ASHA American Speech-Language-Hearing Association

AJA

Research Article

Effect of Realistic Test Conditions on Perception of Speech, Music, and Binaural Cues in Normal-Hearing Listeners

Yang-Soo Yoon,[a] Priyanka Jaisinghani,[a] and Raymond Goldsworthy[b]

[a] Department of Communication Sciences and Disorders, Baylor University, Waco, TX [b] Department of Otolaryngology – Head and Neck Surgery, Keck School of Medicine, University of Southern California, Los Angeles

ARTICLE INFO

Article History:
Received August 2, 2022
Revision received October 3, 2022
Accepted October 12, 2022

Editor-in-Chief: Ryan W. McCreery
Editor: Erin M. Picou

https://doi.org/10.1044/2022_AJA-22-00143

ABSTRACT

Purpose: The purpose of this study was to determine the feasibility of online testing in a quiet room for three auditory perception experiments in normal-hearing listeners: speech, music, and binaural cue.

Method: Under Experiment 1, sentence perception was measured using fixed signal-to-noise ratios (SNRs: +10 dB, 0 dB, and −10 dB) and using adaptive speech reception threshold (SRT) procedures. The correct scores were compared between quiet room and soundproof booth listening environments. Experiment 2 was designed to compare melodic contour identification between the two listening environments. Melodic contour identification was assessed with 1, 2, and 4 semitone spacings. Under Experiment 3, interaural level difference (ILD) and interaural time differences (ITD) were measured as a function of carrier frequency. For both measures, two modulated tones (400-ms duration and 100-Hz modulation rate) were sequentially presented through headphones to both ears, and subjects were asked to indicate whether the sound moved to the left or right ear. The measured ITD and ILD were then compared between the two listening environments.

Results: There were no significant differences in any outcome measures (SNR- and SRT-based speech perception, melodic contour identification, and ITD/ILD) between the two listening environments.

Conclusions: These results suggest that normal-hearing listeners may not require a controlled listening environment in any of the three auditory assessments. As comparable data can be obtained via the online testing tool, using the online auditory experiments is recommended.

This study is an extension of the study by Yoon et al. (2021). The authors in the previous study examined the effect of listening environments and testing procedures on temporal and spectral cue perception of individuals with normal hearing (NH). They compared across listening environments (soundproof vs. quiet room), testing orders (ascending vs. descending vs. counterbalanced), and number of sessions (multiple trials vs. single trial). The study findings imparted evidence for remote audiology administration since the above factors did not significantly impact the results. Therefore, it was opined that a quiet room and a single trial of measurement with testing order counterbalanced for spectral and temporal assessment of individuals with NH would suffice (Yoon et al., 2021). However, as the authors acknowledged as a limitation, further research is warranted to expand the same knowledge base for more complex stimuli, speech, and in more realistic conditions with the presence of noise. In addition, due to COVID-19, in-person testing has substantially been limited. In those scenarios, control data of remote audiological testing in NH listeners for commonly used test protocols can be highly beneficial.

Speech perception at various signal-to-noise ratios (SNRs) is the most common outcome measure to quantify ability of listeners for processing real-world stimuli at

Correspondence to Yang-Soo Yoon: yang-soo_yoon@baylor.edu.
Disclosure: The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.

everyday listening environments (Plapous et al., 2006). This metric is particularly useful to capture nonlinear patterns of performance that can occur at specific SNRs. The nonlinear performance-intensity function is not uncommon in listeners with both NH and hearing loss. For example, consonant recognition improved linearly in NH listeners when SNR improved from −18 dB to 0 dB but remained flattened after then (G. A. Miller & Nicely, 1955). Individuals with hearing loss typically show the opposite pattern. Speech perception recognition remains little or almost no changes below 0 dB SNR but improved linear after then (Phatak et al., 2009). Similar trends are observed in cochlear implant users (Balkany et al., 2007; Choi et al., 2017; Ma et al., 2016) and hearing aid users (Armstrong et al., 1997; C. W. Miller et al., 2016). SNR-based speech perception is useful to capture overall patterns of performance-intensity functions but takes time, which makes it less favorable metric at clinics. In contrast, a speech reception threshold (SRT), one of the most commonly used clinical testing protocols, is a quick way to assess patient's speech perception in noise (Naylor, 2016; Panday et al., 2018). SRT finds a single SNR needed for patients to reach 50% performance that results in ignoring SNRs needed for other points of percent correct. The SRT measure avoids floor and ceiling effects usually encountered when testing at fixed SNR levels. The SRTs obtained from patients with degree of hearing loss can directly be compared, whereas percent correct scores obtained at different SNRs cannot be compared (Smits et al., 2021). Current literature supports online testing for hearing screenings and pure-tone audiometry. However, there is a need to validate the use of most encountered speech stimuli in noise for remote assessment and treatment of hearing loss. In this study, we intended to generate data for both approaches for speech perceptions under quiet room and soundproof booth listening environments so that researchers and clinicians could use them as a control.

Apart from speech perception in noise, there has been a recent surge of research interest in music perception of individuals with hearing loss and hearing device users. It is well known that optimal speech recognition can be achieved through temporal envelope cues even without the spectral and temporal fine structure cues (Shannon et al., 1995; Xu & Pfingst, 2008). However, for melody recognition, spectral and temporal fine structure cues are essential (Pantev et al., 2003; Sares et al., 2018; Smith et al., 2002; Varnet et al., 2015). Individuals with hearing loss or auditory prostheses have a poor spectral resolution, affecting their music perception and appreciation (Galvin et al., 2007, 2009; Gfeller et al., 2006). It would be interesting to know if the music perception in a remote environment, as in a quiet room is like that of a soundproof room. The knowledge base would aid future research in studying spectral and temporal fine structure cues tapped by the musical pitch perception experiments to be conducted in a remote setup.

Another important aspect of daily listening is binaural hearing that aids in sound localization (Grieco-Calub & Litovsky, 2010; Van Deun et al., 2010; Zirn et al., 2019) and better speech perception in adverse listening situations (Choi et al., 2017; Veugen et al., 2017). Two major cues of binaural hearing are the interaural time difference (ITD) and the interaural level difference (ILD; Akeroyd, 2006). However, compared to SNR- and SRT-based speech perception measures, binaural hearing though important for daily living are less discussed and are not often studied in a clinical setup (Shafiro et al., 2020). In this study, our intention was to generate data of ITD and ILD measures with NH listeners in a quiet room so that the use of online testing would increase accessibility for more audiological evaluations.

Hence, as shown in the previous quiet-room study (Yoon et al., 2021), a quiet room listening environment allowed us to obtain comparable data in spectral and temporal processing. However, the validity of quiet rooms needs to be proven in more clinically relevant measures. Thus, this study was formulated to determine whether the speech perception (Experiment 1), music perception (Experiment 2), and binaural cue perception (Experiment 3) performances of individuals with NH are comparable in the two listening environments: quiet room and soundproof booth. Suppose this study findings report the efficacy of a quiet room as that of a soundproof booth. Furthermore, future studies should also apply these methods to listeners with hearing loss or those with the auditory prosthesis. Such a test module will aid in faster and comprehensive evaluations at a remote setup without an expensive soundproof room, resulting in enhancing teleaudiology at large.

General Method

Three experiments were administered in two listening environments, a quiet room and a soundproof booth with an online testing protocol: sentence perception (Experiment 1), music perception (Experiment 2), and binaural cue perception (Experiment 3). Each experiment was conducted on a separate day. A total of 52 NH undergraduate and graduate students at Baylor University participated, and each subject participated in only one experiment but for both listening environments of the experiment. A NH group was purposefully selected to control any effect of intrinsic factors (hearing loss of participants) or minimize the confounding variables as much as possible. All had thresholds better than 20-dB hearing level at audiometric frequencies from 0.25 to 8 kHz. No participants reported prior ear diseases and experience in psychoacoustic experiments. All subjects received course credit for their participation. All the subjects provided informed consent, and the Baylor University Institutional Review Board approved all procedures.

For each experiment and procedure for determining most comfortable level (MCL), step-by-step video instructions were provided to all participations. For the quiet room listening condition, subjects administered the tasks by themselves. The quiet room listening condition had three specific requirements: (a) presentation level for all tasks should be at their MCL (see Procedures in Experiment 1 for details), (b) they use their own choice of headphones including earphones and earbuds and (c) all tasks should be conducted in their room with no obvious ambient noise (e.g., television, music, fan, air conditioner sound, dog in the room). Two experienced graduate assistants were also available for technical difficulties for the quiet room test setups.

For the soundproof booth listening condition, the different subjects from those who participated in the quiet room condition were recruited to repeat the same tasks in the quiet room condition but in the soundproof booth. The graduate assistants provided the verbal instructions and controlled the whole testing procedure.

Any participants who had outlier data in either listening condition were asked to repeat the experiments with the help of a graduate assistant because these psychoacoustic experiments were new to all subjects. The outliers were determined by finding first (F), third (T), and inter (I) quartiles from the entire data. The upper and lower limits were calculated by T+(1.5*I) and F−(1.5*I), respectively (Ferrari et al., 2020). Any data out of these upper and lower limits were considered outliers.

Experiment 1: Speech Perception

Subjects

Twenty-two adults with NH participated in sentence perception measures under the quiet room and soundproof booth listening environments. Sentence perception was administered using two testing procedures: at a specific SNR and using an adaptive SRT. One group of 11 adult subjects participated in the SNR sentence perception for both listening environments (10 women and one man; mean age = 21 ± 0.7). Another group of 11 adult subjects participated in the SRT sentence perception for both listening environments (11 women; mean age = 21 ± 0.61). All the subjects were native speakers of American English.

Stimuli

The speech perception procedure used the Revised Speech Perception in Noise Test (R-SPIN; Bilger, 1984; Wilson et al., 2012) to measure SNR- and SRT-based speech perception. Only the low-predictability sentences from the R-SPIN materials were used. The R-SPIN materials were selected since they allow for a sentence construction of the test materials yet are amenable to closed-set computerized testing with forced-choice procedures. The R-SPIN materials are organized in sets of 25 sentences and an example sentence from the low-predictability set is, “Miss White won’t think about the crack,” where the last word of the sentence is always the scored keyword. A male talker produced the sentences. Sentences were presented to the left ear via headphones, and subjects were instructed to respond using a computer interface to indicate the last word of the sentence. The closed set sentence identification task was preferred in this study since the lexical factors impacting the word recognition and word identification for a closed set become like an open set sentence identification task when a greater set size is chosen for the closed set task (Clopper et al., 2006; G. A. Miller et al., 1951). Moreover, in a closed-set test, despite the different language expertise of participants, the response selection is from the common linguistic pool (Jerger et al., 1968).

For the noise conditions, speech-spectrum noise was generated by filtering random noise drawn from a uniform distribution through a speech-spectrum shaping filter. This filter was generated by estimating the power spectral density of the corresponding speech materials using Welch’s periodogram method and converting this density to an eighth-order IIR filter using Prony’s method (Oppenheim, 1978). A 20-s sample of noise was generated and played continuously throughout the speech recognition procedures. Brief 20-ms attack and release ramps were applied at the beginning and end of the noise sample to avoid snap, crackle, and pop artifacts between loops. The SNR was specified based on the root-mean-square value of the speech and noise samples. For a given SNR, the speech and noise samples were combined and scaled such that the total output power was set equal to the subject-specified comfortable loudness level.

Procedure

Both SNR- and SRT-based speech perception was conducted with monaural presentation (left ear) using headphones to simulate a standard testing protocol at most clinics. Before a formal test, MCL was determined by presenting one R-SPIN sentence, “The woman knew about the lid,” in quiet to the left ear via headphones (Sennheiser HAD 280 supra-aural headphones were used for the soundproof booth listening condition). This sentence was produced by the same male talker who produced the R-SPIN sentences used for the formal test. Each subject adjusted computer volume to be “comfortable, but slightly loud” or “comfortable,” based on Cox’s Loudness Scale (Cox et al., 1997). The MCL was used for both SNR and SRT procedures. Each spoken sentence was shown at the bottom of the computer screen with the last word missing. Subjects were asked to choose the correct word from the 25 possible words on a computer screen.

For the SNR procedure, sentence perception was measured at −10 dB, 0 dB, and +10 dB SNRs (speech-shaped noise). The three SNRs represented bad, better, and good listening environments. For each SNR, 25 sentences were randomly presented twice (a total of 50 presentations per SNR) to obtain more reliable outcomes. Percent correct score was computed by dividing the number of correct responses by 50. For the SRT procedure, the initial SNR of a measurement run was set to 12 dB. The SNR was decreased by 2 dB following correct answers and was increased by 2 dB following mistakes. The procedure continued until the subject made eight mistakes, and the SRT was taken as the average of the last four reversals. Correct-answer feedback was provided on all trials, and participants could use a repeat button for stimulus replay and were instructed to do so if necessary.

Results

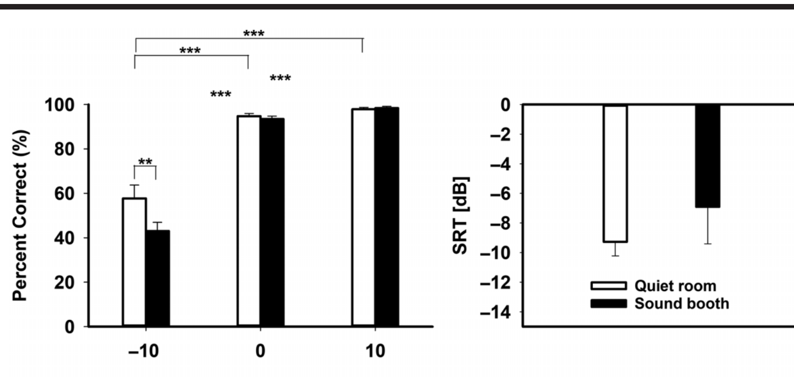
Figure 1 shows the average percent correct, along with standard errors as a function of SNR (left panel) and average SRT (right panel). Mean percent correct data of SNR procedure were analyzed with two-way repeated-measures analysis of variance (ANOVA) using two independent-within variables: listening condition with two levels (i.e., quiet room and soundproof booth) and SNR with three levels (i.e., −10, 0, 10 dB). Mean SRT data were analyzed with a paired *t* test to compare a quiet room versus a soundproof booth.

For the percent correct data across SNRs (left panel), the listening environment had a significant impact on sentence perception (open vs. filled bars), $F(1, 60) = 4.26$, $p = .04$, indicating that percent scores were significantly better under the quiet room listening condition than under the soundproof booth listening condition. The statistical analysis yields a significant main effect of SNR, $F(2, 60) = 149.90$, $p < .001$, indicating that subject's speech perception is significantly affected by the noise level. Analysis also showed significant interaction between the two independent variables, $F(2, 60) = 3.63$, $p = .03$. Pairwise multiple comparisons with Bonferroni correction showed a significant listening environment effect at −10 dB SNR ($p = .001$), as indicated by asterisks. Pairwise multiple comparisons also showed a significant SNR effect between −10/0 dB ($p < .001$) and −10/10 dB ($p < .001$) for both listening environments, as indicated by asterisks. For the SRT data (right panel), the listening environment had no significant impact on sentence perception, $t(20) = -0.89$, $p = .39$, indicating that SRT was not significantly different between the listening environments.

Discussion

Experiment 1 aimed at comparing the speech perception in the noise of individuals with NH across a quiet versus soundproof booth. The results indicated no significant difference across the two listening environments in noise at 0 dB and +10 dB SNRs. No significant difference across the two listening environments were also obtained for SRT measure. These findings indicate that a quiet room would be sufficient to conduct experiments involving fixed SNR testing at 0 dB and +10 dB SNRs and SRT measurements in individuals with NH. This study, thereby, imparts evidence for conducting an experiment or clinical tests in a quiet room set up effectively with results as good as that of a soundproof booth when studying the influence of noise on speech perception. These findings support the previously reported findings of no significant difference in pure tone thresholds across the quiet room and soundproof booth (Maclennan-Smith et al., 2013; Margolis & Madsen, 2015).

Figure 1. Mean percent correct scores as a function of signal-to-noise ratio (SNR; left panel) and mean speech reception threshold (SRT; right panel) with standard errors in quiet room and soundproof booth listening environments. Significant pairwise multiple comparison outcomes are indicated by asterisks, indicating that $^{**}p < .01$ and $^{***}p < .001$.



This study found that the percent correct scores were significantly different across the quiet versus soundproof booth at −10 dB SNR but not at 0 dB or +10 dB SNRs. This result suggests effective utilization of a quiet room for speech perception testing of NH participants at +10 dB and 0 dB SNRs except at −10 dB or negative SNR. At −10 dB SNR, significantly better percent correct scores in quiet room compared to the soundproof booth can be explained by the possibility of negative psychological factors or a sense of claustrophobia sensed by the participants in a soundproof booth (Behar, 2021). In contrast to the soundproof booth, the participants are at much ease and comfort for testing in their quiet room. In addition, the subject is presumed to have greater familiarity and habituation with personal headphones used in a quiet room testing instead of supra-aural headphones in the soundproof booth. Therefore, the familiarity and habituation may provide significant advantage over less familiar supra-aural headphones in quiet room at −10 dB SNR. The occurrence of SNR in the daily listening situation is a crucial factor in deciding the SNRs to study in research practice. Since the negative SNRs have a lower occurrence in real-world listening environments (Smeds et al., 2015; Smits et al., 2021), a quiet room can be recommended for speech perception testing of individuals with NH.

Moreover, this study showed a significant difference in the speech perception scores across SNRs. It was observed that the speech perception scores significantly varied between −10 dB and 0 dB SNR, −10 dB, and +10 dB SNR under both listening environments. In contrast, no significant difference was noted between the 0 dB and +10 dB in either listening condition. These findings suggest that the quiet room has similar potential as that of a soundproof booth to capture an SNR effect on speech perception for clinical testing. Therefore, the quiet room listening environment can be recommenced for experimental designs and clinical evaluation setups.

The SRT of NH adults in the existing literature ranges from −0.10 dB with Bamford–Kowal–Bench Speech-in-Noise test (BKB-SIN) American English sentences (Fowler et al., 2021) to approximately −10 dB with Finnish sentences (Dietz et al., 2014), German sentences (Dietz et al., 2014; Kollmeier & Wesselkamp, 1997), and American English Hearing in Noise Test sentence (Nilsson et al., 1994). All these studies were conducted with open set approach. The SRT obtained in this study in a soundproof room with a closed set approach (−7 dB) is compatible with some of these data sets. We expected lower (i.e., better) SRT scores with our closed set approach because a closed set task is lesser demanding than the open set task (Holmes et al., 1988). Our SRTs were better than SRT obtained with BKB-SIN sentences. A potential reason for the difference is that the speech spectrum-shaped noise used in this study is known to have extremely low-intensity fluctuations over time (Shukla et al., 2018), so weaker masker than competing multitalker babble was used in the BKB-SIN test.

One methodological concern of this study was using a single type of sentence (R-SPIN). Reports have shown that certain lists of R-SPIN sentences have psychometric functions for low predictability and high predictability sentences bisecting each other at low SNRs rather than maintaining reasonable separations (Wilson et al., 2012). Greater insight would have been obtained if multiple test materials, such as BKB-SIN, Hearing in Noise Test, and Institute of Electrical and Electronics Engineers sentences, were assessed, including even word-level tests to confirm the findings. Moreover, only 25 closed-set sentences are available in R-SPIN tests. Due to the limited availability of sentences for all the conditions, all the sentences were repeated 2 times in this study. This would have led to carryover effects, resulting in a threat to the internal validity of the study.

Overall, the study outcomes are promising and support the use of soundproof booth–free, subject's own computer-based speech perception in the noise assessment of individuals with NH. This method offers an effective alternative when circumstances do not allow standard speech assessment in a soundproof booth. Furthermore, it would also be interesting to study whether a quiet room setup would suffice even for speech perception in noise assessment of individuals with hearing loss or those using hearing devices.

Experiment 2: Music Perception

Subjects

Ten adult NH listeners (10 women; mean age = 22 ± 0.4) participated in a melodic contour identification in quiet for both quiet room and soundproof booth listening environments.

Stimuli

The stimuli for melodic contour identification were rendered using Sibelius - Music Notation Software (Avid Technology). The musical notes were rendered using the "Piano" note setting in Sibelius and each note was rendered as 400 ms in duration. The nine contours consisted of five-note patterns including "rising," "falling," "flat," "rising-flat," "falling-flat," "rising-falling," "falling-rising," "flat-rising," and "flat-falling" (Galvin et al., 2007). The notes within a pattern were each separated by 200 ms. Melodic contour identification was assessed using contours having a center note of 220 Hz, which corresponds to A3 in Western music notation.

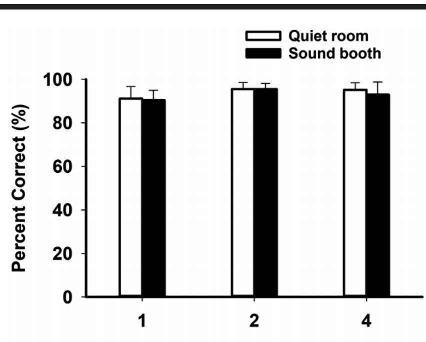
Procedure

Melodic contour identification was assessed with a nine-alternative forced-choice paradigm as a function of tone spacing: 1, 2, and 4 semitones. 1 semitone is the smallest interval used in most music systems (Bell & Jedrzejczak, 2017; Trehub et al., 1986). It is also near the average discrimination thresholds for cochlear implant users (Ping et al., 2012). The use of the 2- and 4-semitone spacings allows us to generate data sets that will be compared with performance for cochlear implant users with poorer discrimination (Ping et al., 2012). The melodic contours were presented to the left ear through headphones at the subjects' MCL, determined in Experiment 1. With each tone spacing setting, percent correct was determined on 27 presentations (9 contours × 3 repetitions). The participants were instructed to select the correct melody contour heard from the options displayed on the screen. Participants could use a repeat button for stimulus and were instructed to do so if necessary.

Results

Figure 2 shows mean percent correct, along with standard errors for the quiet room (open bars) and soundproof booth (filled bars) environments. These mean scores were analyzed with two-way repeated-measures ANOVA using two independent variables: listening condition with two levels (i.e., quiet room and soundproof booth) and note spacing with three levels (i.e., 1, 2, 4 semitones). The statistical analysis showed no significant difference between the listening environments, $F(2, 54) = 0.07$, $p = .79$ and no significant main effect of the note spacing, $F(2, 54) = 0.64$, $p = .53$. Analysis also showed no significant interaction between independent variables, $F(2, 54) = 0.03$, $p = .97$.

Figure 2. Mean percent correct scores with standard errors as a function of note frequency spacing in semitone in quiet room and soundproof booth listening environments.



Discussion

The results indicated no significant difference across a quiet room and a soundproof booth. The mean percent score (greater than 85%) was similar to that in existing literature, 82.4 ± 17.6% (Jung et al., 2010). The percent correct scores greater than 85% across the three-tone spacing settings are also in synchrony with the existing literature of successful melody recognition more than 70% of the time (Jung et al., 2010; Kang et al., 2009).

Furthermore, similar performances between the two listening environments can be explained by the fact that the testing is conducted at the MCL, of the participant rather than the threshold level. The testing at the threshold level, in contrast to the MCL, is expected to be influenced by slight ambient noise. Along the same lines, melody contour identification is similar to auditory processing tests as both are generally carried out above threshold levels. Similar findings of no significant difference in the results of auditory processing measures across the test settings either quiet room or soundproof booth are reported (Lucker, 2017). Our result along with other study suggests that melody contour identification is also less likely to require a soundproof room for administration.

Consequently, it can also be understood that the perception of spectral and temporal fine structure cues critical for music perception is minimally affected by the room acoustics once it is devoid of any obvious ambient noise. There have been related reports of no significant influence on temporal fine structure for the range of level of testing over the range of 20–50 dB sensational level or SL (Moore & Sek, 2009). Since this study involved testing at the MCLs (almost 30–40 dB SL), it can be reasoned that no profound consequences led to the melody identification scores with minimal ambient noise in the quiet room setup. Moore and Sek (2009) also postulated that temporal fine structure tests can be run on any personal computer with a good quality soundcard with no specialized equipment. This study supports these findings and suggests the possibility of music perception testing to be carried out in a quiet room devoid of any obvious ambient noise.

Contrastingly, there have been reports of altered music perception with change in the room acoustics, primarily reverberation. The better musical sound quality of NH in nonreverberant listening environments when compared to reverberant conditions is widely reported in the literature (Roy et al., 2015; Sayles & Winter, 2008). Therefore, it implies that the room acoustics, especially reverberation, has profound consequences on music perception. This study noted no significant difference in melody identification between the two listening environments. It can be explained by the possibility of nonreverberant quiet rooms used in this study. Hence, it is opined that melody contour identification can be carried out in quiet rooms in school buildings or offices by ensuring that they are nonreverberant, with background noises minimal, and prior addressed. It would further aid in easily evaluating subjects or participants at their location, rather than requiring them to visit the soundproof lab.

Furthermore, no significant difference was noted in melody contour identification across the three-tone spacing settings (1, 2, and 4). These tone space settings refer to the pitch difference between notes in a pattern. This feature highly depends on the pitch discrimination ability of a listener. It is reported that just noticeable difference needed for a NH individual to identify a pitch difference between two notes correctly is about one semitone (Kang et al., 2009). Therefore, since this study incorporated one semitone (just noticeable difference) and more as the tone spacing, no significant difference was noted in the melody identification percent correct scores. The findings were common across the tone spacing settings for quiet rooms and soundproof booths. Moreover, it is reasonable to state that a quiet room can be utilized even for studies investigating the just noticeable difference for pitch discrimination or evaluating melody contour identification with very low tone spacings.

Experiment 3: Binaural Cue Perception

Subjects

A total of 20 adult listeners with NH participated in the binaural cue threshold measures (ILD and ITD) in quiet. One group of 10 subjects (nine women and one man; mean age = 21 ± 0.4) participated in the ILD task for both listening environments. Another group of 10 adult NH subjects (nine women and one man; mean age = 25.1 ± 9.7) participated in the ITD task for both listening environments.

Stimuli

For both ILD and ITD threshold measures, stimuli were 400 ms in duration with 20-ms raised-cosine attack and release times. Stimuli were amplitude-modulated sinusoids with the modulator being a half-wave rectified (i.e., transposed) envelope. The modulation frequency of the modulator was 100 Hz, which was selected as representing a typical fundamental frequency of a male talker. The rationale behind using modulated tones rather than unmodulated ones was that they are ecologically valid as natural sounds are typically modulated. Moreover, the modulation envelope facilitates ITD discrimination. We used common carrier frequencies that are known to be effective to ILD and ITD measures of NH listeners. For the ILD threshold measure, carrier frequencies of 1500 Hz, 2500 Hz, and 4500 Hz were used to generate small, medium, and large effects of ILD (Jones et al., 2015; Koka et al., 2011; Stecker, 2014). For the ITD threshold measure, carrier frequencies of 350 Hz, 750 Hz, and 1150 Hz were used to generate small, medium, and large effects of ITD (Kuwada et al., 2010; Stecker, 2014).

Procedure

For both ILD and ITD threshold measures, two modulated tones with a modulation rate of 100 Hz were sequentially presented to both ears through headphones in quiet. The modulation rate of 100 Hz is near the average fundamental frequency of a male talker (Kovacić & Balaban, 2009). So, it is a relevant and ecologically valid modulation frequency to use as a probe. The subjects were asked to indicate which way the sound moved to the left or right ear. Before formal testing, loudness balancing was carried out using the Cox's Loudness Scale (Cox et al., 1997). For loudness balancing, a 1-kHz tone was presented to the left ear at the MCL of the listener, determined in Experiment 1. The participants were instructed to adjust the level at the right ear with 1-dB step using a button given on the computer screen until the loudness was perceptually equal. For both ILD and ITD measures, the threshold was determined with each of the three carrier frequencies in a block testing paradigm.

The procedure to estimate ILD was a two-interval, two-alternative, forced-choice paradigm with three carrier frequencies: 1500, 2500, and 4500 Hz. The three carrier frequencies are known to generate small, medium, and large effect of ILD threshold measure in NH listeners (Akeroyd, 2014; Hartmann & Wittenberg, 1996; Kalcioglu et al., 2003; Willert et al., 2006). The intervals were 400 ms long with 200 ms of silence between intervals. On a given trial, an ILD was applied to either the left or right channel of the stereo signal for the first stimulus interval. The same ILD was applied to the opposite channel for the second interval. For example, if the ILD was 24 dB and the left channel was reduced by 24 dB in the first interval, then the right channel would be reduced by 24 dB in the second interval. For well-balanced stimuli, this process creates a sensation of the sound internally moving across the midline of the head. The channel of the stereo signal for reducing in the first interval was randomly assigned with 50% probability. Adaptive procedures were used to measure the ILD threshold. The initial ILD was set as 24 dB; with each correct answer, it was decreased by a factor of 2, and with each wrong answer, it was increased by a factor of 2. This adaptive rule with a step size 3 times larger (on an exponential scale) following mistakes converges to 75% discrimination accuracy (Kaernbach, 1991). The procedure continued until the subject made eight mistakes, and the ILD threshold was taken as the average of the last four reversals.

Using a similar procedure used for ITD threshold measure, ITD was assessed with three carrier frequencies: 350, 750, and 1150 Hz. The three carrier frequencies are known to generate small, medium, and large effect of ITD threshold measure in NH listeners (Brughera et al., 2013; Hartmann & Wittenberg, 1996; Wightman & Kistler, 1992; Willert et al., 2006). On a given trial, the first interval of the left or right channel of the stereo signal was delayed by the ITD with a 50% probability. Following this, the second interval of the opposite channel was delayed by the ITD. The initial ITD was set to 100 ms, then decreased by two following correct answers and increased by two following mistakes. The procedure continued until the subject made eight mistakes, and the ITD threshold was taken as the average of the last four reversals.

Results

Figure 3 shows the mean ILD threshold (left panel) and ITD threshold (right panel), along with standard error. Both ILD and ITD thresholds were analyzed with two-way repeated-measures ANOVA using two independent variables: listening condition and carrier frequency. For the ILD thresholds, no significant main effects of the listening environment, $F(1, 36) = 0.06$, $p = .80$ and of the carrier frequency, $F(2, 36) = 0.72$, $p = .50$ were observed. Analysis also showed a nonsignificant interaction between independent variables, $F(2, 36) = 0.13$, $p = .88$. For the ITD thresholds, no significant main effects of the listening environment, $F(1, 36) = 0.37$, $p = .55$, but the main effect of the carrier frequency was significant, $F(2, 36) = 5.90$, $p = .006$. There was no significant interaction between independent variables, $F(2, 36) = 0.03$, $p = .98$. Pairwise multiple comparisons with Bonferroni correction showed a significant difference between 350 Hz and 750 Hz for the quiet room ($p = .40$) and the soundproof booth ($p = .38$), as indicated by asterisks.

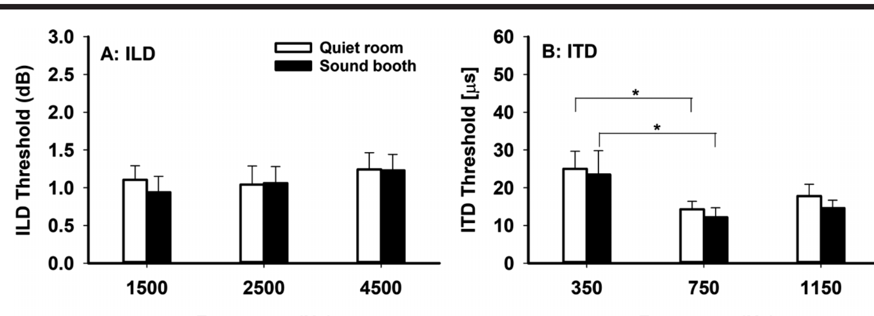
Discussion

This study findings indicated no significant difference in both ILD and ITD thresholds under the two listening environments for all the three carrier frequencies. However, it was noted that the ITD for 350-Hz carrier frequency was significantly larger than the ITDs obtained for 750-Hz carrier frequency in both listening environments.

The information embodied in ITD and ILD cues is known to aid in horizontal plane localization and speech recognition for NH listeners in adverse listening situations, such as cocktail party (Blauert, 1997; Yost et al., 1996). Due to this advantage, ITD and ILD cues have been applied to individuals with bimodal fittings (Choi et al., 2017; Veugen et al., 2017) and bilateral cochlear implants (Grieco-Calub & Litovsky, 2010; Van Deun et al., 2010; Zirn et al., 2019). Zeng et al. (2004) reported that the ILD cues reflect the level differences embedded in the temporal envelope. On the other hand, the ITD cues reflect the differences in the embedded temporal fine structure. Hence, with this information, it can be inferred that in this study the music perception as well as ITD cues both rely on the temporal fine structure cues. Henceforth, the same justification of no significant influence on temporal fine structure for the range of level of testing 20–50 dB SL (Moore & Sek, 2009) as mentioned for music perception experiment also holds good for ITD experiment.

This study provides evidence for using a quiet room for ITD and ILD measurements in individuals with NH. Our results suggest that an online home-based assessment of localization abilities in a quiet room has the potential

Figure 3. Mean interaural level difference (ILD) threshold (left panel) and interaural time difference (ITD) threshold (right panel) with standard errors as a function of carrier frequency in quiet room and soundproof booth listening environments. Significant pairwise multiple-comparison outcomes are indicated by asterisks, indicating that *p < .05.



to be used for individuals with hearing loss and hearing devices.

General Discussion and Future Plans

This study provided data sets of NH listeners for speech perception, music perception, and binaural cue perception experiments under quiet room and soundproof booth listening environments. The findings indicated no significant difference in the two listening environments. Hence, an online testing mode in a quiet room for these three auditory experiments is recommended for research and clinical practice involving NH listeners. Furthermore, with evidence of a quiet room being sufficient for all the three experiments tested, a test battery can be designed for online home-based testing. Such a test battery would result in a comprehensive auditory assessment with added benefits such as low travel and clinic use cost, greater flexibility in accessing the subjects, and flexibility in administering multiple tests. The prerequisite for online administration of auditory experiments may include using a quiet, nonreverberant room with no obvious ambient noise.

One technical concern is that the frequency responses of the different headphones used by each participant may vary; the outcomes of all experiments may be affected. Medwetsky et al. (2020) reported that a significant effect of different earphones on speech perception in noise in listeners with high-frequency hearing loss. Specifically, Sennheiser and Blue Ever Blue earphones provided significantly better speech perception than Apple EarPods (Medwetsky et al., 2020). These findings indicate that our three experiments were affected. We acknowledge this potential effect; however, in this study, no significant difference between quiet and soundproof booths was observed, so the effects of using different headphones are marginal. In the future, a larger study with sufficient sample size under each headphone models can help study the effect of different frequency responses on the outcome measures.

Future studies must also validate these findings in those with hearing loss or with hearing devices, which would aid in catering to the auditory assessment needs of many patients, even in the remote. Remote testing, in contrast to in-lab testing, has several benefits, such as it supports ecological validity since the subjects are tested while they are immersed in real-world environments (Bottalico et al., 2020; Parker et al., 2020). Further, it overcomes geographical and transportation barriers by permitting the subjects to undergo tests in the comfort of their homes (Bobb et al., 2016; de Graaff et al., 2018). Greater flexibility in scheduling appointments and reduced stress associated with the travel planning for in-lab testing is also assured. The approaches for executing remote testing in audiological practice are beyond the scope of the current article and are illustrated in depth by (Peng et al., 2020).

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

This work was supported by the National Institutes of Health under NIH NIDCD R15DC019240 awarded to Yang-Soo Yoon. The authors would like to thank their participants for their patience and continuous support.

References

Akeroyd, M. A. (2006). The psychoacoustics of binaural hearing. *International Journal of Audiology, 45*(Suppl. 1), 25–33. https://doi.org/10.1080/14992020600782626

Akeroyd, M. A. (2014). An overview of the major phenomena of the localization of sound sources by normal-hearing, hearing-impaired, and aided listeners. *Trends Hearing, 18*. https://doi.org/10.1177/2331216514560442

Armstrong, M., Pegg, P., James, C., & Blamey, P. (1997). Speech perception in noise with implant and hearing aid. *The American Journal of Otology, 18*(Suppl. 6), S140–S141.

Balkany, T., Hodges, A., Menapace, C., Hazard, L., Driscoll, C., Gantz, B., Kelsall, D., Luxford, W., McMenomy, S., Neely, J. G., Peters, B., Pillsbury, H., Roberson, J., Schramm, D., Telian, S., Waltzman, S., Westerberg, B., & Payne, S. (2007). Nucleus freedom North American clinical trial. *Otolaryngology—Head & Neck Surgery, 136*(5), 757–762. https://doi.org/10.1016/j.otohns.2007.01.006

Behar, A. (2021). Audiometric tests without booths. *International Journal of Environmental Research and Public Health, 18*(6), 3073. https://doi.org/10.3390/ijerph18063073

Bell, A., & Jedrzejczak, W. W. (2017). The 1.06 frequency ratio in the cochlea: Evidence and outlook for a natural musical semitone. *PeerJ, 5,* e4192. https://doi.org/10.7717/peerj.4192

Bilger, R. (1984). Speech recognition test development. *ASHA Reports, 14,* 2–7.

Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization*. MIT press.

Bobb, M. R., Van Heukelom, P. G., Faine, B. A., Ahmed, A., Messerly, J. T., Bell, G., Harland, K. K., Simon, C., & Mohr, N. M. (2016). Telemedicine provides noninferior research informed consent for remote study enrollment: A randomized controlled trial. *Academic Emergency Medicine, 23*(7), 759–765. https://doi.org/10.1111/acem.12966

Bottalico, P., Codino, J., Cantor-Cutiva, L. C., Marks, K., Nudelman, C. J., Skeffington, J., Shrivastav, R., Jackson-

Menaldi, M. C., Hunter, E. J., & Rubin, A. D. (2020). Reproducibility of voice parameters: The effect of room acoustics and microphones. *Journal of Voice, 34*(3), 320–334. https://doi.org/10.1016/j.jvoice.2018.10.016

Brughera, A., Dunai, L., & Hartmann, W. M. (2013). Human interaural time difference thresholds for sine tones: The high-frequency limit. *The Journal of the Acoustical Society of America, 133*(5), 2839–2855. https://doi.org/10.1121/1.4795778

Choi, J. E., Moon, I. J., Kim, E. Y., Park, H. S., Kim, B. K., Chung, W. H., Cho, Y. S., Brown, C. J., & Hong, S. H. (2017). Sound localization and speech perception in noise of pediatric cochlear implant recipients: Bimodal fitting versus bilateral cochlear implants. *Ear and Hearing, 38*(4), 426–440. https://doi.org/10.1097/aud.0000000000000401

Clopper, C. G., Pisoni, D. B., & Tierney, A. T. (2006). Effects of open-set and closed-set task demands on spoken word recognition. *Journal of the American Academy of Audiology, 17*(05), 331–349. https://doi.org/10.3766/jaaa.17.5.4

Cox, R. M., Alexander, G. C., Taylor, I. M., & Gray, G. A. (1997). The Contour Test of loudness perception. *Ear and Hearing, 18*(5), 388–400. https://doi.org/10.1097/00003446-199710000-00004

de Graaff, F., Huysmans, E., Merkus, P., Theo Goverts, S., & Smits, C. (2018). Assessment of speech recognition abilities in quiet and in noise: A comparison between self-administered home testing and testing in the clinic for adult cochlear implant users. *International Journal of Audiology, 57*(11), 872–880. https://doi.org/10.1080/14992027.2018.1506168

Dietz, A., Buschermöhle, M., Aarnisalo, A. A., Vanhanen, A., Hyyrynen, T., Aaltonen, O., Löppönen, H., Zokoll, M. A., & Kollmeier, B. (2014). The development and evaluation of the Finnish Matrix Sentence Test for speech intelligibility assessment. *Acta Oto-Laryngologica, 134*(7), 728–737. https://doi.org/10.3109/00016489.2014.898185

Ferrari, E., Bosco, P., Calderoni, S., Oliva, P., Palumbo, L., Spera, G., Fantacci, M. E., & Retico, A. (2020). Dealing with confounders and outliers in classification medical studies: The autism spectrum disorders case study. *Artificial Intelligence in Medicine, 108,* 101926. https://doi.org/10.1016/j.artmed.2020.101926

Fowler, S. L., Calhoun, H., & Warner-Czyz, A. D. (2021). Music perception and speech-in-noise skills of typical hearing and cochlear implant listeners. *American Journal of Audiology, 30*(1), 170–181. https://doi.org/10.1044/2020_AJA-20-00116

Galvin, J. J., III, Fu, Q. J., & Nogaki, G. (2007). Melodic contour identification by cochlear implant listeners. *Ear and Hearing, 28*(3), 302–319. https://doi.org/10.1097/01.aud.0000261689.35445.20

Galvin, J. J., III, Fu, Q. J., & Shannon, R. V. (2009). Melodic contour identification and music perception by cochlear implant users. *Annals of the New York Academy of Sciences, 1169*(1), 518–533. https://doi.org/10.1111/j.1749-6632.2009.04551.x

Gfeller, K. E., Olszewski, C., Turner, C., Gantz, B., & Oleson, J. (2006). Music perception with cochlear implants and residual hearing. *Audiology and Neuro-Otology, 11*(Suppl. 1), 12–15. https://doi.org/10.1159/000095608

Grieco-Calub, T. M., & Litovsky, R. Y. (2010). Sound localization skills in children who use bilateral cochlear implants and in children with normal acoustic hearing. *Ear and Hearing, 31*(5), 645–656. https://doi.org/10.1097/AUD.0b013e3181e50a1d

Hartmann, W. M., & Wittenberg, A. (1996). On the externalization of sound images. *The Journal of the Acoustical Society of America, 99*(6), 3678–3688. https://doi.org/10.1121/1.414965

Holmes, A. E., Kricos, P. B., & Kessler, R. A. (1988). A closed-versus open-set measure of speech discrimination in normally hearing young and elderly adults. *British Journal of Audiology, 22*(1), 29–33. https://doi.org/10.3109/03005368809077795

Jerger, J., Speaks, C., & Trammell, J. L. (1968). A new approach to speech audiometry. *Journal of Speech and Hearing Disorders, 33*(4), 318–328. https://doi.org/10.1044/jshd.3304.318

Jones, H. G., Brown, A. D., Koka, K., Thornton, J. L., & Tollin, D. J. (2015). Sound frequency-invariant neural coding of a frequency-dependent cue to sound source location. *Journal of Neurophysiology, 114*(1), 531–539. https://doi.org/10.1152/jn.00062.2015

Jung, K. H., Cho, Y.-S., Cho, J. K., Park, G. Y., Kim, E. Y., Hong, S. H., Chung, W.-H., Won, J. H., & Rubinstein, J. T. (2010). Clinical assessment of music perception in Korean cochlear implant listeners. *Acta Oto-Laryngologica, 130*(6), 716–723. https://doi.org/10.3109/00016480903380521

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics, 49*(3), 227–229. https://doi.org/10.3758/BF03214307

Kalcioglu, M. T., Miman, M. C., Toplu, Y., Yakinci, C., & Ozturan, O. (2003). Anthropometric growth study of normal human auricle. *International Journal of Pediatric Otorhinolaryngology, 67*(11), 1169–1177. https://doi.org/10.1016/s0165-5876(03)00221-0

Kang, R., Nimmons, G. L., Drennan, W., Longnion, J., Ruffin, C., Nie, K., Won, J. H., Worman, T., Yueh, B., & Rubinstein, J. (2009). Development and validation of the University of Washington Clinical Assessment of Music Perception test. *Ear and Hearing, 30*(4), 411–418. https://doi.org/10.1097/AUD.0b013e3181a61bc0

Koka, K., Jones, H. G., Thornton, J. L., Lupo, J. E., & Tollin, D. J. (2011). Sound pressure transformations by the head and pinnae of the adult chinchilla (Chinchilla lanigera). *Hearing Research, 272*(1–2), 135–147. https://doi.org/10.1016/j.heares.2010.10.007

Kollmeier, B., & Wesselkamp, M. (1997). Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. *The Journal of the Acoustical Society of America, 102*(4), 2412–2421. https://doi.org/10.1121/1.419624

Kovacić, D., & Balaban, E. (2009). Voice gender perception by cochlear implantees. *The Journal of the Acoustical Society of America, 126*(2), 762–775. https://doi.org/10.1121/1.3158855

Kuwada, C. A., Bishop, B., Kuwada, S., & Kim, D. O. (2010). Acoustic recordings in human ear canals to sounds at different locations. *Otolaryngology–Head & Neck Surgery, 142*(4), 615–617. https://doi.org/10.1016/j.otohns.2009.12.014

Lucker, J. R. (2017). Auditory processing testing: In the booth versus outside the booth. *Journal of the American Academy of Audiology, 28*(08), 679–684. https://doi.org/10.3766/jaaa.15116

Ma, N., Morris, S., & Kitterick, P. T. (2016). Benefits to speech perception in noise from the binaural integration of electric and acoustic signals in simulated unilateral deafness. *Ear and Hearing, 37*(3), 248–259. https://doi.org/10.1097/aud.0000000000000252

Maclennan-Smith, F., Swanepoel, D. W., & Hall, J. W., III. (2013). Validity of diagnostic pure-tone audiometry without a sound-treated environment in older adults. *International Journal of Audiology, 52*(2), 66–73. https://doi.org/10.3109/14992027.2012.736692

Margolis, R. H., & Madsen, B. (2015). The acoustic test environment for hearing testing. *Journal of the American Academy of Audiology, 26*(09), 784–791. https://doi.org/10.3766/jaaa.14072

Medwetsky, L., Kelly, K., & Bakke, M. (2020). Earphone models for iPhones: Surprising results when used with a hearing app. *Hearing Review, 27*(1), 22–25.

Miller, C. W., Bates, E., & Brennan, M. (2016). The effects of frequency lowering on speech perception in noise with adult hearing-aid users. *International Journal of Audiology, 55*(5), 305–312. https://doi.org/10.3109/14992027.2015.1137364

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41*(5), 329–335. https://doi.org/10.1037/h0062491

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America, 27*(2), 338–352. https://doi.org/10.1121/1.1907526

Moore, B. C. J., & Sek, A. (2009). Development of a fast method for determining sensitivity to temporal fine structure. *International Journal of Audiology, 48*(4), 161–171. https://doi.org/10.1080/14992020802475235

Naylor, G. (2016). Theoretical issues of validity in the measurement of aided speech reception threshold in noise for comparing nonlinear hearing aid systems. *Journal of the American Academy of Audiology, 27*(07), 504–514. https://doi.org/10.3766/jaaa.15093

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America, 95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Oppenheim, A. V. (1978). *Applications of digital signal processing*. Prentice-Hall. https://ui.adsabs.harvard.edu/abs/1978ph...bookR....O

Panday, S., Kathard, H., Pillay, M., & Wilson, W. (2018). The internal and external consistency of a speech reception threshold test for isiZulu speakers with normal hearing sensitivity. *The South African Journal of Communication Disorders, 65*(1), e1–e8. https://doi.org/10.4102/sajcd.v65i1.556

Pantev, C., Ross, B., Fujioka, T., Trainor, L. J., Schulte, M., & Schulz, M. (2003). Music and learning-induced cortical plasticity. *Annals of the New York Academy of Sciences, 999*(1), 438–450. https://doi.org/10.1196/annals.1284.054

Parker, A., Slack, C., & Skoe, E. (2020). Comparisons of auditory brainstem responses between a laboratory and simulated home environment. *Journal of Speech, Language, and Hearing Research, 63*(11), 3877–3892. https://doi.org/10.1044/2020_JSLHR-20-00383

Peng, Z. E., Buss, E., Shen, Y., Bharadwaj, H., Stecker, G. C., Beim, J. A., Bosen, A. K., Braza, M., Diedesch, A. C., Dorey, C. M., Dykstra, A. R., Freyman, R., Gallun, F. J., Goldsworthy, R. L., Gray, L., Hoover, E. C., Ihlefeld, A., Koelewijn, T., Kopun, J. G., & Waz, S. (2020). Remote testing for psychological and physiological acoustics: Initial report of the P&P Task Force on Remote Testing. *Proceedings of Meetings on Acoustics, 42*(1), 050009. https://doi.org/10.1121/2.0001409

Phatak, S. A., Yoon, Y. S., Gooler, D. M., & Allen, J. B. (2009). Consonant recognition loss in hearing impaired listeners. *The Journal of the Acoustical Society of America, 126*(5), 2683–2694. https://doi.org/10.1121/1.3238257

Ping, L., Yuan, M., & Feng, H. (2012). Musical pitch discrimination by cochlear implant users. *The Annals of Otology, Rhinology, and Laryngology, 121*(5), 328–336. https://doi.org/10.1177/000348941212100508

Plapous, C., Marro, C., & Scalart, P. (2006). Improved signal-to-noise ratio estimation for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing, 14*(6), 2098–2108. https://doi.org/10.1109/TASL.2006.872621

Roy, A. T., Vigeant, M., Munjal, T., Carver, C., Jiradejvong, P., & Limb, C. J. (2015). Reverberation negatively impacts musical sound quality for cochlear implant users. *Cochlear Implants International, 16*(Suppl. 3), S105–S113. https://doi.org/10.1179/1467010015Z.000000000262

Sares, A. G., Foster, N. E. V., Allen, K., & Hyde, K. L. (2018). Pitch and time processing in speech and tones: The effects of musical training and attention. *Journal of Speech, Language, and Hearing Research, 61*(3), 496–509. https://doi.org/10.1044/2017_jslhr-s-17-0207

Sayles, M., & Winter, I. M. (2008). Reverberation challenges the temporal representation of the pitch of complex sounds. *Neuron, 58*(5), 789–801. https://doi.org/10.1016/j.neuron.2008.03.029

Shafiro, V., Hebb, M., Walker, C., Oh, J., Hsiao, Y., Brown, K., Sheft, S., Li, Y., Vasil, K., & Moberly, A. C. (2020). Development of the basic auditory skills evaluation battery for online testing of cochlear implant listeners. *American Journal of Audiology, 29*(3S), 577–590. https://doi.org/doi:10.1044/2020_AJA-19-00083

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303–304. https://doi.org/10.1126/science.270.5234.303

Shukla, B., Rao, B., Saxena, U., & Verma, H. (2018). Measurement of speech in noise abilities in laboratory and real-world noise. *Indian Journal of Otology, 24*(2), 109–113. https://doi.org/10.4103/indianjotol.INDIANJOTOL_134_17

Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *Journal of the American Academy of Audiology, 26*(2), 183–196. https://doi.org/10.3766/jaaa.26.2.7

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature, 416*(6876), 87–90. https://doi.org/10.1038/416087a

Smits, C., De Sousa, K. C., & Swanepoel, D. W. (2021). An analytical method to convert between speech recognition thresholds and percentage-correct scores for speech-in-noise tests. *The Journal of the Acoustical Society of America, 150*(2), 1321–1331. https://doi.org/10.1121/10.0005877

Stecker, G. C. (2014). Temporal weighting functions for interaural time and level differences. IV. Effects of carrier frequency. *The Journal of the Acoustical Society of America, 136*(6), 3221–3232. https://doi.org/10.1121/1.4900827

Trehub, S. E., Cohen, A. J., Thorpe, L. A., & Morrongiello, B. A. (1986). Development of the perception of musical relations: Semitone and diatonic structure. *Journal of Experimental Psychology: Human Perception and Performance, 12*(3), 295–301. https://doi.org/10.1037//0096-1523.12.3.295

Van Deun, L., van Wieringen, A., Scherf, F., Deggouj, N., Desloovere, C., Offeciers, F. E., Van de Heyning, P. H., Dhooge, I. J., & Wouters, J. (2010). Earlier intervention leads to better sound localization in children with bilateral cochlear implants. *Audiology & Neurotology, 15*(1), 7–17. https://doi.org/10.1159/000218358

Varnet, L., Wang, T., Peter, C., Meunier, F., & Hoen, M. (2015). How musical expertise shapes speech perception: Evidence from auditory classification images. *Scientific Reports, 5*(1), 14489. https://doi.org/10.1038/srep14489

Veugen, L. C. E., Chalupper, J., Mens, L. H. M., Snik, A. F. M., & van Opstal, A. J. (2017). Effect of extreme adaptive frequency compression in bimodal listeners on sound localization and speech perception. *Cochlear Implants International, 18*(5), 266–277. https://doi.org/10.1080/14670100.2017.1353762

Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America, 91*(3), 1648–1661. https://doi.org/10.1121/1.402445

Willert, V., Eggert, J., Adamy, J., Stahl, R., & Körner, E. (2006). A probabilistic model for binaural sound localization. *IEEE*

Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 36(5), 982–994. https://doi.org/10.1109/tsmcb.2006.872263

Wilson, R. H., McArdle, R., Watts, K. L., & Smith, S. L. (2012). The Revised Speech Perception in Noise test (R-SPIN) in a multiple signal-to-noise ratio paradigm. *Journal of the American Academy of Audiology, 23*(8), 590–605. https://doi.org/10.3766/jaaa.23.7.9

Xu, L., & Pfingst, B. E. (2008). Spectral and temporal cues for speech recognition: Implications for auditory prostheses. *Hearing Research, 242*(1–2), 132–140. https://doi.org/10.1016/j.heares.2007.12.010

Yoon, Y. S., Boren, C. M., & Diaz, B. (2021). Effect of realistic test conditions on spectral and temporal processing in normal-hearing listeners. *American Journal of Audiology, 30*(1), 160–169. https://doi.org/10.1044/2020_aja-20-00120

Yost, W. A., Dye, R. H., & Sheft, S. (1996). A simulated "cocktail party" with up to three sound sources. *Perception & Psychophysics, 58*(7), 1026–1036. https://doi.org/10.3758/BF03206830

Zeng, F.-G., Nie, K., Liu, S., Stickney, G., Rio, E. D., Kong, Y.-Y., & Chen, H. (2004). On the dichotomy in auditory perception between temporal envelope and fine structure cues (L). *The Journal of the Acoustical Society of America, 116*(3), 1351–1354. https://doi.org/10.1121/1.1777938

Zirn, S., Angermeier, J., Arndt, S., Aschendorff, A., & Wesarg, T. (2019). Reducing the device delay mismatch can improve sound localization in bimodal cochlear implant/hearing-aid users. *Trends in Hearing, 23*. https://doi.org/10.1177/2331216519843876