# PERSPECTIVES ON HOW ACOUSTICAL, NON-ACOUSTICAL, AND USER CHARACTERISTICS SHOULD BE CONSIDERED IN MULTIMODAL VIRTUAL REALITY RESEARCH AND APPLICATION

**Jennifer L. Campos [⋆1,2], Darryl McCumber [†3], Brian Chapnik [‡3], Gurjit Singh [♦1,4,5], Sin-Tung Lau [°1], Karen Z. H. Li [✧6], Victoria Nieborowska [✦6] and M. Kathleen Pichora-Fuller [✹1,2]**
[1]Toronto Rehabilitation Institute - University Health Network, Toronto. Ontario
[2]Department of Psychology, University of Toronto, Toronto, Ontario
[3]HGC Engineering, Mississauga, Ontario
[4]Department of Speech-Language Pathology, University of Toronto, Toronto, Ontario
[5]Phonak Canada Ltd, Mississauga, Ontario
[6]Department of Psychology, Concordia University, Montréal, Québec

---

## Résumé

L'utilisation des environnements de réalité virtuelle (RV) est de plus en plus répandue dans la recherche et les milieux cliniques/appliqués, y compris dans le contexte de la recherche sur l'audition. Cela est, en partie, dû à la capacité de recréer des défis réalistes et quotidiens. En tant que tel, il devient de plus en plus important de caractériser les différences entre les propriétés acoustiques des sondes traditionnelles et les nouveaux environnements de test en RV. Bien qu'il existe des normes internationales spécifiant les propriétés acoustiques nécessaires aux environnements sonores hautement contrôlés, tels que les cabines de son (*soundbooths)*, il n'n'existe, actuellement, pas de pratiques optimales pour mesurer et contrôler les propriétés acoustiques des systèmes de RV multimodaux. Dans le présent article, nous fournissons une perspective générale sur comment les caractéristiques acoustiques ou non acoustiques (ex. affichages visuels, dispositifs interactifs) et les caractéristiques des utilisateurs (ex. âge) sont importantes à considérer dans la conception et l'utilisation de systèmes en RV multimodaux. Les mesures ont été effectuées dans des conditions dans lesquelles a) aucun équipement de laboratoire ne fonctionnait, b) l'équipement de laboratoire (ordinateurs, ventilateurs, matériel de projection, tapis roulants) fonctionnait, et c) des stimuli expérimentaux (discours cible, parole concurrente et autres bruits de fond tels que des bruits de circulation simulés) étaient présents ou absents. Comme preuve de concept, nous rapportons ici un protocole d'acquisition de mesures acoustiques (c.-à-d. temps de réverbération, niveau de bruit et rapport signal sur bruit) pour caractériser les propriétés acoustiques d'une cabine de son standard en comparaison aux données obtenues dans un laboratoire de RV multimodal représentatif (*StreetLab à l'Institut de réadaptation de Toronto*). Les mesures ont été effectuées dans des conditions dans lesquelles a) aucun équipement de laboratoire ne fonctionnait, b) l'équipement de laboratoire (ordinateurs, ventilateurs, matériel de projection, tapis roulants) fonctionnait, et c) des stimuli expérimentaux (discours cible, parole concurrente et autres bruits de fond tels que des bruits de circulation simulés) étaient présents ou absents. Nous discutons ensuite des conséquences potentielles et uniques de ces résultats sur la perception auditive et la performance chez des jeunes utilisateurs et des personnes plus âgées. Nous considérons également les implications pour la mise en œuvre du contenu auditif dans les systèmes de RV multimodaux de façon plus générale. Dans l'ensemble, il est très utile d'étendre les connaissances acquises par la recherche sur l'audition conduite dans les cabines de son en utilisant des conditions d'évaluation plus écologiques et plus réalistes offertes par les technologies de RV qui progressent rapidement. En effet, de telles technologies pourraient changer le paysage de la recherche auditive et les approches de pratiques en réadaptation en audiologie. Cependant, comme ces opportunités et technologies évoluent, il est nécessaire d'établir des lignes directrices et des normes appropriées pour la conception, la mesure et la comptabilisation des propriétés acoustiques des environnements des évaluations en RV pour la recherche et d'autres applications à travers des populations d'utilisateur diverses.

**Mots-clés :** simulation, environnements virtuels, audition, auditif, vieillissement, cabine de son, acoustique, réaliste

## Abstract

The use of Virtual Reality (VR) environments is becoming more widespread in research and clinical/applied settings, including in the context of hearing research. This is in part due to the ability to recreate realistic, everyday challenges. As such, it is becoming increasingly important to characterize the differences between the acoustical properties of traditional soundbooths and new VR test environments. While there are international standards specifying the necessary acoustical properties of highly controlled sound environments, such as soundbooths, there are no currently specified best practices for the measurement and control of the acoustical properties of multimodal VR systems. In the present paper, we provide a general perspective on how acoustical, non-acoustical (e.g., visual displays, interactive devices), and user (e.g., age) characteristics are important to consider in developing and using multimodal VR systems. As a proof of concept, we report here a protocol for acquiring acoustical measurements (reverberation time, noise level, and signal-to-noise ratio (SNR)) to characterize the acoustical properties of a standard soundbooth and compare these measurements to a representative multimodal VR laboratory (StreetLab at the Toronto Rehabilitation Institute). Measurements were made under conditions in which a) no lab equipment was operating, b) lab equipment (computers, fans, projection equipment, treadmills) was operating, and c) experimental stimuli (target speech, competing speech and other background noise such as simulated traffic sounds) were present or absent. We subsequently discuss the potential and unique consequences of these results to auditory perception and performance in younger and older user populations. We also consider the implications for implementing auditory content within multimodal VR systems more broadly speaking. Overall, there is great value in extending the knowledge that has been amassed from hearing research conducted in soundbooths by using the more ecological and realistic testing conditions afforded by rapidly advancing VR technologies. Indeed, such technologies could change the landscape of auditory research and approaches to practice in rehabilitative audiology. However, as these opportunities and technologies evolve, there is a need to establish appropriate guidelines and standards for designing, measuring, and accounting for the acoustical and non-acoustical properties of VR testing environments for research and other applications across various user populations.

**Keywords:** simulation, virtual environments, hearing, auditory, aging, sound booth, acoustics, realistic

## 1 Introduction

Over the last several decades VR technologies have improved dramatically. Higher quality, more accessible (i.e., more widely available and easier to use), and less expensive systems now provide novel ways in which to create realistic, controlled, and safe testing conditions [1-4]. Current VR systems also offer the opportunity to create more integrated multimodal simulations in which multiple sensory inputs can be presented with high fidelity and can be controlled systematically. Here we refer to multimodal VR systems as those that include simulated content presented via two or more sensory inputs. For instance, VR systems can be comprised of an immersive visual display (e.g., a large-screen projection display or a head-mounted display), a method of presenting auditory stimuli (e.g., headphones or loudspeakers), and/or interactive devices (e.g., treadmill, haptic glove, joystick, vehicle consoles in driving simulators, cockpits in flight simulators).

The content of the virtual environments and unique testing scenarios can be highly customized to the research question of interest. These systems allow investigators to evaluate human perception and performance under complex, multisensory conditions that more closely resemble conditions encountered in real world environments and interactions. They also permit control over the environmental content and the presence/absence and properties of individual sensory stimuli as might be done in more traditional lab-based experiments conducted in simpler and more artificial, unisensory test environments. While there have been marked improvements in the quality and implementation of visual displays and interactive devices, less careful consideration has been given to the widespread implementation and incorporation of realistic auditory displays within many multimodal VR systems [5-7]. There is great utility in considering the importance of acoustical and auditory stimulus properties across all VR applications for which they are implemented. VR could provide a unique tool for research and applications focused on auditory perception in complex environments.

### 1.1 Hearing research: Moving out of the soundbooth

Traditionally research in fields such as psychoacoustics and audiology has been conducted in sound-attenuating booths or anechoic rooms where it is possible to precisely control environmental conditions. Specifically, experimental conditions have been considered ideal if they minimize acoustical interference (e.g., reverberation or background noise), distracting multisensory stimulation (e.g., complex or dynamic visual or motor inputs), and attentional distractions (e.g., multi-tasking). Limiting these factors has many advantages if the tester wants to precisely evaluate the abilities of individuals to detect, perceive, and interpret auditory stimuli as a function of the physical properties of sound signals, to define neurophysiological responses to auditory stimuli, to characterize different types and magnitudes of hearing loss, and/or to evaluate some basic benefits of using technologies such as hearing aids. Nevertheless, questions may be raised as to the functional significance of the results obtained in such artificial testing environments that lack the typical demands of the multisensory (e.g., auditory, visual), mobility-related, and cognitive conditions that people often encounter in the real world. Thus, researchers are beginning to develop new methods to move from testing *hearing* in acoustically ideal soundbooth conditions to testing *listening* in more realistic and often adverse conditions [8, 9]. These new approaches could enable researchers to study the complex interactions among sensory and cognitive processes that have functional implications for listening in daily life [10]. Therefore, in the context of hearing research, multimodal VR systems can offer a valuable middle ground between controlled laboratory/clinical soundbooth environments and the real world.

---

[*] jennifer.campos@uhn.ca
[†] dmccumber@hgcengineering.com
[‡] bchapnik@hgcengineering.com
[♦] gurjit.singh@phonak.com
[°] sin-tung.lau@uhn.ca
[✧] karen.li@concordia.ca
[✦] victoria.nieborowska@alum.utoronto.ca
[✸] k.pichora.fuller@utoronto.ca

Several decades ago, methods were developed to simulate purely auditory scenes using loudspeaker arrays within acoustically controlled lab settings. By the mid-1990's, investigators began adapting commercially available signal processing software using head-related transfer functions (HRTFs) to simulate spatial displays of sound sources, with the idea being that VR could be applied in rehabilitative audiology (e.g., [11]). The early auditory VR simulations were typically implemented without corresponding simulated visual inputs (or any other concomitant sensory inputs). These simulated audio techniques were subsequently adopted to test auditory perceptual processes, assess the nature of hearing loss, and evaluate new hearing aid technologies [12]. Such algorithms have continued to become more and more sophisticated [13]. There has been a growing awareness and interest in these techniques (see Figure 1 for a historical timeline), including explorations of how they can be applied in clinical settings. For example, VR holds promise as a tool for optimizing hearing aid fittings by testing different hearing aid functions in more realistic scenes [14, 15].

Currently, auditory simulations are being incorporated into even more complex, multimodal VR systems developed to represent specific sensory-cognitive-motor interactions during tasks like those encountered in everyday life, such as in walking simulators and driving simulators [16-19]. From the perspective of hearing researchers, there is a growing awareness that testing the performance of people who have hearing loss within these contexts is very important given that hearing loss affects not only speech intelligibility, but also non-auditory domains such as cognition and mobility [48]. Testing in simulated VR environments could provide new knowledge regarding the effects of hearing loss on perception and performance in everyday environments. For instance, hearing loss is associated with higher rates of falls [20, 48] and driving errors, particularly when individuals are distracted [21, 48]. Importantly, being on the brink of potentially widespread implementation of multimodal VR systems, it is now a critical time for investigators to establish standards and guidelines surrounding the design, measurement, and implementation of auditory simulations within multimodal VR applications. Such guidelines should include considerations of the acoustics of the VR environment (e.g. reverberation time, noise level), the characteristics of the user (here we focus on age), and the nature of additional sensory inputs (e.g., visual environment). Below we reflect on why each of these factors is important to consider in the context of the development of multimodal VR and we describe a proof of concept approach by characterizing the acoustics of a representative multimodal VR research laboratory and the potential consequences to auditory performance across different user populations. Indeed, we have been motivated by our own experiences in attempting to use a listening task traditionally conducted in a soundbooth [22] within a multimodal, VR environment [17, 23]. In order to understand the reasons underpinning the clear differences in word recognition accuracy we observed within these two spaces for younger and older adults, we needed to compare the acoustical properties of the two test environments.

## 1.2 Auditory displays and acoustical factors

Acoustical considerations are important because a potential limitation of using multimodal VR systems for some hearing research is that simulations of auditory scenes may be contaminated by other background sounds inside or outside the test environment (e.g., computers, fans, interactive devices), especially if there is inadequate attenuation of ambient noise provided by the walls of the test room. There may also be excessive reverberation within the test environment. Even with these shortcomings, VR may still be more advantageous than artificial soundbooth testing conditions for some research and clinical purposes.

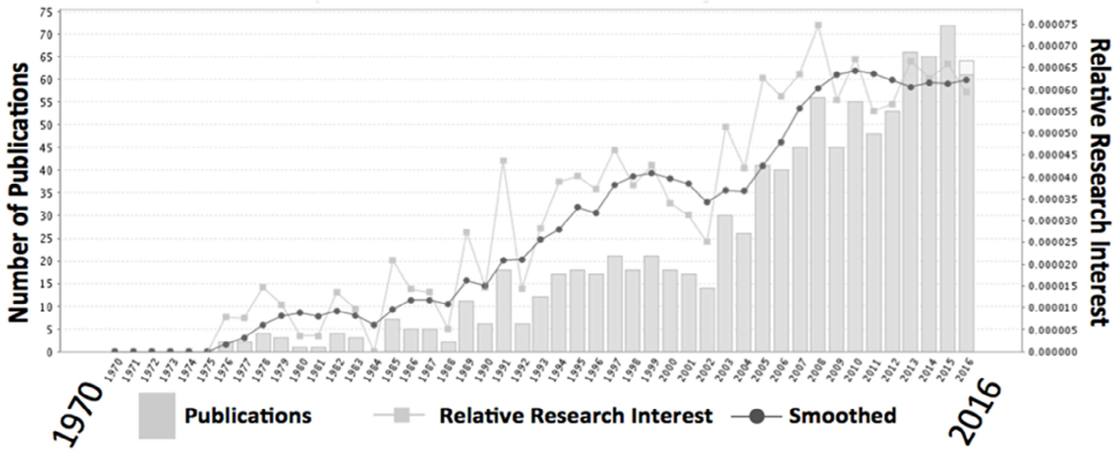

**Figure 1:** A historical timeline reflecting the growing interest in auditory simulations. Values based on a GoPubMed search in December, 2016, using search terms ((((((((auralization*) OR auditor*) OR ((hearing AND loss*))) OR Audiology[MeSH Terms]) OR Hearing Loss[MeSH Terms]) OR "Head-Related Transfer Function*"))) AND (((computer simulation[MeSH Terms]) OR Virtual realit*) OR Simulation*). Bar graphs represent the total number of publications and the line graphs represent the relative research interest (i.e. the relative growth in comparison to the growth of whole PubMed).

In audiology, progress has been made using HRTFs to simulate spatialized sound in clinical tests conducted under headphones (e.g., the Listening in Spatialized Noise-Sentences; [24, 25]). Simulated auditory displays presented over headphones, however, cannot be used to test performance when conventional hearing aids are worn. In future, methods using computational corrections to overcome limitations in acoustical simulations presented over loudspeakers could provide important new tools that would enable VR to be used more extensively in rehabilitation applications [15].

## 1.3 Age of the user

Another factor that is rarely considered in the context of VR system development, application, and evaluation, is the unique characteristics of the users or research participants. Well-documented sensory and cognitive changes occur over the lifespan [26]. Age-related changes may affect performance across a range of basic behavioral tasks, and may interact with the characteristics of the testing environment. Specifically, when considering the effects of auditory and cognitive aging, differences in the acoustical and non-acoustical properties of test environments will likely lead to differences on task performance that may be proportionally greater for older compared to younger adults. For instance, in terms of acoustical properties, highly controlled soundbooth environments allow for better control over parameters such as reverberation, sound level, and signal-to-noise ratio (SNR). In VR environments (and in the real world), these parameters are often more difficult to control, and the consequences to performance may be more apparent in older than in younger adults [18]. It is not uncommon for VR laboratories to have additional sources of noise (e.g., interactive devices such as treadmills, computers, safety devices and ventilation systems). These types of sounds that are not designed as part of the simulation may have differential effects on performance outcomes depending on the abilities of the listener. For example, it may be important to customize acoustical stimuli to individuals or groups of listeners to ensure audibility and minimize disproportionate masking effects. Investigators developing and using these systems must be cognizant of the potential need to adjust the level and other characteristics of the presentation of acoustical signals according to listeners' abilities to compensate for unwanted or unintended interactions between the person and the test environment that are introduced in the VR simulation but that would not be present in the real world conditions being simulated (e.g., the noise of the computers used to produce the VR simulation).

## 1.4 Multisensory and multimodality factors

A non-acoustical property that can differ between soundbooth and VR environments is the presence and complexity of visual input. Visual input that is deliberately incorporated into simulated VR content can be physically or semantically related to the auditory input (e.g., the coupling of dynamic visual and auditory inputs such as wind, tire, engine sounds generated during simulated driving), or unrelated (e.g., the simultaneous presentation of a dynamic visual driving scene with non-informative or even distracting auditory input such as music, radio commentary, or telephone communication). The former case may be beneficial, while the latter case may be detrimental to performance. For some tasks, older adults are thought to demonstrate a heightened integration of redundant and congruent sensory inputs compared to younger adults [27-29], such that congruent multisensory conditions may provide proportionally greater benefit than reduced sensory conditions for older compared to younger adults. In contrast, older adults may be less able to inhibit irrelevant or incongruent multisensory inputs [30], suggesting that the presence of non-informative and potentially distracting multisensory feedback may be more disruptive for older adults.

In summary, testing perception and performance under ecological and realistic simulated conditions using VR may have advantages depending on the question at hand. However, performance will vary with the properties of the stimuli (e.g., unisensory versus multisensory), the testing environment (e.g., impoverished versus enriched), the task, and the age of the user/participant. Thus, it is important to account for and report these factors in VR research and applications. While it is common practice to thoroughly describe and compare test stimuli and user characteristics across studies, testing environments are seldom compared. Hearing researchers typically comply with acoustical standards for auditory testing environments [31], but VR researchers who are not focused on audition per se are more likely to employ highly variable, non-standardized, and poorly characterized auditory testing environments. Thus, standards for auditory VR developed by experts in acoustics and hearing will need interdisciplinary adoption.

## 1.5 Objectives of the current study

In this exploratory study, we characterize the acoustical properties (e.g., noise level, SNR, reverberation time) of a standard soundbooth and compare these measured properties to those of a representative, high-fidelity, immersive VR environment (StreetLab at the Toronto Rehabilitation Institute's Challenging Environment Assessment Laboratory (CEAL)). Measurements were made for the following conditions: a) without any lab equipment operating (i.e., computers, fans, projection equipment, interactive devices), b) with the lab equipment operating, and c) with auditory stimuli (target speech, competing speech and other background noise such as simulated traffic sounds) present or absent. We subsequently discuss the potential consequences of these results to auditory perception and performance across different user populations and the broader implications of implementing auditory content within multimodal VR systems.

# 2 Method

## 2.1 Stimuli and Apparatus

### Stimuli

The speech stimuli present during some acoustical measurement conditions (see below) were the Coordinate Response Measure (CRM) sentences developed for research concerning listening in multi-talker displays [32].

### Soundbooth

The soundbooth used to collect benchmark acoustical measurements was a 3.3 m (l) x 3.3 m (w) x 2.01 m (h) (21.89 m$^3$) single-walled sound-attenuating booth (Industrial Acoustics Company, New York) located at the Human Communication Lab at the Mississauga Campus of the University of Toronto. An array of three Grason-Stadler loudspeakers (No. 1761-9630) was used to present the speech stimuli. The stimuli were presented from each loudspeaker at 60 dB A. The three loudspeakers were positioned in the soundbooth at approximately the head height of a seated person and at a distance of 1.6 meters, with one loudspeaker in front (0° azimuth), one to the right, and one to the left (+/-90° azimuth). All loudspeakers were activated when speech stimuli were presented.

### Multimodal VR Laboratory

Acoustical measurements were collected in StreetLab located within Toronto Rehabilitation Institute's CEAL (Figure 2).

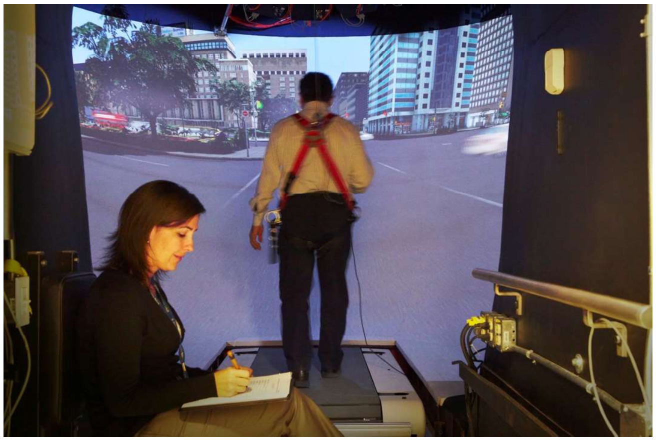

**Figure 2:** StreetLab Virtual Reality environment within the Challenging Environment Assessment Laboratory (CEAL) at the Toronto Rehabilitation Institute.

In StreetLab, an array of seven loudspeakers (Meyersound MP-4XP, Meyersound Laboratories, Inc.) and a subwoofer (Meyersound MP-10XP) were located behind a curved visual projection screen (the screen is made of a thin, sound-permeable material). The center loudspeaker is positioned at 0° azimuth at approximately head height for a standing person and the other six loudspeakers are distributed in an array in the same horizontal plane at +/-28° (right front and left front), +/-90° (right side and left side), and +/-127° (right rear and left rear). The subwoofer is located under the floor, below the center loudspeaker, in front of a treadmill. Each loudspeaker was positioned at a distance of 2.14 m from the listener. StreetLab is teacup shaped and the interior spatial volume is 31.66 m$^3$. Visual simulations within StreetLab were presented using a high-resolution, 240° field-of-view horizontal x 110° field-of-view vertical projection screen with a calibrated six projector system (Eyevis ESP, Reutlingen, Germany). Sound dampening foam panels are installed behind the screen surface, on the surrounding walls, and on parts of the floor and ceiling to provide sound attenuation (BasoTect Melamine sound insulation foam). For this study, the simulated scene was a six-lane, two-directional traffic intersection in downtown Toronto (see Figure 2), which was simulated using a customized OpenSceneGraph application (http://www.openscenegraph.org). The sentence stimuli used for the speech recordings (described above) were presented with no corresponding visible talker in the simulated visual scene. All loudspeakers were activated when speech stimuli and traffic stimuli were presented. StreetLab also has the capability of introducing mobility-related tasks, such as walking on a treadmill, balancing on a force platform, driving a car, or maneuvering a wheelchair. In this study, acoustical measurements were also made during the operation of the treadmill device.

## 2.2 Procedures

Acoustical measurements were conducted in the soundbooth at the Human Communication Lab and in StreetLab by the same acoustical engineer (DM). The sound measurements were conducted using a Norsonic NOR140 sound level meter, serial number 1405033. The measurements were conducted at a height and location approximating the head height and position of a typical participant in the two testing environments. Measurements and analyses included reverberation times (RT), background sound levels, SNR calculations and speech intelligibility calculations.

The criteria discussed below have been developed based on the guidelines provided in the ANSI S12.60 [33] American National Standard, "Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools", which provides guidelines for the acoustical design of teaching spaces. This well-known standard was chosen as a benchmark for evaluating the VR environment because classrooms are real-world communication environments that are expected to meet strict acoustical criteria to enable listeners to achieve acceptable performance when confronting the cognitive demands of verbal communication during learning. This is therefore an important benchmark with respect to acceptable *minimal* standards when simulating realistic, but quiet listening conditions. For simulations that require noisier conditions (e.g., city street), it is easier to titrate up from the minimal standards for a quiet space than it is to attempt to make an unavoidably noisy environment quieter.

### Reverberation time

Reverberation is characterized by the time it takes sound to decay by 60 dB ($RT_{60}$). In this study, $RT_{60}$ was measured over the entire frequency range from 100 to 5,000 Hz in each test environment. While reverberation can be measured across all frequencies, the $RT_{60}$ measured in the octave band around 500 Hz is often referenced as a simple comparative measure. For example, in teaching spaces, the $RT_{60}$ at mid-frequencies (e.g., at 500 Hz) should be kept below 0.5 seconds and values greater than 0.5 seconds can reduce speech intelligibility, especially for individuals with hearing loss [34].

### Background noise levels

Background sound can reduce speech intelligibility. The most widely accepted criteria for recommended levels of background sound are based on overall A-weighted sound levels and/or Noise Criterion (NC) curves [35]. ANSI S12.60 [33] indicates that core teaching spaces should have background sound levels of approximately 35 dB A (or less), corresponding to roughly NC-30. Higher background sound levels may be acceptable if the level of the speech is high enough, but this range is considered reasonable to ensure good intelligibility of source signals at normal voice output levels in a typical space.

### Signal-to-Noise ratio

Once reverberation is reasonably well controlled, the main acoustical factor contributing to speech intelligibility is the SNR; that is, the levels of speech reaching a listener (and particularly the speech peaks) relative to the background sound levels at the listener's position. To illustrate this, calculations can be performed to determine the Speech Intelligibility Index (SII), as described by ANSI/ASA S3.5 [36], Standard Methods for Calculation of the Speech Intelligibility Index. In these calculations, the measured background sound levels are compared to the measured speech peaks. An SII value of 1 indicates that all speech cues reach the listener, whereas an SII value of 0.0 indicates no intelligibility. A value of 0.5 indicates that half of the speech cues are intelligible.

### Conditions tested

In the present study, background sound levels were measured in each of the testing environments under several different conditions that varied as a function of the additional sources of background sounds that could be generated in the test environments during experiments. Specifically, the soundbooth was measured with a) no equipment operating and b) equipment operating (lights, computer, touch screen monitor). StreetLab was measured with a) no equipment operating, b) equipment operating (lights, projectors), c) treadmill on, but not moving, d) treadmill on, moving at 1m/s, e) traffic noise simulation added with no additional equipment operating, f) traffic noise added with the treadmill moving at 1 m/s. Speech stimuli were introduced during Conditions b (soundbooth and StreetLab), c and d (StreetLab only) to measure signal-to-noise ratios and the SII.

## 3 Results

The $RT_{60}$ was lower in the soundbooth than in StreetLab at all frequencies (Figure 3). Nevertheless, the $RT_{60}$ was less than 0.5 seconds for all but the lowest frequencies in StreetLab, which is considered acceptable for communication and unlikely to result in reduced speech intelligibility in an environment such as a classroom [34]. Figure 4 shows the noise levels across frequencies and Table 1 shows the average sound levels for the selected sample of possible testing conditions that could be used during experiments conducted in StreetLab.

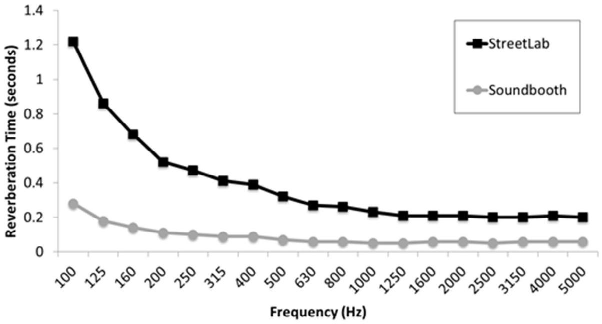

**Figure 3:** Reverberation times ($RT_{60}$) measured from 100 Hz to 5 kHz in both the soundbooth and StreetLab testing environments.

In the soundbooth, when no equipment was operating, the background sound level (23 dB A) was well within the targets specified in ANSI standard S12.60 [33] for classroom environments. As expected, the levels across frequencies also approximated the stricter criteria concerning permissible levels for audiometric testing specified in ANSI S3.1 [37]. In StreetLab, when no equipment was operating, the background sound level (43 dB A) was higher than in the soundbooth and higher than the targets referenced for ANSI S12.60 [33]. Turning on the basic equipment in Condition b in the soundbooth made very little difference to the measured sound levels (1 dB), whereas turning on the basic equipment in StreetLab resulted in a greater increase in the sound level (8 dB). Furthermore, as shown in Figure 4, the frequency response of the noise produced by turning on basic equipment differed between the two environments, with higher noise levels at the mid-high frequencies in StreetLab than in the soundbooth. While simply turning on the treadmill in StreetLab introduced no additional increase in sound level (Condition c), the sound level increased by an additional 6 dB when the treadmill motors were operating (Condition d), and another 5 dB when simulated traffic noise was added (Condition e). Therefore, it would be expected that speech intelligibility performance would be poorer in StreetLab than in a conventional soundbooth because of the elevated levels of background noise, with the differences being greater as more realistic conditions were used in StreetLab (e.g., walking on a moving treadmill or adding traffic noise).

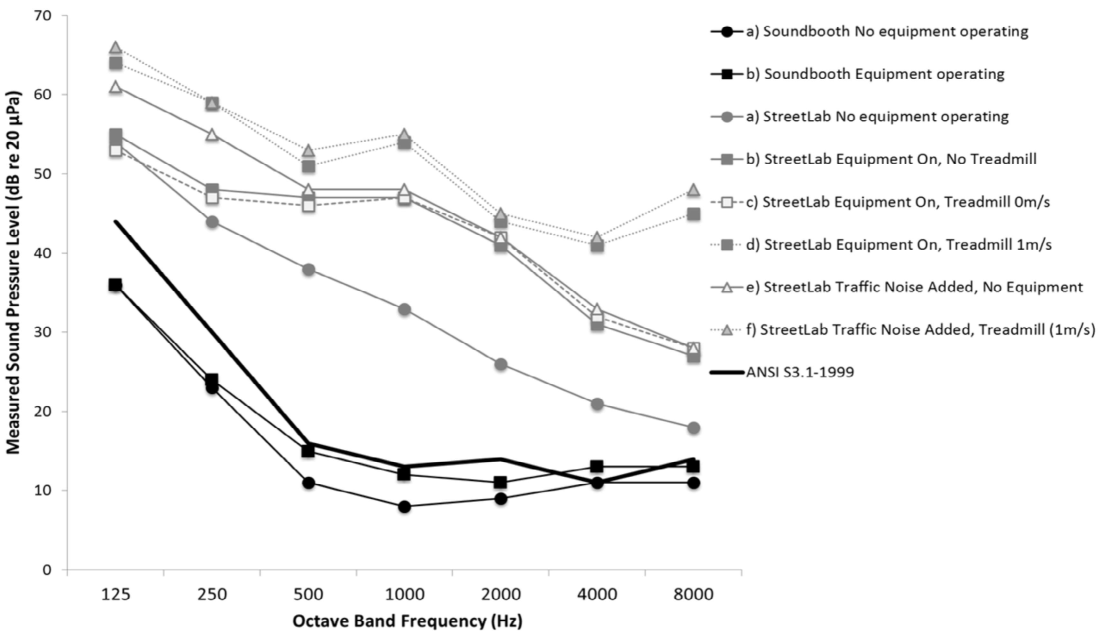

**Figure 4:** Octave-band sound levels measured in the soundbooth and in StreetLab. Results are shown for conditions without and with basic equipment operating in each test environment (Condition a and b respectively). Results are also shown for four additional possible test conditions in StreetLab, including: treadmill on but not moving (Condition c), treadmill moving at 1m/s (Condition d), with traffic noise but no equipment on (Condition e), or with traffic noise and the treadmill moving at 1 m/s (Condition f). The corresponding maximum permissible background sound levels for audiometric testing specified in ANSI S3.1-1999 are also shown for comparison.

**Table 1:** Average background sound levels across a sample of different testing conditions in dB A and NC.

| **Location** | **Condition** | **Background Sound Level** | |
|---|---|---|---|
| | | **dB A** | **NC** |
| Soundbooth | a) No equipment operating | 23 | 15 |
| | b) Equipment operating | 24 | 17 |
| StreetLab | a) No equipment operating | 43 | 37 |
| | b) Equipment operating | 51 | 46 |
| | c) Equipment On, Treadmill 0 m/s | 50 | 46 |
| | d) Equipment On, Treadmill 1 m/s | 57 | 53 |
| | e) Traffic Noise Added, No Equipment | 56 | 65 |
| | f) Traffic Noise Added, Treadmill (1 m/s) | 60 | 65 |

Speech intelligibility calculations were conducted for the CRM sentences presented in the soundbooth and StreetLab under some of the background noise conditions listed in Table 1 (Condition b in both environments and Conditions c and d in StreetLab). In addition to the conditions listed in Table 1, another condition was evaluated in which the intensity level of the CRM sentences was strategically increased in an attempt to bring the SNR in StreetLab closer to the recommended target SNR for classrooms. The speech peaks of the CRM sentences were measured as characterized by the 10th percentile, $L_{10}$, of the sound level in each band occurring over each sentence. Notably, as shown in Figure 5, the frequency responses of the speech stimuli differed between the two test environments, with higher levels of speech in the mid-high frequencies in the soundbooth than in StreetLab. The speech peak measures were used in conjunction with the background sound levels occurring at the same time to calculate the SII.

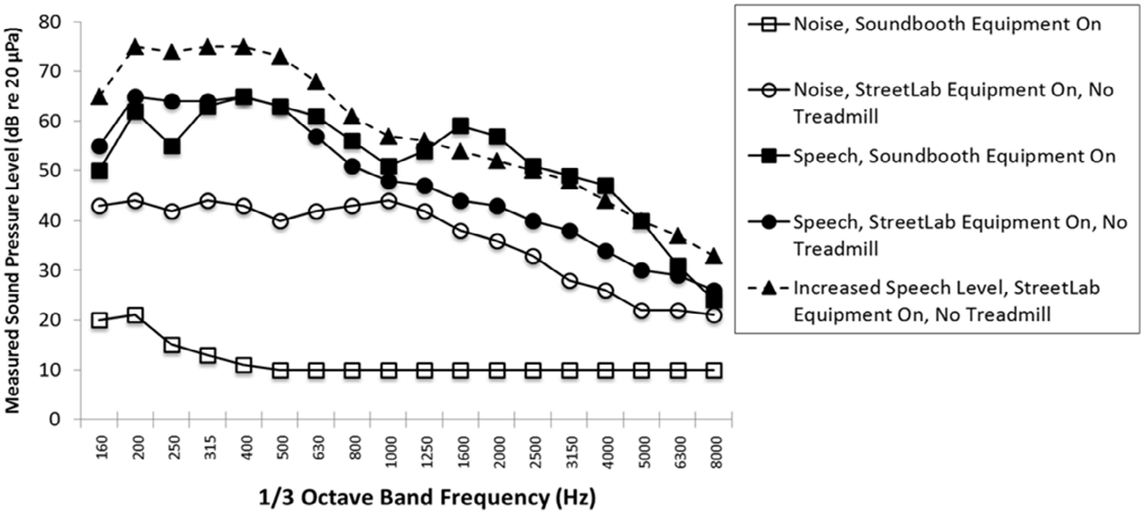

**Figure 5:** 1/3 octave-band sound levels measured in the soundbooth and in StreetLab as used for calculating the SII. Results are shown for noise measured with basic equipment operating (Condition b) in each test environment. Results are also shown for speech stimuli in each environment and at an increased level in StreetLab.

**Table 2:** Comparison of speech peak to background sound and calculated SII.

| **Location** | **Condition** | **Background Sound Level (dB A, $L_{EQ}$)** | **Level of Speech Peaks (dB A, $L_{10}$)** | **SNR** | **Speech Intelligibility Index (SII)** |
|---|---|---|---|---|---|
| Soundbooth | Equipment Operating | 24 | 61 | 37 | 0.98 |
| StreetLab | Equipment Operating | 51 | 66 | 15 | 0.74 |
| | Equipment On, Treadmill 0 m/s | 50 | 66 | 16 | 0.73 |
| | Equipment On, Treadmill 1 m/s | 57 | 66 | 9 | 0.64 |
| | Increased Speech Peaks | 51 | 76 | 25 | 0.96 |

The overall A-weighted levels and the corresponding calculated SII results are summarized in Table 2. The SII in the soundbooth was .98. In contrast, there were much poorer SII values in StreetLab; e.g., with basic equipment operating the SII value was .74, with the treadmill on but not moving it was .73, and when the treadmill was moving it was even lower (.64). These SII values reflect the slightly lower average speech peak levels in StreetLab compared to the soundbooth (5 dB difference) and the much higher noise levels (28 dB difference) in StreetLab with basic equipment operating compared to the soundbooth. However, increasing the level of the speech was successful in achieving an SII value (.96) that was closer to that achieved in the soundbooth.

# 4 Discussion

There were several clear differences between the acoustical properties of the soundbooth and StreetLab, including reverberation, background noise level, and the SNR under different experimental conditions. Below we discuss how these differences in acoustical properties might affect listening performance depending on the nature of the task and the population tested.

## 4.1 Acoustical properties

### Reverberation times

StreetLab was more reverberant than the soundbooth. The reverberation level within StreetLab, however, was still within the range deemed to be acceptable according to the ANSI 12.60 [33] standards for a classroom environment. The effects of increased reverberation on perceptual and behavioral outcomes largely depend on the scenario being tested. For instance, when simple identification tasks are the main outcome of interest, increased reverberation may be less consequential; however, if precise sound localization or speech intelligibly in noise is being evaluated, more reverberant testing environments may be more deleterious [6]. Such deleterious consequences would likely be greater for older compared to younger adults given that age and pure-tone thresholds are independently correlated with ability to recognize words in reverberant and noisy environments (e.g., [38]).

**Background sound levels**

The background sound levels in StreetLab were higher than those in the soundbooth, even in the most acoustically controlled condition (i.e., no equipment operating, no interactive devices activated, and no simulated ambient traffic sounds). When no equipment was operating, the noise level in StreetLab was 8 dB higher than the maximum recommended average level for good speech intelligibility (35 dB A; [34]). Furthermore, the maximum recommended background sound level was exceeded by 16 dB when the basic equipment was operating, by 15 dB when the treadmill was also operating, and by 25 dB when all equipment was operating and the simulated street and traffic noise was turned on. Clearly, for experiments requiring a quiet environment, it would be more appropriate to test in a soundbooth than in a multimodal simulation lab like StreetLab. For controlled experiments examining performance under realistic noisy conditions, sound levels should be matched to those of the naturalistic conditions of interest. For example, with the simulated traffic noise turned on, the sound level in StreetLab (60 dB A) was still significantly lower than real world background sound levels under city traffic conditions (approximately 80 dB A, [39]), but could be systematically increased as appropriate if the purpose of the test was to evaluate performance in higher levels of traffic noise as might be encountered in the real world. Thus, it would be feasible and justifiable to test behavior during non-quiet conditions in StreetLab.

**SNR**

The lower background sound levels in the soundbooth resulted in a higher SNR compared to the SNRs found in StreetLab across all conditions with the various types of equipment and devices operating and/or with street sounds. By intentionally increasing the level of the sentence stimuli from 66 to 76 dB A (typical of increasing speech from a raised to a loud voice; [40], pg. 35), we were able to increase the SNR to 25 dB in StreetLab. This SNR is not as large as the SNR measured within the soundbooth (37 dB), but it is an SNR at which speech intelligibility would be very high for most people (SII = .96). Importantly, when using a VR lab such as StreetLab, the presentation level of target sounds such as speech may need to be adjusted to compensate for the additional extraneous noise introduced by equipment (e.g., interactive devices such as the treadmill included here). In general, appropriate adjustment of the SNR necessitates the acoustical measurement of both the intended experimental target stimuli, as well as the intended and unintended background sound levels. Calibration based on room properties would be warranted given the differences in the frequency responses of the speech and the noise in the two test environments. Furthermore, the absolute and relative levels of the target and background sounds may need to be adjusted based on audiometric thresholds when participants with hearing loss are tested (see [41]). While these considerations would be obvious and intuitive to most hearing researchers, they may not be commonly considered by many researchers who are using simulated auditory scenes, but who are not expert in acoustics.

## 4.2 Factors to consider in developing more naturalistic multimodal testing protocols

The present study demonstrates the need to describe the acoustical properties of test environments and take them into account when designing studies and comparing results across studies in which testing environments differ along a continuum that varies from the highly artificial, controlled and standardized environment of the soundbooth to more realistic and less standardized VR environments and natural environments. While there are standards for audiometric testing and soundbooths used for audiometry, no such standards exist for VR environments or when testing is conducted in more natural environments. It is encouraging that some researchers are starting to develop well-documented, calibrated naturalistic auditory stimuli for auditory research using naturalistic background sounds (e.g., ICRA Natural Sound Library; https://tspace.library.utoronto.ca/handle/1807/66299). As the use of VR becomes more widespread in hearing research, aging research, and beyond, standards should evolve to characterize the properties of testing environments in a systematic manner.

When developing multimodal VR protocols, appropriate baseline perceptual tasks should be incorporated because individuals and groups of listeners (e.g., younger vs. older listeners; listeners with normal hearing vs. listeners who are hard of hearing) may differ in their SNR thresholds or in their auditory processing abilities that are critical for listening in reverberant or complex scenes (for a review see [42]). The presentation of stimuli can then be adjusted according to the properties of the environment in relation to the abilities of the participants if the research question depends on equating the difficulty of listening conditions for all participants. This approach avoids the risks associated with making assumptions about equivalence across environmental conditions and participants without explicitly addressing these factors (e.g., to isolate differences due to age from differences due to hearing loss). A better understanding of the effects of environmental factors on performance and their interactions with individual factors such as age and hearing loss is needed to advance theories pertaining to how people listen in adverse environments (e.g., [8]). Such knowledge of the effects of environmental factors could also be applied by rehabilitative audiologists in training clients on how to reduce listening effort, as well as to improve architectural and engineering designs of communication spaces and technologies for special subpopulations of listeners [10].

## 4.3 Implications for hearing rehabilitation

In the real world, not only are there multiple and changing visual and auditory environmental inputs, but people are typically dynamic (standing, walking, reaching, turning, etc.) and are performing more than one task at any given time (e.g., listening, talking, walking and remembering past experiences, planning what to do next). Compared to testing in a typical soundbooth, conducting studies using a multimodal VR system provides a controlled, yet more proximate estimate of how these factors might be associated with real-world performance. To fully take advantage of the possibility of manipulating aspects of the VR environment that could not be controlled in real-world testing conditions, procedures for measuring and adjusting for the acoustical properties of VR test environments need to be developed.

In the context of hearing rehabilitation, when evaluating the effectiveness of hearing aid technologies, a disconnect has been reported between the benefits observed during laboratory testing (e.g., measuring word-recognition accuracy on speech-in-noise tests administered in soundbooths) and self-reported benefit and satisfaction in everyday usage of hearing aids [43]. It is possible that results obtained in soundbooths may lead to overestimations or misinterpretations of the benefits associated with using hearing aids in the real world because in everyday life observers perform activities with much more complex auditory stimulation, additional sensory stimulation (e.g., visual, tactile, kinesthetic), and with varying cognitive task demands. Therefore, while improvements associated with hearing aid technologies may be observed in highly controlled, but artificial lab environments, the magnitude of this advantage may not necessarily generalize to functioning in everyday life.

In the real world, listeners use their hearing for purposes other than understanding speech. Auditory abilities (e.g., localization) can support mobility and navigation [48]. Indeed, individuals with hearing loss have identified their most commonly reported limitations to be related to “mobility” and “agility” (65%) compared to communication (12%), memory (12%), or learning (11%) [44]. Introducing controlled VR testing conditions that also allow for common, mobility-related tasks to be conducted (e.g., walking), may provide additional insights into the effects of hearing loss and the benefits of hearing aids in more realistic and demanding conditions compared to testing conditions that are limited to standing or sitting in place. There are also many types of hearing aid technologies that introduce a variety of signal processing options and control features (e.g., directional microphones, multichannel compression, noise reduction, bilateral information exchange, etc.). Each of these variations may differentially benefit everyday, real-world behaviors in unique ways. For example, it is possible that specific features of hearing aids (e.g., microphone directionality) may work particularly well when having a conversation with a dinner partner in a noisy restaurant, but may not be as useful (or may possibly be detrimental) when navigating a busy intersection in heavy traffic. Comparing performance with hearing aid technologies across a range of challenging and realistic conditions simulated by VR could provide a richer understanding of their advantages and limitations.

The effects of age-related changes in auditory processing can also be compounded by an increased prevalence of other sensory, motor and cognitive declines that affect older adults. Specifically, much more needs to be learned about how age-related auditory declines interact with age-related declines in other domains of functioning (e.g., vision, mobility, cognition), especially when complex and cognitively demanding tasks are performed in realistic conditions that are often unfavorable, if not adverse. Introducing novel VR methods that allow for the systematic manipulation of sensory inputs and the strategic modification of perceptual and cognitive demands can help to further our understanding of how these factors interact with age. Immersive, multisensory, VR technologies show great promise in addressing these gaps.

VR may also provide an opportunity to improve or extend the possible range of outcome measures. The outcome measures most commonly used to assess auditory abilities and hearing aid effectiveness in the soundbooth are not necessarily the same outcome measures that would be most relevant for everyday listening. Many older adults who have normal or near-normal audiograms have little difficulty in ideal listening conditions. Amplification can restore the audibility of speech for those who have hearing loss. Nevertheless, older adults, regardless of their audiometric thresholds, report poorer functioning in everyday listening conditions than younger adults [45]. Some of the variation across individuals in speech understanding in noise can be explained by measures of supra-threshold auditory temporal processing and cognitive processing [46]. Furthermore, once speech or other sounds have been heard, cognitive resources are required for the person to comprehend, evaluate, remember and respond appropriately to sound input and to integrate it with other incoming signals and stored knowledge. Cognitive measures such as working memory can be used to assess inter-individual differences in the cognitive capacity deployed in specific listening situations and to assess intra-individual differences in the allocation of cognitive capacity in response to varying demands across changing listening environments [47]. The growing interest in how to conceptualize and measure listening effort and aspects of auditory cognition (memory and attention) reflects recognition by audiologists that both auditory and cognitive processing contribute to everyday listening experiences [10]. As behavioral and physiological measures of listening effort continue to be developed, the complex and demanding conditions that can be simulated using VR may be extremely useful for the evaluation of performance across a range of conditions more representative of those encountered in everyday life.

## Acknowledgments

We thank Bruce Haycock, Susan Gorski, and Larry Crichlow for their technical assistance and Jessica Babineau for constructing the historical timeline of auditory VR.

## References

[1] Campos JL, Bülthoff HH. Multisensory integration during self-motion. In: Ed. Wallace M, Murray M, *Front Neural Bases Multisensory Processes,* pages 603-628. London: Taylor and Francis, 2012.

[2] Rizzo A, Parsons TD, Buckwalter, J. Using virtual reality for clinical assessment and intervention. In *Handbook of Technology in Psychology, Psychiatry, and Neurology: Theory, Research, and Practice,* pages 277-318. Nova Science Publishers Inc, 2012.

[3] Hale KS, Stanney KM. *Handbook of Virtual Environments: Design, Implementation, and Applications*, *Second Edition*. CRC Press, Taylor and Francis, Florida, 2015.

[4] Parsons TD. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front Hum Neuro.* 2015:11(9):660.

[5] Shilling RD, Shinn-Cunningham BG. Virtual auditory displays. In *Handbook of Virtual Environments* Technology, K. Stanney (ed.), pages 65-92. Erlbaum, 2002.

[6] Vorlander M, Shinn-Cunningham BG. Virtual auditory displays. In *Handbook of Virtual Environments, 2nd edition: Design, Implementation, and Applications*, K Hale, K Stanney (Eds), pages 87-114. Erlbaum, 2012.

[7] Carlile S, Leung J. The perception of auditory motion. *Trends Hear*. 2016;20:1-19.

[8] Mattys SL, Davis MH, Bradlow AR et al. Speech recognition in adverse conditions: A review. *Lang Cogn Process*. 2012;27(7–8):953–978.

[9] Keidser G. Introduction to special issue: Towards ecologically valid protocols for the assessment of hearing and hearing devices. *J Am Acad Audiol.* 2016;27(7):502-503.

[10] Pichora-Fuller MK, Kramer SE, Eckert M et al. Hearing impairment and cognitive energy: The Framework for Understanding Effortful Listening (FUEL), *Ear Hear.* 2016;37(Suppl 1):5S-27S.

[11] Pichora-Fuller MK, Hodgson M, Schneider BA et al. Potential applications of auralization in audiological rehabilitation. *Proc Int Symp Simulation, Visualization and Auralization for Acoustic Research and Education*, pages 393-394. Tokyo, Japan, 1997.

[12] Seeber BU, Kerber S, Hafter ER. A system to simulate and reproduce audio-visual environments for spatial hearing research. *Hear Res.* 2010;260(1-2):1-10.

[13] Mihelj M, Novak D, Begus S. Acoustic modality in virtual reality. In *Virtual Reality Technology and Applications*, *pages* 131-159. Springer, Netherlands, 2014.

[14] Punch JL, Robb R, Shovels, AH. Aided hearing aid user preferences in laboratory versus real-world environments. *Ear Hear.* 1994;15:50-61.

[15] Oreinos C, Buchholz JM. Evaluation of loudspeaker-based virtual sound environments for testing directional hearing aids. *J Amer Acad Audiol.* 2016;7:541-556.

[16] Wu Y-H, Nazan A, Rizzo M et al. (2014). Measuring listening effort: driving simulator vs. simple dual-task paradigm. *Ear Hear.* 2014;35(6):623-632.

[17] Lau ST, Pichora-Fuller MK, Li KZH et al. Effects of hearing loss on dual-task performance in an audiovisual virtual reality simulation of listening while walking. *J Amer Acad Audiol.* 2016;27(7):567-587.

[18] Ramkhalawansingh R, Keshavarz B, Haycock B et al.. Age differences in visual-auditory self-motion perception during a simulated driving task. *Front Psych.* 2016;7(595):1-12.

[19] Campos JL, Bédard M, Classen S, Delparte JJ, Hebert DA, Hyde N, Law G, Naglie G, and Yung S. Guiding framework for driver assessment using driving simulators. *Front Psychol.* 2017;8: 1428,

[20] Lin FR, Ferrucci L. Hearing loss and falls among older adults in the United States. *Arch Intern Med.* 2012;172(4): 369-371.

[21] Hickson L, Wood J, Chaparro A et al. Hearing impairment affects older people's ability to drive in the presence of distractors. *J Ameri Geriatrics Soc.* 2010;58: 1097-1103.

[22] Singh G, Pichora-Fuller MK, Schneider BA. Auditory spatial attention in conditions of real and simulated spatial separation by younger and older adults. *J Acous Soc Amer.* 2008;124:1294-1305.

[23] Nieborowska V, Lau ST, Campos JL et al. Effects of age on dual-task walking while listening. *J Mot Beh.* (in press) .

[24] Cameron S, Dillon H. Development of the Listening in Spatialized Noise-Sentences Test (LISN-S). *Ear Hear*. 2007:28;196–211.

[25] Cameron S, Dillon H. *Listening in Spatialized Noise-Sentences test (LiSN-S)*, Murten, Switzerland: Phonak Communications AG, 2009.

[26] Albers MW et al. At the interface of sensory and motor dysfunctions and Alzheimer's disease. *Alzhem Demen*. 2015; 11:70-98.

[27] Laurienti PJ, Burdette JH, Maldjian JA et al. Enhanced multisensory integration in older adults. *Neurobio Aging*. 2006;27(8):1155–1163.

[28] Mozolic JL, Hugenschmidt CE, Peiffer AM et al. Multisensory integration and aging. In Murray MM & Wallace MR (Eds), *The Neural Bases of Multisensory Processes* (Chapter 20). Boca Raton (FL): CRC Press, 2012.

[29] Freiherr J, Lundström JN, Habel U et al. Multisensory integration mechanisms during aging. *Front Hum Neuro*. 2013;7:863.

[30] Weeks JC, Hasher L. The disruptive - and beneficial - effects of distraction on older adults' cognitive performance. *Front Psych*. 2014;5:133.

[31] Frank T. ANSI update: maximum permissible ambient noise levels for audiometric test rooms. *Am J Audiol.* 2000:9(1);3-8.

[32] Bolia RS, Nelson WT, Ericson MA et al. A speech corpus for multitalker communications research. *J Acous Soc Amer.* 2000;107:1065-1066.

[33] American National Standards Institute. Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools, Part 1: Permanent Schools (ANSI 12.60 I). New York, 2010.

[34] Bradley JS. Acoustical design of rooms for speech. Construction Technology Update No. 51, Institute for Research in Construction, National Research Council of Canada: Ottawa, 2002.

[35] Beranek LL, Blazier WE, Figwer JJ. Preferred Noise Criterion (PNC) curves and their application to rooms. *J Acous Soc Amer.* 1971;50:1223-1228.

[36] American National Standards Institute. Standard Methods for Calculation of the Speech Intelligibility Index (ANSI/ASA S3.5 (R2012)). New York, 2012.

[37] American National Standards Institute. Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms (ANSI S3.1). New York, 1999.

[38] Helfer KS, Wilber LA. Hearing loss, aging, and speech perception in reverberation and noise. *J Speech Hear Res.* 1990;33(1);149-155.

[39] National Institute on Deafness and other Communication Disorders http://www.nidcd.nih.gov/staticresources/health/education/teachers/CommonSounds.pdf

[40] Pearson KS, Bennett RL, Fidell S. Speech levels in various noise environments. Environmental Health Effects Research Series. EPA-600/1-77-025. Report from the Office of Health and Ecological Effects, U.S. Environmental Protection Agency, Washington, 1977.

[41] Humes LE, Dubno JR. Factors affecting speech understanding in older adults. In Gordon-Salant S, Frisina RD, Popper AN and Fay RR (Eds), Aging Audit Syst., *Springer Handbook of Auditory Research*, pages 211–257. Springer, New York, 2010.

[42] Pichora-Fuller MK, Alain C, Schneider B. Older adults at the cocktail party, pages 227-259. In Middlebrooks J, Simon J, Popper A, Fay RR (Eds), *The auditory system at the cocktail party*, Springer Handbook of Auditory Research. Springer: Berlin, 2017.

[43] Gnewikow D, Ricketts T, Bratt GW et al. Real-world benefit from directional microphone hearing aids. *J Rehab Res and Devlop.* 2009;46:603-618.

[44] Statistics Canada. Participation and activity limitation survey: Facts on hearing limitations. 2006.

[45] Banh J, Singh G, Pichora-Fuller MK. Age affects responses on the Speech, Spatial, and Qualities of Hearing Scale (SSQ) for adults with minimal audiometric loss. *J Amer Acad Audiol.* 2012;23: 81-91.

[46] Humes LE. The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *J Am Acad Audiol.* 2007;18(7):590-603.

[47] Smith S, Pichora-Fuller MK, Alexander G. Development of the Word Auditory Recognition and Recall Measure (WARRM): A working memory test for use in rehabilitative audiology. *Ear Hear*. 2006;37(6):e360-e376.

[48] Campos JL, Ramkhalawansingh R, Pichora-Fuller MK. Hearing, self-motion perception, mobility and aging. *Hear Res.* 2018. https://doi.org/10.1016/j.heares.2018.03.025


WE CAN MAKE WHATEVER YOU MAKE
BETTER.
SAFER.
QUIETER.


EFFECTIVE
SOUND
MASKING