Modifying Distracting Headphone Audio to Increase Situation Awareness

A Thesis
Presented to
The Academic Faculty

by

Keenan R. May

In Partial Fulfillment
of Requirements for the Degree
Master of Science in Psychology in the
School of Psychology

Georgia Institute of Technology
May 2018

Modifying Distracting Headphone Audio to Increase Situation Awareness

Approved by:

Dr. Bruce N. Walker, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Richard Catrambone
School of Psychology
Georgia Institute of Technology

Dr. Francis T. Durso
School of Psychology
Georgia Institute of Technology

Date Approved: June 29, 2016

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

SA	Situation Awareness
BC	Bone Conduction
AC	Air Conduction
NHTSA	National Highway Transportation Safety Administration

SUMMARY

Listener situation awareness (SA) was assessed in a dynamic auditory-only simulated roadway environment created with a multi-speaker setup. The ability of the listener to accurately report the presence, current location, and future location of auditory "vehicles" was measured. This was done in the presence of different presentation methods for distracting music played over headphones, in order to assess which combination of methods was least detrimental to SA. The chief manipulation was whether distracting music was virtually spatialized, which was expected to increase the ease of auditory stream segregation and ultimately improve SA. Also manipulated were two common safety measures that interact with spatialization quality: (a) whether bone conduction or air conduction headphones were used; and (b) whether sounds were presented to one or two ears. Spatialization of distracting music had positive effects on hazard localization under some conditions, but negative effects on hazard presence awareness. Using one ear and bone conduction headphones each had positive effects on SA. Results indicate that pedestrians and cyclists should utilize bone conduction headphones and/or listen with one ear, and that designers and developers should consider spatializing distracting sounds as a safety measur

CHAPTER 1

INTRODUCTION

1.1 Auditory Distraction in the World

As we continue to integrate computing tasks into our daily lives, we sometimes struggle to maintain awareness of our immediate surroundings. While the salient image of this is the pedestrian staring down at a smartphone, oblivious to possible dangers, it is also common for pedestrians or cyclists to listen to audio while conducting these everyday movement tasks. Portable audio playback devices come in a variety of forms, such as earbuds, over-ear headphones, smartphone speakers, and one-ear "hands-free headsets." These devices help us stay connected, informed and entertained while we move about the world during activities such as cycling, walking, or jogging. These tasks all involve navigation in dynamic spaces that present emergent, unpredictable hazards. For sighted individuals, safely performing these tasks is based to a large extent on visually scanning the environment, which should not be disrupted by distracting audio (Wickens, 1991). However, while the issue is often given less attention, much awareness of the environment and its potential hazards also comes from sounds. The human auditory system is well adapted to the task of identifying and orienting attention toward potential threats, particularly those outside of our visual field of view. Unfortunately, using portable audio devices can impede the auditory system's ability to perform this key function, which has led to a rise in adverse events. Lichenstein, Smith, Ambrose, and Moody (2012) identified the trend of pedestrians wearing headphones and not hearing auditory warning signals as a common cause of motor vehicle collisions with pedestrians. Listening to music through earbuds lowers correct detection rate for cyclists attempting to hear auditory warnings (De Waard, Edlinger, & Brookhuis, 2011). Kuzel (2008) assessed cases in which pedestrians were distracted via audio,

and found a number of cases involving collision with salient roadway hazards at road crossings. Pedestrian fatalities, previously on the decline, increased from 2009 to 2012 (NHTSA, 2014). Stelling-Kończak, van Wee, Commandeur, & Hagenzieker (2017) found a majority of polled Dutch cyclists felt that phone conversations and music impaired their ability to perceive traffic sounds. While frequency of these behaviors was not significantly correlated with accident risk, respondents reported needing to undertake a variety of compensatory behaviors to offset their impaired perception, such as listening at low volume or listening with only one earbud. Potentially compounding this problem going forward is the increasing prevalence of relatively quiet electric vehicles: Stelling-Kończak, van Wee, Commandeur, & Agterberg (2016) found that electric vehicles tended to be localized more poorly than conventional vehicles.

In spite of the dangers, these behaviors continue to be common. Goldebeld et al. (2012) surveyed Dutch cyclists (which comprised around 50% of the population of the Netherlands in 2003 (Daniel, 2003)) and found that over a third of cyclists aged 12-34 and around a fifth of those aged 35+ "always" or "nearly always" listened to music while cycling, even in the high-risk situations of riding at night, crossing an intersection, or riding in heavy traffic.

While the pervasiveness of cycling, as well as cycling behaviors, is different in the U.S., there is evidence distracted cycling is a growing problem there as well. A 2012 NHTSA report found that around 1/5 of U.S. cyclists at least sometimes used electronic devices while they were cycling, and that 3% reported a bicycle-related injury in the last two years. Meanwhile, more respondents reported that they were cycling "more often" than "less often" in the previous year (comparing 2012 to 2011) than when this question was asked in 2002 (comparing to 2001); in 2002, "more often" and "less often" replies were equally common (Schroeder & Wilbur, 2012).

There are a variety of possible benefits to both society and the individual that can be expected to occur with increased travel using bicycles or by foot, such as reduced emissions, health benefits, reduction in vehicle-related expenses, and urban space reclamation. However, the confluence of increased interest in cycling and walking in urban areas, an increased number in electric vehicles, and the increasing ubiquity of portable audio devices means that the issue of auditory distraction is likely to get worse in the near future.

1.2 Possible Solutions

From a public health perspective, common approaches to addressing this issue has been to run awareness campaigns to promote safer behaviors, or make such behaviors illegal. The majority of government-issued guidelines relating to auditory distractions for pedestrians relate to either turning the volume down far enough so that a person can still hear their environment, or to not wear headphones at all in specific situations such as at crosswalks (Mwakalonge, Suihi, & White, 2015).

There is also a growing effort to approach the problem through the development of new technologies – specifically, technologies that modify the playback method. Several approaches to maintaining awareness of the environment while hearing digital content have been proposed. Lindeman, Noma, and de Barros (2007) characterized these two approaches as “mic-through” and “hear-though.”

“Mic-through” or “pass-through” devices record environment sounds through in-ear microphones and play them back in real-time through headphones. Mic-through devices were first patented in 2006 (Lee & Arthur, 2006). At present, mic-through devices have gotten some traction amongst consumer devices, but the current implementation is cruder than it will likely be in the future; at present, these devices allow adjustment of mic-through volume as a whole, and

there are limitations on sound fidelity and timeliness. This type of system also has been applied effectively when auditory SA is needed alongside hearing protection devices (Killion, Monroe & Drambarean, 2011). Mic-through devices may in the near future offer a great deal of control and flexibility to listeners by intelligently allowing only certain sounds to be presented to the listener.

"Hear-through" or, more literally, "non-ear-covering" (Mwakalonge, White & Siuhi, 2014) devices remove as much of the physical obstruction posed by the device as possible, and allow environment audio to pass unimpeded into the listener's ear. This can be accomplished by mounting a small directional speaker on a person's head, aimed at their ear, but is at present most often accomplished by using bone conduction (BC) headphones. BC headphones vibrate sound through bones of a listener's skull (generally the cheek bones or the mastoid process) rather than the air in the ear canal. These sounds then propagate into the cochlea, thus bypassing the outer and middle ear. This leaves the ear canal unobstructed, and allows sounds from the environment to stimulate the tympanic membrane without degradation.

Figure 1. Promotional material for Trekz Titanium BC headphones[1]

[1] Trekz Titanium [Two joggers wearing bone conduction headphones]. (2015). Retrieved from http://aftershokz.com/products/trekz-titanium

BC headphones (see Figure 1) are currently in use by cyclists, runners, professionals, and persons with low vision. There are three main differences between AC and BC devices. First, since they do not reduce the intensity of environment sounds, and generally cannot operate at high volume without resorting to undesirable AC "leakage" to augment sounds, the ratio of environment sounds to headphone content is often higher than with AC devices. However, this of course depends on the volume at which a person chooses to listen.

Second, leaving the outer-ear open means that environment sounds are not subjected to muffling or distortion associated with earbuds or over-ear devices. While an AC earbud is still permeable to sound, the sound that passes through is not the same as the sound that initially struck the earbud, which means that the listener has a degraded input source from which to reconstruct a model of their environment.

Third, the frequencies that are transmittable from BC devices (i.e., transducer response), as well as the equal-loudness curves for BC sound itself (Walker & Stanley, 2005), are not the same between AC and BC. Generally, BC sounds are characterized by inadequate high and low ranges, with over-represented midrange. While this negatively impacts sound quality and speech intelligibility (Gripper, McBride, Osafo-Yeboah, & Jiang, 2007), it can also be expected to lead to lessened simultaneous-masking of environment sounds.

Thus, while BC devices are likely to be at least marginally better for SA compared to moderately obstructing AC earbuds, the extent to which this is true, and the form the advantages take, is worthy of exploration. Several studies have made steps toward addressing this question, but have not utilized a realistic task environment. Chang-Geun, Lee and Spencer (2011) found that participants had impaired performance on a simple SA task (detection of a sound while

walking on a treadmill) while listening to music over BC headphones, compared to not listening to music, but improved performance relative to AC earbuds. May and Walker (2017) evaluated participants' accuracy in localizing static targets, in the presence of different types of BC distractors (speech only versus speech + music) and under two types of secondary task instructions (instruction to ignore distractors and instruction to attend to distractors). They found that localization performance was worse for all conditions, compared to a condition with no distractors, and that localization performance was worse in the music+speech condition compared to the speech only condition. No differences were found between "attend" and "ignore" conditions. These findings suggested that simultaneous masking can still be expected to have significant detrimental effects on SA-supporting processes such as localization, even when BC headphones are used, environmental sounds are unimpeded and clearly audible, and the listener is explicitly instructed to attend to environment sounds.

"Mic-through" devices will have analogous issues. Since no microphone is perfect (especially one small enough to fit inside an earbud), and no speaker is perfect, environment sounds will always be reproduced in a degraded form.

Thus, current consumer trends toward the development of "hear-through" and "mic-through" devices do not go far enough; distracting audio needs to be further adjusted so that it is less disruptive to SA-supporting processes. The present study is aimed at evaluating one such intervention, and how it interacts with a listener's choice of device type (standard AC or hear-through BC), as well as their choice to use both ears or just one.

1.3 Formation of Situation Awareness

When assessing additional interventions that could improve SA, it is crucial to understand how auditory SA is achieved. SA in general was characterized by Endsley (1987) as

"the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future." The process by which SA is generated is known as situation assessment (Endsley, 1995). Endsley characterized SA as having three levels, each corresponding with both increased computational difficulty and increased understanding of one's environment as it relates to the task at hand.

While situation assessment ultimately requires higher order cognitive processing, it first requires that certain essential *elements*– bits of information crucial to task performance– are perceived successfully (Endsley, 1987). As such, Level 1 SA was characterized as *perception*, or the process of noticing the presence of elements in the environment, or the presence of changes in those elements, without necessarily being able to attach meaning to said elements.

Maintenance of Level 1 SA can be achieved to some extent through stimulus salience effects and attentional capture by key elements (Salvendy, 2012), but may also require that selective attention is shifted through an environment with a pattern that allows a person to notice key elements. While the least cognitively complex of the SA levels, Jonas and Endsley (1996) found that the bulk of aircraft accidents were attributable to a failure at this level– i.e., often times people simply fail to notice things.

To generate level 1 SA, preattentive processes, such as feature detection and stimulus-driven ("bottom-up") pop-out effects (Triesman & Paterson, 1984), act on a scene, and then a person may notice these, or not, depending on the salience of the stimuli and their choice of selective attention allocation. Endsley (1995) noted that, at this stage, cue salience has a large impact on the extent to which a given element is processed as a distinct object, separate from its perceptual "background," and is thus can be represented properly in a person's situation model. These preattentive distinctions assist subsequent active application of selective attention in

drawing useful information into working memory. Stronger preattentive distinctions can make subsequent application of selective attention (perhaps as part of a visual scanning pattern) more information-lucrative (Salvendy, 2012).

Level 2 SA was characterized as *comprehension*, which is the process of relating elements that have been perceived to the task at hand, imbuing them with meaning, and integrating them into a mental model. Endsley (1987) characterized this mental model as an internal representation of the environment, while Durso et al. (1995) highlighted that it can also be a distributed model comprised of references to information stored in the environment. Level 2 highlights the difference between perceiving elements, and comprehending their meaning. This may be due to degraded perception, or a failure to properly match information about an element to known schemas and the task at hand (Salvendy, 2012). Maintaining this level of SA hinges upon correct perception of key aspects of SA elements as well as attaining an understanding of how those elements relate to future behavioral choices (Endsley, 1995).

Finally, Level 3 SA, known as *projection*, describes a person's ability to integrate Level 1 and 2 information with existing knowledge and schemas in order to produce a mental model that can be "played forward" in time. Achieving this can require extensive manipulation of objects in working memory. The cognitive workload associated with creating and maintaining a comprehensive mental model, as well as matching the current situation with related schema stored in long term memory and doing the "playing forward" itself may be extensive. A person's working memory capacity (Wickens et al., 1987) as well as their experience level and (relatedly) whether they have appropriate pre-existing knowledge structures that can be brought to bear on a situation and are properly activated (Fracker, 1988) all affect Level 3 SA.

In summation, SA proceeds from basic perception toward complex, higher order understanding of the environment, and is more than simply being able to detect alerts and warnings. Failures or slight inaccuracies at lower levels can propagate up to higher levels.

1.4 Forming SA from Sound

SA derived from a person's auditory environment operates in a similar fashion to the visual-centric mode in which SA is typically discussed, but there are a few important differences. Many of these differences stem from the unique perceptual challenges of perceiving auditory objects relative to visual objects. Consideration of visual-centric SA may focus on constraints derived from the need to serially shift the eyes between different well-formed perceptual objects in the environment, in addition to the aforementioned difficulties with working memory. Often, each item is relatively clear when attended to (instrument gauges, icons on a display, etc.), but there is simply not enough time to keep them all current in a person's internal/distributed representation. By contrast, when generating SA from auditory environment features, there is a significant bottleneck present in terms of creating the perceptual objects themselves. In dynamic auditory environments, it is rare for only a single sound to be playing at one time. As such, finding out which components of sound input correspond to which objects in the environment is a core problem of auditory perception, similar to how Gestalt principles and figure-ground discrimination are important for visual perception (Bregman, 1994). *Auditory scene analysis* is the perceptual organization process through which the undifferentiated input waveform is parsed into a number of useful auditory objects (Bregman, 1994). While some of separation can be achieved through analysis of a single moment in time, many of these discriminations must be built up over time into separate *auditory streams* via the process of *auditory stream segregation* (Bregman, 1994). When these streams correspond to auditory objects (Bizley & Cohen, 2013),

they can be incorporated into a person's situation model. As such, the quality of stream segregation that is able to be achieved is crucial to supporting SA in dynamic auditory environments. There are a variety of features of sound sources that the auditory system will utilize to build up a greater number of streams. Designers of auditory displays often craft display elements so that each dimension gives the listener a variety of these cues, to support simultaneous comprehension of multiple data dimensions (Nees & Walker, 2009). Before discussing these cues, it is important to note that there is evidence that stream segregation also depends on person's allocation of selective attention. Cusack, Deeks, Aikmen and Carlyon (2004) proposed the "hierarchical decomposition" model of stream segregation, in which preattentive processes automatically split the input waveform along basic, high level categorical divisions present in an auditory scene, and attention allocation is required to create further splits. For example, restaurant noise could be preattentively divided into "band playing" and "people talking" without requiring that a person be attending to either of those stream clusters. However, there is evidence that allocation of attention is necessary to split elements within those high-level categories. In the restaurant example, attention would need to be allocated toward "people talking" in order to create streams for specific conversations of interest.

Regardless of the allocation of attention, if sound sources are distinct enough in terms of the various stream segregation cues, the hierarchical decomposition model predicts that these can be separated in an automated, pre-attentive manner. This is similar to the visual perceptual grouping and "pop-out" effects described by Triesman and Paterson (1984). If preattentive segregation is not possible, increasing stream distinctness still increases the ease with which a listener brings their selective attention to bear on one aspect of the environment and further segregate those streams.

Another issue specific to auditory SA are the spatial computation problems of (a) localizing auditory objects and (b) tracking multiple spatial objects over time. Sound source localization is itself a complex computational issue (Neuhoff, 2004). Frequency-selective masking can disrupt two of the main binaural cues, front-back reversals are common, and reverberations can lead to decreased localization accuracy (Hafter, Saberi, Jensen & Briolle, 1992).

Finally, the difficulty associated with tracking multiple auditory objects over time has not been studied extensively on its own. Often, the problem of stream segregation or localization of individual objects becomes a perceptual bottleneck before any perceptual or working memory limitation on spatial object tracking would come into play.

1.5.1 Sound-based SA for Cyclists and Pedestrians

Stelling- Kończak, Hagenzieker, and van Wee (2015) provided an overview of auditory perception and SA formation issues that are faced by distracted cyclists. These issues center on task difficulty associated with stream segregation; as such, stream segregation was targeted as the area in need of technological intervention.

In multi-task scenarios such as riding a bicycle and listening to music, two scenarios may occur. In one, the cyclist is focusing selective attention on their music. Here, when pre-attentive stream segregation is able to produce a useful auditory object (such as a car coming up to their side), this object may be represented in their situation model. Technological interventions that decrease the computational difficulty of the stream segregation problem may lead to preattentive segregation processes delineating an oncoming vehicle sooner rather than later.

In the other scenario, the cyclist is deliberately shifting selective attention between music or other computing sounds, and the “road sounds” stream cluster. Since stream segregation is a

process that happens over time (i.e., streams take time to build up), this scenario – that is, one in which selective attention switches back and forth between stream clusters without lingering on one for very long – is not ideal. Here, interventions that increase the ease of stream segregation will decrease the amount of time it takes the listener to become aware of traffic sounds after a shift in selective attention.

Setting aside the presence of distracting digital sounds, the specific listening task of the cyclist is to keep aware of the position and direction of movement of the vehicles in their immediate vicinity. If the cyclist is following a vehicle, they may use their sight to keep track of the lead vehicle. If they are riding in a bike lane, audio may be needed to maintain awareness of the vehicle behind them and to one side.

The task faced by the pedestrian crossing the street or walking down the sidewalk is similar. The sighted pedestrian generally looks one way, then the next, and must keep track of one or more vehicles that may approach unexpectedly from either direction.

In both cases, the number of objects to be tracked and modeled is relatively small-perhaps one to four vehicles. The bulk of the difficulty in this task stems from correctly perceiving the presence, location and direction of travel of the vehicles, in the presence of extensive background city noise as well as noise generated by other vehicles. In isolation, this could be a moderately difficult exercise in spatial auditory object tracking and selective attention. However, in the presence of the aforementioned ambient and competing vehicle noises, stream segregation may become the key source of difficulty. Layering additional noise stemming from digital content over this can be expected to add to the inherent difficulty of the stream segregation task.

Thus, the present work focuses on technological interventions that could improve the quality of auditory stream segregation, which may then alleviate deficits in higher level SA and, ultimately, task performance, that may stem from the inherent difficulty of the stream segregation problem in these task contexts. The focus is on correct perception of elements critical to forming SA, rather than the presence of more the higher level-understanding more typical to operationalizations of SA.

1.5 Auditory Stream Segregation Cues

Bregman (1994), and Cusack and Carolyn (2004) described the set of features that, when present, make the stream segregation task easier.

Many known stream segregation cues relate to an analysis of the frequencies present in a sound. First, the overall frequency content of a given sound is a cue for stream segregation. Sounds that contain intensity in different bands of the spectrum are more likely to be segregated. Relatedly, pitch, or the fundamental frequency of a sound, is also used as a cue. Harmonicity is also used as a cue: if one sound could be within the harmonic complex of another sound, then that sound is less likely to be segregated. Also, the timbre of a sound, which stems from the harmonic profile as well as details such as the attack/decay rate and presence of vibrato, can be used as a cue (Iverson, 1995). Finally, the variance of pitch over time in the form of a melodic pattern is known to be used by the auditory system to segregate streams (Szalardy et al., 2014)

The auditory system also uses several cues relating to sound intensity. The first is the overall intensity of a sound: if two sounds vary greatly in intensity, they are likely to be segregated (Cusack & Carolyn, 2004). Relatedly, patterns in amplitude variations over time (tempo) can also be used to delineate sounds (Szalardy et al., 2014).

There is also strong evidence that spatial separation increases the ease of stream segregation (Bohm et al., 2013; Cusack et al., 2004; Denham et al., 2010; Szalardy et al., 2013). Further, the greater the apparent difference in the angle of origin for two incoming sound sources, the more likely the two sound sources will be rendered as separate perceptual objects (Bonebright et al., 2001; Brown et al., 2003). Middlebrooks and Onsan (2012) found that stream can be segregated with as little as 8 degrees of spatial separation. There is also some evidence that the reflections and echoes of a sound are used as cues, separate from the apparent angle of a stream (Blauert, 1999).

Several of these cues have been shown to lose their effectiveness in the presence of distractors or other degradation. Frequency-specific hearing impairments have been associated with deficits in stream segregation ability, by disrupting frequency-related cues (Oxenham, 2008). Cusack and Carolyn (2004) noted that masking noise, on principle, can effectively remove several of the aforementioned cues.

1.5.1 Stream Segregation Cues in the Roadway Environment

When it comes to motor vehicle sounds, some of these cues can be expected to be more useful than others. Each vehicle may have a slightly different timbre. Pitch, frequency content and harmonicity may differ slightly as well. One might utilize the "melodic" pattern of a vehicle's acceleration as a cue, as well. However, to a certain extent, motor vehicles sound alike in these regards, and may be similar in the aforementioned ways to the environmental ambience as well (more-so in an urban environment). However, intensity, amplitude variations over time, and the spatial cues of location and reflection/echoes are likely to be more useful, because only these are truly unique to each vehicle.

1.6 Spatial Auditory Processing

Unlike other stream segregation cues, the use of location/ spatial separation as a cue may require increased load on a separate neural pathway and, possibly, a separate processing resource pool. In effect, locating a sound in space may be a separate cognitive-perceptual task and source of difficulty as compared to delineating an auditory stream from undifferentiated input.

Based on studies of macaque brains, Rauschecker (1998) proposed that auditory information processing is divided into a parietal-lateral prefrontal "dorsal" pathway and an anterior temporal-inferior frontal "ventral" pathway. In this theory, the dorsal or "where" pathway processes primarily spatial and location information, whereas the ventral "what" pathway processes primarily information relating to sound identity. Neuroimaging studies have provided support for the existence of separate neurological processing areas that fit this division. Recanzone (2000) found populations of neurons in dorsal areas –but not ventral areas– that were strongly location-tuned. Lomber and Malhotra (2008) used a reversible cooling method to deactivate parts of a cat non-primary auditory cortex. They found that deactivation of dorsal stream areas led to deficits in sound localization tasks, while deactivation of ventral stream areas led to deficits in the ability to identify sound patterns. Ahveninen et al. (2006) conducted an fMRI study on humans and found that activation in dorsal auditory areas increased when the location of a spoken phoneme changed, but not when the phoneme itself changed. Conversely, if only the phoneme itself changed, but the location remained the same, activity increased in the ventral stream but not in the dorsal stream. Lesion studies have provided similar results. Adriani et al. (2003) studied patients who had lesions in the dorsal or ventral pathways and found that those with dorsal lesions tended to have deficiencies in sound localization and motion perception, while the group with ventral lesions tended to have difficulties with sound recognition.

It is worth noting that some recent work has suggested that dorsal and ventral pathways may be more tightly integrated. Bizley and Cohen (2013) proposed that the formation of auditory objects occurs through an integrated process in which dorsal and ventral areas collaborate. Cloutman (2013) reviewed literature on visual, linguistic, and auditory dual-stream processing divisions and characterized proposed models into three categories: models that propose full independent processing until higher cortical integration; models that postulate feedback flowing up one system and down the other; and "continuous cross-talk" models in which the two systems communicate laterally in a feed-forward fashion. Of import here is that the dorsal auditory stream has been shown to precede the ventral stream. Leavitt et al. (2011) found that EEG response increased in dorsal areas around 90ms after auditory stimulus onset, whereas response in ventral areas increased around 100ms around stimulus onset. Chen et al. (2007) suggested that this time precedence (for visual pathways, in this case), as well as the observation of concurrent activation in some dorsal and ventral areas, was evidence of lateral activation flowing directly from the dorsal to the ventral stream. Alternatively, Jaaskelainen et al. (2004) found evidence suggesting that the (visual) dorsal stream acts to influence the (visual) ventral stream in a top-down fashion through the prefrontal cortex, with feedback flowing back down to orient ventral stream processing toward features key to the ventral stream's task of object identification.

Finally, it is worth nothing that some of the processing that can be characterized as "stream segregation" occurs at very low level in the auditory pathway (during the first ~90ms after transduction, before any dorsal/ventral split), and that streaming-related computations that occur at this level are propagated up to the cortical level (Yao, Bremen & Middlebrooks, 2015).

1.6.1 Spatial and Sound Identity Processing: Practical Distinctions

While the literature is clear on the presence of a functional and neurological split, it is less clear whether spatial and non-spatial sound identity processing streams should be treated as separate resource pools (Wickens, 2007). However, regardless of the extent to which stream segregation and spatial location determination are truly separate and parallel tasks, it is apparent that the task of keeping track of the location of auditory objects is different from the task of keeping auditory objects distinct. One can imagine a scenario in which a listener must keep track of the location of 3-5 well-formed, distinct auditory objects, each moving about rapidly. This task would be difficult for spatial processing faculties, but workable for sound identity processing. Alternatively, one can imagine a scenario in which two auditory objects that have near-identical frequency content, intensity, etc. are unmoving, and located at the same location in space. The listener would find this taxing on their stream segregation capacities, but not on their ability to process the spatial locations of the object(s).

Thus, the performance impact of virtually spatializing computing audio depends on the relative impact of these two processing considerations, and which processing subsystem is the "bottleneck" in this type of task. Regardless of the extent to which virtually spatializing distracting audio would increase load specifically on a spatial audio "resource pool", it is inarguably providing more information for spatial processing faculties to process, even as it provides an additional cue for stream segregation. Adding spatial audio effects is providing additional, helpful information, but also increasing the complexity of the auditory space in other ways; as such, the effects on SA cannot be predicted. In light of this, investigating how distractor sounds could be given a virtual spatial component, as well as how such an effect might interact with other common safety-related listening manipulations, is the focus of the present work.

1.7 Virtually Spatialized Audio

1.7.1 Definitions of Key Terms

When considering spatial audio, it is helpful to understand the relevant terms and perceptual phenomena. Plenge (1974) demonstrated that there is a qualitative difference between headphone sounds that are *lateralized*– that is, tending to be perceived as emanating from a certain intracranial location, and headphone sounds that are *localized*– that is, tending to be perceived as originating from a specific *extracranial* location. In this paper, localization is considered to be the combination of *lateralization* (the "specific direction" component) and *externalization* (the "extracranial location" component, to be contrasted with *internalization*). *Intracranial* sounds that seem to have a "direction of origin" (Iwaki & Chigira, 2016) can be described as lateralized; not all intracranial sounds are lateralized (such as stereo headphone sounds in which both channels are the same). Similarly, an extracranial sound need not be localized; identical sounds that come from speakers placed at equal distance to the left and right will be externalized, but not necessarily localized.

While lateralization, externalization, and localization are perceptual phenomena, *spatialization* refers to the process of taking an artificial sound and manipulating it to induce the perception of *localization*.

1.7.2 Spatial Audio Overview

Virtual spatialization effects are currently uncommon in portable electronics. Content delivered through stereo headsets typically does not include consistent cues to the location of virtual sound sources, aside from occasional panning effects used in music, movies and games. Most stimuli that listeners currently experience while navigating in the world are presented in (mostly) mirrored stereo and thus can be said to have a minimal spatial component. However, the

concept of "Mobile Augmented Reality Audio (MARA)" (Härmä et al., 2004) has been around for some time.

Virtually spatialized auditory objects have been used in auditory menu interfaces designed for low vision users (Zhao et al., 2007) and for drivers who need to keep their eyes on the road (Sodnik et al., 2008). Härmä et al. (2003) described "wearable augmented reality audio" devices that blend virtual sound sources with real sound sources. Systems such have these have been evaluated by Loomis, Golledge, and Klatzky (2001) as well as Walker and Lindsay (2006) and Wilson et al. (2007). These systems use spatialized auditory beacons as navigation aids for blind or situationally blind users.

However, not all virtually spatialized sounds have a clearly corresponding location or direction in real space. Härmä et al. (2003) called sounds that lack a multimodal correspondence "freely floating acoustic events," which they defined specifically as "[audio] events that are not connected to objects in the user's environment." They enumerated possible uses of such information; these include alerts, news, and music listening. To date, freely-floating acoustic events, and their potential impact on SA compared to traditional "non-spatial" acoustic events, has not been considered.

1.7.3 Implementing Virtually Spatialized Audio

Härmä et al. (2003) proposed five prerequisite aspects of a headphone sound signal that may be required to spatialize a sound. These can be grouped into effects that primarily support lateralization, and those that support primarily externalization.

1.7.3.1 Creating Lateralization

Two of the perquisites laid out by Härmä et al. (2003) that correspond primarily with lateralization are the implementation of Head-Related Transfer Functions (HRTFs) and changes

due to head tracking. HRTFs work by applying spectral distortion to audio depending on its virtual location in order to simulate the differential effects that the human pinnae, head, and shoulders have on incoming sounds, based on the direction from which they come. These HRTFs can be generalized across humans or can be customized for a given person's pinnae shape. While customized versions are preferable, and are more likely to induce externalization, generalized HRTFs can be effective at inducing at least lateralization. Unlike other effects, HRTFs reduce front-back reversals and allow for elevation discrimination, as well as providing an additional azimuth cue (Begault, Wenzel & Anderson, 2001).

Two other effects help with left-right discrimination, and have been traditionally used to lateralize sounds. These are sometimes included as components of the HRTF, while in other cases (and in this paper) the HRTF refers only to spectral manipulations. Interaural Level/Intensity difference (ILD) between the two sources can be simulated by modifying the volume received by the left or right ears (or in the single ear, if only one ear is being used) based on which ear is closer to the virtual sound source. Interaural Time/Phase Difference (ITD) can be simulated by introducing a brief delay between when a sound is played in one ear compared to the other (Bernstein, 1997). For continuous sounds such as music, this amounts to slightly phase-shifting the left and right audio.

1.7.3.2 Creating Externalization

Härmä et al. (2003) also described cues that relate specifically to externalization. Acoustic environment simulation or, at least, parameter-based reverberation effects, were given as a prerequisite for creating audio that is externalized, over and above being lateralized. Blauert (1999) characterized the common phenomena of headphone users experiencing sounds as being intracranial. Zahorik (1998) demonstrated that externalization was achievable over headphones

using in-ear microphone recordings, which inherently incorporated realistic acoustic sound diffusion. Simulating this through programmatic effects requires implementing several features. A basic cue for sound source externalization is loudness. Coleman (1963) found that the sound level of a source that retains consistent intensity falls off by 6dB each time the distance of the source is doubled (setting aside reverberations). However, Mershon & King (1975) showed that intensity is a cue primarily for *relative* distance if a sound source is presented multiple times in sequence. Reverberation, however, is an absolute distance cue. A low ratio of initial sound volume to subsequent reverberations of that sound indicates that a source is far from the listener (Mershon & King, 1975). The fidelity of reverberations required to induce an externalization illusion may be moderate. Begault, Wenzel, and Anderson (2001) found no differences between an "early reflection" reverberation model using either 800ms of simulated room-model reverberation or a full aural reflection condition with 2200ms of simulated room-model reverberation. Higher fidelity sound diffusion models, commonly implemented using finite-difference time-domain (FDTD) waveform simulations, model environmental features such as occlusion and diffraction (Savioja, Lokki, & Väänänen, 1999). In the current use case of mobile computing, it is currently prohibitively difficult to calculate realistic reverberation and waveform propagation effects, with an accurate world model and in real-time, but broad matches to a listener's acoustic space may be feasible. Finally, Loomis, Klatzky, and Golledge (1999) described several other minor externalization cues including spectral changes due to transmission through the air, and motion parallax. For the present study, complex acoustic simulation was not attempted due to resource limitations.

1.7.3.3 Matching Expectations and Creating Multimodal Connections

The final prerequisites of Härmä et al. (2003) were multimodal connections and user expectations. For freely-floating acoustic events, multimodal connections do not exist in the world. Such connections could be created in a mobile computing context through visual augmented reality rendering of elements that coincide spatially with the virtually spatialized audio elements. User expectations are difficult to characterize in a mobile computing context. Expectations about the listening environmentf will change moment-to-moment, and expectations about the virtual sounds themselves may not yet exist. Users could, however, be told to expect a spatialized audio signal, or could be given visual accompaniment such as an inactive physical speaker to take advantage of these effects.

1.7.3.4 Combining Effects to Achieve Spatialization

When the aforementioned effects are implemented with sufficient accuracy, listeners can be made to perceive a spatialized sound. Begault and Wenzel (1991), using many of the aforementioned effects, first demonstrated how externalization could be achieved using headphones and simulated sound properties, instead of in-ear recordings. Begault, Wenzel, and Anderson (2001) found that a combination of generalized HRTFs, reverberation and head tracking led participants to report that virtual sounds were externalized and located at a distinct location in space 79% of the time (compared to only 40% with HRTFs alone with no reverberation effects). Loomis, Klatzky, and Golledge (1999) found that effective reverberation and ILDs were most important to achieving localizability.

1.8 Common Safety-Related Playback Choices That Affect Spatialization Quality

Several device choices that a person may make in order to minimize auditory distraction, interact with how virtual spatialization of distracting audio content is likely to be perceived.

1.8.1 Presenting to One or Two Ears

First, whether a device plays audio to one ear or both will influence the extent to which spatialization effects may be implementable. Listening with one ear is a common practice used by cyclists (and recommended by some governments [Mwakalonge, Suihi, & White, 2015]) for reducing distraction. However, virtual spatialization effects may require two ear-input to function effectively, due to the cues of ILD and ITD not operating when only one ear is used. Regardless of whether spatialization effects are present, one-ear presentation is a scenario in which distractor sounds are being presented at a constant location within the listener's head (lateralized), while presenting to both ears can produce a sound that does not appear to have a point of origin. As such, while the use of one ear versus two obviously affects simple audibility, it also affects whether distractor sounds have a salient spatial component (one-ear presentation) or not (two-ear presentation).

1.8.2 Use of AC or BC Headphones

A person who is striving to avoid distraction may also elect to use BC headphones. Whether an AC or BC device is used can be expected to have some impact on the degree to which virtual spatialization effects are effective. Until relatively recently, even lateralization was not thought to be feasible via BC devices. There are three main difficulties to spatial audio implementations over bone conduction: (a) degradation of spectral cues due to over-representation of midrange frequencies (Walker & Stanley, 2005); (b) reduction of ITD due to the physics of bone conduction; and (c) limited ILD due to crosstalk within the skull (MacDonald, Henry, & Letowski, 2006).

Walker and Stanley (2005) demonstrated that lateralization was possible with BC headphones, but found that performance was inferior to AC headphones. Walker and Stanley (2005) found that participants were less accurate at identifying the location of BC sounds that

were virtually lateralized along the left-right plane. MacDonald, Henry, and Letowski (2006) compared localization performance for virtually spatialized sounds played through BC or AC devices. The authors noted that BC devices typically cannot effectively convey higher frequencies- in particular those associated with interaural level differences, which are most effective for high frequency sounds. While they found that successful lateralization was possible over BC, they did not measure perceptions of externalization. Lindeman, Noma, and de Barros (2007) compared HRTF-spatialized sounds presented through BC and AC, and found that localization accuracy (naming the correct speaker) was lower when BC headphones were used.

Compensation methods have been developed for the three aforementioned difficulties introduced by BC when attempting to present spatialized audio.

Stanley (2009) developed "bone adjustment functions" (BAFs) to compensate for the misrepresentation of HRTF spectral cues when BC is used (again, broadly speaking due to over-emphasis of the midrange). Stanley (2009) found that localization performance was improved for HRTF-spatialized sounds presented over BC when a BAF was applied; however, this difference was only significant for elevation judgments (in which spectral cues are most important). Iwaki and Chigara (2016) used an adjustment procedure to produce compensation filters for correcting ILD and ITD differences when using BC. Localization error relative to AC-spatialized sounds was reduced when these two (participant-specific) compensatory filters were applied.

A main issue with these compensatory methods is that (a) workable generalizable functions have not yet been developed and (b) convenient calibration methods have not been developed. Stanley (2009) wrote that BAFs were quite different between individuals; Iwaki and Chigara (2016) noted the same for the ILD and ITD compensations. As such, for the present study, no compensations were made when BC headphones were used. In the present study,

generic spatial audio effects were applied to BC and AC, without compensation for BC. Thus, one research question was simply whether uncompensated spatial audio effects, degraded as they were by BC presentation, might be equally effective in increasing the ease of stream segregation and ultimately SA formation.

Setting aside the question of whether spatial audio is feasible through BC, BC devices have also not been evaluated in a realistic listening context with multilevel measures of SA. Mwakalonge, White, and Siuhi (2014) pointed to BC devices as a possible technological intervention to distracted cycling, and recommended work be done to verify the extent to which this may be true. There are several advantages that BC devices have over AC, but these may be relatively minor in terms of their practical impact. BC devices (1) preserve the original frequency content of environment sounds, and (2) lead to a different perceived frequency content of computing distractor sounds– generally, sounds are more limited in frequency to midranges (Walker & Stanley, 2005).

However, it remains to be seen whether these differences are impactful enough to create a significant difference between BC and AC earbuds when distractors are played at matched loudness.

1.9 Current Study

The present study investigated whether virtual spatialization of distracting music could be used to improve listener SA in a dynamic environment. Three factors were manipulated: (1) whether presentation occurred in one ear or both ears; (2) whether spatialization effects were present or absent; and (3) whether AC earbuds or BC headphones were used. The target context was that of using headphones to listen to music while attempting to maintain SA and navigate a roadway environment.

CHAPTER 2

METHOD

2.1 Overview

Participants followed a procedure similar to May and Walker (2017), but extended to include direct assessment of SA in a dynamic auditory environment. Participants sat in the middle of a circular array of speakers. They performed a visual-motor navigation task administered via four screens, while tracking the sounds of two simulated vehicles played over the speakers, and hearing distracting music in one of eight possible conditions. Periodically, all sounds ceased playing and participants responded to questions reflecting their SA.

2.2 Participants

Participants were 62 undergraduates from a major southeastern technical university, 28 male and 34 female, aged 18-27 (M = 19.97, SD = 1.73). When asked to report the frequency of relevant activities on a scale of 1(not often) to 6 (very often), participants reported having little experience using bone conduction headphones (M = 1.04, SD = 1.87), or riding a bike while listening to headphones (M = 0.79, SD = 1.35). However, participants reported that they fairly frequently (1) walked while listening to music (M = 4.20, SD = 1.69) and (2) jogged while listening to music (M = 3.20, SD = 1.88). Seventy-nine percent of participants had music experience (M = 6.88 years, SD = 4.80). Eighty percent of participants had experience with video games (M = 6.40 years, SD = 5.07).

Participants reported using a variety of strategies for safely listening to music while out and about. Ten participants mentioned that they keep the volume low, and ten participants reported generally leaving one headphone out. Three participants reported that they tried to use

vision to be more aware of their surroundings, and three said they took special precautions when crossing the street.

2.3 Apparatus

2.3.1 Study Environment

Figure 2 shows the study environment. Study sessions were carried out in room with partial sound-proofing and low levels of ambient noise. Participants sat on a swiveling chair so they could quickly and comfortably rotate to respond to questions. A frame was built around this chair, from which hung a circular curtain of acoustically transparent fabric. This purpose of this was to render participants unable to see the speakers or use their vision to augment responses. Blocking vision in this manner was recommended by Letowski and Letowski (2012) for auditory localization testing.

Figure 2. Experiment environment.

2.3.2 Speakers

Sixteen Eris E5[2] studio monitor speakers were used to render environment sounds (Figure 3). These speakers were designed for sound design, and had a relatively flat frequency response. Sounds were programmatically panned between these speakers to create an arbitrary number of apparent sound directions of origin, and the illusion of a continuous soundscape.

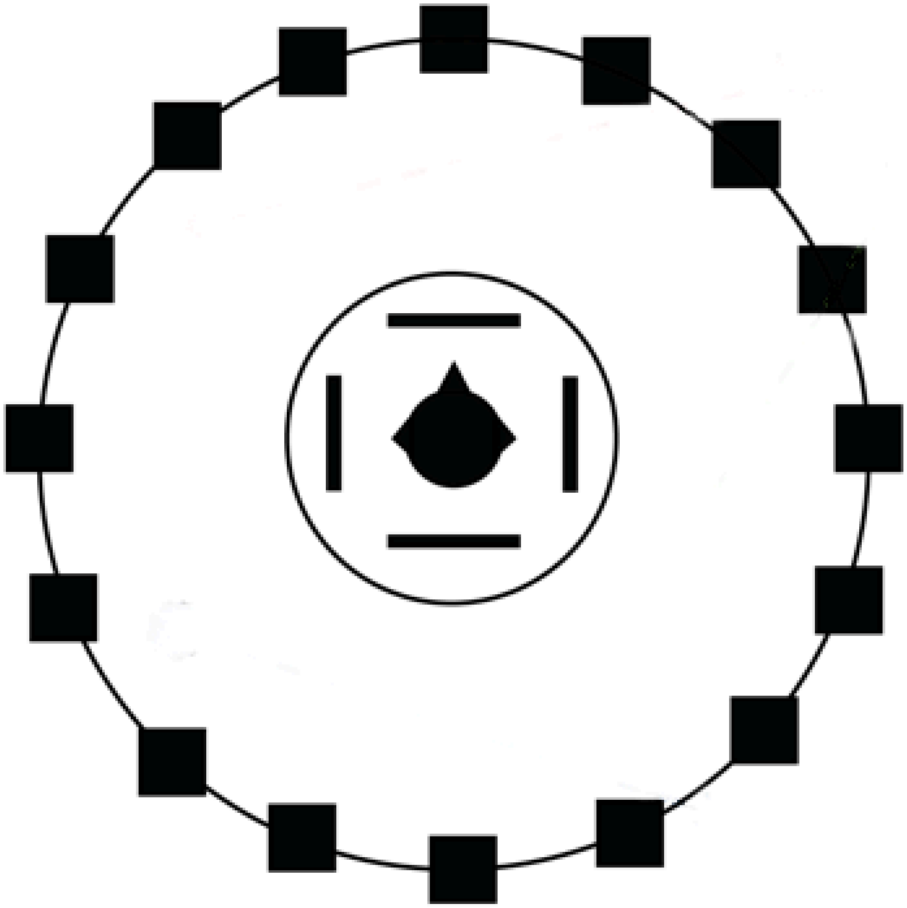

Figure 3. Configuration of speakers, circular curtain, four screens, and seated participant.

[2] Eris E5 | Tech Specs. (n.d.). Retrieved April 06, 2017, from http://www.presonus.com/products/Eris-E5/tech-specs

2.3.3 Headphone Devices

The BC device used was a pair of Trekz Titanium[3] BC headphones (Figure 4, left). The AC device was a set of wired Apple earbuds. These were routed audio wirelessly via a Bluetooth transmitter and receiver.

2.3.4 Head Tracker

Head tracking was done by streaming gyroscope data from a head-mounted iPhone 5s over Wi-Fi to the control computer (Figure 5). Magnetometer data was not used, because it was subject to interference from the speakers. In order to preserve gyroscope accuracy over time and combat "sensor drift," the participant's head angle was automatically recalibrated each time they re-acquired a navigation task fixation cross after responding to a SAGAT probe, since their head angle was known at this point. The head mount was a construction helmet that participants were able to adjust to get a comfortable fit.

Figure 4. BC (left) and AC (right) devices with smartphone head-tracker helmet.

[3] Trekz Titanium. (n.d.). Retrieved from https://cdn.shopify.com/s/files/1/0857/5574/files/Trekz_Spec_Sheet_53a8bcb5-c8e5-445d-8e4a-ee946fcc1139.pdf

3.2.5 Input device

Participants gave their responses by pressing the face buttons on a Bluetooth PlayStation 4 controller[4] (see Figure 5).

Figure 5. Wireless Playstation 4 controller used for input.

2.3.6 Audio Software and Hardware

The software was written in Python 2.7. Visual output was handled by the Kivy graphic user interface module[5], and was mirrored to four screens via a VGA splitter box. Part of each screen was covered so that the participant saw the appropriate visual items. The software listened to input from an iPhone head tracker application that transmitted gyroscope data via UDP to the control computer. Audio output was handled by the PYO digital signal-processing Python library[6]. Output was routed out of a computer into two networked 10-channel USB audio interfaces, providing one channel of output for each monitor speaker, as well as two channels (left and right) for the headphones. These last two channels were routed through two headphone amplifiers (to achieve sufficient volume) and then broadcasted over a Bluetooth transmitter to either the BC headphones or a Bluetooth receiver plugged into the AC earbuds.

---

[4] DualShock 4 Wireless Controller. n.d. Retrieved from https://www.playstation.com/en-us/explore/accessories/dualshock-4-wireless-controller-ps4/

[5] Kivy (Version 1.8). (2016). Retrieved from https://kivy.org/

[6] PYO. (2014). Retrieved May 18, 2016, from http://ajaxsoundstudio.com/software/pyo/

2.4 Materials

2.4.1 Auditory Objects

Critical SA elements when cycling or walking include the location of nearby vehicles which may pose a threat to one's person. Thus, SA questions were inquiries about two auditory objects that could pose a danger to a pedestrian or cyclist– a motor scooter and a car. These sounds were intentionally similar, and might fail to be segregated from each other or the overall soundscape if segregation was degraded, just as would be true in a real-world listening scenario. Schuett and Walker (2013) concluded that participants can be expected to monitor and comprehend up to three continuous auditory streams simultaneously, with exceptions and caveats. Thus, the task of monitoring the positions of the two virtual vehicles should have been possible for participants, but difficult to achieve in the presence of distraction without degradation.

2.4.2 Distractor Music

The music used, shown in Figure 6, was the song "As Colorful as Ever" by artist "Broke for Free" (Cascino, 2012). This song was used under an Attribution-Noncommercial 3.0 International License. The song had no vocals, and could be described as a "groove" drum beat with prominent strings and clean guitar throughout. This song was selected because it was homogenous, lacking sudden onset features, and loop-able without the looping being too noticeable, without being repetitive in a way that was unpleasant to listen to for an extended period of time. Participants spent a significant amount of time listening to the song, so the song was selected to minimize their annoyance. Importantly, participants were not given any instructions regarding the distractor music, aside from being told that there would be music present.

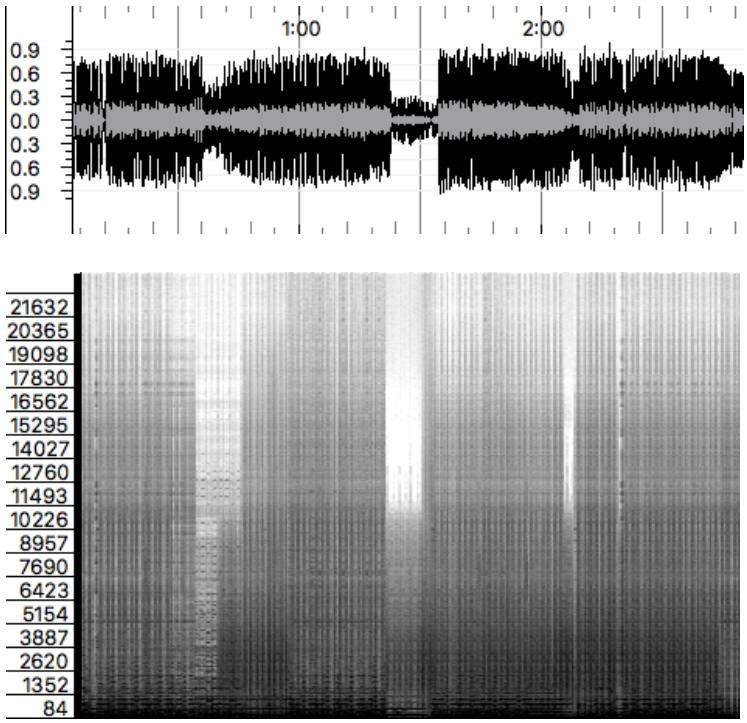

Figure 6. Waveform (above) and spectrogram (below) for distractor music. The ordinate represents time and abscissa the overall intensity (above) or the intensity within each frequency band, in hz (below). Visualizations generated using Sonic Visualizer[7].

2.4.3 Implementation of Spatialization Effects

For the present study, a set of virtual spatialization effects was developed. These effects used a combination of a generalized head-related transfer function (HRTF) simulation of spectral cues, simulation of ILDs by manipulating left and right channel volume, and simulation of ITDs by slightly desyncing left and right audio signals. All of these elements changed in real-time as the participant turned their head. A simple, pre-processed reverberation effect was applied using the

[7] Cannam, C. Landone, C., and Sandler, M. (2010). *Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files*, in Proceedings of the ACM Multimedia 2010 International Conference.

program Audacity[8]. This reverberation effect was tuned to match the size of the experiment room, but was a generic effect rather than an acoustic simulation that accounted for sound travel throughout the room in a realistic manner. The reverberation effect did affect the overall sound intensity and frequency content, so it was kept for the "no effects" sound as well, in the interest of keeping the sound intensity the same. The HRTF function that was used was the open source Python Module Headspace[9]. This module used the publicly available left and right ear impulse tables from Gardner and Martin (1994), who used a KEMAR dummy head for recording.

All of these dynamic changes were turned off for the No Effects conditions. One headphone was also turned off for the "one-ear" conditions. Whether this was the left or right ear was randomized. Additionally, for the AC-one-ear conditions, that headphone was removed from the participant's ear after the software played a prompt indicating which earbud to remove. For the "no effects" conditions, the HRTF function was applied constantly as if the sound were at 0 degrees (straight ahead). This was done to better match the average volume and frequency content between the No Effects and With Effects conditions.

2.4.4 Simulated Environment and Target Sounds

The environmental soundscape consisted of a persistent background sound with up to two overlaid target vehicle sounds. The background sound was played throughout each experimental block, and consisted of calm city ambience with a minimum of individually distinguishable vehicle sounds (Figure 7). The two vehicle sounds were seamlessly loop-able recordings of the engine/driving sounds of a motor scooter (Figure 8) and a car (Figure 9). Each sound was a

[8] http://manual.audacityteam.org/man/reverb.html
[9] https://github.com/crabl/HeadSpace

combination of a combustion engine sustaining a steady speed (there were no audible acceleration patterns or changes over time), combined with the sound of rushing wind.

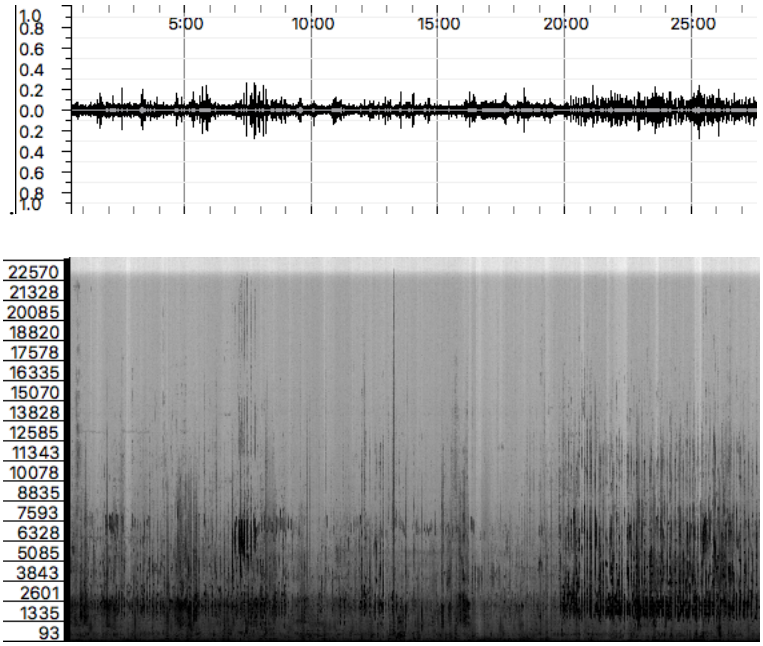

Figure 7. Waveform (above) and spectrogram (below) for ambient noise sound. The ordinate represents time and abscissa the overall intensity (above) or the intensity within each frequency band, in hz (below).

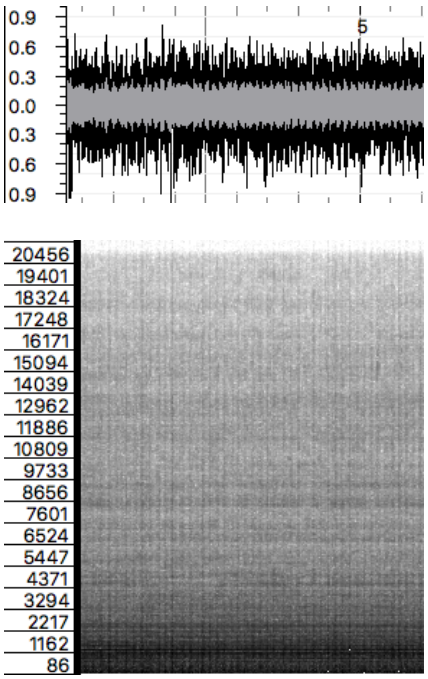

Figure 8. Waveform (above) and spectrogram (below) for "scooter" target sound. The ordinate represents time and abscissa overall intensity, or intensity within each frequency band, in hz.

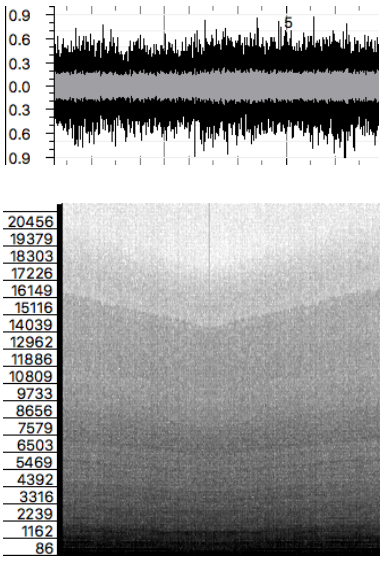

Figure 9. Waveform (above) and spectrogram (below) for "car" target sound. The ordinate represents time and abscissa the overall intensity or the intensity within each frequency band, in hz.

Each vehicle had a simulated position, speed and direction of movement in 2D space. These parameters were generated anew each time a vehicle was reset after fading into the distance. Position, speed and direction were all generated in a pseudo-random fashion (normally distributed about a mean, with bounds, see Figure 10). The position of origin, direction, and speed were generated such that that each vehicle would 'pass by' the listener, without passing directly through them or necessarily getting close, and then recede into the distance over the course of, on average, 15 seconds. For example, if a car was randomly generated such that it entered from in front of the listener and to their right, the parameter bounds were set so that it would move towards the rear and left of the listener. Within those constraints, parameters were generated in the aforementioned bounded Gaussian-random fashion.

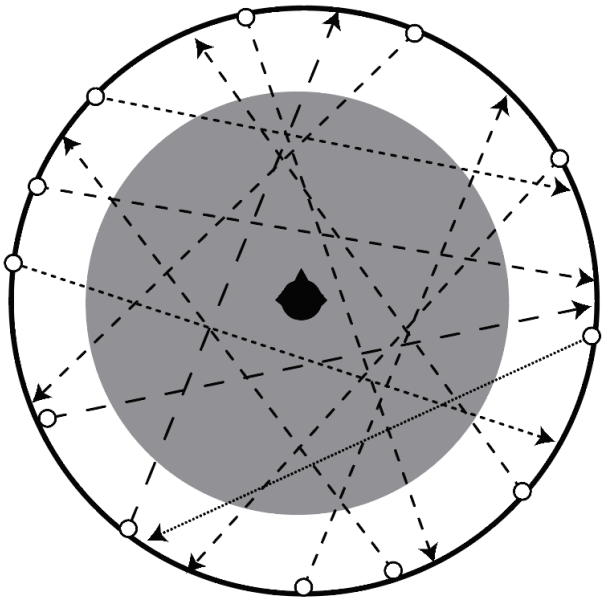

Figure 10. Example vehicle paths. The bold exterior circle indicates the maximum allowable vehicle distance. Vehicles always de-spawned when reaching this outer circle, and were re-spawned elsewhere along the circle. The grey interior circle represents the clearly audible range for vehicles. Vehicles were 'present' when they were inside of this inner circle. Dashed lines represent different vehicle speeds.

Once generated, vehicles moved in a straight line, with a constant speed and direction, until they exited the defined bounds of the simulation space (Figure 10). Speed and direction were represented as a two-item vector. These values were bounded such that each vehicle took around 10-20 seconds to pass by the listener and fade away completely. Thus, rather than moving rapidly by the listener, vehicles were relatively stable entities, and sometimes persisted across multiple SA probes. This feature of the task was tuned to (a) reflect real-world listening conditions, (b) allow the listener time to perceive vehicle locations and build up auditory streams, and (c) nonetheless require that the participant keep track of and continually refresh vehicle locations, rather than relying on memory.

The position of each vehicle was represented through smooth panning amongst the 16 speakers ('ambisonics' were not used), adjustment of sound intensity based on the distance from the participant, and a simulated Doppler effect that modified the frequency of the vehicle sound based on its current distance and velocity relative to the listener (Hiebert, 2005). Sound intensity was calibrated so that each vehicle sound reached approximately 65 dBA at its closest distance. Volume was calibrated so that: (a) distractor music did not drown out target vehicles entirely; (b) conversely, vehicle sounds did not overpower distractor music; and of course (c) target vehicles, music and ambient sounds averaged well below OSHA standards for 1-2 hours of exposure (U.S. Department of Labor, 1996). The fact that calibration was performed in this way is important to interpreting the results. Results may be generalizable to situations in which a person is listening with moderate volume relative to target sounds that have moderate difficulty overpowering ambient noise. This intentionally aligns with the task of the cyclist or pedestrian, and with what many participants said about their habits when cycling, jogging or walking (i.e., that they "keep the volume turned down"), but it should be noted that extremely loud distractor music would be

expected to have different outcomes. The purpose of this study was to investigate scenarios in which target sounds were not entirely masked, but were sometimes difficult to segregate into auditory streams or keep track of if the vehicle was close and moving quickly; simulation parameters were calibrated to create these kinds of difficulty for participants.

Once a vehicle moved far enough away (far past the point of audibility), it spent a period of time being inaudible (for most of this time, it had an actual volume of zero). During the time in which the vehicle was inaudible, if the participant was asked if that vehicle is present, the correct answer would have been 'no.' If a vehicle was currently rendered with a volume that should be clearly audible to a participant with normal or corrected to normal hearing, the correct answer would have been 'yes.' For a brief period in which vehicle volume was greater than zero but less than a volume an unimpeded listener could reasonably be expected to hear (a threshold determined by the experimenter), no SA probes were given. Thus, if a person responded with 'not present' when a vehicle was present, it was not because that vehicle had faded out of earshot, or was about to fade into earshot, but likely due to inattention, failures in stream segregation, etc.

After a Gaussian-random interval, bounded by 5 and 10 seconds, vehicle parameters were reset, and the vehicle was re-created with a new position just outside the edge of the audible simulation space with a new speed, position and direction. A few seconds after this, the vehicle would re-enter audible range (see Figure 10).

2.5 Procedure

2.5.1 Navigation Task

For the majority of the study session, participants performed a simple visual-manual task referred to as the "navigation task." This task was included to simulate the rotation that a person

would undertake in order to move through the world, and thus give the head-tracked virtual spatialization effects a chance to have an impact. Additionally, this task reflected, in a broad sense, some of the visual and cognitive actions a person might take to navigate in an urban environment: looking ahead, evaluating whether a "turn" was required, and executing a turn if appropriate.

To perform this task, participants were asked to look at one of four displays that were situated at 0, 90, 180, and 270 (-90) degrees, close to eye level. At all times, three of the displays were blank, while the fourth showed a fixation cross. Participants were instructed to look at the fixation cross. At intervals of four to eight seconds, the display with the fixation cross changed to show either a right or left arrow. Upon seeing one of these arrows, participants rotated their chair 90 degrees to the left or right, in accordance with the direction of the arrow. They then began looking at this new display, which now showed the fixation cross. Performance on this navigation task was not analyzed. However, data from the head tracker was used to confirm that participants were complying with these instructions.

2.5.2 Listening Task

Participants were also instructed to perform a listening task. They were explicitly instructed to keep track of the current position of two vehicle sounds: a car and a scooter. During the instruction period, they were given repeated exposure to these sounds, and were given the option of continued training until they verbally indicated that they were able to tell them apart and recall which was which.

2.5.3 SA Probes

Periodically, all sounds ceased and the fixation cross display changed to show an SA-related question about the vehicles. Questions were displayed visually, without any

accompanying audio. This method of blanking the environment and then asking questions about it was adapted from the Situation Awareness Global Assessment Technique (SAGAT) method, commonly used to measure internal SA (Endsley, 2000). Alkhanifer and Ludi (2015) used a similar adaptation of SAGAT to study the SA of blind users of an assistive orienting device. The three types of probes were inspired by Endsley's three levels of SA and reflect different levels of understanding of the participant's auditory environment. However, they did not directly correspond to the three SA levels, due to limitations of the simulated environment making it difficult to ask higher-level synthesis questions.

2.5.3.1 Vehicle Presence Probes

If the prompt read "Press the green triangle button if the [car/scooter] is currently present and the red circle button if the [car/scooter] is currently absent" the participant needed to respond "yes" if the given vehicle's sounds were audible just before the audio was cut, and "no" if not. Responses were later categorized as "correct" or "incorrect." This question addressed having a bare minimum knowledge of what was in the environment, similar to Endsley's characterization of Level 1 SA.

2.5.3.2 Vehicle Localization Probes

If the prompt read "Turn your head toward the current location of the [car/scooter] and press the green triangle button" they needed to turn their head (often their body as well) toward the target sound source, then press a button on the gamepad to indicate their response direction. While knowing the location of an object undoubtedly reflects a more thorough understanding of the situation compared to knowing which vehicles were present, it does not require synthesis of elements, or relating objects back to the task at hand, that would make this a Level 2 SA question. For this and the final question type, the participant's response angle was recorded, in

addition to the true angle of the vehicle at the moment just before the question was asked. Later, these values were used to compute hit rates, as well as a continuous RMSE score.

2.5.3.3 Predicted Vehicle Location Probes

Finally, if the prompt read "Turn your head to where you would expect the [car/scooter] to be in five seconds and press the green triangle button" the participant would point their head toward where they thought the target sound would be five seconds into the future and press the response button. This question was inspired by Endsley's Level 3 SA. However, while it did require predicting where an object would be in the future, it again did not require schema-matching, synthesis, or sense of what future action would need to be taken, and as such was not a Level 3 SA probe. However, once again, performance on this question did reflect a more complete understanding of SA elements compared to the other two probes.

2.5.3.4 Pattern of Probes

When a participant responded to each probe, or a 20 second cutoff was reached, the participant heard a confirmation tone, and voice instructions telling them to look back at the fixation cross display and wait for sounds to resume, which would occur after a brief interval so that they had time to get situated. This interval was 10 seconds if it was within the first five trials in a condition, because for these initial trials, a voice recording played to remind the participant to look back at the fixation cross. For the remaining trials in each condition, the duration was 5 seconds, and the voice recording did not play.

Each trial had, on average, a 12 second time between the sounds starting up again, and administration of the next trial, during which participants returned to performing the navigation and listening tasks. These wait times were randomly generated as an integer between 10 and 14

seconds. For each condition, 8 trials were administered for each SAGAT prompt type, for a total of 24 trials.

2.5.4 Self-Report Questions

After each condition, participants completed the NASA TLX (Hart & Staveland, 1988) and a set of questions regarding their perception of spatialization of the distractor sounds. Both were administered using an iPad using an online survey tool. The questions about spatialization effectiveness were Likert items on a scale of 1-6. The first two related to lateralization and localization, while the second two inquired specifically about externality (Appendix C).

2.6 Experiment Design

While performing the navigation and SA listening tasks, each participant listened to distractor music in one of four distracted conditions, and additionally in a No Distractor condition included to assess what typical performance might look like on the SA listening task. "Device type" was varied in a between-subjects fashion: some participants used AC headphones while others used BC headphones (Table 1). Within their assigned device type, each participant experienced four distractor conditions (one ear or two ears, and no virtual spatialization effects vs. virtual spatialization effects) as well as the No Distractor condition, in counterbalanced order using a Latin Square. Participants experienced 24 trials for a condition in a single sequence, before stopping to answer self-report questions and then moving on to the next condition.

Table 1.

Conditions experienced by each participant

Ears Used	No Spatialization Effects	With Spatialization Effects
No Ears	(No Distractor)	(No Distractor)
One Ear	One Ear + No Effects	Two Ears + With Effects
Two Ears	Two Ears + No Effects	Two Ears + With Effects

Note. Each participant was also assigned to either AC or BC headphone type condition.

2.6.1 Dependent Variables

2.6.1.1 Accuracy

Accuracy (hit rate) for the vehicle presence questions was operationalized as the rate of correct responses– that is, the rate at which participants correctly stated that a vehicle was present or absent in the environment. Accuracy for the localization and probes was operationalized as the rate at which participants were less than 40 degrees off from the true location of the target. This accuracy metric was designed to indicate whether the participant generally knew where the vehicle was, or not. The 40-degree threshold was created to reflect this reality. A response outside of this 80-degree arc was likely to be either a guess, a reversal, or confusion with the other vehicle. Accuracy for predicted location questions was defined as the rate at which participants were less than 40 degrees off from the true future location of the target.

2.6.1.2 Localization Error

Localization error was used as a continuous measure of error magnitude for location and predicted location questions. This was defined as the root-mean-square-error (RMSE) of the difference scores, in degrees, between a participant's response angle and the true angle of a target at the time of the probe. RMSE was used in order to provide greater weight to more egregious errors, reflecting the reality of distracted navigation tasks in which larger errors are

more likely to lead to adverse events, while small errors may be less consequential in terms of allowing informed avoidance of hazards. Reversals were not removed before calculating RMSE.

2.6.1.3 Workload.

Self-reported workload was measured via the NASA TLX survey instrument, which produced a composite score on a 0-100 scale. Specific subscales were not analyzed.

2.6.1.4 Self-Report Questions

Scores on the five self-report questions (four relating to perceptions of spatialization, and one relating to perceived audibility of target sounds) were recorded as a value from 1 to 6 indicating the participant's selected response.

2.6.2 Analyses

For each of the task performance metrics as well as the NASA TLX scores, a Hyunh-Feldt 3-way mixed within-between ANOVA was conducted, with Bonferroni corrected post-hoc t-tests to consider simple effects if interactions were found. The three independent variables were (1) device type, with levels "bone conduction (BC)" and "air conduction (AC);" (2) "ears used" with levels "one ear" and "two ears;" and (3) "spatialization effect presence" with levels "no effects and "with effects."

The self-report spatialization variables were analyzed via nonparametric methods to make select comparisons of interest– those that spoke to the subjective effectiveness of the spatialization effects. Specifically, Wilcoxon signed-rank tests were used to assess differences in self-reported spatialization variables when either one or both ears were presented to.

2.6.3 Hypotheses

It was hypothesized that there would be higher task performance and lower self-reported workload when spatialization effects were used, compared to when they were not used, across all task performance metrics.

It was hypothesized that there would be higher task performance and lower self-reported workload when one ear was used, compared to two, across all task performance metrics. It was hypothesized that there would be a performance decrease when one ear - effects was used, compared to one ear-no effects, due to the effects being confusing when only one ear was used.

It was hypothesized that there would be higher task performance and lower self-reported workload when BC headphones were used, compared to AC headphones. It was hypothesized that there would be a performance decrease when BC-effects was used, compared to BC-no effects, due to uncompensated spatialization effects being less effective over BC.

Finally, it was hypothesized that spatialization effects would lead to higher scores on self-reported directionality, localizability, and externality, but only when two ears were used.

CHAPTER 3

RESULTS

3.1 Task Performance and Workload

3.1.1 Spatialization Effect Presence

Adding spatialization effects had mixed consequences for task performance, depending on the type of probe. Participants had lower accuracy on presence probes when effects were present (see Figure 12). However, they had lower localization probe error and higher accuracy when effects were present. An explanation for this pattern is explored in the discussion.

There was a significant main effect of spatialization effect presence on vehicle presence task accuracy, $F(1,53) = 5.38$, $p = .024$, $\eta_p^2 = .09$. Figure 11 illustrates that participants had a higher presence task accuracy when effects were absent ($M = 0.67$, $SD = 0.12$) compared to when effects were present ($M = 0.66$, $SD = 0.02$).

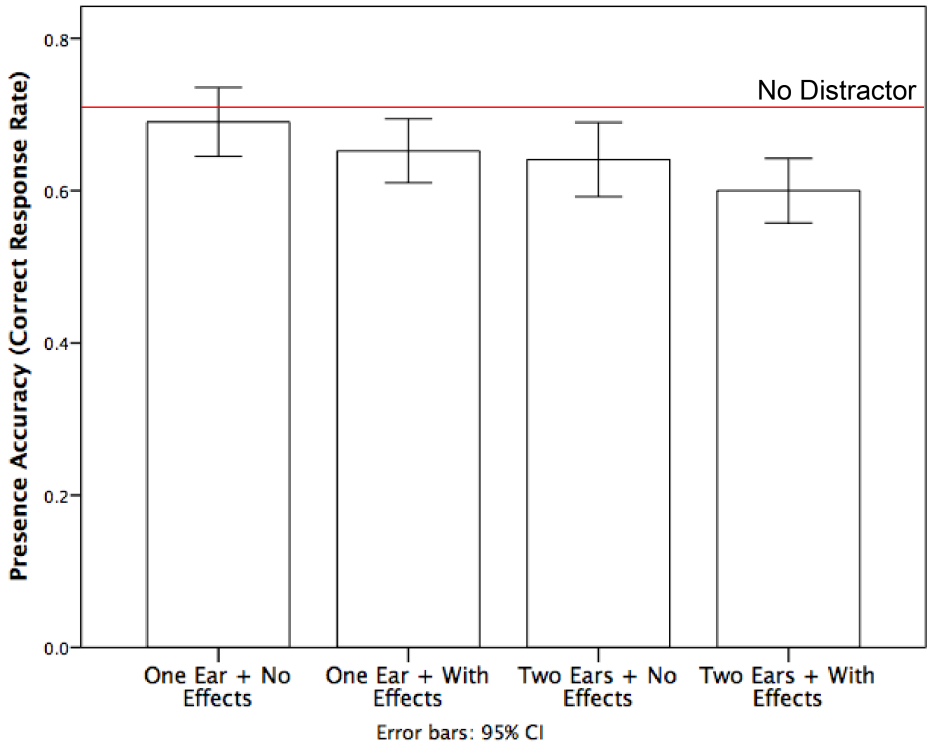

Figure 11. Vehicle presence accuracy by number of ears and effect presence (headphone type collapsed). Line reflects No Distractor performance.

While the effects led to a performance decrease for vehicle presence questions, for localization probes, the opposite was true (see Figure 12). There was a significant main effect of spatialization effect presence on localization probe error, $F(1,53) = 6.21$, $p = .016$, $\eta_p^2 = .11$. Localization probe error was smaller when effects were present, ($M = 83.71$, $SD = 17.47$), compared to when effects were absent, ($M = 89.64$, $SD = 14.80$). Relatedly, there was a significant main effect of spatialization effect presence on localization accuracy, $F(1,53) = 7.02$, $p = .011$, $\eta_p^2 = .12$. Localization accuracy was greater when effects were present, ($M = 0.41$, $SD = 0.18$), compared to when effects were absent, ($M = 0.35$, $SD = 0.05$). For all other task performance variables, the effect of spatialization effect presence was not significant.

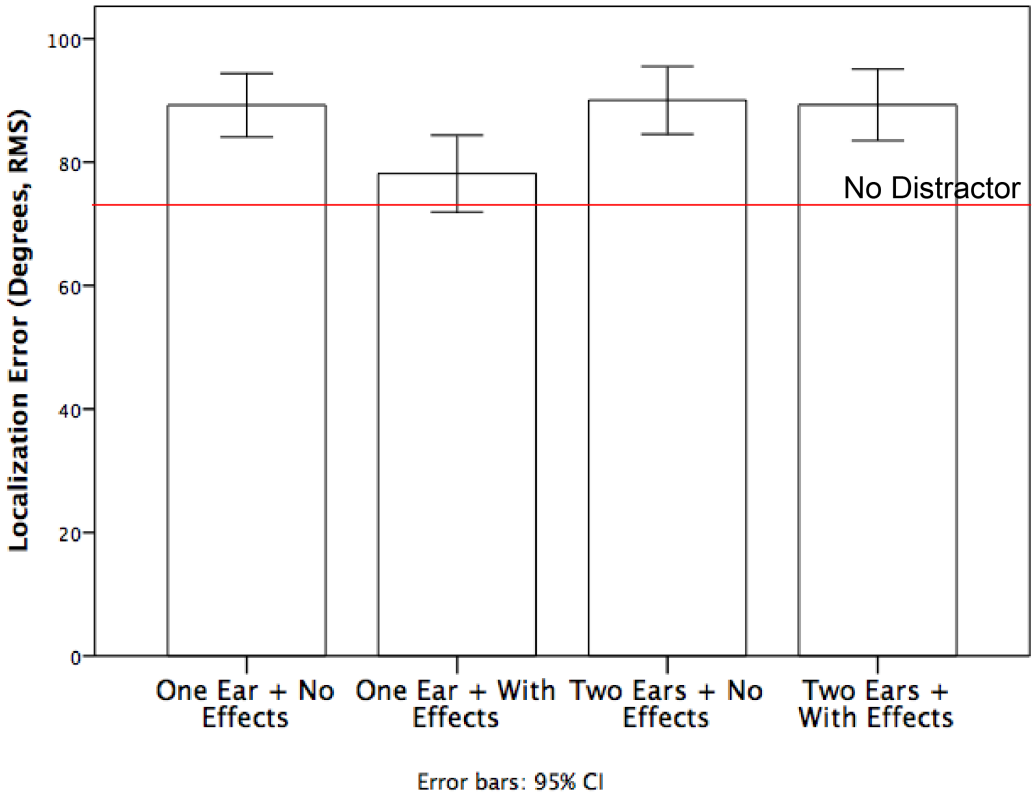

Figure 12. Localization RMSE by number of ears and effect presence (headphone type collapsed). Line reflects No Distractor performance.

3.1.2 Ears Used

Overall, participants performed better when hearing music in one ear than when hearing music in both ears. This comparison was the only one that led to significantly different NASA TLX scores, indicating that participants were consistently able to self-report this difference. This is an especially notable difference in light of the fact that all other main effects and interactions were nonsignificant for NASA TLX scores (Appendix B).

There was a significant main effect of ears used on presence task accuracy, $F(1,53) = 4.29$, $p = .043$, $\eta_p^2 = .08$. Participants had higher accuracy when music was presented in one ear ($M = 0.67$, $SD = 0.13$) compared to two ears ($M = 0.62$, $SD = .14$). In addition, there was a significant main effect of ears used on localization task error, $F(1,53) = 4.56$, $p = .037$, $\eta_p^2 = .08$. Participants had lower mean localization task error when music was presented in one ear ($M = 83.73$, $SD = 17.04$) compared to two ears ($M = 89.63$, $SD = 16.96$). For all other task performance variables, the main effect of ears used was nonsignificant (see Appendix A). Finally, there was a significant main effect of ears used on NASA TLX scores, $F(1,53) = 12.05$, $p = .001$, $\eta_p^2 = .19$. Participants had lower self-reported workload when music was presented in one ear ($M = 40.15$, $SD = 12.32$) compared to two ears ($M = 45.56$, $SD = 13.42$).

3.1.3 Headphone Type

There were no differences between AC and BC for vehicle presence or localization questions (keeping in mind that the volume was kept moderate and subjectively matched between device types, and the AC earbuds used were not heavily obstructing), but for the

predicted location questions, the minor advantages inherent to the BC device had a significant impact.

There was a significant main effect of headphone type on predicted location task error, $F(1,53) = 7.80$, $p = .007$, $\eta_p^2 = .04$. Figure 13 illustrates that participants who heard music through BC headphones ($M = 87.69$, $SD = 11.11$) had lower mean error compared to participants who heard music through AC headphones ($M = 96.06$, $SD = 11.11$). For all other task performance variables, the main effects of headphone type were nonsignificant (see Appendix A).

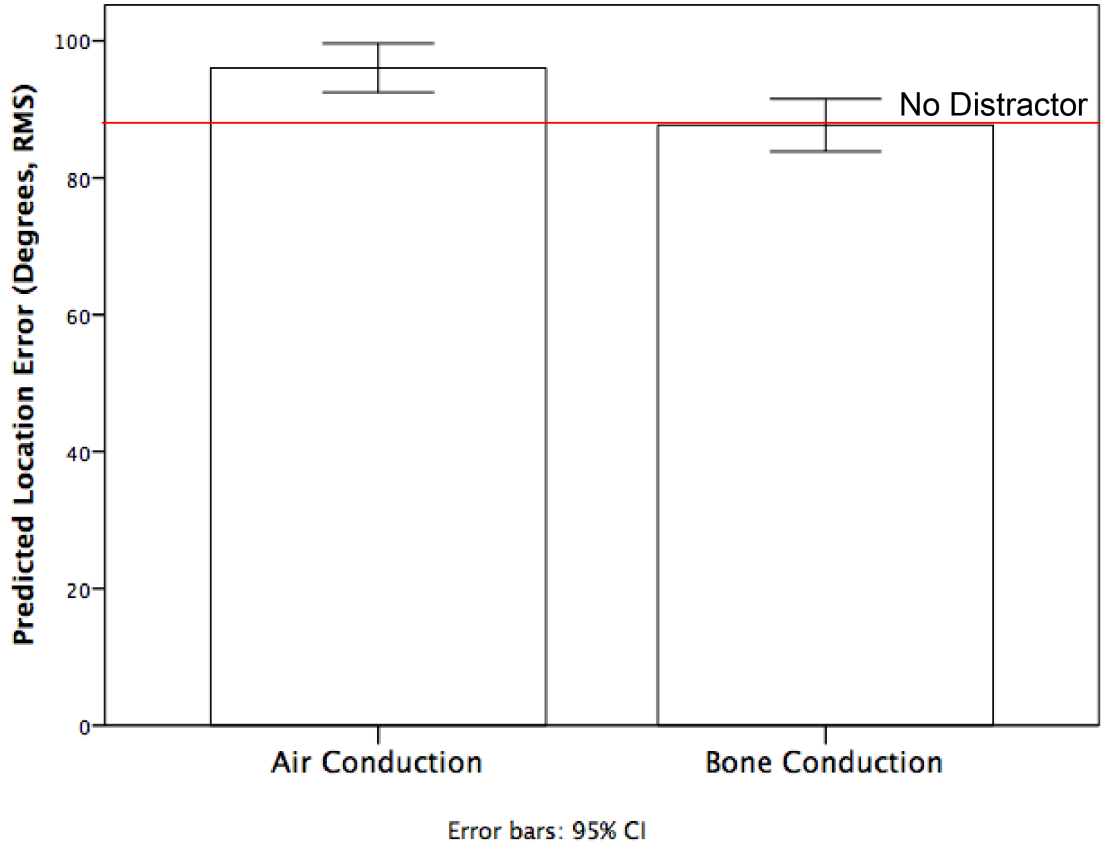

Figure 13. Predicted location RMSE by device type. Line reflects No Distractor performance.

3.1.4 Ears Used by Headphone Type

For all other task performance variables, the interaction between ears used and headphone type was nonsignificant (Appendix A).

3.1.5 Spatialization Effect Presence by Headphone Type

The interaction between spatialization effect presence and headphone type was non-significant across all task performance variables (see Appendix A).

3.1.6 Spatialization Effect Presence by Ears Used

The interaction between spatialization effect presence and ears used on localization task error was significant, $F(1,53) = 4.70$, $p = .035$, $\eta_p^2 = .08$. To analyze this interaction effect, scores were first collapsed across headphone type.

When one ear was used, the effect of spatialization effects on vehicle localization error was significant, $t(54) = 3.30$, $p = .016$. When participants heard music in one ear, they had a lower localization error when they heard music with spatialization effects ($M = 78.13$, $SD = 23.03$) compared to when they heard music without spatialization effects ($M = 89.24$, $SD = 78.13$). However, when two ears were used, the effect of spatialization effects on localization probe error was not significant (see Figure 14). This pattern was initially surprising, but a likely explanation was forthcoming (see discussion section).

When spatialization effects were absent, the effect of ears used on localization error was not significant. When spatialization effects were present, the effect of ears used on localization task error was significant $t(54) = -2.96$, $p = .040$. Overall, when participants heard music with spatialization effects, they had lower localization error when hearing music in one ear ($M = 78.13$, $SD = 23.03$) than when they heard it in two ears ($M = 89.28$, $SD = 21.42$). For all other

task performance variables, (including localization accuracy, see Figure 15) the interaction between spatialization effect presence and ears used was not significant.

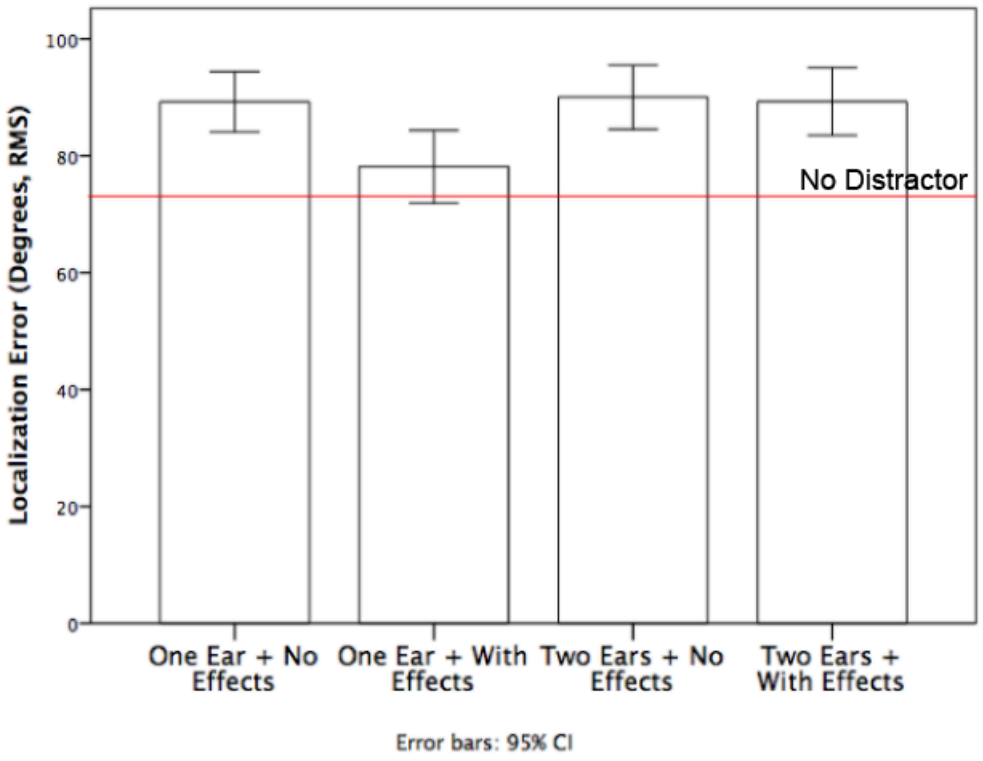

Figure 14. Localization RMSE by number of ears and effect presence (headphone type collapsed). Line reflects No Distractor performance.

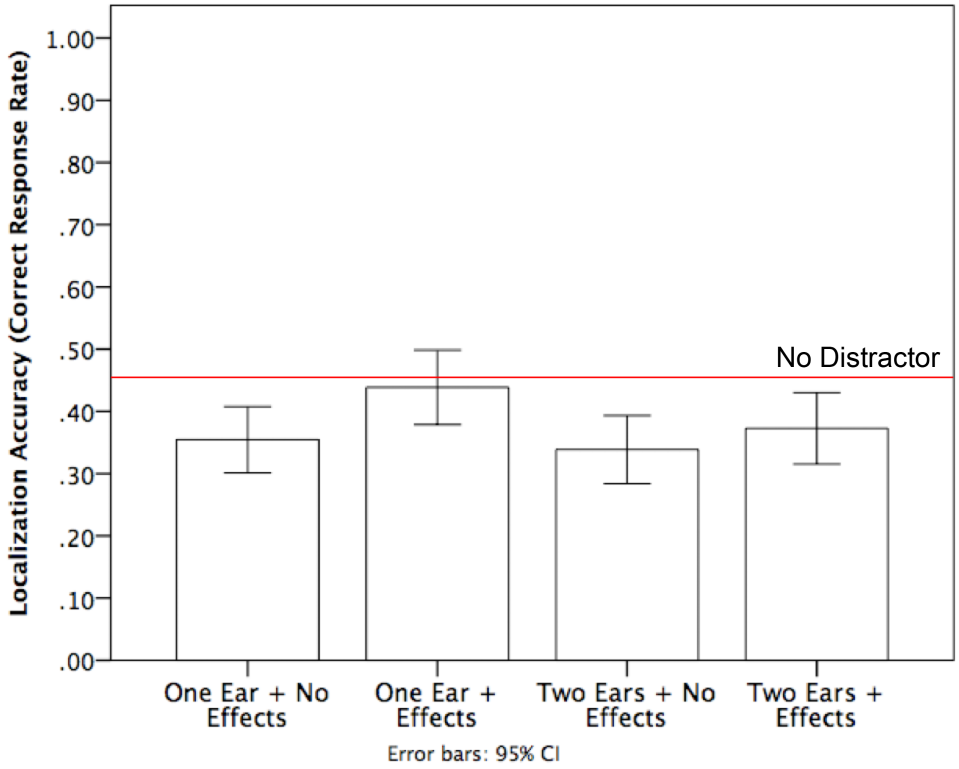

Figure 15. Localization accuracy by number of ears and effect presence (headphone type collapsed). Line reflects No Distractor performance.

3.1. Three-way Interactions

The three-way interaction between spatialization effect presence, ears used, and headphone type on presence probe response time was nonsignificant for all task performance variables.

3.2 Self-Report Spatialization Questions

To the extent that participants were able to self-report, the spatialization effects were only moderately effective, and may have induced a sense of lateralization, rather than externalization.

When one ear was used, the effect of spatialization effect presence on self-reported "directionality" was nonsignificant, $Z = -0.56$, $p = .573$ (Appendix B). However, when two ears were used, the effect of spatialization effect presence on self-reported "directionality" was

significant, Z = -2.10, p = .036. As shown in Figure 16, when spatialization effects were absent (M = 2.07, SD = 1.00), participants gave lower self-reported music "directionality" ratings than when effects were present (M = 2.63, SD = 1.37).

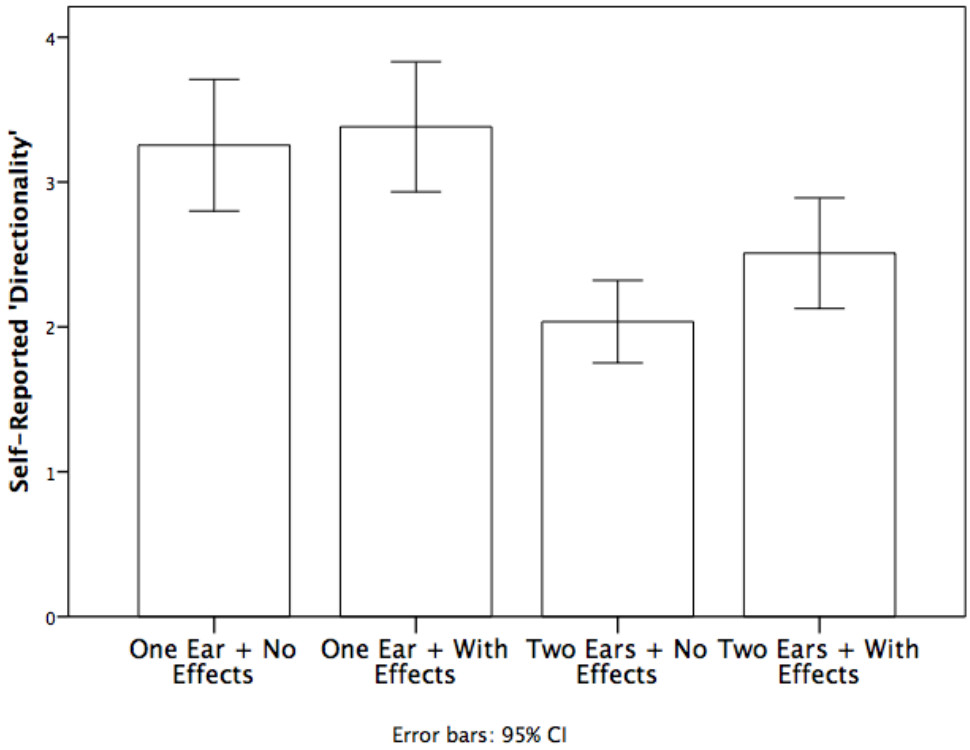

Figure 16. Self-reported "directionality" (collapsed across headphone type).

The effect of spatialization effect presence was nonsignificant for self-reported "localizability," whether one ear, Z = -0.80, p = .422, or two ears were used, Z = -1.53, p = .124. There were no significant differences in "externality" for either one ear, Z = -0.33, p = .745, or two ears, Z = -0.52, p = .602. Similarly, the effect of spatialization effect presence on "all-around-ness" was nonsignificant for one ear, Z = -0.68, p = .494, as well as two, Z = -0.98, p = .327.

Although differences were not significant, observed "externality" and "localizability" means were greater in the "effects" conditions, and observed "all-around-ness" means were smaller. It is also worth noting that the scores for "directionality" and "localizability" were less

than half of the maximum possible score when two ears and effects were used; regardless of differences between “effects” and “no effects”, ratings were relatively low throughout for these measures. Relatedly, responses for “externality” were close to or below the midpoint of the scale when two ears + effects were used.

CHAPTER 4

DISCUSSION

Whether using a BC or AC device, listening to music in one ear led to performance increases at multiple levels of probe difficulty. As such, the common practice (according to participants) of leaving one earbud "out" can be recommended. This effect was present for BC as well, indicating that the effect is not solely a matter of leaving one ear physically unobstructed. Unsurprisingly, masking is more impactful– a finding in line with May et al. (2017). The performance increase from listening in only one ear was present for both presence and localization questions, and participants reported lower workload when using one ear. This last finding is especially notable in light of the fact that no other significant effects were found on NASA TLX scores.

Performance differences between BC and AC headphones, subjectively loudness-matched as they were in this study, and with moderately sound-blocking plastic earbuds, were only detected for the predicted location questions. In the real world, being able to predict the path nearby vehicles will take is important to being able to respond appropriately to possible dangers- the other two question types are necessary but not sufficient for informed action. The presence of this difference was most likely attributable to the different audibility thresholds of BC hearing/ limited device transduction capabilities, leading to decreased masking of low and high frequencies, and additionally to the fact that target sounds were not subjected to any frequency-selective "muffling." The fact that no advantages were observed for the lower-level SA probe types indicates that the slight perceptual degradation introduced by ear-obstructing earbuds may only be detrimental for more complex environmental judgments (motion perception, in this

case). It may be that slight degradations to early auditory object perception "snowballs" and can lead to significant impairments when such information is aggregated to build higher-level SA.

From a practical standpoint, this is an important difference. These results provide evidence that BC headphones can be recommended for pedestrian and cyclist safety, with the important caveat that choosing a low volume and using only a single ear (if possible) both can be expected to remain more impactful than the choice of headphone type.

Spatialization effects had mixed effects on participant SA. They led to decreased performance on presence probes, but led to increased performance on localization probes. Unexpectedly, the increased localization probe performance was found exclusively when one ear was used, which removed the ability of the listener to perceive the ITD and ILD components of the spatialization effects, but preserved a noticeable level change with head movement (a "pseudo ILD") as well as the generalized HRTF spectral manipulations.

In seeking an explanation for this pattern, it was observed that when the spatialization effects were applied, music was rendered with an 18.75% average increase in sound intensity. This intensity difference was due to an error in the way the ILD effect was implemented. The effect boosted intensity in the proximal ear but never dropped intensity in the distal ear below its baseline volume, which was what both ears were set to in the "no effects" conditions. This likely explains why performance was lower for the "presence" probes: it was harder for participants to detect faint vehicle sounds, due to increased simultaneous-masking caused by the increase in mean sound intensity. However, the presence of this intensity differences makes the positive impact of effects-present playback on localization probe performance (when one ear was used) more notable. In this scenario, participants performed better on SA tasks when a *louder* distractor was played to them. This loudness difference likely had a slight detrimental effect on localization

probe performance, which was evidently offset by advantageous properties of the spatialization effects. As such, one might expect an even larger performance increase to localization performance if sound intensity were properly matched between effects and no-effects conditions.

When two ears were presented to, the ILD implementation led to an intensity increase in both ears, which would have increased masking of vehicle sounds and decreased task performance. However, when both ears were used, this intensity increase may have been more detrimental to performance because the average intensity increase was, in effect, twice as large. Additionally, there were issues with the HRTF implementation that introduced problems exclusively in the two ears + effects conditions. Instead of continuously updating, HRTFs updated at the boundaries of ~20 degree arcs. Within these arcs, there was no interpolation. As such, the majority of the time, the HRTF effect did not precisely agree with the ILD and ITD effects. Since these latter effects were nonfunctional in the one-ear conditions, any confusion resulting from this disagreement would not have been present in those conditions.

Research presently underway will correct the issues with the "with effects" conditions by correcting the volume imbalance and adding interpolation to the HRTF effect, and additionally investigate the addition of an adaptive component to the spatialization. It will evaluate (a) whether the disadvantages to vehicle presence performance go away with these corrections and (b) whether vehicle localization or predicted location advantages appear for two-eared presentation.

There are still conclusions to be drawn from the present data, even in light of this methodological issue. From a practical standpoint, despite the fact that task difficulty is slightly increased from a spatial auditory processing standpoint, listeners may benefit from virtual distractor sounds being spatialized. In the present study, these benefits were seen across

headphone types. While BC devices introduce difficulties to achieving convincing spatial audio, in this case, spatialization effects were effective in increasing localization probe performance with each headphone type, and even in the one ear-BC condition.

The presence of this effect implies that, for the task of tracking several vehicles while listening to music, the ventral task of keeping distractors segregated from vehicle sounds may be a greater source of difficulty/ bottlenecking compared to the dorsal task of spatially modeling 2-3 auditory objects. This broader implication is to some extent specific to scenarios with a difficulty profile similar to the one used for this study. The vehicles were (1) two in number; and (2) similar sounding, but distinguishable under ideal listening conditions. This reflects a common scenario for cyclists, in which they need to remain aware of several nearby cars that might veer into their lane, turn across them, or stop suddenly. Pedestrians at street crossings face a similar issue. A person crossing the street essentially needs to be able to detect and localize the nearest car coming from the left, and the nearest car coming from the right. These results indicate that, in this common type of situation, stream segregation is the more difficult cognitive-perceptual problem compared to spatial processing and tracking, and as such, stream segregation should be targeted by new technologies that will endeavor to keep us aware of the world around us as we receive information and entertainment via headphones.

From the perspective of dual processing stream theory, it is worth noting that this study was a case in which an intervention that was designed to decrease primarily-ventral task difficulty while moderately increasing primarily-dorsal task difficulty, led to an advantage in vehicle localization, which is a primarily-dorsal stream task. As such, these results provide some evidence for either feedback or continuous cross-talk models. Facilitating the formation of

clearer auditory streams seems to have influenced proficiency in the dorsal task of spatially mapping the soundscape.

Finally, in understanding the meaning of these results, it is important to make the distinction between what was and was not achieved with the spatial audio effects used in this study. While the effects used in this study did lead to significantly increased self-reported "directionality," it is apparent that they were not successful at creating the perception of externality. As such, results may reflect "lateralization plus" rather than full spatialization. However, while this distinction is not ordinarily drawn, the majority of research on stream segregation has, in fact, essentially tested the efficacy of "direction of origin" as a stream segregation cue, rather than "apparent point of origin in 3D space." Essentially, lateralization, in this case also able to be characterized as "directionality," is the stream segregation cue that was utilized in this research, which reflects the manner in which spatial cues have been evaluated in the past.

More successful spatialization effects could contain additional channels of information that could be used as segregation cues, but these could also add difficulty to the multi-tasking listener's task, and may in fact not be necessary. First, it is possible that externality itself provides a separate segregation cue, and that keeping computing sounds lateralized rather than externalized is actually preferable. This cue may or may not be distinct from the cue of apparent distance, which is itself confounded with sound intensity, but nonetheless could be delineated as a separate cue from a practical/technology design standpoint.

Second, the manner in which a sound reverberates, including the timing and angle of early and late reflections, has been suggested by Cusack and Carlyon (2004) as yet another stream segregation cue that would be present with high-fidelity, externalization-inducing spatial

audio effects. However, this facet of externalized audio could be a source of great difficulty to the listener, if the simulated acoustic space was complex and highly echoic; highly echoic spaces have been shown to double localization error versus anechoic spaces (Hafter, Saberi, Jensen & Briolle, 1992). In summation, there may be a balance to be struck between realistic acoustic simulation needed to induce the perceptual illusion of externalization and the need to create a simple, clear virtual auditory environment to minimize auditory processing demands and facilitate SA.

Thus, while externalized spatial audio has clear uses in augmented reality, in which sounds ought to correspond with real or virtual objects perceived in other sensory modalities, for environment-agnostic computing audio, externalized spatial audio may not be necessary to facilitate stream segregation and increase SA. The results of this study suggest that relatively crude, lateralized audio, even presented in one ear and/or through an uncompensated BC device, is sufficient for this purpose. It may even be true that lateralized audio is *preferable*, due to (a) the potential for externality itself to be a segregation cue and (b) the processing difficulty that would be introduced by accurately simulating room echoes and other complex acoustic elements. Future work should compare these alternatives, as well as evaluate whether improvements to spatial audio effects could mitigate the adverse effects on object presence detection that were observed in this study and/or offer advantages to higher-level SA.

Appendix A. Task Performance Results

Table A1.

Descriptive statistics (*M, SD)*- Vehicle Presence Accuracy.

	Condition Means				
	AC		**BC**		**No Distractor**
	No Effects	**Effects**	**No Effects**	**Effects**	.71, .19
One Ear	.67, .18	.63, .17	.71, .16	.67, .14	
Two Ear	.64, .22	.59, .15	.64, .14	.61, .17	
	Collapsed				
One Ear	**Two Ears**	**No Effects**	**Effects**	**AC**	**BC**
0.67, 0.13	.62, .14	.67, .12	.66, .02	.63, .09	.66, .10
	Collapsed across Ears		**Collapsed across Effect Presence**		
	No Effects	**Effects**	**One Ear**	**Two Ears**	
AC	.66, .11	.61, .11	.69, .12	.65, .11	
BC	.67, .12	.64, .11	.64, .13	.60, .11	
	Collapsed across Headphone Type				
	No Effects	Effects			
One Ear	.69, .17	.65, .16			
Two Ears	.64, .18	.60, .16			

Table A2.

Descriptive statistics (*M, SD)*- Vehicle Localization RMSE.

	Condition Means				
	AC		**BC**		**No Distractor**
	No Effects	**Effects**	**No Effects**	**Effects**	74.87, 22.03
One Ear	90.64, 19.12	81.59, 23.13	87.89, 12.20	74.81, 22.88	
Two Ears	89.59, 22.08	86.34, 21.57	90.47, 18.90	92.12, 21.28	
	Collapsed				
One Ear	**Two Ears**	**No Effects**	**Effects**	**AC**	**BC**
83.73, 17.04	89.63, 16.96	89.65, 14.80	83.71, 17.47	87.04, 13.56	86.32, 13.56
	Collapsed across Ears		**Collapsed across Effect Presence**		
	No Effects	**Effects**	**One Ear**	**Two Ears**	
AC	90.12, 14.80	83.96, 17.46	89.27, 13.42	78.195, 16.11	
BC	89.18, 14.80	83.46, 17.46	90.03, 14.64	89.23, 15.29	
	Collapsed across Headphone Type				
	No Effects	Effects			
One Ear	89.24, 19.03	78.13, 23.03			
Two Ears	90.04, 20.33	89.28, 21.42			

Table A3.

Descriptive statistics (*M, SD)*- Predicted Vehicle Location RMSE.

		Condition Means			
	AC		**BC**		**No Distractor**
	No Effects	**Effects**	**No Effects**	**Effects**	87.70, 19.48
One Ear	91.35,17.38	98.76, 16.69	90.58, 18.46	85.73, 23.43	
Two Ears	99.16, 18.96	94.97, 21.46	89.02, 24.69	85.33, 14.54	
		Collapsed			
One Ear	**Two Ears**	**No Effects**	**Effects**	**AC**	**BC**
91.63, 12.29	92.12, 15.04	92.55, 16.22	91.20, 13.85	96.06, 11.11	87.69, 11.11
	Collapsed across Ears		**Collapsed across Effect Presence**		
	No Effects	**Effects**	**One Ear**	**Two Ears**	
AC	95.25, 16.22	96.87, 13.85	91.01, 12.57	92.25, 14.30	
BC	89.85, 16.22	85.53, 13.85	94.09, 15.75	90.15, 13.03	
	Collapsed across Headphone Type				
	No Effects	**Effects**			
One Ear	91.01, 17.77	92.13, 21.26			
Two Ears	94.00, 22.45	90.06, 18.74			

Table A4.

Descriptive statistics (*M, SD)*- Vehicle Localization Accuracy (Hit Rate).

		Condition Means			
	AC		**BC**		**No Distractor**
	No Effects	**Effects**	**No Effects**	**Effects**	.45, .19
One Ear	.32, .18	.44, .22	.38, .21	.44, .23	
Two Ear	.31, .21	.38, .23	.37, .19	.37, .20	
		Collapsed			
One Ear	**Two Ears**	**No Effects**	**Effects**	**AC**	**BC**
.40, .16	.36, .18	.35, .15	.41, .18	.36, .14	.39, .14
	Collapsed across Ears		**Collapsed across Effect Presence**		
	No Effects	**Effects**	**One Ear**	**Two Ears**	
AC	.32, .15	.41, .18	.38, .17	.34, .18	
BC	.38, .15	.40, .18	.41, .16	.37, .18	
	Collapsed across Headphone Type				
	No Effects	**Effects**			
One Ear	.36, 0.20	.44, 0.22			
Two Ears	.34, .20	.37, .21			

Table A5.

Descriptive statistics (*M, SD)*- Predicted Vehicle Location Accuracy (Hit Rate).

		Condition Means			
	AC		**BC**		**No Distractor**
	No Effects	**Effects**	**No Effects**	**Effects**	.29, .17
One Ear	.28, .13	.26, .16	.28, .19	.33, .23	
Two Ears	.24, .21	.30, .13	.34, .19	.29, .15	
		Collapsed			
One Ear	**Two Ears**	**No Effects**	**Effects**	**AC**	**BC**
.29, .11	.29, .13	.28, .15	.29, .13	.27, .10	.31, .10
	Collapsed across Ears		**Collapsed Across Effect Presence**		
	No Effects	**Effects**	**One Ear**	**Two Ears**	
AC	.26, .15	.28, 0.13	.27, .11	.27, .13	
BC	.31, .15	.31, 0.13	.31, .12	.31, .13	
	Collapsed across Headphone Type				
	No Effects	**Effects**			
One Ear	.28, .162	.29, .20			
Two Ears	.29, .21	.29, .14			

Appendix B. Self-Report Responses

Table B1.

Descriptive statistics (*M, SD)*- Self-Reported Directionality.

		Condition Means			
	AC		**BC**		
	No Effects	**Effects**	**No Effects**	**Effects**	
One Ear	2.95, 1.46	3.09, 1.44	3.87, 1.60	3.83, 1.61	
Two Ear	2.41, 0.96	2.82, 1.18	1.75, .944	2.46, 1.53	
		Collapsed			
One Ear	**Two Ears**	**No Effects**	**Effects**	**AC**	**BC**
3.44, 1.37	2.36, 1.08	2.75, 1.04	3.05, 1.11	2.69, 1.36	2.92, 1.72
	Collapsed across Ears		**Collapsed across Effect Presence**		
	No Effects	**Effects**	**One Ear**	**Two Ears**	
AC	2.68, 1.05	2.96, 1.12	3.02, 1.38	2.61, 1.10	
BC	2.81, 1.03	3.15, 1.09	3.85, 1.35	2.10, 1.07	
	Collapsed across Headphone Type				
	No Effects	**Effects**			
One Ear	3.39, 1.15	3.48, 1.56			
Two Ears	2.07, 1.00	2.63, 1.37			

Table B2.

Descriptive statistics (*M, SD)*- Self-Reported Externality.

		Condition Means			
	AC		**BC**		
	No Effects	**Effects**	**No Effects**	**Effects**	
One Ear	3.55, 1.06	3.32, 1.29	3.25, 1.42	3.25, 1.33	
Two Ear	3.00, 1.35	3.05, 1.25	2.88, 1.23	2.96, 1.30	
		Collapsed			
One Ear	**Two Ears**	**No Effects**	**Effects**	**AC**	**BC**
3.34, 1.22	2.97, 1.24	3.17, 1.04	3.14, 1.25	3.168, 1.04	3.14, 1.25
	Collapsed across Ears		**Collapsed across Effect Presence**		
	No Effects	**Effects**	**One Ear**	**Two Ears**	
AC	3.27, 1.05	3.18, 1.27	3.43, 1.24	3.023, 1.26	
BC	3.06, 1.02	3.10, 1.23	3.25, 1.21	3.023, 1.26	
	Collapsed across Headphone Type				
	No Effects	**Effects**			
One Ear	3.39, 1.26	3.28, 1.29			
Two Ears	2.93, 1.27	3.00, 1.27			

Table B2.

Descriptive statistics (*M, SD)*- Self-Reported Localizability.

	Condition Means				
	AC		**BC**		
	No Effects	**Effects**	**No Effects**	**Effects**	
One Ear	3.45, 1.10	3.23, 1.31	3.46, 1.74	3.04, 1.68	
Two Ear	2.32, 1.13	2.73, 1.24	1.96, 1.33	2.46, 1.53	
	Collapsed				
One Ear	**Two Ears**	**No Effects**	**Effects**	**AC**	**BC**
3.30, 1.38	2.37, 1.19	2.80, 1.16	2.86, 1.21	2.80, 1.16	2.86, 1.21
	Collapsed across Ears		**Collapsed across Effect Presence**		
	No Effects	**Effects**	**One Ear**	**Two Ears**	
AC	2.89, 1.18	2.98, 1.23	3.34, 1.40	2.30, 1.91	
BC	2.71, 1.15	2.75, 1.20	3.25, 1.37	2.52, 1.21	
	Collapsed across Headphone Type				
	No Effects	**Effects**			
One Ear	3.46, 1.46	3.13, 1.50			
Two Ears	2.13, 1.24	2.59, 1.39			

Table B3.

Descriptive statistics (*M, SD)*- NASA TLX Composite Subjective Workload.

	Condition Means				
	AC		**BC**		**No Distractor**
	No Effects	**Effects**	**No Effects**	**Effects**	34.51, 13.44
One Ear	41.57, 12.93	43.26, 12.78	37.72, 10.50	38.03, 12.49	
Two Ear	46.77, 14.55	49.44, 14.44	44.06, 11.89	41.97, 10.20	
	Collapsed				
One Ear	**Two Ears**	**No Effects**	**Effects**	**AC**	**BC**
40.15, 12.32	45.56, 13.42	42.53, 11.99	43.17, 12.64	45.26, 12.13	40.45, 12.31
	Collapsed across Ears		**Collapsed Across Effect Presence**		
	No Effects	**Effects**	**One Ear**	**Two Ears**	
AC	44.17, 12.14	46.35, 12.79	42.42, 12.47	48.10, 13.59	
BC	40.89, 11.84	40.00, 12.47	37.88, 12.16	43.01, 13.25	
	Collapsed across Headphone Type				
	No Effects	**Effects**			
One Ear	39.56, 11.75	40.53, 12.76			
Two Ears	45.35, 13.15	45.54, 12.83			

Appendix C. Self-Report Spatialization Questions

1. To what extent did you perceive the music to be coming from a **specific direction**?

(1: clearly did not come from a specific direction)- (6: clearly did come from a specific direction)

2. To what extent did you perceive the music as coming from a **specific location in space**?

(1: clearly did not come from a point in space) --- (6: clearly did come from a point in space)

3. To what extent did you perceive the music as coming from **inside or outside of your head**?

(1: clearly inside) ---------- (6: clearly outside)

4. To what extent did you perceive the music as coming from **all around you**?

(1: Clearly did not come from all around) ---------- (6: Clearly did come from all around)

REFERENCES

Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., & Witzel, T. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proceedings of the National Academy of Sciences*, *103*(39), 14608-14613.

Adriani, M., Maeder, P., Meuli, R., Thiran, A. B., Frischknecht, R., Villemure, J. G., ... & Thiran, J. P. (2003). Sound recognition and localization in man: specialized cortical networks and effects of acute circumscribed lesions. *Experimental brain research*, *153*(4), 591-604.

Alkhanifer, A. A., & Ludi, S. (2015, October). Developing SAGAT Probes to Evaluate Blind Individuals' Situation Awareness when Traveling Indoor Environments. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (pp. 305-306). ACM.

Begault, D. R., & Wenzel, E. M. (1991) Headphone Localization of Speech Stimuli, *In Proceedings of the Human Factors Society 35th Convention, pp. 82-86*

Begault, D. R., Wenzel, E. M., & Anderson, M. R. (2001). Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Audio Engineering Society*, *49*(10), 904-916.

Bernstein, L. (1997). *Detection and discrimination of interaural disparities: Modern earphone-based studies* (pp. 117-138). Lawrence Erlbaum.

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience*, *14*(10), 693-707.

Blauert, J. (1999). Spatial Hearing: The psychophysics of human sound localization. Cambridge, MA, USA: The MIT Press, 1999.

Bőhm, T. M., Shestopalova, L., Bendixen, A., Andreou, A. G., Georgiou, J., Garreau, G., & Winkler, I. (2013). The role of perceived source location in auditory stream segregation: separation affects sound organization, common fate does not. *Learning & Perception*, *5*(Supplement 2), 55-72.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Broke for Free. (2012). *As Colorful as Ever* [MP3]. Tom Cascino.

Chang-Geun, O., Lee, K. and Spencer, P. (2011). Effectiveness of Advanced BC Earphones for People Who Enjoy Outdoor Activities. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* September 2011 55: 1788-1792

Chen, C.-M., Lakatos, P., Shah, A. S., Mehta, A. D., Givre, S. J., Javitt, D. C., et al. (2007).

Cloutman, L. L. (2013). Interaction between dorsal and ventral processing streams: where, when and how? *Brain and language*, *127*(2), 251-263.

Cusack R, Deeks J, Aikman G, Carlyon, R.P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. J Exp Psychol Hum Percept Perform. 2004;30(4):643-56.

Cusack, R., & Carlyon, R. P. (2004). Auditory perceptual organization inside and outside the laboratory. In J.G Neuhoff (Ed.), *Ecological psychoacoustics* (pp. 15-48). San Diego, California: Elsevier Academic Press.

Daniel, V., 2003. Explaining differences in bicycle use among Dutch municipalities. Vrije Universiteit, Amsterdam.

De Waard, D., Edlinger, K. M. & Brookhuis, K. A. (2011). Effects of listening to music, and of using a handheld and handsfree telephone on cycling behavior. *Transportation research part four- Traffic psychology and behaviour*. 14, 6, 626-637.

Driver, J., & Spence, C. (1998). Cross–modal links in spatial attention. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *353*(1373), 1319-1331.

Durso, F. T., Truitt, T. R., Hackworth, C. A., Crutchfield, J. M., Nikolic, D., Moertl, P. M., & Manning, C. A. (1995). Expertise and chess: A pilot study comparing situation awareness methodologies. *Experimental analysis and measurement of situation awareness*, 295-303.

Endsley, M. R. (1987). SAGAT: A methodology for the measurement of situation awareness (NOR DOC 87-83). *Hawthorne, CA: Northrop Corporation*.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32-64.

Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. *Situation awareness analysis and measurement*, *10*.

Fracker, M. L. (1988, October). A theory of situation assessment: Implications for measuring situation awareness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 32, No. 2, pp. 102-106). SAGE Publications.

Functional anatomy and interaction of fast and slow visual pathways in macaque monkeys. Cerebral Cortex, 17, 1561–1569.

Gardner, B., & Martin, K. (1994). HRTF Measurements of a KEMAR Dummy-Head Microphone. Retrieved from http://sound.media.mit.edu/resources/KEMAR.html

Goldenbeld, C., Houtenbos, M., Ehlers, E., & De Waard, D. (2012). The use and risk of portable electronic devices while cycling among different age groups. *Journal of safety research*, *43*(1), 1-8.

Goodale, M. A., & Murison, R. C. (1975). The effects of lesions of the superior colliculus on locomotor orientation and the orienting reflex in the rat. *Brain Research*, *88*(2), 243-261.

Gripper, M., McBride, M., Osafo-Yeboah, B., & Jiang, X. (2007). Using the Callsign Acquisition Test (CAT) to compare the speech intelligibility of air versus bone conduction. *International journal of industrial ergonomics*, *37*(7), 631-641.

Gutzwiller, R. S., & Clegg, B. A. (2013). The role of working memory in levels of situation awareness. *Journal of Cognitive Engineering and Decision Making*, *7*(2), 141-154.

Hafter, E. R., Saberi, K., Jensen, E. R., & Briolle, F. (1992). Localization in an echoic environment. *Adv Biosci*, *83*, 555-561.

Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., & Nironen, H. (2003, March). Techniques and applications of wearable augmented reality audio. In *Audio Engineering Society Convention 114*. Audio Engineering Society.

Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Hiipakka, J., & Lorho, G. (2004). Augmented reality audio for mobile and wearable appliances. *Journal of the Audio Engineering Society*, *52*(6), 618-639.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology, 52, 139-183.

Hiebert, G. (2005). Openal 1.1 specification and reference.

Iverson, P. (1995). Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(4), 751.

Iwaki, M., & Chigira, Y. (2016, October). Compensation of sound source direction perceived through consumer-grade bone-conduction headphones by modifying ILD and ITD. In *Consumer Electronics, 2016 IEEE 5th Global Conference on* (pp. 1-4). IEEE.

Jaaskelainen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levanen, S., et al. (2004). Proceedings of the National Academy of Sciences USA, 101, 6809–6814.

Killion, M. C., Monroe, T., & Drambarean, V. (2011). Better protection from blasts without sacrificing situational awareness. *International journal of audiology*, *50*(sup1), S38-S45.

Kuzel, M. J., Heller, M. F., Sala, J. B., Ciccarelli, L., & Gray, R. (2008). A review of real-world collision involving distracted pedestrians. In *Proceedings of the Annual International Occupational Ergonomics and Safety Conference*.

Langendijk, E. H., Kistler, D. J., & Wightman, F. L. (2001). Sound localization in the presence of one or two distracters. *The Journal of the Acoustical Society of America*, *109*(5), 2123-2134.

Leavitt, V. M., Molholm, S., Gomez-Ramirez, M., & Foxe, J. J. (2011). "What" and "Where" in auditory sensory processing: a high-density electrical mapping study of distinct neural processes underlying sound object recognition and sound localization. *Frontiers in integrative neuroscience*, *5*, 23.

Lee, M. M., & Arthur, J. (2006). *U.S. Patent Application No. 11/349,619*.

Letowski, T. R., & Letowski, S. T. (2012). *Auditory spatial perception: Auditory localization* (No. ARL-TR-6016). Army Research Laboratory. Aberdeen Proving Ground, MD.

Lichenstein, R. Smith, D. C., Ambrose, J.L., Moody, L.A. (2012). Headphone use and pedestrian injury and death in the United States: 2004–2011. *Injury Prevention*

Lindeman, R.W., Noma, H., de Barros, P.G. (2007). Hear-Through and Mic-Through Augmented Reality: Using Bone Conduction to Display Spatialized Audio, Proc. of Int'l Symposium on Mixed and Augmented Reality (ISMAR) 2007.

Lomber, S. G., & Malhotra, S. (2008). Double dissociation of 'what' and 'where' processing in auditory cortex. *Nature neuroscience*, *11*(5), 609-616.

Loomis J.M., Golledge R.G., Klatzky R.L. (2001). GPS-based navigation systems for the visually impaired. In: Barfield W, Caudell T, editors. *Fundamentals of wearable computers and augmented reality*. Lawrence Erlbaum; Mahway, NJ: 2001. pp. 429–446.

MacDonald, J. A., Henry, P. P., & Letowski, T. R. (2006). Spatial audio through a bone conduction interface: Audición espacial a través de una interfase de conducción ósea. *International journal of audiology*, *45*(10), 595-599.

May, K. R., & Walker, B. N. (2017). The effects of distractor sounds presented through bone conduction headphones on the localization of critical environmental sounds. Applied Ergonomics, 61, 144-158.

Mershon, D. H., & King, L. E. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception & Psychophysics*, *18*(6), 409-415.

Middlebrooks JC, Onsan ZA. (2012) Stream segregation with high spatial acuity. J Acoust Soc Am 132:3896–3911, doi:10.1121/1.4764879, pmid:23231120.

Mwakalonge, J., Siuhi, S., & White, J. (2015). Distracted walking: examining the extent to pedestrian safety problems. *Journal of traffic and transportation engineering (English edition)*, *2*(5), 327-337.

National Highway Transportation Safety Administration. (2014) Traffic Safety Facts 2012 Data. U.S. Department of Transportation, National Highway Transportation Safety Administration, Publication DOT HS 811 888, Washington DC.

Nees, M. A., & Walker, B. N. (2009). Auditory interfaces and sonification. *The universal access handbook*, 507-521.

Neuhoff, J. (2004). Auditory motion and localization. In J.G Neuhoff (Ed.), *Ecological psychoacoustics* (pp. 87-111). San Diego, California: Elsevier Academic Press.

Oxenham, A. J. (2008). Pitch Perception and Auditory Stream Segregation: Implications for Hearing Loss and Cochlear Implants. *Trends in Amplification*, *12*(4), 316–331.

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual review of neuroscience*, *35*, 73.

Plenge, G. (1974). "On the differences between localization and lateralization," J. Acoust. Soc. Am., vol. 56, pp. 944–951, October 1974.

Rauschecker, J. P. (1998). Parallel processing in the auditory cortex of primates. *Audiology and Neurotology*, *3*(2-3), 86-103.

Recanzone, G. H. (2000). Spatial processing in the auditory cortex of the macaque monkey. *Proceedings of the National Academy of Sciences*, *97*(22), 11829-11835.

Salvendy, G. (2012). *Handbook of human factors and ergonomics*. John Wiley & Sons.

Santangelo, V., & Spence, C. (2008). Is the exogenous orienting of spatial attention truly automatic? Evidence from unimodal and multisensory studies. *Consciousness and cognition*, *17*(3), 989-1015.

Savioja, L., Huopaniemi, J., Lokki, T., & Väänänen, R. (1999). Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society*, *47*(9), 675-705.

Schroeder, P., & Wilbur, M. (2013). *2012 National Survey of Bicyclist and Pedestrian Attitudes and Behavior. Volume 1: Summary Report* (No. DOT HS 811 841 A).

Sodnik, J., Dicke, C., Tomažič, S., & Billinghurst, M. (2008). A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies*, *66*(5), 318-332.

Stanley, R. M. (2009). Measurement and validation of bone-conduction adjustment functions in virtual 3D audio displays.

Stelling-Kończak, A., Hagenzieker, M., & van Wee, B. (2015). Traffic sounds and cycling safety: The use of electronic devices by cyclists and the quietness of hybrid and electric cars. *Transport Reviews*, *35*(4), 422-444.

Stelling-Kończak, A., Hagenzieker, M., Commandeur, J. J., Agterberg, M. J., & van Wee, B. (2016). Auditory localisation of conventional and electric cars: laboratory results and implications for cycling safety. *Transportation Research Part F: Traffic Psychology and Behaviour*, *41*, 227-242.

Stelling-Kończak, A., van Wee, G. P., Commandeur, J. J. F., & Hagenzieker, M. (2017). Mobile phone conversations, listening to music and quiet (electric) cars: Are traffic sounds important for safe cycling? *Accident Analysis & Prevention*, *106*, 10-22.

Sussman, E., Winkler, I., & Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychonomic Bulletin & Review*, *10*(3), 630-637.

Szalárdy, O., Bendixen, A., Böhm, T. M., Davies, L. A., Denham, S. L., & Winkler, I. (2014). The effects of rhythm and melody on auditory stream segregation. *The Journal of the Acoustical Society of America*, *135*(3), 1392-1405.

Szalárdy, O., Bendixen, A., Tóth, D., Denham, S. L., & Winkler, I. (2013). Modulation-frequency acts as a primary cue for auditory stream segregation. *Learning & Perception*, *5*(Supplement 2), 149-161.

Treisman, A., & Paterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(1), 12.

U.S. Department of Labor. (1996). Occupational noise exposure- standard 1910.95. *Washington, DC: Occupational Safety and Health Administration.*

Walker, B. N., & Stanley, R. M. (2005). Thresholds of audibility for bone-conduction headsets. Georgia Institute of Technology.

Walker, B. N., Stanley, R. M., Iyer, N., Simpson, B. D., & Brungart, D. S. (2005). Evaluation of bone-conduction headsets for use in multitalker communication environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, No. 17, pp. 1615-1619). SAGE Publications.

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.

Walker, B. N., & Nees, M. A. (2011). Theory of sonification. In Hermann, T., Hunt, A., Neuhoff, J. G., editors, The Sonification Handbook, chapter 2, pages 9–39. Logos Publishing House, Berlin, Germany.

Wickens, C. D. (1991). Processing resources and attention. *Multiple-task performance*, 3-34. U.S.

Wickens, C. D. (2007). How many resources and how to identify them? Commentary on Boles et al. and Vidulich and Tsang. *Human Factors*, *49*(1), 53.

Wickens, C. D., Stokes, A., Barnett, B. & Davis, T. (1987). Componential analysis of pilot decision making: Final Report (Report No. SCEEE-HER/86-6). ChamDaien. &" IL: Aviation Research Laboratory, University of Illinois.

Wickens, C.D., McMarley, J.S. (2008). Applied Attention Theory. CRC Press.

Wilson, J., Walker, B. N., Lindsay, J., Cambias, C., & Dellaert, F. (2007). SWAN: System for Wearable Audio Navigation. Proceedings of the 11th International Symposium on Wearable Computers (ISWC 2007), Boston, MA (11-13 October).

Yao, J. D., Bremen, P., & Middlebrooks, J. C. (2015). Emergence of spatial stream segregation in the ascending auditory pathway. *Journal of Neuroscience*, *35*(49), 16199-16212.

Zahorik, P. A. (1998). *Experiments in auditory distance perception*. University of Wisconsin--Madison.