

Published in final edited form as:

Ear Hear. 2014 ; 35(6): 623–632. doi:10.1097/AUD.000000000000079.

Measuring listening effort: driving simulator vs. simple dual-task paradigm

Yu-Hsiang Wu¹, Nazan Aksan², Matthew Rizzo², Elizabeth Stangl¹, Xuyang Zhang¹, and Ruth Bentler¹

¹Department of Communication Sciences and Disorders, The University of Iowa

²Department of Neurology, The University of Iowa

Abstract

Objectives—The dual-task paradigm has been widely used to measure listening effort. The primary objectives of the study were to (1) investigate the effect of hearing aid amplification and a hearing aid directional technology on listening effort measured by a complicated, more real world dual-task paradigm, and (2) compare the results obtained with this paradigm to a simpler laboratory-style dual-task paradigm.

Design—The listening effort of adults with hearing impairment was measured using two dual-task paradigms, wherein participants performed a speech recognition task simultaneously with either a driving task in a simulator or a visual reaction-time task in a sound-treated booth. The speech materials and road noises for the speech recognition task were recorded in a van traveling on the highway in three hearing aid conditions: unaided, aided with omni directional processing (OMNI), and aided with directional processing (DIR). The change in the driving task or the visual reaction-time task performance across the conditions quantified the change in listening effort.

Results—Compared to the driving-only condition, driving performance declined significantly with the addition of the speech recognition task. Although the speech recognition score was higher in the OMNI and DIR conditions than in the unaided condition, driving performance was similar across these three conditions, suggesting that listening effort was not affected by amplification and directional processing. Results from the simple dual-task paradigm showed a similar trend: hearing aid technologies improved speech recognition performance, but did not affect performance in the visual reaction-time task (i.e., reduce listening effort). The correlation between listening effort measured using the driving paradigm and the visual reaction-time task paradigm was significant. The finding showing that our older (56 to 85 years old) participants' better speech recognition performance did not result in reduced listening effort was not consistent with literature that evaluated younger (approximately 20 years old), normal hearing adults. Because of this, a follow-up study was conducted. In the follow-up study, the visual reaction-time dual-task experiment using the same speech materials and road noises was repeated on younger adults with normal hearing. Contrary to findings with older participants, the results indicated that the directional technology significantly improved performance in both speech recognition and visual reaction-time tasks.

Conclusions—Adding a speech listening task to driving undermined driving performance. Hearing aid technologies significantly improved speech recognition while driving, but did not significantly reduce listening effort. Listening effort measured by dual-task experiments using a simulated real-world driving task and a conventional laboratory-style task was generally consistent. For a given listening environment, the benefit of hearing aid technologies on listening effort measured from younger adults with normal hearing may not be fully translated to older listeners with hearing impairment.

Keywords

listening effort; hearing aid; driving; dual-task paradigm; directional processing

INTRODUCTION

Understanding speech involves not only auditory but also cognitive factors (e.g., Kiessling et al. 2003; Worrall & Hickson 2003; Pichora-Fuller & Singh 2006; Humes 2007). When speech signals are degraded (e.g., by noise), listeners engage “top-down” cognitive processing to understand speech. As a result, even though the level of speech understanding can be maintained, listening will be effortful. Although the research community has not reached a consensus regarding the definition of listening effort, it has been suggested that listening effort could be conceptualized as the cognitive resources allocated for speech processing (Hick & Tharpe 2002; Fraser et al. 2010; Gosselin & Gagné 2010; Zekveld et al. 2010).

Different methodologies have been used to quantify listening effort, including subjective measures (e.g., Gatehouse & Noble 2004), physiological measures (e.g., Zekveld et al. 2010), and behavioral measures (e.g., Sarampalis et al. 2009). Among these, the dual-task paradigm is one of the most widely-used behavioral measures (Gosselin & Gagné 2010). In this paradigm, listeners perform a primary speech recognition task while executing a secondary task. Speech recognition is the primary task because listeners are instructed to maximize speech recognition performance. Secondary tasks include visual reaction-time (e.g., watching for a visual stimulus and pushing a button) or recall (e.g., memorizing heard speech) tasks. Primary task difficulty is systematically varied (e.g., sentence recognition at different noise levels). Change in secondary task performance at different primary task difficulties reflects a shift in cognitive resources for speech processing, i.e., listening effort. This interpretation assumes that: (1) performance on each task requires allocation of some common cognitive resources to each task, and (2) cognitive resources are limited.

Dual-task paradigms have been used to investigate the effect of age (Larsby et al. 2005; Gosselin & Gagné 2011; Desjardins & Doherty 2013), hearing loss (Hick & Tharpe 2002; Larsby et al. 2005), visual cues (Larsby et al. 2005; Fraser et al. 2010; Picou et al. 2013), hearing aid amplification (Downs 1982; Gatehouse & Gordon 1990; Hällgren et al. 2005; Hornsby 2013; Picou et al. 2013), and noise reduction algorithms (Sarampalis et al. 2009; Pittman 2011; Ng et al. 2013; Desjardins & Doherty, 2014) on listening effort. For example, Picou et al. (2013) used a visual reaction-time task to investigate the effect of hearing aid amplification on listening effort. The results indicated that listeners’ reaction time in the

aided condition was shorter than in the unaided condition by 9 msec, suggesting that hearing aid use could reduce listening effort. Sarampalis et al. (2009) investigated the effect of noise reduction algorithms on listening effort. With the algorithm enabled, listeners responded approximately 50 msec faster to the visual stimuli, suggesting that the noise reduction algorithms freed up cognitive resources for speech processing and reduced listening effort.

Although research using dual-task paradigms in laboratories has indicated that hearing aid technologies could release cognitive resources from speech processing, the clinical meaning of dual-task paradigm results is less clear. For example, it is unknown if the 9-msec benefit reported by Picou et al. (2013) or the 50-msec benefit reported by Sarampalis et al. (2009) is perceived as meaningful by hearing aid users in the real world. Determining the clinical impact of these data is difficult because listening effort is not directly measured in the dual-task paradigm; but is inferred indirectly from change in secondary task performance.

The clinical significance of dual-task paradigm results can be examined by comparing laboratory and real-world data. Hornsby (2013) investigated the effects of hearing aid amplification and advanced hearing aid features on listening effort and mental fatigue (mental fatigue can be conceptualized as the negative consequence of increased listening effort; see Hornsby 2013 for details). Participants with hearing impairment acclimated to a given hearing aid condition (unaided or aided with basic or advanced features) for one to two weeks in the field. Participants were then tested in the laboratory using the dual-task paradigm at the end of their work day. Participants also estimated listening effort and mental fatigue that they experienced during the work day using a self-report inventory. Laboratory dual-task experiment results suggested that hearing aid amplification could reduce listening effort, yet participants did not report a significant difference in listening effort during daily activities across the three hearing aid conditions. Also, laboratory objective measures of mental fatigue did not correlate with listening effort and fatigue experienced during the day.

This mismatch between laboratory and real-world listening effort could be due to the nature of the measure. The dual-task paradigm is an objective measure while real-world listening effort is often estimated using subjective (self-report) inventories. A large body of literature has shown the weak link between subjective and objective measures of listening effort, suggesting that they might be measuring different aspects (Larsby et al. 2005; Fraser et al. 2010; Zekveld et al. 2010).

Another factor in the mismatch might reflect the ecological validity of the dual-task paradigm. Real world listeners often multitask, e.g., listening while writing or walking. However, most secondary tasks used in the dual-task paradigm (such as pushing a button when seeing a light) are uncommon in the real world and lack natural complexity and context, i.e., less ecologically valid. Since different tasks could tax various cognitive resources and/or require different sensory/motor systems, it is unknown whether listening effort measured using less ecologically valid dual-task experiments are generalizable to the real world. Based on the assumption that more ecologically valid laboratory testing would better predict real-world results, for decades researchers have used speech recognition tests that imitate real-world listening tasks (e.g., audio-visual listening tests) and simulated real-world environments (e.g., simulated classrooms) to examine the benefit of hearing aid

technologies. To date, listening effort has not been measured using more ecologically valid dual-task paradigms.

The specific objectives of this study were to (1) investigate the effect of hearing aid amplification and an advanced hearing aid feature on listening effort, measured using a dual-task paradigm that was comprised of a more natural task, and (2) compare those results to a simpler, laboratory-style dual-task paradigm. The more natural task used in the study was driving. Listening while driving is a common dual-task scenario in the real world. Wu and Bentler(2012) found that the duration of listening activities involving automobile/traffic was approximately 1 hour per day. Survey data indicate approximately forty percent of drivers reported talking on phones at least a few times per week (Braitman & McCartt 2010). It has been well established that conversations during driving result in longer reaction times in braking response or more misses in fast reaction response (Strayer & Johnston 2001; Consiglio et al. 2003; McCarley et al. 2004; Morel et al. 2005). Hearing aid technologies can make speech understanding easier and less effortful and, therefore, have the potential to improve driving performance.

This paper includes three experiments. Experiment 1 examined the effect of hearing aid amplification and a directional technology on listening effort using a driving simulator. Experiment 2 investigated the relationship between listening effort measured using the driving simulator and a dual-task paradigm that was comprised of a simple visual reaction-time task. Older adults with hearing impairment were tested in these two experiments. Because the results of Experiment 2 were not quite consistent with the literature, younger adults with normal hearing were tested in Experiment 3 to verify the methodologies of Experiment 2.

EXPERIMENT 1

The objective of this experiment was to investigate the effect of hearing aid amplification and directional technology on listening effort for older adults with hearing impairment using a driving simulator.

Materials and Methods

Participants—Thirty-four adults were recruited. Participants were eligible for inclusion in this study if their hearing loss met the following criteria: (1) postlingual bilateral downward-sloping sensorineural hearing loss (air-bone gap < 10 dB); (2) hearing thresholds no better than 20 dB HL at 500 Hz, and no worse than 85 dB HL at 3 kHz (ANSI 2010); and (3) hearing symmetry within 15 dB for all test frequencies. Participants were current drivers with valid driver's licenses. Participants were considered current drivers if they drove a car within the last 12 months and would drive that day, if needed. Among the 34 participants, 5 participants did not complete the study because of "simulator sickness." The simulator sickness was due to a mismatch between visual cues of movement (which were plentiful) and inertial cues (which were lacking) in the simulator. For the 29 participants (13 males, 16 females) who completed the experiment, ages ranged from 56 to 85 years with a mean of 72.7 years ($SD = 7.9$). The mean pure tone thresholds are shown in Figure 1 as gray circles. Of the 29 participants, 25 were experienced hearing aid users. Participants were considered

experienced users if they used hearing aids no less than 6 hours per day for at least six months during the past 12 months. The processing scheme/gain/output of the user's previous hearing aids was not tested or considered a factor in this investigation because the effect of acclimatization on sentence recognition for listeners with mild-to-moderate hearing loss is minimal (Saunders & Cienkowski 1997; Bentler et al. 1999). The participants were asked to rate the quality of their own driving using a 5-point scale (1: excellent; 2: good; 3: average; 4: fair; 5: poor) and estimated the miles that they drove in an average seven-day week. The reported driving quality was good (mean = 2.2, SD = 0.6) and on average they drove 92.5 miles per week (SD = 107.5).

As no previous research was available to estimate the effect size of hearing aid technologies on driving performance, the sample size of Experiment 1 was determined based on the literature in driving research. The literature indicated that, with 10 to 30 participants in each subject group, the driving task used in Experiment 1 (see below) was able to detect the between-subject effect of age (Ni et al. 2010), poor visibility due to fog (Broughton et al. 2007), psychoactive substance (Dastrup et al. 2010), or mobile phone use (Alm & Nilsson 1995) on driving performance. Therefore, it was estimated that, if the effect size of hearing aid technologies on driving performance was not smaller than that of the factors listed above, a sample size of 29 participants would have enough power to detect the within-subject effect of hearing aid technologies.

Speech recognition task—In order to increase the ecological validity of the dual-task experiment, the speech materials and automobile/road noises were recorded in a van moving on a highway. The detailed methodologies used to record and process the stimuli were described in the study by Wu et al. (2013b). This section gives a summary of the methodologies.

A pair of prototype Siemens Pure hearing aids was used to record the stimuli. This model was chosen because it is equipped with a backward-facing directional microphone system, which employed an anti-cardioid directivity pattern to enhance speech coming from behind the listener. See Wu et al. (2013a) for more details of this technology. The hearing aids were mini behind-the-ear aids with receiver-in-the-canal and dome tips. This style was chosen because of its frequent use by individuals with mild-to-moderate sloping hearing loss.

To increase ecological validity, ideally the speech stimuli would have been recorded for each participant separately using hearing aid settings fitted to their hearing loss. However, doing so would have required recording the stimuli on the highway multiple times. This would result in a large variation in the noise level because the amount of traffic on the highway changes considerably from time to time. Therefore, a different approach was used to record speech stimuli. The hearing aids were programmed to fit a pre-determined bilateral, symmetrical sloping hearing loss (thresholds 25, 30, 35, 45, 65, and 65 dB HL for octave frequencies from 250 Hz to 8000 Hz). These were the average participant hearing thresholds from several of our previous studies with identical inclusion criteria. The recorded stimuli were then processed to compensate for each participant's hearing loss. Using a pre-determined hearing loss to fit hearing aids allowed us to start the data collection before recruiting all participants. The hearing aids were programmed to have two programs

with different settings in each. In the first program the omnidirectional microphone mode (referred to as the OMNI condition) was used. Backward-facing directional processing was enabled in the second program (the DIR condition).

The hearing aids were then coupled to the ears of a manikin (Knowles Electronics Manikin for Acoustic Research; KEMAR) using closed domes. Closed domes were chosen to ensure adequate low frequency gain for the given hearing loss. In situ responses were measured using a probe-microphone hearing aid analyzer (Audioscan Verifit) with a 65 dB SPL speech signal (the “carrot passage”) presented from the manikin’s 0° azimuth. The gains of each program were adjusted to produce real-ear aided responses equivalent (± 3 dB, from 250 to 6000 Hz) to that prescribed by the NAL-NL1 (National Acoustics Laboratory - Nonlinear version 1; Dillon 1999). For both programs, the compression processing, the maximum output limits, and the feedback suppression system were set to the fitting software’s default without further modification. The noise reduction algorithms were deactivated.

The automobile/road noises and the Connected Speech Test (CST; Cox et al, 1987; Cox et al, 1988) sentences were recorded bilaterally with and without hearing aids coupled to KEMAR’s ears using closed domes, while KEMAR was in the passenger seat of a 2009 Ford 150E van. To simulate the conversation between the driver and a talker sitting in the back seat, a loudspeaker was placed 180° azimuth relative to KEMAR to present the speech stimuli. The distance between the center of the manikin’s head and the loudspeaker was 50 cm. The CST was chosen because it was designed to approximate everyday conversations.

The recordings were made on Iowa Highway I-80 between Exit 249 to 284. The van’s speed was set to 70 miles/hour. Before recording, the speech and noise levels were calibrated. Specifically, while the van drove on the highway, the noise level was measured from both of KEMAR’s ears without hearing aids. The noise levels at the right and left ears were approximately 75 and 78 dBA, respectively. After the noise level had been determined, the speech level was set to achieve a -1 dB signal-to-noise ratio (SNR) at the ear with the better SNR (right ear). This SNR was chosen in accordance with the data of Pearsons et al. (1976), which indicated that the typical SNR in 75-dBA environments (e.g., trains and aircraft) was -1 dB. See Wu et al.(2013b) for the spectra of the recorded speech and noise.

After calibration, the van was driven on the highway to record the stimuli. All of the CST sentences were recorded in three conditions: Unaided condition (without hearing aids in KEMAR’s ears), aided OMNI condition, and aided DIR condition. In an effort to make the noise level equal across conditions, the driver kept a constant speed using cruise control and stayed in the right lane whenever possible. The same section of the road was used across conditions such that, between Exit 249 and 284, the recording was always made when driving eastbound. In addition to the speech-plus-noise stimuli recording, a noise-only recording (i.e., no speech) was made in the Unaided condition.

While recording, the speech signals were presented from a computer with an M-Audio ProFire 610 sound interface, routed via a Stewart Audio PA-100B amplifier, and then presented from a Tannoy i5 AW loudspeaker, which was located 180° azimuth relative to KEMAR. The hearing aids’ outputs were routed through a G.R.A.S. Ear Simulator Type

RA0045, a Type 26AC preamplifier, a Type AR12 power module, and the M-Audio ProFire 610 sound interface, to the computer. The recorded signals were digitized at a 44.1 kHz sampling rate and 16-bit resolution.

The speech recognition stimuli were then processed for laboratory testing. Before presenting the stimuli to participants in the driving simulator through earphones, the stimuli were first processed to eliminate the effect of the playback system and earphones to ensure that the sound levels and spectra heard by the listeners were identical to what occurred in the manikin's ear in the van. To do so, the stimuli were played to KEMAR using the playback system and earphones, which would be used in the experiment, and were re-recorded from KEMAR's ears using the same recording equipment used in the van. The levels and spectra of the re-recorded stimuli were then compared to those of the original recorded stimuli. The difference between the two recordings was used to design an inverse filter. The filter was then applied to the original recorded stimuli to make the playback system acoustically transparent.

The stimuli were further processed to compensate for each participant's hearing loss. For the 65 dB SPL speech input, the NAL-NL1 targets generated by the AudioscanVerifit for an individual's hearing loss and the targets for the sloping hearing loss used to program the hearing aid were compared. The differences were then used to create a filter, one for each ear, to shape the spectra of the stimuli such that the outputs of the hearing aids met an individual's NAL-NL1 targets within ± 3 dB from 250 to 6000 Hz for a 65 dB SPL speech input.

During the dual-task experiment, the speech-plus-noise stimuli were generated by a laptop computer and a Creative Labs Extigy sound interface, and then presented through a pair of Sennheiser IE8 insert earphones. The listener's task was to repeat the CST sentences. One pair of CST passages was presented and scored in each condition. Scoring was based on the number of key words repeated by the listener out of 25 key words per passage, totaling 50 key words per condition.

Driving task—Driving performance was assessed using SIREN (Simulator for Research in Ergonomics and Neuroscience), which is a fixed-based, interactive driving simulator (Rizzo et al. 2005a; Rizzo et al. 2005b). SIREN is comprised of a 1994 GM Saturn with the running gear removed, embedded electronic sensors, and miniature cameras for recording driver performance, four LCD projectors with image generators, and computers for scenario design, control, and data collection. SIREN can generate a range of driving environments including roadway types and grades, traffic, signal controls, and light and weather conditions, for differing clinical and experimental needs.

A car-following scenario was used in the experiment. This task was chosen because (1) it is a sensitive task (Dastrup et al. 2010), (2) the skills required to successfully complete this task are essential for driver safety and to avoid rear-end collisions, and (3) the scenario of this task is similar to the highway traffic that was used to record the speech recognition stimuli. In this task, the driver was asked to follow a lead vehicle at a distance of two-car lengths without crashing. The velocity of the lead vehicle was programmed to vary across

time. The scenario started with the lead vehicle 18 m in front of the participant vehicle. As the participant accelerated to 55 miles/hour, the lead vehicle started moving. After 500 m, the lead vehicle began to vary its velocity.

Driving performance was quantified as the distance between the lead vehicle and participant vehicle (i.e., the following distance). The sampling rate was 60 Hz. Three metrics were derived based on the following distance: mean, SD, and the interquartile range (IQR) of the following distance. In general, the shorter the mean following distance and the smaller the SD and IQR of the following distance, the better the driving performance.

Procedures—The study was approved by the Institutional Review Board of the University of Iowa. After agreeing to participate in the study and signing the consent form, participants' pure tone thresholds were measured and the auditory speech-plus-noise stimuli were processed to compensate individual hearing loss. The participant then sat in the driver's seat of SIREN. The auditory stimuli were presented using earphones. The participants were asked about their perception of the loudness of the stimuli. All participants reported that the loudness was appropriate. Participants were instructed to follow the lead vehicle and maintain a two-car length distance, while repeating the sentences. Note that for most dual-task measures of listening effort, listeners are instructed to give the priority to speech recognition performance (i.e., the primary task). However, in the real world it is not reasonable for drivers to treat driving as the secondary task. Therefore, participants were instructed to do both the speech recognition and the driving tasks simultaneously, paying equal amounts of attention to the two in order to complete them to the best of their abilities.

Each participant was tested in seven conditions. Three of them were dual-task conditions wherein participants drove the simulator and listened to/repeated the sentences recorded using different hearing aid settings (Unaided, OMNI, and DIR). Participants also completed a baseline condition wherein they drove the simulator listening to road noise without the speech (the Baseline condition). In this condition, the auditory stimuli were the noise-only stimuli recording. The remaining three conditions were single-task conditions wherein listeners sat in the simulator and repeated the sentences recorded in different hearing aid settings (Unaided, OMNI, and DIR) without driving the vehicle. To minimize the possibility of simulator sickness, the driving (dual-task conditions and Baseline) and non-driving conditions (single-task conditions) were interleaved. Within the driving and non-driving conditions, the test order of the condition was randomized. Prior to beginning the experiment, a warm-up and training session was held to familiarize participants with the vehicle controls and speech recognition.

Data analysis—For speech recognition performance, the original CST scores (in percent correct) were first transformed into rationalized arcsine units (RAU) to homogenize the variance (Studebaker 1985). For driving performance, the three metrics (mean, SD, and IQR of the following distance) were combined to a composite score to facilitate data analysis. To derive the composite score, T-score standardization (a type of linear transformation) was carried out for each original metric in each condition. In order to ensure meaningful comparisons of means across conditions using the composite score, the original metric of the Baseline condition was first T-score transformed to have a mean of 50 and a SD of 10. The

transformation of a given hearing aid condition was then completed using the mean and SD of the Baseline condition as references. For example, if in the original metric the mean of the Unaided condition was higher than the mean of the Baseline condition by twice the Baseline SD, the transformation would ensure that the T-score mean of the Unaided condition would be higher than the T-score mean of the Baseline condition by twice the Baseline T-score SD. Finally, for a given condition the T-score averaged across the three driving metrics (mean, SD, and IQR of following distance) served as the composite score of that condition. The arcsine transformed speech recognition score and the driving performance composite score were then analyzed using the repeated measures analyses of variance (ANOVA). Post-hoc analyses were performed using the Holm–Sidak test, adjusting for multiple comparisons.

Results

Figure 2A shows the mean CST scores in the single-task (black bars) and dual-task (gray bars) conditions as a function of hearing aid condition. The two-way (task type: single/dual; hearing aid condition: Unaided/OMNI/DIR) repeated measures ANOVA indicated that both task type ($F_{1, 28} = 4.71$, $p = 0.04$, partial $\eta^2 = 0.14$) and hearing aid condition ($F_{2, 56} = 72.84$, $p < 0.001$, partial $\eta^2 = 0.72$) had a significant effect on CST score. No interaction between task type and hearing aid condition was found ($F_{2, 56} = 0.042$, $p = 0.96$, partial $\eta^2 = 0.002$). Post-hoc analysis indicated that, averaged across single- and dual-tasks, all differences among Unaided, OMNI, and DIR were significant. These results suggested that adding the driving task to speech understanding decreased speech performance, and that hearing aid amplification and the directional technology improved speech performance.

Figure 2B shows the mean driving performance composite score (wider light gray bars) as a function of test condition (Baseline and the three hearing aid conditions). Note that higher scores reflect poorer driving performance. The three original metrics (mean, SD, and IQR of the following distance) are also shown in the graph as narrower bars to illustrate their relationship with the composite score. Two correlation analyses were first conducted to examine the relationship between the Baseline composite score and self-reported (1) quality of driving and (2) distance that participants typically drove in a week. Results indicated that neither of the correlations were significant (quality: $r = 0.085$, $p = 0.67$; distance: $r = -0.056$, $p = 0.78$).

To examine if additional speech recognition task would compromise driving performance, a one-way repeated measures ANOVA was first performed to examine the effect of test condition (Baseline/Unaided/OMNI/DIR) on composite score. The result indicated that test condition had a significant effect ($F_{3, 84} = 3.35$, $p = 0.02$, partial $\eta^2 = 0.11$). The contrast between Baseline and the dual-task conditions (Unaided, OMNI, DIR) were significant ($F_{1, 28} = 9.94$, $p = 0.004$). However, the composite score across the three dual-task conditions (Unaided: 59.0; OMNI: 58.4; DIR: 59.5) did not differ significantly from each other. To reduce variability in the dual-task driving performance associated with the difference in an individual's motor function and driving skills, the participant's composite score of the Baseline condition was subtracted from the score of each dual-task condition. The difference score was then analyzed using the one-way (hearing aid condition: Unaided/

OMNI/DIR) repeated measures ANOVA. The result revealed that hearing aid condition did not have an effect on driving performance ($F_{2, 56} = 0.05$, $p = 0.95$, partial $\eta^2 = 0.002$).

Discussion

Consistent with Wu et al. (2013b), the CST results revealed that listening in a traveling automobile was a very difficult task. At the unfavorable SNR (-1 dB) that occurred in the van, the participants could only understand approximately 42% and 47% of the speech in the Unaided and OMNI conditions, respectively. Therefore, unless talkers constantly raise their voices or drivers turn their heads toward to the talkers, drivers with hearing impairment would have great difficulty following the conversation. Also in line with Wu et al. (2013a), the backward-facing directional technology was able to improve SNR and increase speech recognition performance by approximately 22%. These results supported the feasibility of using directional technologies to provide a better listening experience for hearing aid users in automobiles.

In line with the literature in driving research, the results indicated that adding the speech recognition task to driving compromised driving performance (Figure 2B). Contrary to most previous studies that used the dual-task paradigm to measure listening effort, adding the driving task to the speech recognition task significantly decreased the speech recognition performance (Figure 2A). This result suggests that some cognitive resources used to process speech in the single-task condition were allocated to perform the driving task in the dual-task conditions. Because the decrement in speech recognition performance did not differ across the three hearing aid conditions (i.e., the non-significant interaction between task type and condition), it seems that the amount of cognitive resources shifted from speech recognition task to the driving task were fairly close across the conditions. Therefore, it is unlikely that the poorer speech performance in the dual-task conditions would affect the interpretation of hearing aid effect on listening effort. Accordingly, the finding that driving performance was essentially the same across the three hearing aid conditions was interpreted as hearing aid amplification and directional processing did not reduce listening effort.

This finding was somewhat surprising because the better SNR and speech recognition performance in the DIR condition should result in less effortful listening (relative to the Unaided and OMNI conditions), which would allow drivers to allocate more cognitive resources to driving and therefore, improve driving performance. This negative finding could be due to the complexity of the task of operating a motor vehicle, which requires a driver to constantly maintain adequate awareness, code inputs from central and peripheral vision and the other senses, and to switch attention between onboard and roadway targets and distracters (Rizzo 2011). Therefore, even if hearing aid technologies could make speech understanding easier and free up some cognitive resources, those resources may not be sufficient to improve performance in this challenging task. If the effect of hearing aid technologies was smaller than what we thought, the experiment sample size, which was estimated based on the effect of several factors such as age and mobile phone use, may not have enough power to detect the subtle change in driving performance. However, the results of Experiment 2 (see below) did not support this speculation. A more plausible explanation for the negative effect of hearing aid technologies on driving performance is the SNR used

in the speech recognition task. This issue will be discussed in the section General Discussion.

EXPERIMENT 2

The objective of this experiment was to examine the consistency between listening effort measured using the driving simulator and a dual-task paradigm that used a simple visual reaction-time task as the secondary task. In short, the participants from Experiment 1 were recruited for this experiment. Participants completed the dual-task experiment described by Sarampalis et al. (2009). The results were then compared to the driving simulator data collected in Experiment 1.

Materials and Methods

Participants—Nineteen out of the 29 participants of Experiment 1 responded to recruitment messages and participated in the study. Eight of them were males. Participants' age ranged from 56 to 85 years with a mean of 71.7 years ($SD = 8.2$). The mean pure tone thresholds are shown in Figure 1 as black diamonds. Of the 19 participants, 18 were experienced hearing aid users. Based on (1) the assumption that the directional technology used in this study can improve SNR by 4 dB (Wu et al. 2013a) and (2) the effect size of a 4-dB SNR improvement on the visual reaction-time task performance estimated from the study by Sarampalis et al. (2009), with 19 participants Experiment 2 would have over 90% power to detect the effect of hearing aid technologies on listening effort.

Dual-task paradigm—The dual-task paradigm used in this experiment required participants to perform a speech recognition task and a visual reaction-time task simultaneously. The speech recognition task, auditory stimuli, and the audio playback equipment were identical to those used in Experiment 1. During the visual reaction-time task, two boxes were displayed on a 19 in. computer monitor, which was placed in front of the listener. A digit between one and eight was shown in either one of the two boxes at random intervals. The participants pressed the arrow button on the computer keyboard according to the digit shown on the screen: pressing the arrow that pointed toward the digit if it was even, and pressing the arrow that pointed away from the digit if it was odd. Each digit remained on the screen for a maximum of 2.5 sec or until the participant pressed the arrow keys. The next digit appeared after a random interval, between 0.5 and 2.0 sec, after the previous digit had disappeared. Accuracy and reaction times (RTs) were recorded for each trial. See Sarampalis et al. (2009) for more details.

The dual-task paradigm by Sarampalis et al. (2009) was chosen because this paradigm and the driving dual-task paradigm were similar in three ways. Firstly, both the driving and the visual reaction-time tasks involved listeners' visual and motor systems. Secondly, participants in Experiment 1 were asked to pay an equal amount of attention to the driving and speech recognition tasks. This instruction was consistent with that used by Sarampalis et al. (2009). Thirdly, in Experiment 1 driving performance was measured continuously throughout the experiment, i.e., during the times when drivers were listening to and repeating the speech. Similarly, in the paradigm described by Sarampalis et al. (2009) the

presentations of auditory and visual stimuli were not synchronized, and RTs of all trials were used to quantify performance in the visual reaction-time task.

Procedures—The participants sat in a sound-treated booth. The auditory stimuli were presented using the playback system that was used in Experiment 1 (i.e., via the Sennheiser IE 8 earphones). Prior to the experiment, a warm-up and training session was given to familiarize participants with the tasks. For the dual-task condition, the presentations of auditory and visual stimuli were started at the same time. The timing of the consequential visual stimulus presentation was uncorrelated with the auditory stimuli. The research participants were asked to concentrate equally on both the auditory and the visual reaction-time tasks, in order to complete them to the best of their abilities. One pair of CST sentences were presented and scored in each condition. For the visual reaction-time task, the trial number varied across conditions and participants, depending on how quickly the participants reacted to the stimuli. In general, 50 to 70 trials were completed in a given condition. Identical to Experiment 1, seven conditions were tested for each participant: dual-task Unaided, OMNI, and DIR conditions, single-task (speech recognition only) Unaided, OMNI, and DIR conditions, and the Baseline condition (visual reaction-time task only). Experiment 2 started two months after the completion of Experiment 1.

Data analysis—The CST scores were converted to RAU and analyzed using the methods described in Experiment 1. For the visual reaction-time task, the accuracy of the task was first examined. Across all conditions, the mean accuracy was 97.5% correct (SD = 2%). Since the accuracy was high, the median of the RTs across all trials in a given condition served as the RT of that condition. The RT data were then analyzed using the repeated measures ANOVA. Post-hoc analyses were performed using the Holm–Sidak test, adjusting for multiple comparisons. The relationship between driving performance (Experiment 1) and visual reaction-time task performance (this experiment) was examined using Pearson’s correlations.

Results

Figure 3A shows the mean single-task (black bars) and dual-task (gray bars) CST scores for each of the hearing aid conditions. The two-way (task type: single/dual; hearing aid condition: Unaided/OMNI/DIR) repeated measures ANOVA indicated that hearing aid condition had a significant effect on CST score ($F_{2, 36} = 51.62, p < 0.001$, partial $\eta^2 = 0.74$). Post-hoc analysis further indicated that, averaged across single- and dual-tasks, all differences among the Unaided, OMNI and DIR conditions were significant. The effect of task type ($F_{1, 18} = 0.69, p = 0.42$; partial $\eta^2 = 0.037$) and the interaction between task type and hearing aid condition ($F_{2, 36} = 0.28, p = 0.76$; partial $\eta^2 = 0.015$) were not significant. These results indicated that hearing aid amplification and directional processing improved speech performance, while the addition of the visual reaction-time task did not affect speech performance.

Figure 3B shows the RT averaged across participants as a function of test condition. Longer RTs represented poorer performance. This figure shows that the RT of the Unaided condition was longer than that of the OMNI condition by 5.2 msec, which is in turn longer

than the RT of the DIR condition by 4.4 msec. This trend was consistent with the literature suggesting that better SNR was associated with better speech intelligibility and less effortful listening (Sarampalis et al. 2009; Zekveld et al. 2010). To take into account individual differences in motor function, the RT of the Baseline condition was subtracted from the RT of the dual-task condition. The differences were then analyzed using a one-way (hearing aid condition: Unaided/OMNI/DIR) repeated measures ANOVA. The result indicated that the effect of hearing aid condition was not significant ($F_{2, 36} = 0.10$, $p = 0.90$, partial $\eta^2 = 0.006$).

To examine the consistency between listening effort measured using the driving and visual reaction-time paradigms, the correlation between driving performance composite score and RT of the visual reaction-time task was examined. Since hearing aid condition did not have an effect on performance in either task, we averaged the three difference scores between each dual-task hearing aid condition and Baseline condition in the driving and the visual reaction-time tasks. These difference scores measure listening effort in the two paradigms. Figure 4 reveals this correlation for the 19 participants to be significantly positive ($r = 0.50$, $p = 0.029$). If the outlier indicated by the arrow in Figure 4 was eliminated from analysis, the correlation coefficient increased to 0.71 ($p = 0.0009$).

Discussion

Experiment 2 replicated Experiment 1 showing that hearing aid amplification and directional processing significantly improved speech recognition performance relative to the Unaided condition, but did not have an effect on listening effort. Therefore, the non-significant effect of hearing aid technologies on listening effort found in Experiment 1 was unlikely to be due to the complexity of the driving task. The significant correlation between listening effort measured using the driving and visual reaction-time paradigms further suggests that the two paradigms seem equivalent.

The result that directional processing considerably improved CST score but did not reduce listening effort, however, is not consistent with the literature. Using the same visual reaction-time paradigm, Sarampalis et al. (2009) found that an increase in SNR of auditory speech stimuli by 4 dB was associated with an increase in speech recognition performance by 15% to 30% and a decrease in RT by 40 to 50 msec (cf. Figure 4 in Sarampalis et al., 2009, the “Unprocessed” data). In contrast, the current experiment revealed that although the SNR improvement provided by directional processing (approximate 4 dB) could increase CST score by 18% (relative to the OMNI condition), the RT was shortened only by a non-significant 4.4 msec. A possible explanation for the inconsistent results between the two studies could be participant characteristics: younger adults with normal hearing were tested in the study by Sarampalis et al. (2009) while older adults with hearing impairment were evaluated in the current study. This speculation was somewhat consistent with the study by Ng et al. (2013), which shows that noise reduction algorithms can reduce listening effort only for listeners with better working memory capacity. In order to test this speculation and to ensure that the methodologies used by Sarampalis et al. (2009) were correctly implemented in the current study, Experiment 2 was repeated using younger adults with normal hearing.

EXPERIMENT 3

Materials and Methods

Twenty-four adults (12 males and 12 females) participated in this experiment. Ages ranged from 20 to 37 years with a mean of 23.4 years ($SD = 3.7$). All participants had normal hearing (pure-tone thresholds < 25 dB HL at 0.5, 1, 2, and 4 kHz). The auditory stimuli, visual reaction-time dual-task paradigm, procedures, and equipment were identical to those used in Experiment 2. Experiment 3 differed from Experiment 2 in that (1) the auditory stimuli were not processed to compensate individual hearing loss, (2) in the OMNI and DIR conditions participants were allowed to adjust the sound level of the auditory stimuli to their most comfortable level, and (3) only three dual-task conditions (Unaided, OMNI, and DIR) and the Baseline (visual reaction-time task only) condition were tested. The single-task conditions (speech recognition task only) were not included in the experiment. Data were analyzed using the methods described in Experiment 2.

Results and discussion

Figure 5A shows the mean CST score in each of the dual-task hearing aid conditions. Although participants' speech performance reached the ceiling level and the difference between conditions was small, the repeated measures ANOVA indicated that hearing aid condition had a significant effect on CST score (converted to RAU) ($F_{2, 46} = 5.88$, $p = 0.005$, partial $\eta^2 = 0.20$). Post-hoc analyses indicated that the mean CST score of the DIR condition was higher than that of the Unaided and OMNI conditions, while the difference between the latter two conditions was not significant. Figure 5B shows the RT as a function of test condition. Similar to Experiment 2, the RT of the Baseline condition was subtracted from the RT of a given hearing aid condition before analysis. The one-way (hearing aid condition: Unaided/OMNI/DIR) repeated measures of ANOVA indicated that the effect of hearing aid condition on RT was significant ($F_{2, 46} = 6.32$, $p = 0.004$, partial $\eta^2 = 0.22$). Post-hoc analyses indicated that the mean RT of the DIR condition was shorter than that of the Unaided and OMNI conditions, while the difference between the latter two conditions was not significant. Lastly, a planned one-tailed t-test was conducted to investigate if directional benefit in listening effort (i.e., the difference in RT between the DIR and OMNI conditions) of younger adults with normal hearing (this experiment; 57.6 msec) was larger than that of older listeners with hearing impairment (Experiment 2; 4.4 msec). The result revealed that this is the case ($t_{41} = 1.97$, $p = 0.028$).

Experiment 3 indicated that although the difference in CST scores between the DIR and OMNI conditions was small (due to the ceiling effect), directional processing considerably reduced listening effort. These results suggested that the inconsistent findings between Experiment 2 and the study by Sarampalis et al. (2009) were due to the characteristics of the listeners, and that the methods used in Experiment 2 were comparable to those applied by Sarampalis et al. The results of Experiments 2 and 3 further suggested that, for a given listening situation and SNR, the benefit of hearing aid technologies on listening effort measured from younger adults with normal hearing may not be fully realized on older listeners with hearing impairment.

GENERAL DISCUSSION

This study assessed the effect of hearing aid amplification and a directional technology on listening effort measured using a dual-task paradigm, wherein participants performed a speech recognition task concurrently with a simulated driving task (Experiment 1) or a simple visual reaction-time task (Experiment 2). Although the two paradigms generated consistent results and the correlation between them was significant, it is premature to conclude that the two dual-task paradigms are equivalent. An observation of note is that the RT of the visual reaction-time task tended to decrease as CST score increased (Figure 3B), but that was not so with driving performance (Figure 2B). Furthermore, although listeners who had better driving performance tended to have better visual reaction-time task performance (Figure 4), individual variation was not small. Therefore, although the current study suggests that laboratory-style dual-task paradigms remain appropriate to measure listening effort, researchers and clinicians alike should not ignore the effect of ecological validity when designing or choosing a dual-task paradigm.

Although considerable efforts were made to increase the ecological validity of the driving dual-task paradigm in Experiment 1, driving the simulator in the laboratory was still very different from driving an automobile in the real world. For example, the perception of the movement and inertia in the simulator differed significantly from that in the van traveling on the road. Because the automobile/road noises were recorded with the van set at a constant speed, the level of automobile (engine) noise was not consistent with the acceleration and deceleration of the vehicle in the simulator. Furthermore, the speech recognition stimuli were recorded using the hearing aids that were not optimized for individual hearing loss. The effect of non-optimized hearing aids and stimulus post-processing (to compensate individual hearing loss) might alter the listeners' auditory perception and listening effort. Because the stimuli and tasks used in the driving dual-task paradigm are still quite different from the real-world listening-while-driving task, it is unknown if listening effort measured using this paradigm is generalizable to the real world.

Of note, directional processing considerably increased CST score but did not reduce listening effort for older adults (Experiment 2, Figure 3). For younger listeners with normal hearing, directional processing slightly improved speech recognition performance but greatly reduced listening effort (Experiment 3, Figure 5). To explain these findings, psychometric functions that describe the performances in the dual-task experiment as a function SNR are proposed (Figure 6). The function of the speech recognition task is sigmoid shaped (solid curves in Figure 6) and the function of the visual reaction-time task (in RT) is assumed to be reverse-sigmoid shaped (dashed curves). The thicker black curves represent the older listeners' psychometric functions. It is likely that, at the -1 dB SNR used in the experiment (indicated by the dark gray rectangle in Figure 6), older listeners' speech recognition performance was at the steep portion of the function (50 to 70% correct), while visual reaction-time performance was at the floor level. This floor effect might occur in the driving task as well. Because younger participants had better speech recognition performance, their psychometric functions would fall to the left side of older listeners' functions (the thinner gray curves; assuming that the psychometric functions of younger and older adults have the same shape). At the same SNR (the dark gray rectangle in Figure

6), younger listeners' speech recognition performance was at the ceiling level ($> 90\%$ correct), while RT was at the steep portion of the function. Therefore, if a better SNR was used in the study, both the driving and visual reaction-time dual-task paradigms might be able to demonstrate the effect of hearing aid technologies on listening effort for older listeners (indicated by the light gray rectangle in Figure 6). The psychometric functions of the dual-task paradigm have not been derived empirically. More research is needed to investigate the characteristics of dual-task paradigm psychometric function, such as the shape of the function and the relationship between the functions of the two tasks, for older and younger listeners.

CONCLUSIONS

Based on the results of this study, three important conclusions can be drawn. First, adding speech listening/repetition to driving compromised driving performance. Hearing aid technologies could not reduce this negative effect at SNRs that typically occur in vehicles traveling on the highway. Second, listening effort measured in dual-task experiments using a simulated real-world driving task and a simple, laboratory-style task was generally consistent. Third, for a given listening environment, the benefit of hearing aid technologies on listening effort measured from younger adults with normal hearing may not be fully translated to older listeners with hearing impairment.

Acknowledgments

Source of Funding:

Yu-Hsiang Wu is currently receiving grants from NIH, ASHA Foundation, and Siemens Hearing Instruments. Nazan Aksan, Matthew Rizzo, and Ruth Bentler are currently receiving grants from NIH; Nazan Aksan, Matthew Rizzo are also receiving grants from Toyota Motor Company. For the remaining authors, none were declared.

This work was in part supported by a research grant from Siemens Hearing Instruments, Inc. and a research grant from the National Institute on Deafness and Other Communication Disorders (R01-DC012769).

This research was supported by a research grant from Siemens Hearing Instruments and NIH/NIDCD R01DC012769. The authors thank Dr. Jeffery Dawson for SIREN data analyses and Dr. Sepp Chalupper for his contributions in research design. Portions of this paper were presented at the annual meeting of the American Auditory Society, March 9th, 2013, Scottsdale, Arizona, USA.

References

- Alm H, Nilsson L. The effects of a mobile telephone task on driver behaviour in a car following situation. *Accident Anal Prev.* 1995; 27:707–715.
- ANSI. Specification for Audiometers (ANSI S3.6). New York: American national standards institute; 2010.
- Bentler RA, Holte L, Turner C. An update on the acclimatization issue. *Hear J.* 1999; 52 (11):44–47.
- Broughton KLM, Switzer F, Scott D. Car following decisions under three visibility conditions and two speeds tested with a driving simulator. *Accident Anal Prev.* 2007; 39:106–116.
- Braitman KA, McCartt AT. National reported patterns of driver cell phone use in the United States. *Traffic Injury Prevention.* 2010; 11:543–548. [PubMed: 21128181]
- Consiglio W, Driscoll P, Witte M, et al. Effect of cellular telephone conversations and other potential interference on reaction time in a braking response. *Acc Anal Prev.* 2003; 35:495–500.
- Cox RM, Alexander GC, Gilmore C. Development of the Connected Speech Test (CST). *Ear Hear.* 1987; 8:119S–126S. [PubMed: 3678650]

- Cox RM, Alexander GC, Gilmore C, et al. Use of the Connected Speech Test (CST) with hearing-impaired listeners. *Ear Hear.* 1988; 9:198–207. [PubMed: 3169400]
- Dastrup E, Lees MN, Bechara A, et al. Risky car following in abstinent users of MDMA. *Accident Analysis and Prevention.* 2010; 42:867–873. [PubMed: 20380914]
- Desjardins JL, Doherty KA. Age-related changes in listening effort for various types of masker noises. *Ear Hear.* 2013; 34:261–272. [PubMed: 23095723]
- Desjardins JL, Doherty KA. The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear Hear.* 2014 Published Ahead-of-Print.
- Dillon H. NAL-NL-1: a new prescriptive fitting procedure for nonlinear hearing aids. *Hear J.* 1999; 52:10–16.
- Downs DW. Effects of hearing and use on speech discrimination and listening effort. *J Speech Hear Disord.* 1982; 47:189–193. [PubMed: 7176597]
- Fraser S, Gagné JP, Alepins M, et al. Evaluating the effort expended to understand speech in noise using a dual-task paradigm: the effects of providing visual speech cues. *J Speech Lang Hear Res.* 2010; 53:18–33. [PubMed: 19635945]
- Gatehouse S, Gordon J. Response times to speech stimuli as measures of benefit from amplification. *Br J Audiol.* 1990; 24:63–68. [PubMed: 2317602]
- Gatehouse S, Noble W. The Speech, Spatial and Qualities of Hearing Scale (SSQ). *Int J Audiol.* 2004; 43:85–99. [PubMed: 15035561]
- Gosselin PA, Gagné JP. Use of a dual-task paradigm to measure listening effort. *Canadian J Speech Lang Patho Audiol.* 2010; 34:43–51.
- Gosselin PA, Gagné JP. Older adults expend more listening effort than young adults recognizing speech in noise. *J Speech Lang Hear Res.* 2011; 54:944–958. [PubMed: 21060138]
- Hällgren M, Larsby B, Lyxell B, et al. Speech understanding in quiet and noise, with and without hearing aids. *Int J Audiol.* 2005; 44:574–583. [PubMed: 16315448]
- Hick CB, Tharpe AM. Listening effort and fatigue in school-age children with and without hearing loss. *J Speech Lang Hear Res.* 2002; 45:573–584. [PubMed: 12069009]
- Hornsby BW. The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear Hear.* 2013; 34:523–534. [PubMed: 23426091]
- Humes LE. The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *J Am Acad Audiol.* 2007; 18:590–603. [PubMed: 18236646]
- Kiessling J, Pichora-Fuller MK, Gatehouse S, et al. Candidature for and delivery of audiological services: special needs of older people. *Int J Audiol.* 2003; 42:2S92–101. [PubMed: 12918635]
- Larsby B, Hällgren M, Lyxell B, et al. Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *Int J Audiol.* 2005; 44:131–143. [PubMed: 15916114]
- McCarley JS, Vais MJ, Pringle H, et al. Conversation disrupts changes detection in complex traffic scenes. *Human factors.* 2004; 46:424–436. [PubMed: 15573543]
- Morel M, Petit C, Bruyas M, et al. Physiological and behavioral evaluation of mental load in shared attention tasks. *Conf Proc IEEE Eng Med Biol Soc.* 2005; 5:5526–5527. [PubMed: 17281505]
- Ni, Rui; Kang, JJ.; Andersen, GJ. Age-related declines in car following performance under simulated fog conditions. *Accident Anal Prev.* 2010; 42:818–826.
- Ng EHN, Rudner M, Lunner T, et al. Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *Int J Audiol.* 2013; 52:433–441. [PubMed: 23550584]
- Pearsons, KS.; Bennett, RL.; Fidell, S. BBN Report # 3281. Cambridge, MA: Bolt, Beranek and Newman; 1976. Speech levels in various environments. Report to the Office of Resources and Development, Environmental Protection Agency.
- Pichora-Fuller MK, Singh G. Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiologic rehabilitation. *Trends Amplif.* 2006; 10:29–59. [PubMed: 16528429]
- Picou EM, Ricketts TA, Hornsby BW. How hearing aids, background noise, and visual cues influence objective listening effort. *Ear Hear.* 2013; 34:e52–e64. [PubMed: 23416751]

- Pittman A. Children's performance in complex listening conditions: effects of hearing loss and digital noise reduction. *J Speech Lang Hear Res.* 2011; 54:1224–1239. [PubMed: 21330646]
- Rizzo M. Impaired Driving From Medical Conditions: A 70-Year-Old Man Trying to Decide if He Should Continue Driving. *J AM Med Assoc.* 2011; 305:1018–1026.
- Rizzo M, Lamers CTJ, Sauer CG, et al. Impaired perception of self-motion (heading) in abstinent ecstasy and marijuana users. *Psychopharmacology.* 2005a; 179:559–566. [PubMed: 15723231]
- Rizzo M, Shi Q, Dawson JD, et al. Stops for cops: Impaired response implementation in older drivers with cognitive decline. *Journal of the Transportation Research Board.* 2005b; 1922:1–8.
- Saunders GH, Cienkowski KM. Acclimatization to hearing aids. *Ear Hear.* 1997; 18:129–139. [PubMed: 9099562]
- Sarampalis A, Kalluri S, Edwards B, et al. Objective measures of listening effort: effects of background noise and noise reduction. *J Speech Lang Hear Res.* 2009; 52:1230–1240. [PubMed: 19380604]
- Strayer DL, Johnston WA. Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychol Sci.* 2001; 12:462–466. [PubMed: 11760132]
- Studebaker GA. A "rationalized" arcsine transform. *J Speech Hear Res.* 1985; 28:455–462. [PubMed: 4046587]
- Worrall, LE.; Hickson, L. *Communication disability in aging from prevention to intervention.* New York: Thomson, Delmar Learning; 2003.
- Wu YH, Bentler RA. Do older adults have social lifestyles that place fewer demands on hearing? *J Am Acad Audiol.* 2012; 23:697–711. [PubMed: 23072962]
- Wu YH, Stangl E, Bentler R. Hearing-aid users' voices: A factor that could affect directional benefit. *Int J Audiol.* 2013a; 52:789–794. [PubMed: 23777478]
- Wu YH, Stangl E, Bentler R, et al. The effect of hearing aid technologies on listening in an automobile. *J Am Acad Audiol.* 2013b; 24:474–485. [PubMed: 23886425]
- Zekveld AA, Kramer SE, Festen JM. Pupil response as an indication of effortful listening: the influence of sentence intelligibility. *Ear Hear.* 2010; 31:480–490. [PubMed: 20588118]

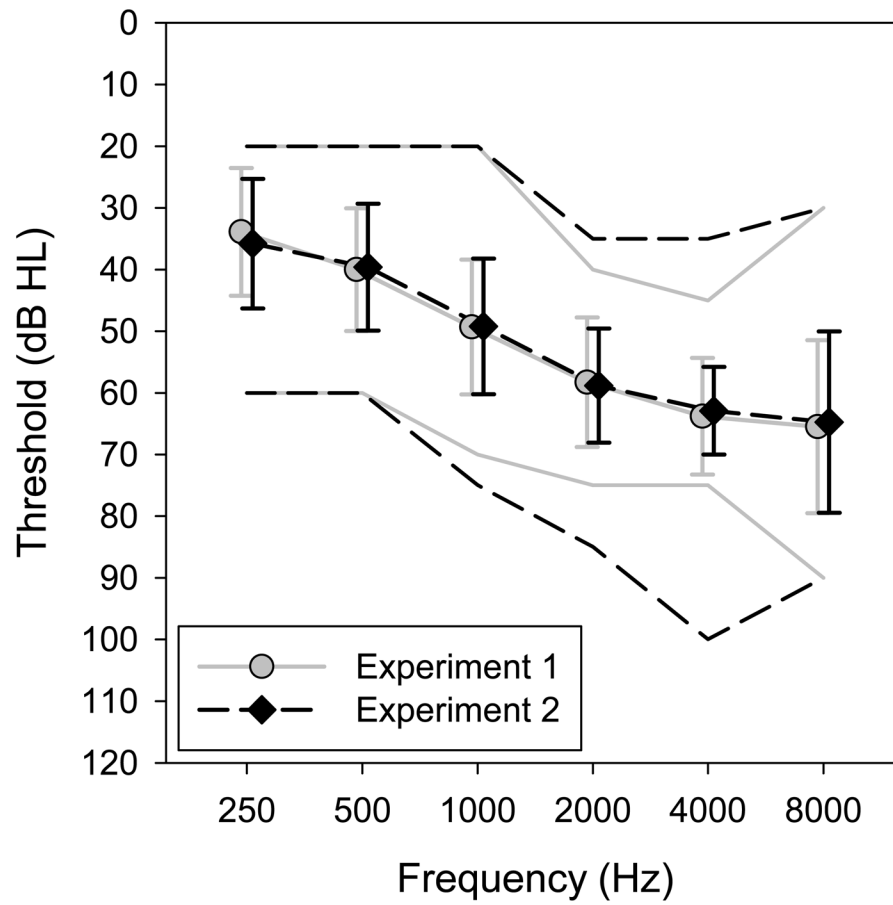


Figure 1. Average audiograms of study participants. Gray circles and black diamonds are for participants in Experiments 1 and 2, respectively. Error bars = 1 SD. The maximum and minimum thresholds are shown by the lines.

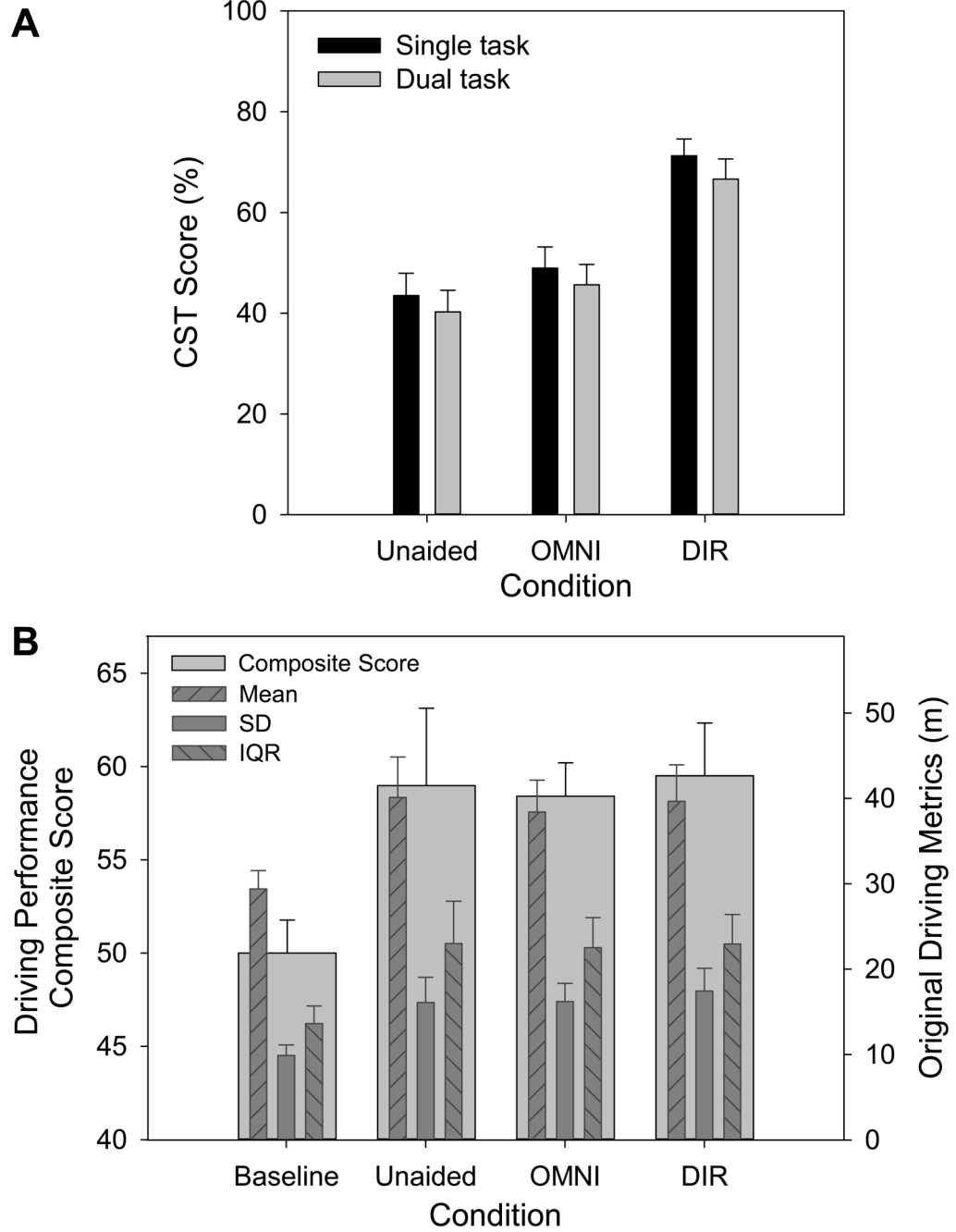


Figure 2. A: CST score as a function of test condition in Experiment 1. Higher scores reflect better speech recognition performance. B: Driving performance composite score (refer to the left y-axis) and the three original driving metrics (mean, SD, and interquartile range (IQR) of the following distance in meters; refer to the right y-axis) as a function of test condition. Higher scores reflect poorer driving performance. Error bars = 1 SE.

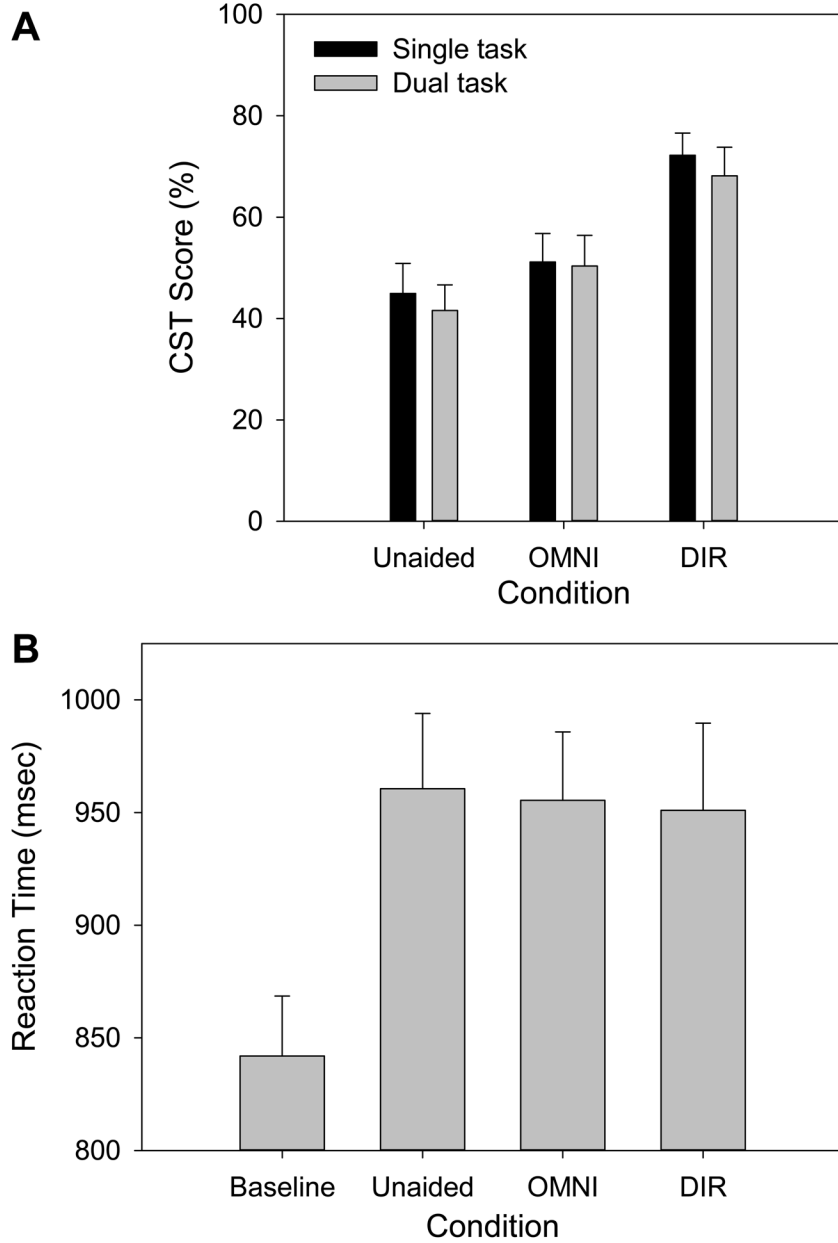


Figure 3.

A: CST score as a function of test condition in Experiment 2. Higher scores reflect better speech recognition performance. B: Reaction time as a function of test condition. Longer reaction times reflect poorer performance. Error bars = 1 SE.

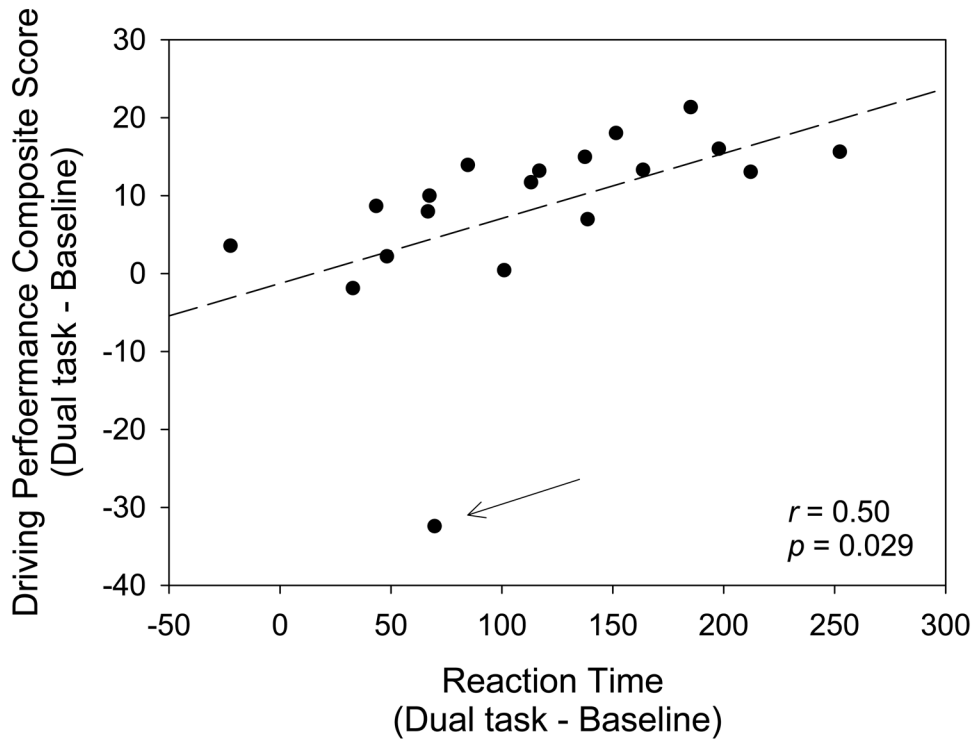


Figure 4. Relationship between listening effort measured using the driving (Experiment 1) and visual reaction-time tasks (Experiment 2). The data are the average of the three difference scores between each dual-task hearing aid condition (Unaided, OMNI, and DIR conditions) and Baseline condition. The arrow indicates an outlier.

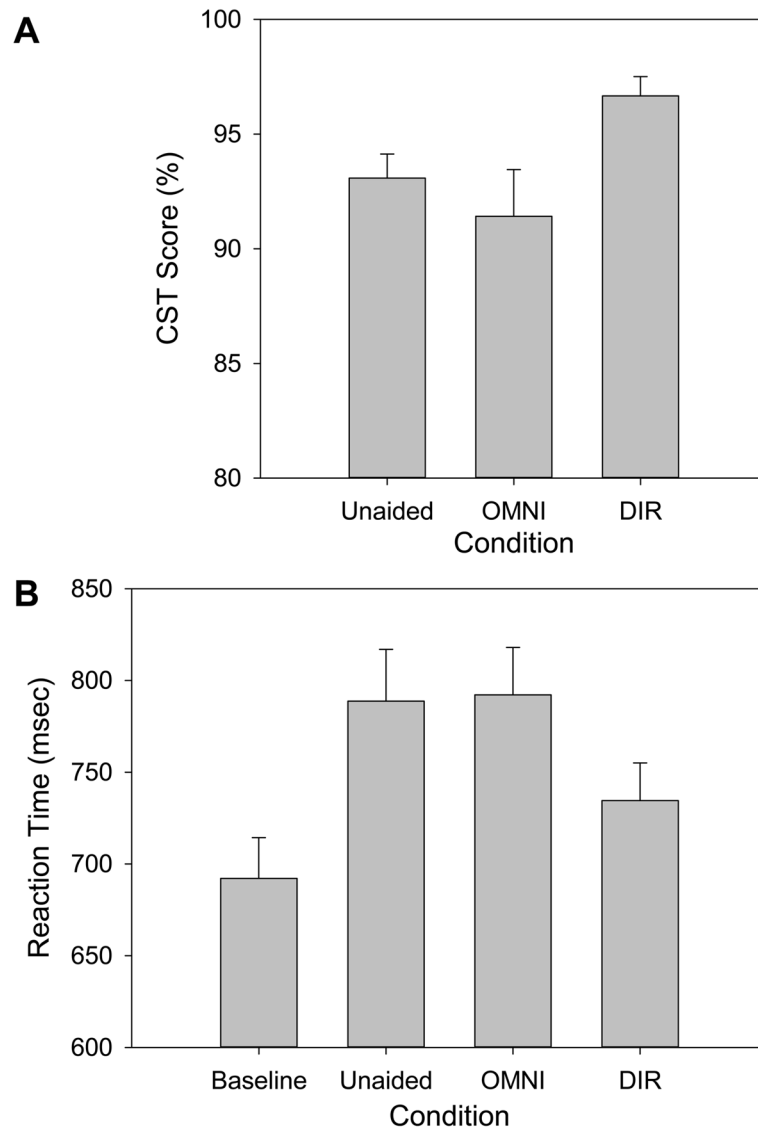


Figure 5.

A: CST score in the dual-task condition as a function of test condition in Experiment 3.

Higher scores reflect better speech recognition performance. B: Reaction time as a function of test condition. Longer reaction times reflect poorer performance. Error bars = 1 SE.

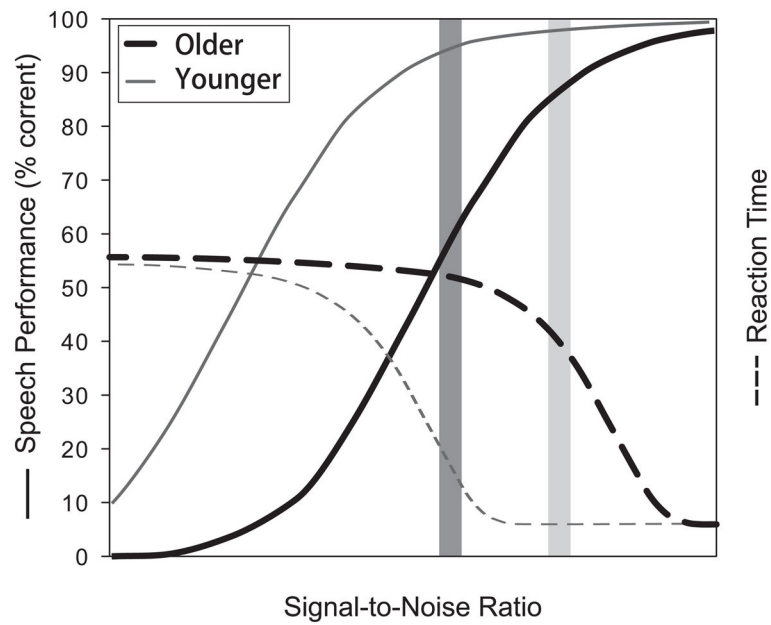


Figure 6. Hypothetical psychometric functions of speech recognition task (solid curves, refer to the left y-axis) and reaction-time task (dashed curves, refer to the right y-axis) for older (thicker black curves) and younger (thinner gray curves) listeners. Dark gray rectangle represents the signal-to-noise ratio used in the study. Light gray rectangle indicates the signal-to-noise ratio that corresponds to the steep portion of older listeners' functions.