# 3D Sound Rendering in a Virtual Environment to Evaluate Pedestrian Street Crossing Decisions at a Roundabout

Haojie Wu* Daniel H. Ashmead† Haley Adams‡ Bobby Bodenheimer§

Vanderbilt University, USA

## Abstract

Crossing streets is a potentially hazardous activity for pedestrians, and it is made more hazardous when the complexity of the intersection increases beyond a simple linear bisection, as it does in the case of a roundabout. We are interested in how pedestrians make decisions about when to cross at such intersections. Simulating these intersections in immersive virtual reality provides a safe and effective way to do this. In this context, we present a traffic simulation designed to assess how pedestrians make dynamic gap affordance judgments related to crossing the street when supplied with spatialized sound cues. Our system uses positional information to generate generic head-related transfer functions (HRTFs) for spatializing audio sources. In this paper we evaluate the utility of using spatialized sound for these gap crossing judgments. In addition, we evaluate our simulation against varying levels of visual degradation to better understand how those with visual deficits might rely on their auditory senses. Our results indicate that spatialized sound enhances the experience of the virtual traffic simulation; its effects on gap crossing behavior are more subtle.

**Keywords:** Virtual reality, spatialized sound, gap affordance, traffic simulation

**Index Terms:** I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtal Reality; J.4 [Computer Applications]: Social and Behavioral Sciences—Psychology

## 1 Introduction

People can determine the location of objects and the distance to objects by auditory cues alone [26,50], although this ability is not as well developed as localization by visual cues. However, sound localization may be more significant in people with low or impaired vision [25]. In this paper, we investigate this ability in the context of street crossing behavior; more specifically, we determine how such ability trades off against vision in the context of selecting gaps in traffic for crossing the street.

We present an immersive virtual environment (IVE) simulation for traffic crossing at a roundabout. This simulation incorporates a system for rendering three-dimensional (3D) spatialized sound. While acoustic virtual environments have been demonstrated for some time, e.g. [3,46], this paper presents our architecture for a general, distributed real-time system for spatialized sound in the context of an HMD-based, immersive virtual environment and application.

We chose a roundabout for our traffic crossing scenario, because roundabouts are seen in the United States as increasingly desirable intersection designs due to improved safety and efficiency features over traditional intersections [7]. Although roundabouts are generally considered safe, particularly for vulnerable populations [1], pedestrian behavior at these intersections has been understudied. Meanwhile, a rich body of literature has developed using IVES to evaluate both pedestrian and cyclist interactions with traffic to improve safety [20,30,33,35]. IVEs are ideal for this type of research since real world traffic studies expose pedestrians to unnecessary risk by involving actual, moving vehicles. IVEs are appealing for traffic studies due to the control they give experimenters and the safety they afford participants.

However, these traffic simulations have largely been limited to linear intersections, such as that discussed in Kearney et al. [21], and typically do not provide spatialized sound. More realistic auditory information may inform how people interact with traffic simulations. Our IVE consists of a system for sound localization, and it generates more complex traffic patterns by using a traffic circle environment similar to that seen in Wu et al. [47].

In addition to the roundabout environment, our setup consists of a controllable traffic simulation and a 3D acoustic subsystem. The roundabout is modeled after a real location, the Pullen-Stinson roundabout on the North Carolina State University campus. It models a single lane traffic circle with crosswalks and splitter islands placed near entry and exit lanes. The traffic simulation generates natural vehicle acceleration and deceleration patterns, based on information taken from the location [40]. To this system, we have added a 3D acoustic subsystem capable of synthesizing the sounds associated with moving vehicles, and we are able to track the sounds' locations in the environment in real-time. Our audio system uses a non-individual head-related transfer function (HRTF), derived from the anthropomorphic audiological research mannequin KEMAR (Knowles Electronics) [10].

We anticipate that the acoustic system may affect our simulation in two ways. First, it may increase a user's sense of presence and immersion in the environment. An enhanced sense of immersion is desirable as it promotes more ecologically valid judgments from participants. Stated in another way, by enhancing the realism of our traffic simulation, we can elicit more natural responses from participants. This realism in turn better informs us about behavior in the real world.

As mentioned above, we are interested in studying the traffic crossing behavior of subjects with visual impairments. It is likely that people with visual impairments use auditory cues to aid with their traffic judgments, although the evidence for this is mixed [11], and vision, even when impaired, seems to dominate our sensory information. However, by adding a sound system to our virtual environment that allows people to localize sounds, we have the ability to test and assess this phenomenon in a manner similar to the evaluation of HRTFs for visually impaired and normal people conducted by Dong et al. [13]. Thus, although our experiments will use normally sighted individuals, we will introduce mock visual impairments to degrade the visual quality of their viewing experience to see if they will then employ the available auditory cues. This is motivated by evidence that such spatial memory cues are independent of the source (visual or auditory) from which they are derived [25].

The remainder of this paper is organized as follows: Section

*e-mail: haojie.wu@gmail.com

†e-mail: daniel.h.ashmead@vanderbilt.edu

‡e-mail: haley.a.adams@vanderbilt.edu

§e-mail: bobby.bodenheimer@vanderbilt.edu

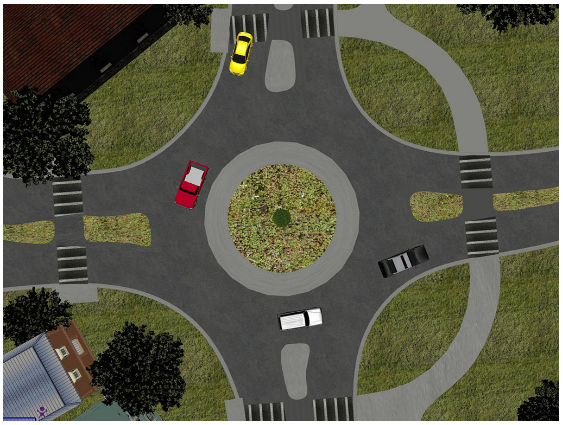

Figure 1: An overview of our virtual environment, modeled on the Pullen-Stinson roundabout, showing traffic circulating. Pedestrians cross at the walkways.

2 reviews prior literature and places our current work in context. Section 3 presents the details of the 3D audio system as well as a description of the traffic simulation and virtual environment. Section 4 presents two psychophysical experiments to evaluate the 3D audio system, and Section 5 discusses our results and presents future goals for the project.

## 2 Related Work

A sound wave undergoes complicated transformations as it travels from its source to the left and right ear canals of a listener. These transformations, called head-related transfer functions (HRTFs), are specific to each individual but can be approximated by non-individual ones, such as the approximation used in Kolarik et. al. [22]. There is some degradation in the quality of the 3D sound caused by this approximation, but measuring individual HRTFs is laborious and impractical for widespread use in IVEs. Overviews of this subject can be found in Begault [5] and Xie [48]. Suarez et al. [42] compared a version of the measured KEMAR HRTFs used here with modelled HRTFs in an immersive VE over several tasks and found no differences. In particular, our work applies measured HRTFs to a dynamic affordance (gap crossing).

The ability to represent large numbers of dynamic sound sources has become more viable with the ability to implement binaural rendering using a GPU (Graphical Processing Unit) to improve computational efficiency [6, 44]. These binaural sounds can then be presented through either headphones [19, 24, 43] or by loudspeaker via crosstalk cancellation systems [23, 27]. To enhance immersion and spatial orientation, HRTF-based binaural sound has been used widely in both augmented reality [37] and virtual reality [41]. Applications using spatialized sound have been developed to help the visually impaired navigate architectural spaces [34] and virtual maps [15]. In this paper, we implement the non-individual HRTF filter coefficients of the KEMAR mannequin published by the MIT Media Lab [14]. For HMD-based virtual environments, HRTF-based systems fed through earphones seem the obvious choice due to their compact form factor.

There has been much work evaluating spatialized audio in desktop virtual environments, e.g. [8, 9], and in immersive VEs [31, 45]. In desktop environments presence has been found to increase with the inclusion of spatialized audio, but quantitative task performance measures, such as task completion time, have been unaffected. Riecke et al. [38] also found that presence can be increased by spatialized sound. However, neither Naef et al. [31] nor Doerr et al. [45] used HRTFs for sound localization.

Virtual traffic crossing experiments often quantify the assessment of immersion by evaluating gap affordance judgments in traffic [18, 32, 35, 36]. When individuals cross a street, whether in reality or in a simulation, they must select a suitable gap in between vehicles to cross. To prevent collisions with oncoming vehicles, this gap must afford them sufficient time to physically locomote across the street before the next vehicle approaches. The assessment of traffic crossing behavior has important ramifications for the design of traffic intersections. For example, this information can be used to better design intersections to accommodate for vulnerable populations, such as children and the visually impaired, who can make poor gap judgments.

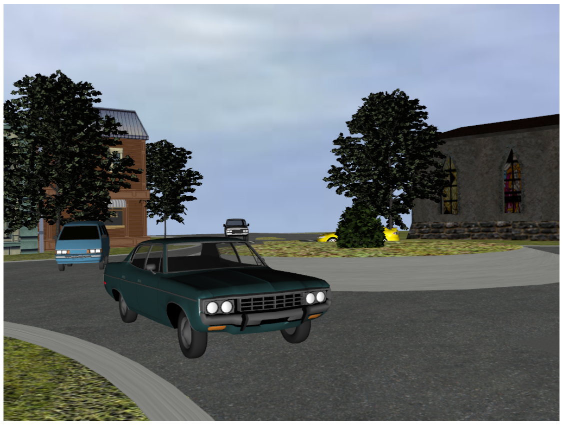

Figure 2: View of car approaching the traffic crossing, i.e., not circulating around the roundabout, as would be seen from a viewer about to cross the street in a a non-blurred ("normal") viewing condition. This view is from the eye-height of one of the authors, and shows the several of the car models used in the simulation.

## 3 System Design

### 3.1 Virtual Environment

The experiments were conducted in a 7.3m by 8.5m laboratory. The virtual environment was presented by a full color stereo NVIS nVisor SX Head Mounted Display (HMD) with 1280 x 1024 resolution per eye, a nominal FOV of 60 degrees diagonally, and a frame rate of 60Hz. As opposed to newer commodity-level HMDs, this HMD was equipped with an Arrington eye-tracker, although it was not employed in these experiments. An interSense IS-900 precision motion tracker was used to update the participant's rotational movement around all three axes. Position was updated using four optical tracking cameras that operated with two LED lights. The virtual environment displayed in the HMD was rendered in Vizard (Worldviz, Santa Barbara, CA).

### 3.2 Traffic Simulation

Roundabouts are traffic junctions where vehicles enter a circulating one-way stream of traffic around a central island. Instead of following a traffic control signal, vehicles must proactively yield to both upstream traffic and pedestrians. Pedestrians never walk through the circulating path or cross the center island. They only access designated crosswalks at the entry and exit lanes. Modern roundabouts have been statistically reported to reduce the severity of vehicle-to-vehicle crashes [12].

Our roundabout environment is an accurate, graphic street model of the Pullen-Stinson roundabout on North Carolina State University campus. Additional environmental features, such as buildings and vegetation, were added to the environment for realism; however, these additions were not based on the real world location. Prior work has utilized this same street model for the evaluation of street

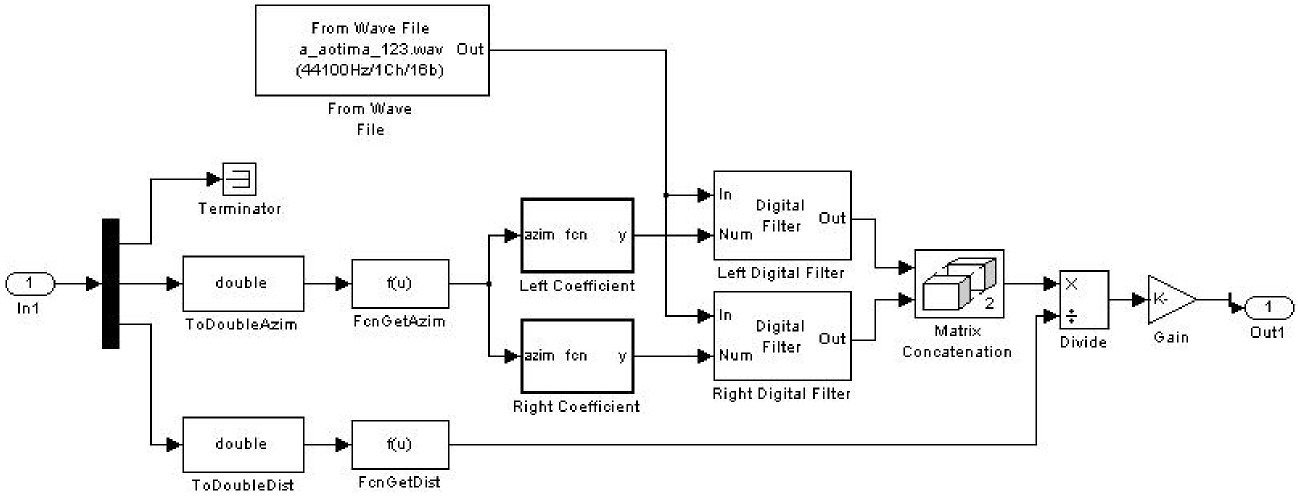

Figure 3: The Simulink implementation of the HRTF filter.

crossing behavior with nonlocalized sound in individuals with normal eyesight [47]. A bird's-eye view of the virtual environment is shown in Figure 1, and the environment seen from normal eye height with traffic approaching is shown in Figure 2. Prior observation of traffic in the real Pullen-Stinson roundabout showed that the average approaching speed of vehicles was 15.6m/s (35 mph) and the circulating speed was 8.0m/s (18 mph) [40]. In our virtual environment, approaching vehicles gradually reduce speed from 13.5m/s (around 32 mph) to 7.5m/s (around 17 mph), maintain 7.5m/s in the circulatory roadway, and then gradually resume to 13.5m/s upon exiting the traffic circle. Complicating this behavior further are provisions for vehicle-vehicle interactions and vehicle-pedestrian interactions. A vehicle avoids collisions with other vehicles by using a predetermined forward sight distance and an intersection sight distance. For vehicle-pedestrian interactions, each vehicle yields to pedestrians in the crosswalk area. More details of the traffic simulation can be found in Wu et al. [47].

### 3.3 Audio Simulation

The 3D acoustic subsystem applied HRTF measurements of KEMAR dummy head microphone [14] to simulate sound source interactions within the acoustic environment. The KEMAR measurements sampled a total of 710 different positions, and yielded impulse responses at a 44.1 kHz sampling rate, 512 samples long, and stored as 16-bit signed integers. There were two concurrent processes in the audio simulation. One process updated each sound source's position relative to the listener. This was accomplished by collecting the positional data of each vehicle in relation to the position of the pedestrian in the virtual environment. The other process linked the individual sound source positions with different sound source signals and applied the digital filters specified by KEMAR data to the sound source signals. This computational system was built in Simulink (MathWorks, Natick, MA).

For our system design, we assumed that sound emanated isotropically from an object. Thus the orientation of the vehicle models had no effect on their sound. Instead of their absolute x, y, and z coordinate values, KEMAR data lookup required the spatial information in azimuth, elevation, and distance related to the listener. To generate a synchronized soundtrack for traffic animations, the azimuth and distance of each moving vehicle were calculated along an animation path and updated in real-time. Values between the sample points were rounded to the nearest sample point. We simplified the elevation by assuming it to be a constant value of 10 degrees, although this was not an intrinsic limitation of the system. This spatial information was packed and sent from Vizard to Simulink over a UDP connection. No significant time delay was found between these two modules, and no packet loss over the local intranet was experienced.

In Simulink, a sound source signal was attached to a model of a vehicle as was its spatial information, which was received from Vizard. Sound source signals came from real engine noise recordings (cf. Baldan et al. [4]). We recorded eight different vehicles for each of the eight vehicle models in our system. These recordings were made of the car from outside the car with the car stationary at a constant engine speed. A sampling frame of recorded sound was selected based on the update rate of the positions. And the appropriate KEMAR filter coefficients for each ear were selected based on the values of the positions. The KEMAR filter and the sound source were then convolved, and the left and right channels were combined into binaural audio.

A significant auditory cue for distance is the sound level [2]. Accordingly, we scaled the sound level as the reciprocal of distance with the measured intensity at a distance of 1m. We used a sound level meter to calibrate each sound volume. Measurements were done with a dB-A scale slow setting using a RadioShack sound level meter (Cat. No. 33-2055A), with the average volume set to 70dB at 5 meters. Figure 3 demonstrates the Simulink implementation for a single vehicle. The current filter pair and loudness scale were replaced whenever new spatial information was received. Input was processed for each vehicle individually, and then all synthesized 3D sounds mixed to a set of shared signal busses. Finally, both visual output from Vizard and audio output from Simulink were then delivered through the HMD and into earbuds simultaneously. We used Klipsch S4 earbuds.

## 4 Experiments And Results

We conducted two experiments to assess our sound system. Because vision so dominates spatial cognition, we blurred the visual

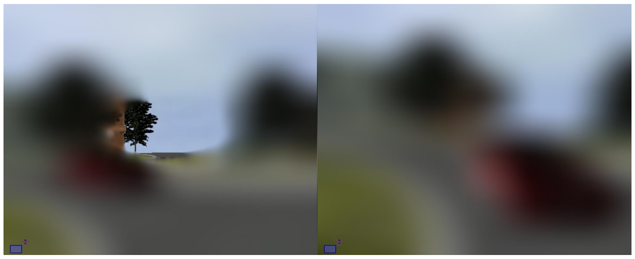

Figure 4: Blurred displays experienced by subjects in Experiment 1 (left) and Experiment 2 (right). In both scenes the view is from the side of the road with a car approaching. Notice that while the lead car can be seen, as in Figure 2, details of further cars are obscured in both cases.

environment severely in some conditions for better assessment of auditory effects. Sixteen total subjects, eight males and eight females, with ages ranging from 20 to 29 years old participated in the experiments with eight subjects participating in each experiment. Experimental groups were gender-balanced, and all participants self-reported normal sight and hearing. In both experiments, all conditions were within subject. No participant reported a significant amount of experience with virtual reality.

## 4.1 Experiment 1

Our first experiment attempted to assess the effect of both spatialized sound and visual field degradation on gap crossing behavior.

### 4.1.1 Procedure

Experiment 1 used a 2x2 design. Subjects experienced four traffic crossing conditions in counter-balanced order. The conditions were as follows: with visual blur and spatialized sound, with blur but no spatialized sound, with spatialized sound but no blur, and with no blur and no spatialized sound. For the blur condition, a simulation of glaucoma-like damage was presented, where the periphery was blurred and 10 degree FOV in the center was clear. The blur was set so that moving objects could be reliably discerned at 8 to 10 meters. An image of the blur effect for Experiment 1 is displayed in Figure 4 on the left-hand side.

Before the experiment began, written consent was obtained from all participants. Subjects were then introduced to the roundabout environment, the crosswalk, and all of the conditions during a brief learning phase within the IVE. During the experiment, subjects performed a series of trials to determine their gap threshold, or minimum safe gap crossing threshold. The gap for each trial was determined using a maximum-likelihood (ML) procedure similar to that seen of Grassi & Soranzo [17]. The participant crossed the street 15 times, which was sufficient for the ML procedure to converge.

For each trial of the experiment, a stream of traffic with a randomly chosen number of cars between 4 and 8 passed through the roundabout. Each adjacent car maintained a gap with less than 2 seconds between it and the next, except for the car that was assigned the designated target gap. This safe gap allowed for a time between 3 and 12 seconds in length. The position of the longer gap was randomly assigned in the traffic stream each time. For each condition, the participant was required to execute 15 street crossings. These limits were all determined through pilot testing. At the velocity the vehicles were traveling, 2 seconds is not sufficient for a pedestrian to safely cross the street, and 12 seconds is more than sufficient. The threshold at which people choose to cross lies somewhere in between. What will happen in this type of ML procedure is that the system will automatically adjust the gap threshold to make the next crossing a more difficult choice for each subject — longer if the prior response was a not to cross the target gap, and shorter if the prior response was to cross — based on the subject's prior history of responses. This procedure will converge in the 15 traffic crossing trials to that subject's gap crossing threshold, the minimum gap at which they are likely to cross. See Wu et al. [47] for further details.

Subjects were instructed to select a safe gap in traffic and to act upon this selection by physically walking across the virtual street. Subjects were told that, although the cars would not hit them, the drivers did not want to yield to them and that they should seek *the first* available safe gap in traffic. In all trials, the true gap, the subject's time to cross, and whether the subject actually made a safe crossing assuming no yielding on the part of a vehicle were recorded. We also administered questionnaires asking subjects' qualitative impressions of the conditions.

Table 1: Mean gap crossing time for Experiment 1 in seconds in each condition.

| | | Sound | |
|---|---|---|---|
| | | *Yes* | *No* |
| Blur | *Yes* | 5.75 | 5.69 |
| | *No* | 5.06 | 4.62 |

### 4.1.2 Results

In Experiment 1, with n=8 and the dependent variable as the gap duration of the converged ML procedure, we ran an analysis of variance (ANOVA) with blur and sound as the factors. The only significant effect was the main effect of blur: $F(1,7) = 5.8136, p < 0.0467$. The mean gap durations were 4.84 seconds with no blur and 5.72 seconds with blur. As expected, a substantial blurring of the visual image led to an increase in the average accepted gap duration. The difference was approximately 1 second, which would be functionally significant in a busy traffic setting.

An interesting possibility suggested by the results of this experiment is that one effect of adding sound is to increase the accepted gap duration. This effect is shown in Table 1, which presents mean gap durations by condition. In the condition with no blur, mean gap duration was actually longer with sound than without (although the difference was not statistically significant). This result is consistent with ratings made by the participants, that the sound added to the realism of the simulation.

So there may be counteracting effects of adding sound. On one hand, the sound may provide information that is useful for perceiving gaps, which could lead to better perception and shorter accepted gap durations. On the other hand, vehicle sounds may make the scenario more realistic, perhaps shifting the criterion for accepting gaps upward, toward longer gaps. Qualitatively, seven of eight subjects preferred the (sound, no blur) environment over all others, the eighth preferring (no sound, no blur).

## 4.2 Experiment 2

Our second experiment focused on determining the effect of spatialized sound on people's performance.

### 4.2.1 Procedure

Experiment 2 followed a similar experimental procedure to that seen in Experiment 1. However, this time a subject only experienced two test conditions, the traffic simulation with and without spatialized sound. For both conditions a subject's entire view was blurred.

The expanse and severity of visual blur was increased to discourage participant dependency on visual information, thereby isolating any effect of spatialized sound. This blur was extended throughout the entire screen, incidentally simulating a severe visual impairment.

The blur factor was also increased so that discernibility was decreased further to approximately 5 meters. An example of the blur effect may be seen in Figure 4 on the right-hand side.

The aspects of the experimental procedure for Experiment 2 not explicitly discussed followed that of Experiment 1. Experimental data and questionnaires were likewise collected in the same manner as in Experiment 1.

### 4.2.2 Results

In Experiment 2, with n=8 and the dependent variable as the gap duration of the converged ML procedure, the mean gap with sound was 5.81 seconds and without sound was 6.63 seconds. A paired sample t-test was marginally significant, $t(7) = 2.154, p < 0.068$. In this experiment there was very substantial blur in both the sound and no sound conditions. Not surprisingly, gap thresholds were generally higher than in the previous experiment.

Figure **??** shows a set of trials for one of the participants. The top chart shows the level of the next stimulus for the trials in the condition without sound and the bottom chart shows the level of the next stimulus for the trials in the condition with sound. Red circles represent failures to cross (no-go) and blue circles represent successful crossing (go). In both conditions, the initial stimulus was 12 seconds of gap separation. Over the course of the experiment, the stimulus selections gradually converged to 7.5 seconds without sound, and to 4.5 seconds with sound.

In this situation, a premium may be placed on extracting acoustic information about gaps, when that information is present. Thus, although the difference in gap thresholds between the sound and no sound conditions was not quite at the 0.05 level of statistical significance, the average thresholds were about 0.8 seconds lower when sound was available. This suggests that participants attempted to use the sound information, but that this information has limited reliability. Qualitatively, six of eight subjects preferred sound over no sound, one expressed no preference, and one preferred no sound.

## 5 Discussion

Our results are consistent with prior work on audio in virtual environments that show an increase in subjective satisfaction with spatialized audio but little improvement in task performance [9, 38]. It may be that, even in the blurred condition, subjects are able to extract enough optic flow information to make vision the dominant modality, and our results may suggest that subjects find audio localization unreliable. Nonetheless, the subjective improvement in the virtual environment from the presence of spatialized sound is clear.

Quantitatively, our results are more ambiguous. It is generally acknowledged that using generalized HRTFs such as the KEMAR ones used in this study can result in degraded localization performance in users [28, 29]. There is some evidence that the localization performance laterally is only minimally impacted by the use of these generalized HRTFs [39], although how this impacts moving targets is unclear. However, the generalized nature of the HRTFs employed in this work represents a limitation of our system and study, and employing more modern computational techniques to personalize the HRTF [16, 49] is a clear area for future work. It may be that isolating the localizing ability of spatialized sound in IVEs requires larger experimental power than employed in this study, another avenue for future work.

## Acknowledgments

This material is based upon work supported by the National Eye Institute, National Institutes of Health, under R01 EY12894-04, and by the National Science Foundation under grant 1526448.

## References

[1] D. Ashmead, D. Guth, R. Wall, R. Long, and P. Ponchillia. Street crossing by sighted and blind pedestrians at a modern roundabout. *Journal of Transportation Engineering*, 131(11):812821, 2005.

[2] D. H. Ashmead, D. Leroy, and R. D. Odom. Perception of the relative distances of nearby sound sources. *Perception & Psychophysics*, 47(4):326–331, 1990.

[3] P. Astheimer. What you see is what you hear-acoustics applied in virtual worlds. In *Proceedings of 1993 IEEE Research Properties in Virtual Reality Symposium*, pp. 100–107, Oct 1993. doi: 10.1109/VRAIS.1993.378256

[4] S. Baldan, H. Lachambre, S. D. Monache, and P. Boussard. Physically informed car engine sound synthesis for virtual and augmented environments. In *2015 IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, pp. 1–6, March 2015. doi: 10.1109/SIVE.2015.7361287

[5] D. R. Begault. *3D Sound for Virtual Reality and Multimedia*. Academic Press Professional, Inc., San Diego, CA, USA, 1994.

[6] J. A. Belloch, M. Ferrer, A. Gonzalez, F. Martinez-Zaldivar, and A. M. Vidal. Headphone-based virtual spatialization of sound with a gpu accelerator. *J. Audio Eng. Soc*, 61(7/8):546–561, 2013.

[7] T. R. Board, E. National Academies of Sciences, and Medicine. *Roundabouts in the United States*. The National Academies Press, Washington, DC, 2007. doi: 10.17226/23216

[8] K. Bormann. Presence and the utility of audio spatialization. *Presence: Teleoper. Virtual Environ.*, 14(3):278–297, June 2005. doi: 10.1162/105474605323384645

[9] K. Bormann. Subjective performance. *Virtual Reality*, 9(4):226–233, Apr 2006. doi: 10.1007/s10055-006-0019-5

[10] M. D. Burkhard and R. M. Sachs. Anthropometric manikin for acoustic research. *The Journal of the Acoustical Society of America*, 58(1):214–222, 1975. doi: 10.1121/1.380648

[11] A. Cheong, D. Geruschat, and N. Congdon. Traffic gap judgment in people with significant peripheral field loss. *Optometry and vision science : official publication of the American Academy of Optometry*, 85:26–36, 01 2008.

[12] S. Daniels, T. Brijs, E. Nuyts, and G. Wets. Externality of risk and crash severity at roundabouts. *Accident Analysis & Prevention*, 42(6):1966 – 1973, 2010. doi: 10.1016/j.aap.2010.06.001

[13] M. Dong, H. Wang, and R. Guo. Towards understanding the differences of using 3d auditory feedback in virtual environments between people with and without visual impairments. In *2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, pp. 1–5, March 2017. doi: 10.1109/SIVE.2017.7901608

[14] W. G. Gardner and K. D. Martin. Hrtf measurements of a kemar. *The Journal of the Acoustical Society of America*, 97(6):3907–3908, 1995. doi: 10.1121/1.412407

[15] M. Geronazzo, A. Bedin, L. Brayda, C. Campus, and F. Avanzini. Interactive spatial sonification for non-visual exploration of virtual maps. *International Journal of Human-Computer Studies*, 85:4 – 15, 2016. Data Sonification and Sound Design in Interactive Systems. doi: 10.1016/j.ijhcs.2015.08.004

[16] M. Geronazzo, A. Carraro, and F. Avanzini. Evaluating vertical localization performance of 3d sound rendering models with a perceptual metric. In *2015 IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, pp. 1–5, March 2015. doi: 10.1109/SIVE.2015.7361293

[17] M. Grassi and A. Soranzo. Mlp: a matlab toolbox for rapid and reliable auditory threshold estimation. *Behavior Research Methods*, 41(1):21–28, 2009.

[18] T. Y. Grechkin, B. J. Chihak, J. F. Cremer, J. K. Kearney, and J. M. Plumert. Perceiving and acting on complex affordances: How children and adults bicycle across two lanes of opposing traffic. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1):23–36, 2013.

[19] M. Hiipakka, T. Kinnari, and V. Pulkki. Estimating head-related transfer functions of human subjects from pressurevelocity measurements. *The Journal of the Acoustical Society of America*, 131(5):4051–4061, 2012. doi: 10.1121/1.3699230

[20] Y. Jiang, E. O'Neal, P. Rahimian, J. P. Yon, J. M. Plumert, and J. K. Kearney. Action coordination with agents: Crossing roads with a computer-generated character in a virtual environment. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '16, pp. 57–64. ACM, New York, NY, USA, 2016. doi: 10.1145/2931002.2931003

[21] J. K. Kearney, T. Grechkin, J. Cremer, and J. Plamert. Traffic generation for studies of gap acceptance. *Driving Simulation Conference*, 01 2006.

[22] A. Kolarik, S. Cirstea, and S. Pardhan. Discrimination of virtual auditory distance using level and direct-to-reverberant ratio cues. *The Journal of the Acoustical Society of America*, 134(5):3395–3398, 2013. doi: 10.1121/1.4824395

[23] Y. Lacouture-Parodi and E. A. P. Habets. Crosstalk cancellation system using a head tracker based on interaural time differences. In *IWAENC 2012; International Workshop on Acoustic Signal Enhancement*, pp. 1–4, Sept 2012.

[24] A. Lindau and F. Brinkmann. Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings. *J. Audio Eng. Soc*, 60(1/2):54–62, 2012.

[25] J. M. Loomis, R. L. Klatzky, B. McHugh, and N. A. Giudice. Spatial working memory for locations specified by vision and audition: Testing the amodality hypothesis. *Attention, Perception, & Psychophysics*, 74(6):1260–1267, Aug 2012. doi: 10.3758/s13414-012-0311-2

[26] J. M. Loomis, Y. Lippa, R. L. Klatzky, and R. G. Golledge. Spatial updating of locations specified by 3-d sound and spatial language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(2):335, 2002.

[27] P. Majdak, B. Masiero, and J. Fels. Sound localization in individualized and non-individualized crosstalk cancellation systems. *The Journal of the Acoustical Society of America*, 133(4):2055–2068, 2013. doi: 10.1121/1.4792355

[28] J. Middlebrooks and D. M. Green. Sound localization by human listeners. *Annual review of psychology*, 42:135–59, 02 1991.

[29] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi. Binaural technique: Do we need individual recordings? *Journal of the Audio Engineering Society*, 44(6):451–469, 1996.

[30] B. A. Morrongiello, M. Corbett, M. Milanovic, S. Pyne, and R. Vierich. Innovations in using virtual reality to study how children cross streets in traffic: evidence for evasive action skills. *Injury Prevention*, 21(4):266–270, 2015. doi: 10.1136/injuryprev-2014-041357

[31] M. Naef, O. Staadt, and M. Gross. Spatialized audio rendering for immersive virtual environments. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '02, pp. 65–72. ACM, New York, NY, USA, 2002. doi: 10.1145/585740.585752

[32] E. E. O'Neal, Y. Jiang, L. J. Franzen, P. Rahimian, J. P. Yon, J. K. Kearney, and J. M. Plumert. Changes in perception–action tuning over long time scales: How children and adults perceive and act on dynamic affordances when crossing roads. *Journal of Experimental Psychology: Human Perception and Performance*, 2017.

[33] E. E. O'Neal, Y. Jiang, L. J. Franzen, P. Rahimian, J. P. Yon, J. K. Kearney, and J. M. Plumert. Changes in perception-action tuning over long time scales: How children and adults perceive and act on dynamic affordances when crossing roads. *Journal of experimental psychology. Human perception and performance*, 2017.

[34] L. Picinali, A. Afonso, M. Denis, and B. F. G. Katz. Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge. *Int. J. Hum.-Comput. Stud.*, 72(4):393–407, Apr. 2014. doi: 10.1016/j.ijhcs.2013.12.008

[35] J. Plumert and J. K. Kearney. How do children perceive and act on dynamic affordances in crossing traffic-filled roads? *Child Development Perspectives*, 8, 10 2014.

[36] J. M. Plumert, J. K. Kearney, and J. F. Cremer. Children's road crossing: A window into perceptualmotor development. *Current Directions in Psychological Science*, 16:255–258(4), October 2007. doi: doi:10.1111/j.1467-8721.2007.00515.x

[37] R. Ranjan and W. S. Gan. Natural listening over headphones in augmented reality using adaptive filtering techniques. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(11):1988–2002, Nov 2015. doi: 10.1109/TASLP.2015.2460459

[38] B. E. Riecke, A. Väljamäe, and J. Schulte-Pelkum. Moving sounds enhance the visually-induced self-motion illusion (circular vection) in virtual reality. *ACM Transactions on Applied Perception (TAP)*, 6(2):7, 2009.

[39] G. D. Romigh and B. D. Simpson. Do you hear where i hear?: isolating the individualized sound localization cues. *Frontiers in Neuroscience*, 8:370, 2014. doi: 10.3389/fnins.2014.00370

[40] N. Rouphail, R. Hughes, and K. Chae. Exploratory simulation of pedestrian crossings at roundabouts. *ASCE Journal of Transportation Engineering*, 131(3):211–218, 2005.

[41] C. Schissler, A. Nicholls, and R. Mehra. Efficient hrtf-based spatial audio for area and volumetric sources. *IEEE Transactions on Visualization and Computer Graphics*, 22(4):1356–1366, April 2016. doi: 10.1109/TVCG.2016.2518134

[42] A. S. Suarez, J. Y. Tissieres, L. S. Vieira, R. Hunter-McHardy, S. K. Sernavski, and S. Serafin. A comparison between measured and modelled head-related transfer functions for an enhancement of real-time 3d audio processing for virtual reality environments. In *2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, pp. 1–9, March 2017. doi: 10.1109/SIVE.2017.7901609

[43] K. Sunder, J. HE, E. L. Tan, and W. S. Gan. Natural sound rendering for headphones: Integration of signal processing techniques. *IEEE Signal Processing Magazine*, 32(2):100–113, March 2015. doi: 10.1109/MSP.2014.2372062

[44] M. Taylor, A. Chandak, Q. Mo, C. Lauterbach, C. Schissler, and D. Manocha. Guided multiview ray tracing for fast auralization. *IEEE Transactions on Visualization and Computer Graphics*, 18(11):1797–1810, Nov 2012. doi: 10.1109/TVCG.2012.27

[45] K. u. Doerr, H. Rademacher, S. Huesgen, and W. Kubbat. Evaluation of a low-cost 3d sound system for immersive virtual reality training systems. *IEEE Transactions on Visualization and Computer Graphics*, 13(2):204–212, March 2007. doi: 10.1109/TVCG.2007.37

[46] E. M. Wenzel and S. H. Foster. Real-time digital of virtual acoustic environments. *SIGGRAPH Comput. Graph.*, 24(2):139–140, Feb. 1990. doi: 10.1145/91394.91431

[47] H. Wu, D. H. Ashmead, and B. Bodenheimer. Using immersive virtual reality to evaluate pedestrian street crossing decisions at a roundabout. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, APGV '09, pp. 35–40. ACM, New York, NY, USA, 2009. doi: 10.1145/1620993.1621001

[48] B. Xie. *Head-Related Transfer Function and Virtual Auditory Display*. J Ross Publishing: Plantation, FL, USA, 2013.

[49] K. Yamamoto and T. Igarashi. Fully perceptual-based 3d spatial sound individualization with an adaptive variational autoencoder. *ACM Trans. Graph.*, 36(6):212:1–212:13, Nov. 2017. doi: 10.1145/3130800.3130838

[50] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst. Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica*, 91(3):409–420, 2005.