Designing Auditory Pointers

Robert Harald Lorenz
Chair for Media Design
Faculty of Computer Science
TU Dresden
Dresden, Germany
Robert.H.Lorenz@gmail.com

Axel Berndt
Chair for Media Design and
Interactive Media Lab Dresden
Faculty of Computer Science
TU Dresden
Dresden, Germany
Axel.Berndt@tu-dresden.de

Rainer Groh
Chair for Media Design
Faculty of Computer Science
TU Dresden
Dresden, Germany
Rainer.Groh@tu-dresden.de

ABSTRACT
Sound can be used to give orientation, drag the listener's attention into a certain direction and provide navigational cues in virtual, as well as in physical environments. In analogy to the concept of visual pointers we call such sounds *Auditory Pointers.* While previous work mainly focused on the spacial localization property of sounds, we would like to complement this by using properties of the sound itself. Properties, like loudness, timbre, and pitch, can be used to sonify distance and direction to a target point. In this paper, we describe an exemplary implementation of respective sound synthesis techniques and investigate the effectiveness of different properties in a user study. The findings reveal big differences between the sound parameters and give clues for functional sound design.

Categories and Subject Descriptors
H.5.1 [**Multimedia Information Systems**]: Audio Input/Output, Evaluation

General Terms
Evaluation, Design

1. INTRODUCTION
In modern media, the visual channel is becoming more and more exhausted. This situation gets even more problematic when screen sizes decrease, for instance, on mobile devices. Providing orientation in such circumstances is hardly possible by more visual cues. This is where audio comes into play. We know that the navigation through a virtual, three-dimensional space is accomplished faster and more precise with a combination of auditory and visual cues. Auditory cues provide directional and distance information and allow us to head towards a target even if we cannot see it. Visual cues help to identify the target when it comes into sight and approach it precisely [7]. In such cases, audio does not only support our visual sense, but expands it. Information that is not visible on screen becomes present even without using additional techniques for off-screen visualization.

AM '13, September 18 - 20 2013, Piteå, Sweden
Copyright 2013 ACM 978-1-4503-2659-9/13/09 $15.00
http://dx.doi.org/10.1145/2544114.2544117

Sounds that are used in human-computer interfaces are, more or less, prerecorded or synthesized speech or simple, inalterable supervisory sounds. However, modern computer hardware and mobile devices are capable to create sounds in realtime. Therefore, it is possible to not only play ready-made sounds but also generate them dynamically to adapt them to the particular situation and to use them as an individual assistance for the user. Thus, the user performance can be improved and the mental workload can be reduced [1, 15]. But there are not many guidelines describing which sounds and sound properties are appropriate to solve which kind of task. There is a big body of work about the impact of pitch, volume and the spatial placement. It also investigates complex sounds, melodies and rhythms or unique applications and requirements. But the design process is strongly limited right at the beginning. Comparative surveys are performed with insufficient numbers of sound parameters. Hence, designers of auditory displays still have to make ad-hoc decisions for both the processes of selection and the creation of sounds [12].

Regarding auditory aid for navigation tasks we see a strong tendency towards two directions: synthesized speech (as in navigation devices) and spatialized sound (especially found in mobile devices). The following Section gives an overview of previous work on auditory navigational cues. We wish to complement this research and show how certain properties of a synthesized sound itself can serve this purpose, likewise. Section 3 introduces the concept of auditory pointers and describes their technical implementation. Four sound properties were investigated in a user study of which Section 4 reports. Section 5 draws a conclusion.

2. NAVIGATION WITH AUDITORY CUES
Auditory displays have many potential aplications. Besides the presentation of EEG-data[1], weather reports or stock-exchange prices there is the support of navigational tasks. Sound is not only relevant for audio-only applications but also when the visual channel is overloaded with information or should not be interrupted by additional hints and interface elements. The task of the auditory display for navigational tasks is twofold. On the one hand, the listener should

[1]Data gained from an electroencephalogram.

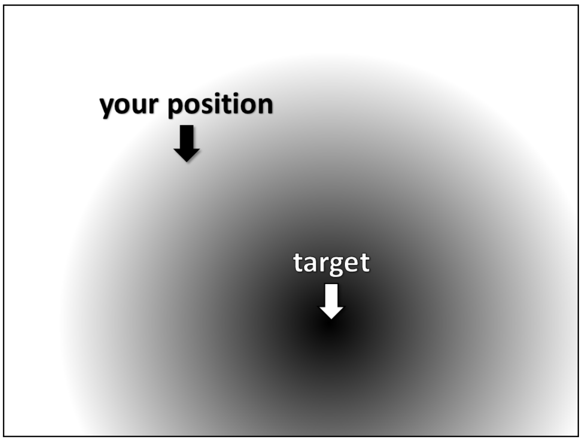
be guided to a specific point, and on the other hand, existing obstacles and threats in the surrounding area should be determined and overcome. The challenge in handling auditory pointers in navigation is the appropriate representation of the virtual or physical environment and the accuracy of localizing the user in the environment [9].

Constant research in the field of spatial perception of virtual cues has lead to numerous mappings by which three-dimensional space can be acoustically represented [5]. Indeed, an *audiovisual* navigation aid is surely more performant than a navigation that consists solely of either acoustic or visual pointers [3, 4, 7]. The direct comparison between solely acoustic or visual pointers showed that both are equally effective. Sound helps to roughly figure out a direction. Visual pointers qualify for the final approach and accurate aiming [10].

Auditory pointers can be a great help for people who are visually impaired. However, not only visually handicapped people benefit from them but also people whose visual sense is temporarily affected or distracted, e.g., business people who have to concentrate on the traffic as well as on the person they are talking to over their mobile phone and navigating in a city they do not know. Navigation in virtual spaces can also be improved or made possible by auditory pointers, e.g., in audio-only games. Three-dimensional virtual worlds in particular would be inaccessible for visually impaired people without auditory cues [10]. However, there are almost no guidelines, rules and metaphors for acoustic navigation or orientation aids. Since similar things are expressed with most different sounds, their symbolism needs to be acquired anew every time [10].

In devices for planning and supporting navigation tasks, spatially arranged synthesized speech is employed almost exclusively to provide the user with pointers to targets and objects [18]. In contrast, mainly non-verbal pointers qualify for navigation, because, on one hand, direction, distance and speech are hard to estimate [13], and, on the other hand, a speaking navigation aid can obstruct tasks carried out in a parallel way, such as the communication with others [16]. Research shows that non-verbal pointers lead to a better performance in situations in which other task are carried out at the same time [6]. The simultaneous rendering of a wider range of information by means of speech is also less useful than by non-verbal means [11].

Alternatives are mainly auditory icons, earcons and parameter mapped sonification. Moreover, according to [18], they are easier to perceive and apply. Different studies have already dealt with the design of auditive navigation pointers [2, 8, 7, 14, 15, 16, 17, 18]. In the course of these studies, auditory icons and earcons with different timbres were contrasted. They were presented as spatially localizable sound events, partly with acoustic effects. The volume was changed, depending on the distance of the listener to the target (the closer, the louder) or sonar-like tempo effects were provided (the closer, the closer in succession the sound is played back). Learning effects can always be observed. Broadband sounds (noise) tend to be more effective in comparison with narrow-band sound (a single sinusoid). In contrast, room acoustics tend to be less useful for orientation in a virtual environment. For the mapping of distance between listener and target, the test subjects preferred volume and volume changes, even though they reacted rather sensitive to discrepancies between real and virtual distance attenuation.



(a) Mapping of sound properties via a two-dimensional scalar field.

(b) Directional mapping of sound properties via a one-dimensional scalar field.

Figure 1: Abstract application scenarios. The shading maps onto sonic properties like pitch, timbre and volume.

Recent research in auditory navigation aids concentrated mostly on spatial localization of sounds. To indicate the direction of the origin from a sound, stereo or surround-panorama are used very often [9, 14, 17]. However, these techniques are not handy in every situation, e.g., the stereo audio output of smartphones is not sufficient enough unless using headphones. Interaction with wall-sized screens requires the user to turn and move the body and to leave the acoustical sweet spot. Less effort was put into the investigation of the sound itself. With modern hardware even complex sound synthesis techniques can be run and modified in realtime. The generated sounds can act as a function of the interaction in manifold ways. But to what extent do, e.g., changes of timbre and pitch help the user to *hear* the distance from a particular point or the direction to head towards it? Is the user moving towards the target or away from it? Can certain sound features, like timbre [19], be effectively employed to display the distance and direction from a target to be found?

3. AUDITORY POINTER

An *Auditory Pointer* maps spatial parameters, like distance and direction, onto sonic parameters. Our software tool synthesizes these sounds in realtime according to the distance between a cursor and a target. In the survey (Section 4), the target was not represented visually, the screen remained black, and the participants had to find it only via listening to the sound. The concept can be transferred to other scenarios. Basically, it is a question of scale, whether it is a mouse pointer on a screen or a geo-referenced position in the real world. The distance calculations can also be done only along the horizontal axis where the target points into the direction to go. Abstract examples for application scenarios are shown in Figures 1(a) and 1(b).

The sound parameters used here are volume, pitch, tempo, LFO, reverberation time, envelope, stereo panning, and, timbre (harmonic, disharmonic, THX). The THX effect derives from the well-known quality label of the Lucas' Films Company and is inspired by their first audiovisual logo, "wings", from 1983.[2] Admittedly, it is not a reproduction of the original. Our THX effect synthesizes a sound consisting of a fundamental frequency and a number of partials. The frequency of the partials is subject to a certain random offset that produces a disharmonic sound. The offset gradually gets smaller the closer the cursor gets to the target until all partials are in a harmonic ratio to the fundamental frequency. The signal chain of the sound synthesis is shown in figure 2.

3.1 Sound Synthesis

The sound of the auditory pointer is generated via additive sound synthesis. Up to 10 sinusoids are added to produce the signal $s_v(t)$. Each of the partials has its frequency in Hertz, amplitude A_n and phase angel φ_n.

$$s_v(t) = V(\text{Fundamental}(t) + \text{Overtones}(t))$$

with

$$\text{Fundamental}(t) = A_1 \sin(2\pi f t + \varphi_1)$$
$$\text{Overtones}(t) = \sum_{n=2}^{10} A_n \sin((n + rand_n)\, 2\pi f t + \varphi_n)$$

The fundamental frequency f can be set as a function of distance. If the cursor approaches the target, the frequency of the fundamental and relatively to this also the frequency of the overtones increases and vice versa. This behavior can also be inverted, as well for all further parameters. Parameter V controls the overall amplitude; as a function of distance it fades the sound in the closer the cursor comes to the target. The parameters $A_{1,\ldots,n}$ are used to control the amplitude balance between the partials. These can also be subject to different distance mappings so that distance changes cause timbral changes (e.g. the closer the brighter).

Normally the frequencies of the overtones are multiple integers of the fundamental with the result that the sound is always harmonic. The THX effect introduces a certain frequency offset to the overtones using nine random numbers $\text{rand}_n \in [-1, 1]$ so that the sound becomes disharmonic. If the cursor moves towards the target, the offsets decrease so that the sound becomes most harmonic directly at the target.

[2]http://www.thx.com/consumer/movies/8492259

In parallel to $s_v(t)$ a white-noise signal is generated and added. Contrary to the harmonic signal the noise gets louder with decreasing distance to the target. At the target the volume of the noise is set to the maximum and the harmonic sound is at its minimum.

In addition to a continuous playback of the sound, impulses can be rendered with the help of a clock. The onsets are triggered in periodic intervals either with a static tempo or depending on the distance (the closer, the faster). The envelope of the signal describes the gradient of its amplitude. The length of the attack and the sustain phase can be altered so that swelling or percussive sounds are generated, again connected to the distance. A low frequency oscilator (LFO) produces a sinusoidal amplitude modulation (tremolo) depending on the distance, too.

Next, the signal can be panned using intensity stereophony. The center is put on the horizontal position of the target point. Thus, the signal can be heard on the right side if the cursor is on the left and vice versa. Finally, the sound is passed to a reverb generator. A long reverberation time creates the impression of a large distance; in target proximity little or no reverb is played.

Every effect can be switched on separately, any combination of effects is possible. Interactions between them were avoided in the design of the synthesis approach. If an effect is switched off, the signal is passed through to the next generator.

3.2 Distance Mapping

Every activated sound parameter is set and updated depending on the distance d of the cursor (x_{cursor}, y_{cursor}) to the target (x_{target}, y_{target}). The mapping of distance to the parameters, except for the panning parameter, is calculated via function $v(d)$.

$$v(d) = \left(\frac{d_{max} - d}{d_{max} - d_{min}}\right)^k (v_{near} - v_{far}) + v_{far}$$

with

$$d = \sqrt{(x_{target} - x_{cursor})^2 + (y_{target} - y_{cursor})^2}$$

The variables v_{near} and v_{far} determine the minimum and maximum values of the actual sound parameter, respectively. For example, if the distance of the cursor is greater or equal d_{max}, the value of the parameter is set to v_{far}. If the distance is no more than d_{min}, the value is set to v_{near}. The course of the decay graph between d_{min} and d_{max} is described by the power function $\left(\frac{d_{max}-d}{d_{max}-d_{min}}\right)^k$. Its exponent k specifies the manner of the decay of the sound parameter. For $k = 1$ the decay is linear, for $k > 1$ the slope gets steeper in the proximity zone of the target. Hence, the auditory representation of the target is more concise. In exchange, the

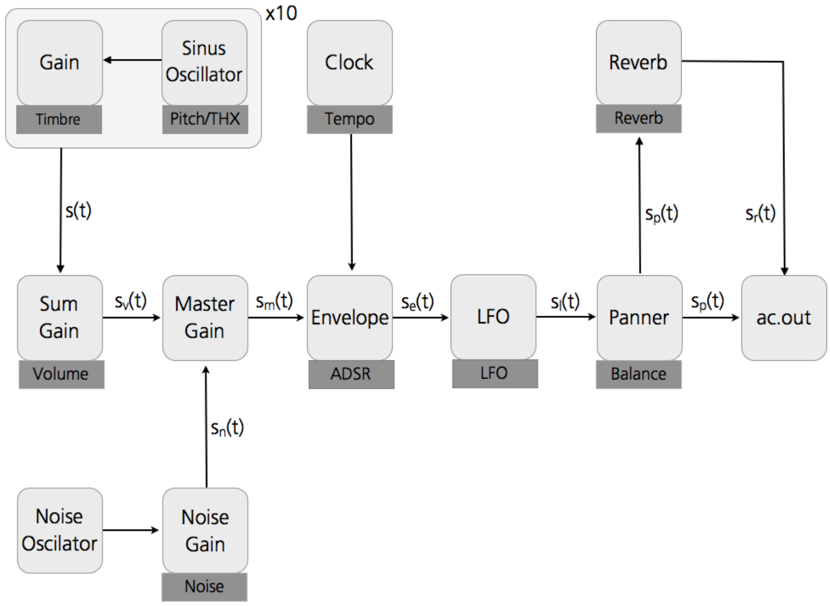


Figure 2: Signal chain

radius where the sound or the parameter change is clearly audible gets smaller. For $0 < k < 1$ the behavior is inverse. This case will no further be considered in the survey.

4. SURVEY

A user study has been carried out to investigate a selection of the parameters and analyze their effectiveness for the task. The study covers the properties volume (VOL), pitch (FRQ), low frequency oscillation (LFO) and timbre (TIM). For the latter, i.e. timbre, we used a signal with a fundamental frequency f and nine harmonics $2f$, $3f$, $4f$, $5f$, $6f$, $7f$, $8f$, $9f$, and $10f$, that fade in gradually when the cursor comes closer to the target. This effect is similar to shifting a low-pass filter's cut-off frequency up.

The task of the participants was to find an invisible target on a black screen solely with the help of a changing sound (just as indicated in Figure 1(a)). By moving the mouse cursor, they scanned the search space (screen) and listened to the sound, and its changes, over their headphones. Each participant was asked to find the target as accurately and quickly as possible even though there was no time limitation and no training in advance. The four sound parameters were combined with four different decay graphs ($k \in \{1, 2, 3, 4\}$). Every combination was presented five times so that in total there were 80 trials per proband. One study consisted of four groups with 20 trials. In one group, one sound parameter and all decay graphs were presented. The order of the groups was randomized. A possible learning effect might distort the results and was therefor excluded by a random target position in each trial. A detailed consideration of learning effects for the different parameters should be addressed by an individual study.

To compare the parameters with each other and to get first insights into the impact of sound parameters, the survey gathered informations about the duration the subject needed to find the target (in seconds), the distance between the real target position and the clicked position (in pixels) and how direct the mouse path approached the target position (difference between target direction and direction of mouse movement per timeframe in degrees). In addition to these objective measures, the participants scored their mental workload with the help of a questionnaire. It asked for (1) the mental demand, (2) the personal assessment for how successful and satisfied they were with their performance, (3) the necessary effort therefore, (4) the uncertainty during the task accomplishment, and (5) the sonic annoyance. Hence, the independent variables of the survey were the sound parameters and the decay graphs. The dependent variables were the duration, the distance, the directional accuracy of the mouse path, and the subjective mental workload.

4.1 Participants

There were 27 probands (7 female, 20 male) that participated in the study. The mean age was 30.2 years with a standard deviation of 8.2 years. Seven of them played an instrument sometimes or often and five were well-versed in sound synthesis. Two had a slightly permanent defective hearing. None of the participants knew about the context of this study.

4.2 Findings

The hypothesis was that the findings would differ significantly in the selection of the sound parameter as well as in the decay graph regarding the duration, distance and the directional accuracy. To determine if those assumptions apply and whether the sound parameters would differ a two-way ANOVA was conducted. The mean values of the examined readings were made up per sound parameter and decay graph over all trials and participants. The mean values are

Table 1: Mean values for *dist* (distance in pixels), *dur* (duration in seconds) and *acc* (directional accuracy in degrees). k is the exponent of the decay graph.

k	1	2	3	4
Mean values for VOL.				
dist	142.0	93.8	80.5	60.6
dur	19.6	16.9	16.3	16.0
acc	86.8	86.3	85.3	84.2
Mean values for FRQ.				
dist	79.8	68.2	53.4	49.9
dur	15.4	15.7	17.4	16.2
acc	83.1	85.4	84.8	85.1
Mean values for LFO.				
dist	171.5	110.8	96.2	77.9
dur	24.7	20.4	19.1	18.2
acc	87.2	86.9	87.3	87.0
Mean values for TIM.				
dist	71.2	54.3	42.1	40.4
dur	15.6	15.4	16.3	16.9
acc	84.1	84.1	84.3	84.7

shown in Table 1 (smaller values stand for better results).

Strikingly, the results show that nearly every value decreased with increasing decay graphs. Hence, the efficiency raises. Furthermore the results reveal a great disparity between the sound parameters. There were main effects for distance ($F(3, 78) = 7.462, p < .001$), duration ($F(3, 78) = 4.693, p = .005$) and directional accuracy ($F(3, 78) = 7.173, p < .001$). The study furthermore confirms the findings of [5]. Both, timbre and pitch were rated as very efficient. The assessment of the relative loudness and the LFO were rated rather insufficient. There were also main effects of the decay graphs for distance ($F(3, 78) = 53.450, p < .001$) and duration ($F(3, 78) = 4.772, p = .004$) but none for directional accuracy ($F(3, 78) = .554, p = .65$).

Furthermore, there were interactions between sound and graph for distance ($F(9, 234) = 3.202, p = .001$), duration ($F(9, 234) = 9.373, p < .001$) and directional accuracy ($F(9, 234) = 2.773, p = .004$). Considerable differences between the sound parameters are observable primarily at small exponents k. Particularly, the combination of volume or LFO with small values of k leads to superior high values for duration, distance, and directional accuracy. Exponential decay graphs lead to an improvement of all sound parameters with respect to distance and duration. While the differences between the sound parameters decreased with increasing exponents, TIM and FRQ were, nonetheless, always significantly better. Concerning the directional accuracy of the mouse path, there were also some differences between the sounds but not between the decay graphs. Especially at lower values of k LFO was worse than FRQ and TIM.

However, the analysis for the mental workload supports the known results. VOL and LFO had significantly higher mean values for all questions of the questionnaire (every $p \leq .05$) except for the sonic annoyance ($p = .006$). Here, the difference between FRQ and TIM was remarkably high so that FRQ was perceived as very annoying. Furthermore, there was a correlation between the duration and the annoyance ($p = .01$); the latter was higher, the longer the search for the target took. But no single condition was significant (every $r < .302$, every $p > .127$) so that one can not tell which sound led to a higher duration.

Differences between VOL and LFO, and between FRQ and TIM, were almost not present. Hence, it is not possible to rate the more efficient parameter within these two pairs. Although, the findings suggest that TIM is the most efficient parameter. In opposition to the hypothesis the decay graphs had no significant influence on the distance between the target and the clicked position. Only the durations increased with increasing exponent for FRQ. It seems that the pitch is a very capable sound property but, due to all exceptions, it is not easy to use.

5. CONCLUSIONS

Sounds are already used in many ways as a unidirectional information channel, e.g., the beeping sounds of a microwave or the output of a Geiger counter. The disadvantage of many such auditory displays is the use of inflexible, static and, often, very simple sounds. In the use of adaptive and dynamic sounds lies a lot of further potential for human-computer interfaces. However, there are almost no guidelines or rules for creating appropriate sounds so that designers are often forced to make ad-hoc decisions and simplify the design process due to time constraints.

Auditory pointers are already in use in a wide domain of auditory displays. Their usefulness is well-known but the representation of objects or target positions in the environment, solely with spatial localization is not always feasible. Thus, the aim of this research is to complement previous research by the investigation of certain sound properties, apart from spatial sound, to give a proposition about their efficiency and functionality. A software tool was implemented that offers up to 10 sound parameters to indicate the distance and direction of a target.

The distance between a cursor position and a target position was mapped onto certain sound properties. In the study we focused on the four properties volume, pitch, LFO and timbre. Additionally, an alternation of the decay graph was introduced that controls the parameter change depending on the target distance. The four sound parameters were combined with four decay characteristics.

The findings suggest that there is a strong impact between the sound alternation used and the duration to find the target, the distance between the real target position and the clicked position, and the directional accuracy of the mouse path toward the target. Both pitch and timbre are more efficient properties than volume and LFO. The good capability of the pitch could be approved. However, the analysis of the questionnaire reveals a stronger annoyance of this parameter. The acoustical disturbance is a crucial factor regarding the user acceptance of such an auditory display. This emphasizes the usefulness of timbre even more. Additionally, the effect of the decay characteristic shows that the higher the exponent of the polynomial function is, the shorter duration and distance are and the better the directional accuracy is.

The findings indicate that the choice of the sound and the mapping of the distance are two essential aspects for the design of auditory navigation and orientation aids. Even the mental workload could be reduced by an informed and aware design process. Furthermore, aside from a spatial presentation of paths, other sound parameters could be used as well. The potential lies in the almost innumerable possible combinations of sounds and distance mappings. For instance, timbre could be used to find the target roughly in a wide surrounding and, for a precise determination of the target, pitch can be used in combination with a higher exponent of the decay graph. The findings provide important clues for such functional sound design. They prove that it is possible to mark invisible or remote objects acoustically without the help of a visual representation and guide the user there. Future work will investigate all further properties not yet studied.

6. ACKNOWLEDGMENTS

We would like to thank our dear college Romy Müller for her assistance in the study evaluation and we like to thank all participants.

7. REFERENCES

[1] S. A. Brewster and V.-P. Raty. Earcons as a method of providing navigational cues in a menu hierarchy. In *Proc. BCS HCI '96*, pages 169–183. British Computer Society, 1996.

[2] R. Etter and M. Specht. Melodious walkabout: Implicit navigation with contextualized personal audio contents. In *Adjunct Proc. Int. Conf. Pervasive Computing*, pages 43–49, 2005.

[3] M. Gröhn, T. Lokki, and T. Takala. Author's comments on Gröhn, Lokki, and Takala, ICAD 2003. *ACM Transactions on Applied Perception*, 2(4):571–573, 2005.

[4] M. Gröhn, T. Lokki, and T. Takala. Comparison of auditory, visual, and audiovisual navigation in a 3d space. *ACM Transactions on Applied Perception*, 2(4):564–570, 2005.

[5] F. Grond and J. Berger. Parameter mapping sonification. In T. Hermann and J. G. Neuhoff, editors, *The Sonification Handbook*, chapter 15, pages 363–398. 2011.

[6] R. L. Klatzky, J. R. Marston, N. A. Giudice, R. G. Golledge, and J. M. Loomis. Cognitive load of navigating without vision when guided by virtual sound versus spatial language. *Journal of Experimental Psychology*, 12(4):223–232, 2006.

[7] T. Lokki and M. Gröhn. Navigation with auditory cues in a virtual environment. *IEEE Multimedia*, 12(2):80–86, 2005.

[8] T. Lokki, M. Gröhn, L. Savioja, and T. Takala. A case study of auditory navigation in virtual acoustic environments. In *Proc. Int. Conf. Auditory Display '00*, pages 145–150, 2000.

[9] J. M. Loomis, R. G. Golledge, and R. L. Klatzky. Navigation system for the blind. *Presence*, 7(2):193–203, 1998.

[10] G. McAllister, G. R. White, and G. Fitzpatrick. Toward accessible 3d virtual environments for the blind and visually impaired. In *Proc. Int. Conf. Digital Interactive Media in Entertainment and Arts '08*, pages 134–141. ACM, 2008.

[11] G. H. Mowbray. Simultaneous vision and audition: The comprehension of prose passages with varying levels of difficulty. *Journal of Experimental Psychology*, 46(5):365–372, 1953.

[12] A. Pirhonen, E. Murphy, G. McAllister, and W. Yu. Non-speech sounds as elements of a use scenario: a semiotic perspective. In *Proc. Int. Conf. Auditory Display '06*, pages 134–140, 2006.

[13] T. V. Tran, T. Letowski, and K. S. Abouchacra. Evaluation of acoustic beacon characteristics for navigation tasks. *Ergonomics*, 43(6):807–827, 2000.

[14] Y. Vazquez-Alvarez, I. Oakley, and S. A. Brewster. Auditory display design for exploration in mobile audio-augmented reality. *Personal and Ubiquitous Computing*, 16(8):987–999, 2011.

[15] B. N. Walker and J. Lindsay. Effect of beacon sounds on navigation performance in a virtual reality environment. In *Proc. Int. Conf. Auditory Display '03*, pages 6–9, 2003.

[16] B. N. Walker and J. Lindsay. The effect of a speech discrimination task on navigation in a virtual environment. In *Proc. Human Factors and Ergonomics Society 2006*, volume 50, pages 1538–1541, 2006.

[17] B. N. Walker and J. Lindsay. Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. In *Proc. Human Factors and Ergonomics Society 2006*, volume 48, pages 265–78, 2006.

[18] J. Wilson, B. N. Walker, J. Lindsay, C. Cambias, and F. Dellaert. SWAN: System for wearable audio navigation. In *IEEE Int. Symposium on Wearable Computers 2007*, volume 1, pages 91–98, 2007.

[19] H. Zhao, C. Plaisant, B. Shneiderman, and R. Duraiswami. Sonification of geo-referenced data for auditory information seeking: Design principle and pilot study. In *Proc. Int. Conf. Auditory Display '04*, pages 1–8, 2004.