Task, User Characteristics, and Environment Interact to Affect Mobile Audio Design

Bruce N. Walker, Raymond M. Stanley, and Jeffrey Lindsay

Sonification Lab, School of Psychology
Georgia Institute of Technology
654 Cherry Street, Atlanta, GA, USA 30332-0170
bruce.walker@psych.gatech.edu,
gte457e@mail.gatech.edu, gtg141n@mail.gatech.edu
http://sonify.psych.gatech.edu/index.html

Abstract. Context-awareness can be a powerful asset to a mobile auditory system. However, in order to fully leverage the information provided by context a system must examine the user, the task and the environment. Auditory display designs must then take into consideration all of these elements. This paper focuses on these facets of context and their application to the System for Wearable Auditory Navigation (SWAN), a context-aware, mobile auditory system for wayfinding.

1 Introduction

Designing auditory interfaces for mobile computing applications requires a blend of functional human-centered computing and aesthetic sound design. Taking the context of the user into consideration is a huge component of any successful design or application. However, in order to discuss context-aware systems it is first critical to establish what is meant by context. We hope to make this point explicit through some examples that have arisen in our own mobile audio designs.

2 Context

Regardless of domain, the context of a system includes three primary elements: the task being completed, characteristics of the user, and the physical environment that the system operates in. All three of these key elements can be dynamic, especially given mobile applications, with each element influencing both of the others. Given the dynamic nature of context, it is important to consider in any implementation of mobile computing. We first provide more detail about what these three components of context are, in general, and then show how we have been forced to address them in our work.

3 Elements of Context Overview

3.1 Task

Audio can be used to assist in the completion of an increasingly large number of categorically different tasks, all within a mobile computing scenario. These tasks include, in no particular order: (1) speech communication; (2) data analysis related to some task; (3) menu navigation; and (4) wayfinding. Each of these requires a different set of information, and thus they merit separate consideration in auditory display design. Moreover, people may switch or alternate tasks, or complete multiple simultaneous tasks in their mobile computing, making the audio design task even more challenging. We will focus mostly on wayfinding in this paper, but we will briefly consider the other kinds of tasks as well.

In early interfaces, "audio" basically meant the presentation of alarms and unprocessed speech communication. Notifications and speech remain an important component of the audio in mobile devices, though both of these types of output have been well studied [e.g., 1, 2, 3]. Thus, for brevity we can simply say that the designer needs to be aware of such research, and ensure that the output takes full advantage of the key findings. An example is the use of more meaningful sounds such as earcons and auditory icons to make more sophisticated notifications and warnings [4], and the use of lateralized or spatialized speech to improve signal detection and comprehension [5].

Second, non-speech audio has been shown to be an excellent display modality for certain tasks that take advantage of the human auditory system's pattern recognition capabilities, and in cases where visual displays are inappropriate or at least suboptimal [e.g., 6]. Examples of more complex tasks that sound can facilitate, especially for a mobile user, include the estimation of specific data values in a data series (i.e., point estimation) and trend analysis. The key here is that the specifics of the user's task places a number of constraints on how the sounds must be designed so that maximal performance can be obtained. For example, audio applications intended to support a point estimation task [e.g., 7] need to make differences between sounds very salient. On the other hand, trend analysis tasks need sounds that can group together to form an auditory stream in some cases, and remain distinct in others [e.g., 8].

Third, given the fact that mobile devices are becoming much more sophisticated and feature-laden, it is often sensible to design the interface as a menu structure. Interacting or controlling the device really means navigating through the menu to the desired command. Appropriate use of adaptive text-to-speech rates, and intelligent sorting of menu options, are examples of how such a menu can be enhanced based on the user's task, and on what the system learns about how those tasks are typically performed by a given user (see www.curointeractive.com for an excellent example of such an adaptive interface).

The fourth kind of task is wayfinding, or getting from one place to another safely, while learning about the surroundings as much as possible to enable the creation of a mental map. Clearly this is a major component of mobile computing systems, not only in the context of GPS-based directions and maps, but in many other applications as well. We will discuss the task of wayfinding in more detail, and some of the implications for auditory display design, shortly.

3.2 User Characteristics

No matter how advanced a computer interface becomes, the user (rather than the designer) will always be the final "judge" of any interface, so their characteristics are important to consider [9]. User characteristics can be divided into two major characteristics. First, **traits** are individual differences that remain relatively stable over time. Examples of traits that affect auditory display design considerations include pitch perception, working memory capacity, and training or experience. Second, **user preferences** are more apt to change depending on other elements of context. For example, a user may prefer sounds from nature, rather than synthetic sounds. This preference may depend on the context that he or she is in. That is, at work the sounds of birdcalls might be preferred as a counterpoint to the office machine sounds, whereas at home a soundscape based on Star Trek sounds might be preferred. Despite the fact that they can definitely be dynamic, preferences tend to be relatively stable in any given context, and thus can be taken into account during system design, without too much effort.

3.3 Environment

Determining whether the user is performing his or her task inside or outside, or in a noisy versus quiet environment, plays a major role in the auditory display design of a mobile computing system. For example, issues of thresholds of audibility and masking come into play. The way information about the environment can be delivered to the system can be classified in two major ways: bottom-up and top-down. Sensors such as a simple microphone can be used to determine direct (bottom-up) measurements of the ambient noise, so that the loudness or spectral components of an audio output are adjusted accordingly. Knowledge-based (top-down) inferences can also be made about the environment, such as using a person's daily schedule to predict where a person is (in the office versus out in the street) to then adjust sound levels accordingly. Since bottom-up and top-down processes each have their respective disadvantages, it is clear that a mixture of bottom-up and top-down processing is needed to make effective decisions about what an auditory display should sound like at any given moment. Light sensors can easily determine whether a person is inside or outside, and this can be cross-referenced with a schedule to improve estimates of location; microphones can then determine the details of the acoustic environment as a last step before audio is rendered and output.

3.4 Interaction of Contextual Elements

Task, user characteristics, and environment all interact together. In some ways, it is impossible to separate these elements. Though it is important to consider the holistic impact of these factors, considering each of these factors separately through this taxonomy provides a useful framework to approach user-centered design of the audio for mobile context-aware systems. The remainder of this paper will discuss specific considerations of these key elements in the implementation of auditory displays, with particular reference to a system for wearable audio navigation [10].

4 The System for Wearable Audio Navigation (SWAN)

An example of a context-aware mobile computing application that uses extensive non-speech audio output (and some audio input) in its interface is the System for Wearable Audio Navigation [10]. The SWAN assists the user in moving from place to place along a planned path, while at the same time providing audio cues about other features in the environment. Users may be anyone for whom vision is not reliable for navigation. This includes blind individuals, as well as, for example, firefighters in a smoke-filled building. Navigation cues are presented as spatialized sounds; the listener simply walks toward the apparent location of the beacon sound, until reaching a waypoint. At the waypoint a new beacon sound appears, to guide the listener to the next waypoint. Along the path, other sounds present information about offices, benches, buildings, and so on, depending of course on the user's location, task(s), characteristics, and preferences. The system can adjust the sounds based on both measurements and inferences about the environmental soundscape for optimal presentation.

4.1 Tasks With the SWAN Affect Audio Design

Identifying the task that a user is completing is a seemingly simple and obvious thing to do, but can be a deceivingly complex endeavor. The task is often what provides the catalyst for developing a particular system. It is important, however, to take an in-depth look into what exactly the task is. When examined more closely, tasks that seemed simple may turn out to be composed of a significant number of important subtasks.

The purpose of the SWAN is to provide people with a way of navigating their world without the use of visual input. At an initial level of analysis, one approach is to turn every visual object into an auditory event. This is not feasible, however, due to the cacophony of sounds that would arise. Furthermore, it is not necessary to perform such a 'one-to-one' translation, because other assistive devices such as a cane can provide information about specific nearby objects. Extraneous sounds simply add to auditory clutter. Therefore, it is critical to first determine what information is most important and appropriate to be displayed to the user based on the real needs of the task at hand.

Rather than translating *all* visual stimuli into sound, the SWAN focuses mostly on the path that the user will travel, and only adds a few extra sounds as appropriate. In a manner similar to following a trail of bread crumbs, the user navigates the series of beacons to complete the route. The general effectiveness of this task-based approach to auditory display design has been demonstrated [10], and is represented in Figure 1.

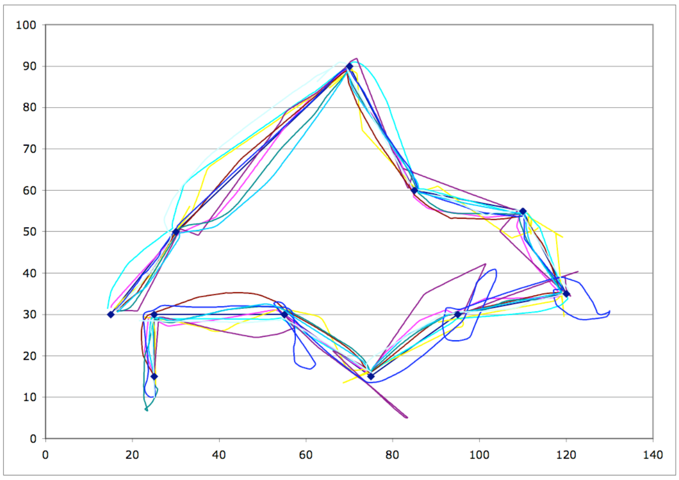

Fig. 1. This figure shows user-traversed paths relative to the "correct" path as created by the SWAN system. The diamonds are the locations of beacons along a preplanned route, and the various colored lines are plots of users' movements as they followed the path. This generally high level of performance was achieved with little or no training. In this system the primary task of wayfinding is supported effectively by carefully designed non-speech audio

As a further example, the task of navigating the world is surprisingly complex, and can be broken down into subtasks. Normal walking during most of a path segment gives way to slightly slower walking, with smaller steps, as one approaches a turning point. This is natural, and even required by the walker, in order to make a smooth and natural curve around the corner. To indicate the impending waypoint via the audio interface (which allows the blind user to prepare for the corner), the tempo of the spatialized beacon increases as the person nears the waypoint. This is an example of how the audio design needs to take into account the user's location, and in some cases adjust the audio attributes in support of the listener's wayfinding task.

The design of the auditory display must also never impede the accomplishment of other secondary tasks that the user may have. For example, in the case of the SWAN the sounds provide assistance in the primary wayfinding task. However, one of the secondary tasks is to learn about the surroundings, and also to monitor the sounds in the environment for salient additional information. That is, getting from point A to point B is primary, but listening for the sound of approaching vehicles is an important secondary task. Thus, wayfinding audio cues must not prevent the detection of other sounds in the world. Here we have a challenge: the headphones that allow for spatialized audio to support the primary task of wayfinding prevent the listener from hearing the important ambient sounds. Our user population has made it clear that these ambient sounds are so important that they would not use the system with hardware that deteriorates the perception of these sounds, such as headphones. Covering the ears is unacceptable, when all of the user's tasks are considered. Regular loudspeakers (even if small) are also unacceptable for reasons of privacy and user-acceptance. For this reason we have had to consider display alternatives such as bone conduction headsets that can provide private audio for the primary task while not impacting too much the secondary task. Certainly, with the use of bone conduction, the nature of the audio output needs to be reconsidered for audibility, quality, etc. [11-13].

Fig. 2. Bonephones support auditory display and simultaneous processing of ambient cues

4.2 Users Characteristics Affect SWAN Audio Design

4.2.1 Traits and Individual Differences

Traits of the user that are particularly relevant to the SWAN system include, among many others, are the ability to utilize spatial audio, perceptual abilities, and the user's experience/training. First, because SWAN utilizes spatialized audio to facilitate essential information to perform the task, the ability to hear spatialized audio is a prerequisite. Despite otherwise normal hearing, a small percentage of people just "don't get" spatial audio. That is, they cannot seem to localize the sounds very well. For these people, a "straightforward" audio interface in which a listener simply walks toward the apparent location of the sound source is useless. Some other way of displaying the waypoint location would need to be devised. If the people who could not hear spatial audio were a larger portion of the population, then the design of SWAN would have to be adjusted appropriately. This echoes the common wisdom to "know your user" when designing a system, especially when non-traditional interfaces are to be employed.

A second example of individual differences influencing audio design involves the perceptual abilities of a listener. For example, listeners differ in how small a pitch or tempo discrimination they can make. If information crucial to a task is mapped onto pitch or tempo, and the listener is unable to distinguish sounds that represent different data, the task will be much harder, or even impossible. In the SWAN system, proximity to a waypoint is mapped to tempo. Failure to hear different tempos might prevent the user from preparing for a turn. This might only lead to a less graceful turn, but it could also lead the user to stride past the waypoint, and possibly off the sidewalk and onto the street.

Third, using any audio interface, including the SWAN, is a completely new experience for most people. Regardless of their inexperience, most people can perform quite well from the start, and improve rapidly as they practice [10]. This is welcome news, indeed. However, there is, still, a learning curve (see Figure 3). If it took much more than a few minutes of practice and training to perform effectively (and safely!) with a mobile audio interface such as the SWAN, then the design would have to be re-evaluated. For someone faced with a lifetime of vision loss, requiring extensive experience or training to perform the task is not problematic or atypical as they often already receive many other types of mobility training. The long-term user might be willing to have a more complex display, with the tradeoff of a longer learning period: power versus practice. On the other hand, if the user simply cannot see due to temporary (or infrequent) environmental factors, then requiring extensive training may be less appropriate. That user might prefer a simpler, stripped-down audio interface instead. Thus, the details of the user's abilities interact with their larger task situation to affect how an audio interface is implemented, and what support such as training must be developed.

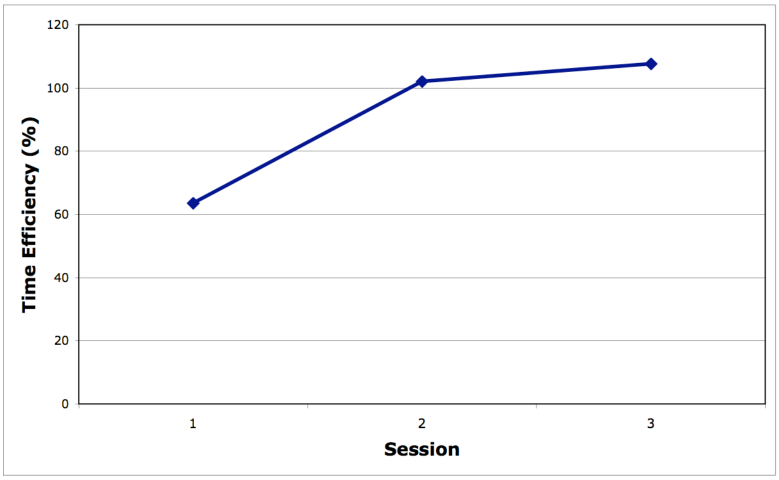

Fig. 3. This graph shows the mean time efficiency participants demonstrated in navigating increasingly complex paths. Time efficiency is a measure of a participant's time to finish a session, normalized for the distance they were required to travel in a given session. As this graph indicates, performance was good immediately and reached asymptotic levels after only two or three five-minute sessions

4.2.2 Users Preferences

In addition to differences between users, a given user will have preferences that differ across contexts. The choice of the sound used for the beacon sounds in SWAN has been shown to have significant effects on performance [10]. The broad-spectrum nature of a noise burst, for example, has proven to be a very good beacon stimulus. However, the lack of aesthetic appeal of the noise burst indicates that this beacon type may not be the optimal choice, all things considered. A sonar ping, which was not quite as good a beacon, but which sounds much more pleasing, might be considered the better choice. Of course, this preference may depend on how important the navigation task is to the user. In wayfinding that requires more precision, a user may be willing to sacrifice some listening pleasure for increased accuracy. A sacrifice of this type may not be necessary, however, if there are sounds that are aesthetically pleasing and yield good performance. In fact, preferences for beacon sounds can be a good predictor of performance with that beacon [14].

Another user preference that we have mentioned is whether the ears may be covered or not. While most users in a real-world wayfinding task would balk at any system that obstructed the ambient audio cues, when that same listener is at home listening to a concerto he or she might be perfectly willing to give up external audio cues in favor of a more high-fidelity audio experience with closed-ear headphones. This shows how preferences, while generally stable, can still vary from situation to situation, within a single user. The audio display needs to be adaptable to such differences in preferences.

5 Environment

Understanding the user's environment is the task of the SWAN hardware and software. Information about the environment can be determined by the SWAN in one of two ways: bottom-up or top-down (or, as we have seen, typically a combination of the two). Top-down estimations of the user's environment can be used to change the types of navigation cues that are used, or change the kinds of features in the surroundings that are sonified. For example, if the system determines that the user is outside, then offices and emergency exits are not useful (or even sensical). On the other hand, upcoming fire hydrants are relevant in that environment. Even with this kind of "knowledge" about the environment, spectral analysis of the ambient sounds (gathered in a bottom-up manner with a simple microphone) can allow the system to avoid masking of the beacon sounds by ambient traffic noise. Although the same information could be inferred via top-down assessment, this measure will be more responsive to unpredictable ambient noise changes in the environment such as a particularly noisy construction project in a normally quiet street. It is important to note, however, that again, elements of context interact. In some cases, the user may want to override an adjustment made by the system, so that ambient sounds can be fully heard, and the wayfinding cues diminished. Those sounds can, after all, become primary navigation cues if the listener should so desire.

6 Conclusion

Context awareness is at the very least an advantage for mobile computing systems, and for some systems, a necessity. When designing for such a system it is critical to emphasize the basic factors of task, user and environment. Each needs to be considered individually, in addition to the interaction that forms the totality of a 'context.' A truly context aware system must not only know the state of these factors, but also be able to account for how changes in one affect the others.

References

1. M. S. Wolgather, M. J. Kalsher, and B. M. Racicot, "Behavioral compliance with warnings: Effects of voice, context, and location," *Saftey Science*, vol. 16, pp. 637-654, 1993.
2. J. Edworthy, S. Loxley, and I. Dennis, "Improving auditory warning design: Relationship between warning sound parameters and perceived urgency," *Human Factors*, vol. 33, pp. 205-232, 1991.
3. R. D. Patterson, "Auditory warning sounds in the work environment," *Phil. Trans. R. Soc. Lond.*, vol. B327, pp. 485-494, 1990.
4. S. Brewster, "Non-speech auditory output," in *The Human Computer Interaction Handbook*, J. Jacko and A. Sears, Eds.: Lawrence Erlbaum Associates, 2002, pp. 220-239.
5. D. S. Brungart, M. A. Ericson, and B. D. Simpson, "Design considerations for improving the effectiveness of multitalker speech displays," presented at 8th International Conference on Auditory Display, Kyoto, Japan, 2002.
6. G. Kramer, B. N. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner, J. Neuhoff, R. Bargar, S. Barrass, J. Berger, G. Evreinov, W. T. Fitch, M. Gröhn, S. Handel, H. Kaper, H. Levkowitz, S. Lodha, B. Shinn-Cunningham, M. Simoni, and S. Tipei, "The Sonification Report: Status of the Field and Research Agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display," International Community for Auditory Display (ICAD), Santa Fe, NM 1999.
7. B. N. Walker, J. Lindsay, and J. Godfrey, "The Audio Abacus: Representing a wide range of values with accuracy and precision," presented at 10th International Conference on Auditory Display (ICAD2004), Sydney, Australia, 2004.
8. B. N. Walker and M. Lowey, "Sonification Sandbox: A graphical toolkit for auditory graphs," presented at Rehabilitation Engineering & Assistive Technology Society of America (RESNA) 27th International Conference, Orlando, FL, 2004.
9. Y. Liu, "Software-user interface design," in *Handbook of Human Factors and Ergonomics*, G. Salvendy, Ed. New York, NY: John Wiley & Sons, Inc., 1997, pp. 1689-1724.
10. B. N. Walker and J. Lindsay, "Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice," *Human Factors*, in press.
11. B. N. Walker and J. Lindsay, "Auditory navigation performance using bone-conduction headphones," presented at International Conference on Auditory Display, Limerick, Ireland, 2005.
12. B. N. Walker and R. Stanley, "Thresholds of audibility for bone-conduction headsets," presented at International Conference on Auditory Display, Limerick, Ireland, 2005.
13. B. N. Walker, R. Stanley, N. Iyer, D. S. Brungart, and B. D. Simpson, "Evaluation of bone-conduction headsets for use in multitalker communication environments," presented at Annual Meeting of the Human Factors and Ergonomics Society, Orlando, FL, 2005.
14. T. V. Tran, T. Letowski, and K. S. Abouchacra, "Evaluation of acoustic beacon characteristics for navigation tasks," *Ergonomics*, vol. 43, pp. 807-827, 2000.