9). The main attraction was that the orientation guide was always visible regardless of map scale, so users could establish the relationship between their orientation and the destination whenever they turned around. Most users, however, thought that the orientation guide was not a key feature but a supplemental cue.

"This orientation ring surely gives me the rough information of the orientation of the destination even when I zoom in to the map. But if I want to listen to the auditory cue of the destination, I have to find the marker to tap it anyway" (P3).

4.6 Discussion

Overall, the main two features helped pedestrians use the system and the moving auditory feedback played a significant role in navigating with satisfactory results. However, there are remaining issues, such as front–back confusion, which were somewhat improved but far from completely resolved. Perhaps additional cues, such as changes in tone, can be applied. The following comment shows both the pros and cons of this system.

"I think that I could take a roundabout way to the destination by just following the sound, particularly when I reach a fork. So I think I have to check the map sometimes even if I am satisfied with this system. But it is quite convenient that there is no need to concentrate on the map throughout the wandering" (P1).

The experiment comes with a few limitations. Although we have allowed the users free head movement while the sound is playing, with continuous updates to the direction to which the auditory cue moves, there was still a front–back confusion. Also, the field study conducted with headphones introduced a safety issue due to users being insulated from ambient sound. It remains to be seen whether devices like bone conduction headphones can provide enough auditory cues without sacrificing accuracy.

5. CONCLUSION

We explored the feasibility of using a moving auditory cue generated by HRTF for spatial acquisition for pedestrian navigation. From a controlled experiment, we found that users can more accurately determine the direction of a moving sound than a static sound. Also, a field study with a navigation system with such a capability revealed that pedestrians can be guided to their destinations without the need to constantly look at a map.

6. ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. NRF-2014R1A2A2A03006998) and Samsung Electronics. Jinwook Seo is the corresponding author.

7. REFERENCES

[1] Brown, B., & Laurier, E. 2005. Designing electronic maps: an ethnographic approach. *In Map-based Mobile Services* (pp. 241-257). Springer Berlin Heidelberg.

[2] Fujimoto, E., & Turk, M. 2014. Non-Visual Navigation Using Combined Audio Music and Haptic Cues. *In Proc. of the 16th Int. Conf. on Multimodal Interaction* (pp. 411-418). ACM.

[3] Heller, F., Krämer, A., & Borchers, J. 2014. Simplifying orientation measurement for mobile audio augmented reality applications. *In Proc. of the 32nd annual ACM conf. on Human factors in computing systems* (pp. 615-624).

[4] Liljedahl, M., Lindberg, S., Delsing, K., Polojärvi, M., Saloranta, T., & Alakärppä, I. 2012. Testing two tools for multimodal navigation. *Advances in Human-Computer Interaction*, 2012, 15.

[5] Minnaar, P., Olesen, S. K., Christensen, F., & Moller, H. 2001. The importance of head movements for binaural room synthesis. *In Proc. of Int. Conf. on Auditory Display*, (Espoo, Finland, 2001 July), pp. 21–25.

[6] Perrott, D. R., & Musicant, A. D. 1977. Minimum auditory movement angle: Binaural localization of moving sound sources. *The Journal of the Acoustical Society of America*, 62(6), 1463-1466.

[7] Pielot, M., & Boll, S. 2010. "In Fifty Metres Turn Left": Why Turn-by-turn Instructions Fail Pedestrians. *In Proc. of HaptiMap, Workshop at MobileHCI.* ACM, New York (pp. 26-28).

[8] Pielot, M., Poppinga, B., & Boll, S. 2010. PocketNavigator: vibro-tactile waypoint navigation for everyday mobile devices. *In Proc. of the 12th Int. Conf. on Human Computer Interaction with Mobile Devices and Services* (pp. 423-426). ACM.

[9] Rieser, J. 2008. Theory and issues in research on blindness and brain plasticity. *Blindness and brain plasticity in navigation and object perception*, 3-12.

[10] Siegel, A. W., & White, S. H. 1975. The development of spatial representations of large-scale environments. *Advances in child development and behavior*, 10, 9.

[11] Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S., & O'Modhrain, S. 2005. GpsTunes: controlling navigation via audio feedback. *In Proc. of the 7th Int. Conf. on Human Computer Interaction with Mobile Devices & Services* (pp. 275-278). ACM.

[12] Wightman, F. L., & Kistler, D. J. 1999. Resolution of front–back ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America*, 105(5), 2841-2853.

[13] Wilson, J., Walker, B. N., Lindsay, J., Cambias, C., & Dellaert, F. (2007, October). Swan: System for wearable audio navigation. *In Wearable Computers, 2007 11th IEEE Int. Symp. on* (pp. 91-98). IEEE.