Characteristics of Real-World Signal-to-noise Ratios and Speech Listening Situations of Older Adults with Mild-to-Moderate Hearing Loss

Yu-Hsiang Wu[1], Elizabeth Stangl[1], Octav Chipara[2], Syed Shabih Hasan[2], Anne Welhaven[3], and Jacob Oleson[3]

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

[2]Department of Computer Science, The University of Iowa

[3]Department of Biostatistics, The University of Iowa

Conflicts of Interest and Source of Funding:

Yu-Hsiang Wu and Octav Chipara are currently receiving grants from the National Institute on Disability and Rehabilitation Research. Jacob Oleson is currently receiving grants from the National Institute on Deafness and Other Communication Disorders, National Heart, Lung, and Blood Institute, Department of Defense, Centers for Disease Control and Prevention, Fogarty International Center, and the Iowa Department of Public Health. The current research was supported by NIH/NIDCD R03DC012551 and the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR, grant number 90RE5020-01-00).

Address correspondence to Yu-Hsiang Wu

125C SHC, Department of Communication Sciences and Disorders, The University of Iowa,
Iowa City, IA 52242, USA.

E-mail: yu-hsiang-wu@uiowa.edu

Fax number: 319-335-8851 Telephone number: 319-335-8728

ABSTRACT

Objectives: The first objective was to determine the relationship between speech level, noise level, and signal-to-noise ratio (SNR), as well as the distribution of SNR, in real-world situations wherein older adults with hearing loss are listening to speech. The second objective was to develop a set of Prototype Listening Situations (PLSs) that describe the speech level, noise level, SNR, availability of visual cues, and locations of speech and noise sources of typical speech listening situations experienced by these individuals.

Design: Twenty older adults with mild-to-moderate hearing loss carried digital recorders for 5 to 6 weeks to record sounds for 10 hours per day. They also repeatedly completed *in-situ* surveys on smartphones several times per day to report the characteristics of their current environments, including the locations of the primary talker (if they were listening to speech) and noise source (if it was noisy) and the availability of visual cues. For surveys where speech listening was indicated, the corresponding audio recording was examined. Speech-plus-noise and noise-only segments were extracted and the SNR was estimated using a power subtraction technique. SNRs and the associated survey data were subjected to cluster analysis to develop PLSs.

Results: The speech level, noise level, and SNR of 894 listening situations were analyzed to address the first objective. Results suggested that, as noise levels increased from 40 to 74 dBA, speech levels systematically increased from 60 to 74 dBA and SNR decreased from 20 to 0 dB. Most SNRs (62.9%) of the collected recordings were between 2 and 14 dB. Very noisy situations that had SNRs below 0 dB comprised 7.5% of the listening situations. To address the second objective, recordings and survey data from 718 observations were analyzed. Cluster analysis suggested that the participants’ daily listening situations could be grouped into 12 clusters (i.e., 12 PLSs). The most frequently occurring PLSs were characterized as having the talker in front of the listener with visual cues available, either in quiet or in diffuse noise. The mean speech level of the PLSs that described quiet situations was 62.8 dBA, and the mean SNR of the PLSs that represented noisy environments was 7.4 dB (speech = 67.9 dBA). A subset of observations (n = 280), which was obtained by excluding the data collected from quiet environments, was further used to develop PLSs that represent noisier situations. From this subset, two PLSs were identified. These two PLSs had lower SNRs (mean = 4.2 dB), but the most frequent situations still involved speech from in front of the listener in diffuse noise with visual cues available.

Conclusions: The current study indicated that visual cues and diffuse noise were exceedingly common in real-world speech listening situations, while environments with negative SNRs were relatively rare. The characteristics of speech level, noise level, and SNR, together with the PLS information reported by the current study, can be useful for researchers aiming to design ecologically-valid assessment procedures to estimate real-world speech communicative functions for older adults with hearing loss.

Key words: hearing loss; hearing aid; signal-to-noise ratio; real world

INTRODUCTION

In order to improve quality of life for individuals with hearing impairment, it is vital for hearing healthcare professionals to decide if a certain hearing aid intervention, such as an advanced feature or a new fitting strategy, provides a better outcome than an alternate intervention. Although evaluating intervention benefit in the real world is important, hearing aid outcomes are often assessed under controlled conditions in laboratory (or clinical) settings using measures such as speech recognition tests. To enhance the ability of contrived laboratory assessment procedures to predict hearing aid outcomes in the real world, researchers aim to use test materials and settings that simulate the real world in order to be ecologically-valid (Keidser 2016). In order to create ecologically-valid test materials and environments, the communication activities and environments of individuals with hearing loss must first be characterized.

Several studies have attempted to characterize daily listening situations for adults with hearing loss (Jensen & Nielsen 2005; Wagener et al. 2008; Wu & Bentler 2012; Wolters et al. 2016). For example, Jensen and Nielsen (2005) and Wagener et al. (2008) asked experienced hearing aid users to record sounds in typical real-world listening situations. The recordings were made by portable audio recorders and bilateral ear-level microphones. In Jensen and Nielsen (2005), the research participants completed *in-situ* (i.e., real-world and real-time) surveys in paper-and-pencil journals to describe each listening situation and its importance using the ecological momentary assessment (EMA) methodology (Shiffman et al. 2008). The survey provided seven listening situation categories (e.g., conversation with several persons). In Wagener et al. (2008), the research participants reviewed their own recordings in the laboratory and described and estimated the importance and frequency of occurrence of each listening situation. The listening situations were then categorized into several groups based on the participants' descriptions (e.g., conversation with background noise, two people). For both studies, the properties of each listening situation category, including importance, frequency of occurrence, and overall sound level, were reported. In another study, Wu and Bentler (2012) compared listening demand for older and younger adults by asking individuals with hearing loss to carry noise dosimeters to measure their daily sound levels. Participants were also asked to complete *in-situ* surveys in paper-and-pencil journals to describe their listening activities and environments. The survey provided six listening activity categories (e.g., conversation in a group more than three people) and five environmental categories (e.g., moving traffic), resulting in 30 unique listening situations. The frequency of occurrence of each listening situation as well as the mean overall sound level of several frequent situations were reported.

More recently, Wolters et al. (2016) developed a Common Sound Scenarios framework using the data from the literature. Specifically, information regarding the listener's intention and task, as well as the frequency of occurrence, importance, and listening difficulty of the listening situation, was extracted or estimated from previous research. Fourteen scenarios, which are grouped into three intention categories (speech communication, focused listening, and non-specific listening), were developed.

Speech listening and signal-to-noise ratio

Among all types of listening situations, it is arguable that speech listening is the most important. Although previous research (Jensen & Nielsen 2005; Wagener et al. 2008; Wu & Bentler 2012) reported the overall sound level of typical real-world listening environments, none provided information regarding the signal-to-noise ratio (SNR) of speech listening situations. SNR is highly relevant to speech understanding (Plomp 1986) and has a strong effect on hearing aid outcome (Walden et al. 2005; Wu & Bentler 2010a). Historically, Pearsons et al. (1977) was one of the first studies to examine SNRs of real-world speech listening situations. In that study audio was recorded during face-to-face communication in various locations including homes, public places, department stores, and trains using a microphone mounted near the ear on an eyeglass frame. Approximately 110 measurements were made. For each measurement, the speech level and SNR were estimated. The results indicated that when the noise level was below 45 dBA, the speech level at the listener’s ear remained at a constant 55 dBA. As noise level increased, speech level increased systematically at a linear rate of 0.6 dB/dB. The SNR decreased to 0 dB when the noise reached 70 dBA. Approximately 15.5% of the measurements had SNRs below 0 dB.

The data reported by Pearsons et al. (1977) has been widely used to determine the SNR of speech-related tests for individuals with normal hearing or with hearing loss. However, the participants in Pearsons et al. were adults with normal hearing. More recently, Smeds et al. (2015) estimated the SNRs of real-world environments encountered by hearing aid users with moderate hearing loss using the audio recordings made by Wagener et al. (2008). The speech level was estimated by subtracting the power of the noise signal from the power of the speech-plus-noise signal. A total of 72 pairs of SNRs (from two ears) were derived. The results were not completely in line with those reported by Pearsons and colleagues (1977). Smeds et al. (2015) found that there were very few negative SNRs (approximately 4.2% and 13.7% for the better and worse SNR ears, respectively); most SNRs had positive values. At a given noise level, the SNRs estimated by Smeds et al. (2015) were 3 to 5 dB higher than those reported by Pearsons et al. (1977), especially in situations with low-level noise. In quiet environments (median noise = 41 dBA), the median speech level reported by Smeds et al. was 63 dBA, which was higher than that reported by Pearsons et al. (55 dBA). Smeds and her colleagues suggested that the discrepancy between the two studies could be due to the difference in research participants (hearing aid users vs. normal-hearing adults) and the ways that recordings were collected and analyzed.

Visual cues and speech/noise location

Other than SNR, there are real-world factors that can impact speech understanding and hearing aid outcome and that should be considered in ecologically-valid laboratory testing. For example, visual cues, such as lip-reading, are often available in real-world listening situations. Visual cues have a strong effect on speech recognition (Sumby & Pollack 1954) and have the potential to influence hearing aid outcomes (Wu & Bentler 2010a, b). Therefore, some speech recognition materials can be presented in an audio-visual modality (e.g., the Connected Speech Test; Cox et al. 1987a). Another example is the location of speech and noise sources. Because this factor can impact speech understanding (e.g., Ahlstrom et al. 2009) and the benefit from hearing aid technologies (Ricketts 2000; Ahlstrom et al. 2009; Wu et al. 2013), researchers have tried to use realistic speech/noise sound-field configurations in laboratory testing. For example, in a study designed to examine the effect of asymmetric directional hearing aid fitting, Hornsby and Ricketts (2007) manipulated the location of speech (front or side) and noise sources (surround or side) to simulate various real-world speech listening situations.

Only a few studies have examined the availability of visual cues and speech/noise locations in real-world listening situations (Walden et al. 2004; Wu & Bentler 2010b). Wu and Bentler (2010b) asked adults with hearing loss to describe the characteristics of listening situations wherein the primary talker was in front of them using repeated *in-situ* surveys. The research participants reported the location of noise and the availability of visual cues in each situation. However, because the purpose of Wu and Bentler (2010b) was to examine the effect of visual cues on directional microphone hearing aid benefit, the descriptive statistics of the listening situation properties were not reported. In a study designed to investigate hearing aid users' preference between directional and omnidirectional microphones, Walden et al. (2004) asked adult hearing aid users to report microphone preference as well as the properties of major active listening situations using *in-situ* surveys. The questions asked in the survey categorized the listening environments into 24 unique situations. The categories were arranged according to binary representations of five acoustic factors, including background noise (present/absent), speech location (front/others), and noise location (front/others). The frequency of occurrence of each of the 24 unique situations was reported. The most frequently encountered type of listening situations involved the speech from in front of the listener and background noise arising from locations other than the front.

Prototype Listening Situations

The term Prototype Listening Situations, or PLSs, refers to a set of situations that can represent a large proportion of the everyday listening situations experienced by individuals. The concept of a PLS was first introduced by Walden (1997). In particular, Walden et al. (1984) conducted a factor analysis on a self-report questionnaire and found that there were four dimensions of hearing aid benefit; one for each unique listening situation. Those unique listening situations included listening to speech in quiet, in background noise, and with reduced (e.g., visual) cues, as well as listening to environmental sounds. Walden (1997) termed these unique listening situations as “PLSs.” Walden and other researchers (Cox et al. 1987b) suggested that hearing aids should be evaluated in PLSs so that test results can generalize to the real world. However, the PLSs specified by Walden (1997) do not describe important acoustic characteristics such as speech level, noise level, and SNR. Further, although previous research has examined the properties of real-world communication situations for adults with hearing loss in terms of SNR (Pearsons et al. 1977; Smeds et al. 2015), availability of visual cues, and speech/noise configuration (Walden et al. 2004), these data were individually collected by different studies. Therefore, no empirical data are available for developing a set of PLSs that can represent typical speech listening situations and can be used to create ecologically-valid speech-related laboratory testing.

Research objectives

The current study had two objectives. The first objective was to determine the relationship between speech level, noise level, and SNR, as well as the distribution of SNR, in real-world speech listening situations for adults with hearing loss, as the data reported by Pearsons et al. (1977) and Smeds et al. (2015) are not consistent. The second objective was to develop a set of PLSs that relate to speech listening and describe the (1) SNR, (2) availability of visual cues, and (3) locations of speech and noise sources in the environments that are frequently encountered by adults with hearing loss. In accordance with the PLSs described by Walden (1997), the PLSs in the current study do not characterize the listener’s intention (e.g., conversation vs. focused listening) or the type of listening environment (e.g., restaurant vs. car).

However, unlike Walden's PLSs that include non-speech sound listening situations, the PLSs in the current study only focus on speech listening situations.

The current study was part of a larger project comparing the effect of noise reduction features in premium-level and basic-level hearing aids. The participants were older Iowa and Illinois residents with symmetric mild-to-moderate hearing loss. The participants were fit bilaterally with experimental hearing aids. During the field trial of the larger study, the participants carried digital audio recorders to continuously record environmental sounds, and they repeatedly completed *in-situ* surveys on smartphones to report the characteristics of the listening situations. SNRs were derived using the audio recordings. SNRs and survey data were then used to develop the PLSs.

MATERIALS AND METHODS

Participants

Twenty participants (8 males and 12 females) were recruited from the community. Their ages ranged from 65 to 80 years with a mean of 71.1 years. The participants were eligible for inclusion in the larger study if their hearing loss met the following criteria: (1) postlingual, bilateral, sensorineural type of hearing loss (air-bone gap < 10 dB); (2) pure-tone average across 0.5, 1, 2, and 4 kHz between 25 and 60 dB HL (ANSI 2010); and (3) hearing symmetry within 20 dB for all test frequencies. The larger study focused on mild-to-moderate hearing loss because of its high prevalence (Lin et al. 2011). The mean pure-tone thresholds are shown in Figure 1. All participants were native English speakers. Upon entering the study, 15 participants had previous hearing aid experience. A participant was considered an experienced user if he/she had at least one year of prior hearing aid experience immediately preceding the study. While 20 participants completed the study, two participants withdrew from the study due to scheduling conflicts (n = 1) or unwillingness to record other people's voices (n = 1).

Hearing aids and fitting

In the larger study, participants were fit with two commercially-available behind-the-ear hearing aids. One model was a more-expensive, premium-level device and the other was a less-expensive, basic-level device. The hearing aids were coupled to the participants' ears bilaterally using slim tubes and custom canal earmolds with clinically-appropriate vent sizes. The devices were programmed based on the second version of the National Acoustic Laboratory nonlinear prescriptive formula (NAL-NL2, Keidser et al. 2011) and were fine-tuned according to the comments and preferences of the participants. The noise reduction features, which included directional-microphone and single-microphone noise reduction algorithms, were manipulated (on vs. off) to create different test conditions. All other features (e.g., wide dynamic range compression, adaptive feedback suppression and low-level expansion) remained active at default settings. The volume control was disabled.

Audio recorder

To derive the SNR, the Language Environment Analysis (LENA) digital language processor (DLP) system was used to record environmental sounds. The LENA system is designed for assessing the language-learning environments of children (e.g., VanDam et al. 2012) and the LENA DLP is a miniature, light-weight, compact, and easy-to-use digital audio recorder. The microphone is integrated into the case of the DLP. During the field trial of the study, the DLP was placed in a carrying pouch that had an opening for the microphone port. The pouch was worn around the participants' necks so that the microphone laid at chest height, faced outward, and was not obscured by clothing. The LENA DLP was selected due to its superior portability and usability. Audio recorders that are easy to carry and use were required because audio data was collected over a longer period (weeks) to better characterize real-world listening situations that differ considerably between and within individuals. Note that although the LENA system includes software that can automatically label recording segments offline according to different auditory categories, the results generated by the LENA software were not used in the current study.

The electroacoustic characteristics of three LENA DLPs, which consisted of 10% of the DLPs used in the study, were examined in a sound-treated booth. A white noise and a pink noise were used as stimuli and both generated similar results. Figure 2A shows the one-third octave band frequency response averaged across the three DLPs relative to the response of a Larson-Davis 2560 ½ inch microphone. Although the response of the DLPs is higher than the reference microphone by 6.3 dB at 6 kHz, the response is fairly flat (± 2 dB) between 100 and 3000 Hz. Figure 2B shows the broadband sound level measured using the DLPs (averaged across two stimuli and three DLPs) as a function of the actual level. It is evident from the figure that the DLP has an output limiting algorithm for sounds higher than approximately 80 dBA and a low-level expansion algorithm for sounds lower than approximately 50 dBA. The expansion ratio is approximately 0.4:1. The effect of the expansion was taken into account when analyzing data (see the data preparation section below). The DLP is fairly linear for sounds between 50 and 80 dBA. Due to the noise floor of the device, the lowest level of sound that the DLP can measure is

40 dBA.

In-situ survey

The EMA (i.e., the ecological momentary assessment) methodology was used to collect the information regarding availability of visual cues and the speech/noise location of real-world listening situations. EMA employs recurring assessments or surveys to collect information about participants' recent experiences during or right after they occur in the real world (Shiffman et al. 2008). In the current study, the EMA was implemented using Samsung Galaxy S3 smartphones. Specifically, smartphone application software (i.e., app) was developed to deliver electronic surveys (Hasan et al. 2013). During the field trial, the participants carried smartphones with them in their daily lives. The phone software prompted the participants to complete surveys at randomized intervals approximately every two hours within a participant's specified time window (e.g., between 8 am and 9 pm). The 2-hr inter-prompt interval was selected because it seemed to be a reasonable balance between participant burden, compliance, and the amount of data that would be collected (Stone et al. 2003). The participants were also encouraged to initiate a survey whenever they had a listening experience they wanted to describe. Participants were instructed to answer survey questions based on their experiences during the past five minutes. This short time window was selected to minimize recall bias. The survey assessed the type of listening activity ("*What were you listening to?*") and provided seven options for the participants to select (conversations ≤ 3 people/conversations > 4 people/live speech listening/media speech listening/phone/non-speech signals listening/not actively listening). The participants were instructed to only select one activity in a given survey. If involved in more than one activity (e.g. talking to friend while watching TV), the participants were asked to select the activity that happened most of the time during the previous five minutes. Selection of only the primary activity when completing a survey stemmed from a goal of the larger study to develop algorithms that can use audio recordings to automatically predict listening activities reported by participants. The survey also assessed the type of listening environment (*"Where were you?"*, home ≤ 10 people/indoors other than home ≤ 10 people/indoors crowd of people > 10 people/outdoors/traffic). The listening activity and environment questions were adapted from Wu and Bentler (2012). Whenever applicable, the survey questions then assessed the location of speech signals ("*Where was the talker most of the time?*", front/side/back), availability of visual cues ("*Could you see the talker's face?*", almost always/sometimes/no), noisiness level ("*How noisy was it*?", quiet/somewhat noisy/noisy/very noisy), and location of noise ("*Where was the noise most of the time?*", all around/front/side/back). In the survey, the participants also answered a question regarding hearing aid use during that listening event (*yes/no*). For all questions, the participants tapped a button on the smartphone screen to indicate their responses. The questions were presented adaptively such that certain answers determined whether follow-up questions would be elicited. For example, if a participant answered "quiet" in the noisiness question, the noise location question would not be presented and "N/A" (i.e., not applicable) would be assigned as the answer. After the participants completed a survey, the answers to the questions and the time information were saved in the smartphone. The survey was designed for the larger study but only the questions that are relevant to the current study are reported in this paper. See Hasan et al. (2014) for the complete set of survey questions.

Procedures

The study was approved by the Institutional Review Board at the University of Iowa.

After agreeing to participate and signing the consent form, the participants' hearing thresholds were measured using pure-tone audiometry. If the participant met all of the inclusion criteria, training regarding the use of the LENA DLP was provided. Attention was focused on instructing the participants on how to wear the DLP, especially regarding the orientation of the microphone and the pouch (e.g., to always keep the microphone facing outward and not under clothing). The participants were asked to wear the DLP during their specified time window in which the smartphone delivered surveys. The storage capacity of a DLP is 16 hours, so the participants were instructed to wear a new DLP each day. Each of the DLPs were labeled with the day of the week corresponding to the day that it was to be worn. If they encountered a confidential situation, the participants were allowed to take off the DLP. The participants were instructed to log the time(s) when the DLP was not worn so these data would not be analyzed.

Demonstrations of how to work and care for the smartphone, as well as taking and initiating surveys, were also provided. The participants were instructed to respond to the auditory/vibrotactile prompts to take surveys whenever it was possible and within reason (e.g., not while driving). Participants were also encouraged to initiate a survey during or right after they experienced a new listening experience lasting longer than 10 min. Each participant was given a set of take-home written instructions detailing how to use and care for the phone, as well as when and how to take the surveys. Once all of the participants' questions had been answered and they demonstrated competence in the ability to perform all of the related tasks, they were sent home with three DLPs and one smartphone and began a three-day practice session. The participants returned to the laboratory after the practice session. If a participant misunderstood any of the EMA or DLP related tasks during the practice session, they were re-instructed on how to properly use the equipment or take the surveys.

Next, the hearing aids were fit and the field trial of the larger study began. In total there were four test conditions in the larger study (2 hearing aid models x 2 feature settings). Each condition lasted five weeks and the assessment week in which participants carried DLPs and smartphones was in the fifth week. After the fourth condition, the participants randomly repeated one of the four test conditions to examine the repeatability of the EMA data, which was another purpose of the larger study. Six participants of the current study, including one experienced hearing aid user, also completed an optional unaided condition. Therefore, each participant's audio recordings and EMA survey data were collected in five to six weeks across all test conditions of the larger study. Even though the data were collected in conditions that varied in hearing aid model (premium- vs. basic-level), feature status (on vs. off), and hearing aid use (unaided vs. aided), it was determined *a priori* that the data would be pooled together for analysis, as the effect of hearing aid on the characteristics of the listening situations was not the focus of the current study. More importantly, pooling the data obtained under rather different hearing aid conditions would make the findings of the current study more generalizable than had they been obtained under just a single condition. Similarly, although the manner by which a survey was initiated varied (app-initiated vs. participant-initiated), the survey data collected using both manners would be pooled. The total involvement of participation in the larger study lasted approximately six to eight months. Monetary compensation was provided to the participants upon completion of the study.

Data preparation

Prior to analysis, research assistants manually prepared the audio recordings made by the LENA DLP and the EMA survey data collected by smartphones. The EMA survey data were inspected first. Surveys in which the participants indicated that they were not listening to speech and surveys of phone conversations (i.e., conversational partner's speech could not be recorded) were eliminated. For the rest of the surveys, the audio recording five minutes prior to the participant conducting the survey was extracted. Research assistants then listened to the 5-min recording and judged if it contained too many artifacts (e.g., the DLP was covered by the clothing and recorded rubbing sounds) and was unanalyzable. If the recording was analyzable, the research assistants then tried to identify the participant's voice and the speech sounds that the participant was listening to. If they judged that the participant was actively engaged in a conversation or listening to the speech, the research assistants identified up to three pairs of recording segments that contained (1) speech-plus-noise and (2) noise-only signals from the 5-min recordings. The criteria for selecting segments were that speech-plus-noise and noise-only segments should be adjacent in time and the duration of each segment must be at least two seconds. Also, the three segment pairs should be spread over the 5-min recording so that the SNR could be more accurately estimated. Each segment was then extracted as its own sound file and saved for further analysis. If it was not possible to find speech-plus-noise or noise-only segments that were longer than two seconds, the 5-min recordings were discarded.

When identifying the speech signals for media listening situations (e.g., TV or radio), a special rule was applied: the speech from the media was not treated as the target signal. Instead, the research assistants attempted to identify if the participants engaged in conversations during the media listening situation. If the participant did, speech from their conversation partners was treated as the target signal and media and environmental sounds were considered noise. In other words, only live-speech listening situations were analyzed. This special rule was used because previous studies of Pearsons et al. (1977) and Smeds et al. (2015) characterized live-speech listening situations. Focusing on similar situations allows comparison of the present study to the literature. If the 5-min recording contained only media sounds, the recording was discarded and no further analysis was conducted.

To estimate the SNR, the power subtraction technique described by Smeds et al. (2015) was used. Specifically, the long-term RMS level of each extracted segment was converted to an absolute sound level using a correction factor that was obtained from the calibration stage of the current study. The calculations were performed on the broadband, A-weighted signals. For segments that had levels lower than 50 dBA, the sound level was adjusted to compensate for the effect of the low-level expansion algorithm of the DLP, using an expansion ratio of 0.4:1 (Figure 2B). Next, for a given pair of speech-plus-noise and noise-only segments, the power of speech was estimated by subtracting the power of the noise-only segment from the power of the speech-plus-noise segment. SNR was then computed from the power of the noise-only segment and the estimated speech power. See Smeds et al. (2015) for more details about the assumptions and limitations of this technique. For a given 5-min recording, up to three sets of speech level, noise level, and SNR were derived. The data across these sets were averaged (each variable individually) and saved with the data of the corresponding EMA survey.

RESULTS

A total of 894 5-min recordings were analyzed and 2,336 pairs of speech-plus-noise and noise-only segments were extracted. The average durations of the speech-plus-noise and noise-only segments were 3.0 sec (SD = 1.5) and 2.9 sec (SD = 2.4), respectively. Among all of the 4,672 segments, 937 segments (20.1%) were adjusted for the effect of the DLP's low-level expansion algorithm, with two-thirds of them (n = 602) being noise-only segments. As mentioned above, the data from the same 5-min recordings were averaged. Therefore, a total of 894 sets of speech level, noise level, and SNR, together with the data from the corresponding EMA surveys, were available for analysis. Among the 894 surveys, 623 (69.7%) were prompted by the phone application software and the remaining 271 (30.3%) were initiated by the participants.

Recall that the data of the current study were collected in various hearing aid conditions of the larger study. The manner that a survey was initiated varied too. Further, 15 participants had previous hearing aid experience while five participants were new users. Although it was determined *a priori* that all data would be pooled together for analysis, it is of interest to examine if hearing aid, survey, and participant characteristics could affect the properties of the listening situations. To this end, a linear mixed-effects regression model that included a random intercept to account for multiple observations per participant (Fitzmaurice et al. 2011) was conducted to examine the effect of hearing aid model (premium vs. basic), hearing aid noise reduction feature setting (on vs. off), use of hearing aids when completing surveys (aided vs. unaided), survey type (app-initiated vs. participant-initiated), and hearing aid experience (experienced users vs. new users) on SNR. The results indicated that the SNR was higher with basic-level (10.0 dB) than premium-level (8.6 dB) models ($p = 0.02$), was higher in the unaided (10.3 dB) than aided (8.7 dB) situations ($p = 0.002$), and was higher in the app-initiated (9.4 dB) than participant-initiated

(8.7 dB) surveys ($p = 0.002$). The effects of feature status (on: 9.9 dB; off: 8.9 dB) and hearing aid experience (experienced users: 8.7 dB; new users: 9.4 dB) were not significant.

Speech level, noise level, and SNR

Gray circles in Figure 3A show speech levels and noise levels of the 894 listening situations. The diagonal solid gray line represents where the speech level was equal to the noise level. To determine the relationship between speech level and noise level, speech level data were fit as the dependent variable using a linear mixed-effects regression model with a random intercept and a random slope for noise level. Both linear and quadratic terms of noise level were included in the model to account for the nonlinear trajectory seen in Figure 3A. The results indicated that the effects of the linear and quadratic terms of noise level were both significant (both $p < 0.0001$), suggesting that speech level systematically increased as noise level increased, and that the effect of noise level on speech level depends on the level of noise. The regression curve estimated by the mixed model is plotted in Figure 3A with a thick solid curve. The curve indicates that when the noise level is between 40 and 50 dBA, the speech level is close to 60 dBA. When the noise is above 74 dBA, the speech level is lower than the noise level. Although the relationship between speech level and noise level is nonlinear, it is of interest to estimate the linear slope of this relationship. To this end, the speech and noise level data were fitted by a 2-segment piecewise linear function in accordance with Pearsons et al. (1977). The fitted function almost overlaps with the nonlinear regression curve (thick solid curve in Figure 3A) and therefore is not plotted in the figure. The piecewise linear function indicates that when the noise is below 59.3 dBA (speech = 66.0 dBA), speech level increases by 0.34 dB for every dB increment of noise. The linear slope is 0.54 dB/dB when the noise is higher than 59.3 dBA.

Regression lines that describe the relationship between speech level and noise level reported by Pearsons et al. and Smeds et al. are also shown in Figure 3A (gray dashed lines) for comparison.

Figure 3B shows SNR as a function of noise level. The linear mixed-effects model indicates that the effects of linear and quadratic terms of noise level on SNR were statistically significant (both $p < 0.0001$). Based on the regression curve estimated by the model shown in Figure 3B, the SNR is approximately 20 dB when the noise is 40 dBA. The SNR systematically decreases to 0 dB as the noise increases to 74 dBA.

The distribution of 894 SNRs is shown in Figure 4 as a bar histogram (refer to the left y-axis). To better illustrate the pattern of the distribution, the histogram data (i.e., frequency of occurrence and bin center value) were fitted by an asymmetric peak function. The fitted distribution curve (r-squared = 0.97) is shown in the figure as the dashed curve. Next, the frequency of occurrence and the bin upper limit value of the histogram were used to calculate cumulative frequency distribution (open circles in Figure 4; refer to the right y-axis), which indicates the frequency of SNRs that are lower than a given SNR. Figure 4 indicates that SNRs between 2 and 14 dB consisted of approximately 62.9% of all SNRs, with the most common SNRs being around 8 dB. Very noisy situations that had SNRs below 0 dB comprised 7.5% of the listening situations.

Although information on the type of listening environment (e.g., home vs. traffic) was collected in EMA surveys, it was not used to develop the PLSs (as mentioned in the Introduction). However, it is of interest to examine the SNRs of different listening environments.

Figure 5 shows boxplots of speech level, noise level (refer to the left y-axis) and SNR (refer to the right y-axis) as a function of self-reported listening environment. The number of the surveys completed in each type of environment is also shown in the figure. It is evident that most surveys were completed at home environments (52%), which had the lowest speech and noise levels (medians = 63.7 and 53.6 dBA, respectively). The median SNRs of “home,” “indoors other than home,” and “outdoors” were very close (9.9, 9.3, and 9.7 dB, respectively), while “traffic” and “indoors crowd” had lower median SNRs (5.6 and 5.3 dB, respectively).

PLSs

To develop the PLSs, speech level, noise level, SNR, and three categorical variables from the EMA surveys were used. The categorical variables were availability of visual cues (three levels: almost always/sometimes/no), talker location (three levels: front/side/back), and noise location (five levels: N/A (quiet)/all around/front/back/side). Recall that a special rule was used to analyze the SNR of the situations that the participants reported as media listening situations in the EMA surveys: Target speech signals were a conversational partner’s speech, rather than the sounds from media such as the television or radio. However, when reporting the characteristics of listening situations in the EMA surveys, the participants’ reports were based on the media listening situation, rather than on the conversation with their partners. In other words, the situation to which the SNR referred (i.e., conversation) differed from the situation reported in the EMA survey (i.e., media listening). Therefore, the media listening situation data (n = 176) were not included in this analysis; the remaining 718 observations were used to develop the PLSs.

To develop the PLSs, cluster analysis was used. The goal of a cluster analysis is to group similar observations together, such that within a cluster there is little difference between observations and there are large differences between clusters. Similarity in the clustering is measured by the distance between observations in the data space. Because the dataset for the cluster analysis contained both continuous variables (e.g., SNR) and categorical variables (e.g., availability of visual cues), Gower's distance (Gower 1971) was used to compute the distance matrix. The Partitioning Around Medoids function of the statistical software R (R Core Team 2016) was then used to identify the optimal number of clusters and to determine the clusters. Twelve clusters were identified. Table 1 shows the size and centroid of each of the 12 clusters. Specifically, the cluster size (the third column of Table 1) represents the number of observations belonging to a cluster, which reflects the frequency (in the parenthesis of the third column) of a certain type of listening situation in the collected data. The fourth to ninth columns of Table 1 further indicate cluster centroids, which describe the mean speech and noise levels, mean SNR, and the most frequent level (i.e., the mode) of the three categorical variables of the observations that belong to a given cluster. Therefore, the cluster centroid reflects the typical characteristics of the cluster and represents the PLS. The 12 clusters shown in Table 1 were referred to as general PLSs (gPLSs) because they were derived using the 718 observations that were collected from all types of speech listening situations ranging from quiet to very noisy. To facilitate data presentation, each gPLS was given a number, which is shown in the second column of Table 1.

In Table 1 the 12 gPLSs are further categorized into three subgroups (see the first column) based on the presence and location of the noise signals. The first subgroup is referred to as Quiet gPLS, because the most frequent observations belonging to these clusters characterized noise as "N/A (quiet)." The second subgroup is Diffuse Noise gPLS, as the most frequent observations characterized noise as “all-around.” The third subgroup is labeled Non-Diffuse Noise gPLS and consists of the two clusters where noise is most frequently located either in front of or to the side of the participants. For each of the three gPLS subgroups in Table 1, the clusters are listed in a descending order based on the cluster size. Two observations can be made. First, in terms of availability of visual cues and talker location, the five clusters in the Quiet gPLSs and in the Diffuse Noise gPLSs share the same characteristics and order. For example, in the most frequent situation the talker is in front of the listener and visual cues are almost always available (gPLS1 and gPLS6), and in the least frequent situation the talker is behind the listener and visual cues are only available sometimes (gPLS5 and gPLS10). Second, the characteristics of visual cues and talker location in the two Non-Diffuse Noise gPLSs are identical to the two most frequent clusters of the Quiet and Diffuse Noise gPLSs. For Quiet gPLSs, the speech level, noise level, and SNR averaged across all observations were 62.8 dBA (SD = 5.6), 50.6 dBA (SD = 5.7), and 12.2 dB (SD = 6), respectively. For Diffuse and Non-Diffuse Noise gPLSs, the mean speech level, noise level, and SNR were 67.9 dBA (SD = 5.2), 60.5 dBA (SD = 7.4), and 7.4 dB (SD = 6.0), respectively.

PLSs for noisy speech listening situations

In addition to gPLSs that represent all types of speech listening situations, it is of interest to develop a set of PLSs that describe noisy situations, as hearing aid users frequently report difficulty in these situations (e.g., Takahashi et al. 2007). To this end, only a subset of the data that were collected in noisy environments were used in cluster analysis to create the PLSs. In order to exclude quiet environments, the SNR data and noisiness ratings reported in the EMA surveys (four levels: quiet/somewhat noisy/noisy/very noisy) were examined. Figure 6 shows the boxplot of SNR as a function of self-reported noisiness. Although a linear mixed-effects model indicated that the participants tended to rate the environments as noisier when the SNR became poorer ($p < 0.0001$), the variation across observations was considerable. Because the SNR and self-reported noisiness were not always consistent with each other, a situation wherein the SNR was higher than 10 dB or the noisiness was reported as “quiet” was defined as a quiet situation and was excluded from the analysis. The 10-dB SNR criterion was selected based on the median SNR of the “quiet” noisiness ratings (10.6 dB, see Figure 6).

After excluding quiet situations, the remaining 280 observations were subjected to cluster analysis. Two clusters were identified (Table 2) and labeled as noisy PLSs (nPLSs). Both nPLSs are characterized by including all-around noise. The visual cues and talker/noise location characteristics of nPLS1 and nPLS2 are identical to the two most frequent Diffuse Noise gPLSs (gPLS6 and gPLS7). The speech level, noise level, and SNR averaged across all 280 observations that belong to the nPLSs are 67.5 dBA (SD = 5.2), 63.3 dBA (SD = 6.1), and 4.2 dB (SD = 3.8), respectively.

DISCUSSION

The current study characterized SNR and real-world speech listening situations for older adults with mild-to-moderate hearing loss. The data were collected from 20 participants over an interval of five to six weeks for each, spread over six to eight months.

Relationship between speech level, noise level, and SNR

Statistical models indicated that as noise level increased from 40 to 74 dBA, speech level systematically increased from 60 to 74 dBA, so SNR decreased from 20 to 0 dB (Figure 3). In order to compare this result to existing literature, the regression lines that describe the relationship between speech level and noise level reported by Pearsons et al. (1977, cf. Figure 20) are reproduced in Figure 3A with long dashed lines. Figure 3A also shows the linear regression lines estimated based on the speech and noise level data reported by Smeds et al. (2015, cf. Figure 5), for the ear with the better SNR (better ear, short dashed line) and the ear with the poorer SNR (worse ear, dash-dotted line) separately. The result of the current study is fairly close to Smeds et al., such that the current study's regression curve is located in between the Smeds et al. study participants' better and worse ears' regressions lines. This is coincident with the positioning of the microphones: in the current study sounds were logged by a chest-level recorder and in Smeds et al. two ear-level microphones were used. Both studies suggest that the speech level is approximately 60 dBA when the noise level is 40 dBA. In contrast, the speech level reported by Pearsons et al. (1977) is 3 to 5 dB lower than Smeds et al. and those in the current study when noise levels are lower than 60 dBA. All regression curves/lines shown in Figure 3A converge around 70 to 75 dBA noise, at which the SNR is close to 0 dB.

The result that the speech at a given noise level reported by Pearsons et al. is lower than Smeds et al. (2015) and the current study may be due to the difference in participants: the former study used adults with normal hearing while the latter two used adults with hearing loss. There are several reasons that the speech may be measured at a higher level in the studies examining individuals with hearing loss. For example, people may speak louder if they are aware that their communication partners have listening difficulty. This is somewhat supported by the finding that the SNR was slightly higher in the unaided (10.3 dB) than aided (8.7 dB) situations. Another potential explanation for the lower speech level reported by Pearsons et al. is related to the SNR analysis technique. For all three studies, the speech-plus-noise segment was used to derive speech power and SNR. The duration of this segment is generally longer in Pearsons et al. (at least 10 sec) than that examined in Smeds et al. (5 sec) and the current study (3 sec; a SNR was derived using up to three segments). As pointed out by Smeds et al., longer speech-plus-noise segments may contain more pauses between speech sounds, resulting in an underestimation of speech power.

Distribution of SNR

To compare the distribution of SNR with existing literature, Figure 7 shows the histograms estimated from the SNR data reported by Pearsons et al. (7A) and Smeds et al. (7B; light gray and dark gray shades represent better and worse SNR ears, respectively) together with the distribution curve of the current study. Compared to Smeds et al. and the current study, Pearsons et al. reported more low-SNR situations. Specifically, approximately 15.5% of the SNRs reported by Pearsons et al. were below 0 dB. In contrast, the frequencies of the situations that had SNRs below 0 dB were 4.2% (the better ear) and 13.7% (the worse ear) in Smeds et al. and 7.5% in the current study. One potential explanation for this difference is that the research participants with hearing loss in Smeds et al. (mean age = 51.4 years) and the current study (71.1 years) avoided low-SNR situations in order to promote successful communication in their daily lives (Demorest & Erdman 1987). The normal-hearing research participants in Pearsons et al. (age was not specified) might encounter more noisy environments in their daily lives. Another explanation involves the sampling strategy. In Smeds et al., participants selected situations that were representative to their daily lives to record sounds. In the current study, the audio was recorded continuously throughout the day and the recordings that were associated with smartphone surveys were analyzed. The timing of the surveys was either determined by the phone application software or by the participants. In contrast, the location of measurement in Pearsons et al. was determined by researchers. It seems that Pearsons and colleagues intentionally selected some very noisy situations, such as trains and aircrafts, resulting in oversampling low-SNR situations. Note that due to its output limiting algorithm, the LENA DLP used in the current study was unable to accurately measure the level of the sounds that are higher than 80 dBA (Figure 2B). However, the limited dynamic range of the DLP is unlikely to be responsible for the infrequency of low-SNR situations observed in the current study, as Smeds et al., whose recording equipment had a dynamic range up to 110 dB SPL, demonstrated a similar result.

The limited dynamic range of the LENA DLP, however, could cause the difference between Smeds et al. and the current study in the frequency of occurrence of high-SNR situations. Specifically, Smeds et al. reported more situations that had SNRs above 20 dB (approximately 22.2% and 19.2% for the better and worse ears, respectively) than the current study (5.5%) (Figure 7B). Among the high-SNR situations reported by Smeds et al., approximately 50% (better ear) and 71.4% (worse ear) occurred in very quiet situations that had noise levels lower than 40 dBA (cf. Figure 5 of Smeds et al.). Because the lower limit of the LENA DLP’s dynamic range is 40 dBA, the noise level of very quiet situations could be overestimated in the current study, resulting in fewer high-SNR observations. The dynamic range of the LENA DLP, however, had little impact on speech level estimation, as the levels of speech signals are often higher than 40 dBA even in very quiet environments (Pearsons et al. 1977;

Smeds et al. 2015).

Relationship between SNR and type of environment

Comparing the SNR of a given type of listening environment (Figure 5) to the literature is less straightforward, as listening environments were categorized differently across studies. Nevertheless, the current study and Smeds et al. (2015) show a similar trend. Specifically, the current study found that the median SNRs of “outdoors,” “traffic” (mainly in cars), and “indoors crowd” were 9.7, 5.6, and 5.3 dB, respectively, and Smeds et al. reported that the median SNRs (two ears averaged) of “outdoors,” “car”, “department store” are 10.9, 3.6, and 2.3 dB, respectively.

PLSs

The cluster analysis suggested that the 718 speech listening situations experienced by the participants in daily life can be grouped into 12 clusters, with little difference between situations within the cluster and large differences between clusters (Table 1). The most frequent situation was characterized as having the talker in front of the listener with visual cues available. This is the same for all three gPLS subgroups (Quiet, Diffuse Noise, and Non-Diffuse Noise). This result is also well aligned with the listening situations reported by Walden et al. (2005). For the Quiet gPLSs, the mean speech level was 62.8 dBA, which is very close to the 63-dBA reported by Smeds et al., while higher than the level suggested by Pearsons et al. (55 dBA, Figure 3A). For noisy listening situations, diffuse (all-around) noise was more common than non-diffuse noise. This is consistent with Woods et al. (2010), who found that most real-world noisy environments are close to a diffuse or semi-diffuse sound field. Note that the 12 gPLSs do not include a configuration that has been widely used in clinical and research settings: both speech and noise come from in front of the listener and visual cues are not available.

The characteristics of visual cue availability and talker location described by the gPLSs warrant more discussion. Specifically, gPLS4 and gPLS9 were characterized as having the talker beside the listener with visual cues almost always available (Table 1). The high availability of visual cues implies that the listeners constantly oriented their heads toward the talkers beside them. Orienting the head toward the talker was also likely to happen, but to a lesser extent, in other PLSs wherein visual cues were reported to be available sometimes. Ricketts and Galster (2008) used video cameras to monitor children's head orientation in actual school settings. They found that although children often oriented their head toward the sound source of interest, considerable individual variability existed. Because objective data regarding the participants' head orientation are not available in the current study, the extent to which how often participants oriented their heads toward the talker in visual cue availability ratings "almost always" and "sometimes" is unknown.

The two nPLSs (Table 2) were generated using observations where the SNR was lower than 10 dB and a noisiness rating other than "Quiet" was selected. Therefore, the nPLS represented speech listening situations that were noisy. The mean SNR of the nPLS (4.2 dB) was 3.2 dB lower than that of the Diffuse and Non-Diffuse Noise gPLSs. For sentence recognition tests like the Connected Speech Test (Cox et al. 1988), a 3-dB difference could result in a 30% change in performance. Note that the mean SNR of the nPLS (4.2 dB) is very close to the test SNRs of the Connected Speech Test used in several randomized clinical trials comparing hearing aid outcomes (e.g., Humes et al, 2017; Larson et al, 2000), although these studies did not include visual cues in the testing.

Limitations

The current study has several limitations concerning its generalizability. First, the LENA DLP, which was selected for its superior portability and usability, has several disadvantages. Specifically, the microphone of the DLP was worn in front of the participant at chest-level, rather than at ear-level. As a result, the SNR at the DLP's microphone port was somewhat different from what would have been measured with ear-level microphones, especially for speech from behind the wearer in environments with less diffuse noise (Byrne & Reeves 2008). Although the estimated speech level and SNR are quite similar to those reported by Smeds et al. (2015) who used ear-level microphones, the results of the current study would be more relevant to the participants' true perception if ear-level microphones had been used. Another disadvantage of the DLP is its limited dynamic range. As discussed earlier, the inability of the DLP to measure sounds lower than 40 dBA could result in the discrepancy between the current study and Smeds et al. in the frequency of occurrence of high-SNR listening situations. Further, the sound level adjustment, which was conducted to compensate for the effect of the low-level expansion algorithm of the DLP, could result in less accurate SNR estimations.

Second, although the current study collected information from 894 situations, the data were provided by 20 older adults with mild-to-moderate hearing loss living in rural and suburban areas. It is unknown if the results of the current study can generalize to populations of different ages, degrees of hearing loss, and geographic areas. It is also unknown if the results of the current study can generalize to different hearing aid settings and models, as (1) the volume control was disabled for the larger study and (2) the SNR was found to be lower with premium-level (8.6 dB) than basic-level (10.0 dB) models (noise reduction feature-on and -off combined). The effect of hearing aid model (basic vs. premium) on SNR could result from the more advanced noise reduction features of the premium-level model increasing users' willingness to spend more time in situations with lower SNRs. However, this statistically significant effect of hearing aid model may not be meaningful because the mean SNR of the feature-on conditions (9.9 dB, premium- and basic-level models combined) was not lower than the feature-off conditions (8.9 dB).

Third, the frequency of very noisy situations might be underestimated. When analyzing the audio recordings, a very poor SNR might preclude the research assistants from identifying the target speech and conducting the subsequent SNR analysis. Further, the auditory/vibrotactile prompt of the smartphone, which occurred approximately every two hours, may not have been detectable by the participants in very noisy environments. If no survey was conducted, the audio recordings were not analyzed. A shorter inter-prompt interval may increase the likelihood for the participants to conduct surveys in very noisy situations. However, too-frequent prompts would interfere with the participant's activities (Stone 2003), which might in turn change the characteristics of listening situations.

Implications

Researchers can use the PLS information reported in Tables 1 and 2 to design sound fields for speech-related laboratory testing. If the three most frequent Quiet and Diffuse Noise gPLSs are simulated in testing (gPLSs 1 to 3 and 6 to 8), these six test environments would represent 71% of daily speech listening situations. If researchers are interested in more difficult situations, the two nPLSs can be used. The PLS data shown in Tables 1 and 2 do not preclude researchers from using very low SNRs or unmentioned speech/noise configurations in testing. However, researchers should be cautious about the real-world generalizability of their data.

Because all of the PLSs in this study have positive SNRs and many of them have visual cues available, it is anticipated that listeners with mild-to-moderate hearing loss will have a speech recognition performance approaching the ceiling level in most PLSs, especially when hearing aids are used. If the ceiling effect occurs, the speech recognition test will no longer have the sensitivity to detect the difference between interventions. From this perspective, it is likely that listening effort would serve as a better metric than speech recognition performance in testing environments that are designed to simulate the real world. Research has shown that listening effort measures are still sensitive to change even when speech recognition performance is at the ceiling level (e.g., Sarampalis et al., 2009; Winn et al. 2015; Wu et al. 2016). Other measures, such as speech quality, could also be appropriate in this regard (Naylor 2016). Future research to investigate if these measures, conducted in the PLSs suggested by the current study, would better predict real-world speech communicative function is warranted.

CONCLUSIONS

The current study characterized real-world speech listening situations for older adults with mild-to-moderate hearing loss. The results indicate that as noise level increased from 40 to 74 dBA, SNR systematically decreased from 20 to 0 dB. Visual cues and all-around (i.e., diffuse) noise were quite common in real-world listening situations, while very low-SNR environments were relatively rare. A wide range of daily speech listening situations can be represented by 12 PLSs and nosier listening situations can be characterized by two PLSs. These results could be useful for researchers to design more ecologically-valid assessment procedures to estimate real-world speech communicative functions for older adults with mild-to-moderate hearing loss.

ACKNOWLEDGMENTS

The current research was supported by NIH/NIDCD R03DC012551 and the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR, grant number 90RE5020-01-00). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS). The contents of this paper do not necessarily represent the policy of NIDILRR, ACL, HHS, and the reader should not assume endorsement by the Federal Government. Portions of this paper were presented at the International Hearing Aid Research Conference, August, 2016, Tahoe City, CA, USA.

REFERENCES

Ahlstrom, J. B., Horwitz, A. R., & Dubno, J. R. (2009). Spatial benefit of bilateral hearing aids. *Ear Hear, 30*, 203-218.

ANSI. (2010). *Specification for audiometers (ANSI S3.6)*. New York: American national standards institute.

Byrne, D. C., & Reeves, E. R. (2008). Analysis of nonstandard noise dosimeter microphone positions. *J Occup Environ Hyg, 5*, 197-209.

Cox, R. M., Alexander, G. C., & Gilmore, C. (1987a). Development of the Connected Speech Test (CST). *Ear Hear*, *8*, 119s-126s.

Cox, R. M., Alexander, G. C., & Gilmore, C. (1987b). Intelligibility of average talkers in typical listening environments. *J Acoust Soc Am*, *81*, 1598-1608.

Cox, R. M., Alexander, G. C., Gilmore, C., et al. (1988). Use of the Connected Speech Test (CST) with hearing-impaired listeners. *Ear and Hearing*, *9,* 198-207.

Demorest, M. E., & Erdman, S. A. (1987). Development of the communication profile for the hearing impaired. *J Speech Hear Disord, 52*, 129-143.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis* (2nd edition). Hoboken, NJ: John Wiley & Sons.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27, 857-874.

Hasan, S. S., Lai, F., Chipara, O., et al. (2013). AudioSense: Enabling real-time evaluation of hearing aid technology in-situ. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems* (pp. 167-172). IEEE.

Hornsby, B. W., & Ricketts, T. A. (2007). Effects of noise source configuration on directional benefit using symmetric and asymmetric directional hearing aid fittings. *Ear Hear*, *28*, 177-186.

Humes, L. E., Rogers, S. E., Quigley, T. M., et al. (2017). The effects of service-delivery model and purchase price on hearing-aid outcomes in older adults: A randomized double-blind placebo-controlled clinical trial. *Am J Audiol*, 26, 53-79.

Jensen, N. S., & Nielsen, C. (2005). Auditory ecology in a group of experienced hearing-aid users: Can knowledge about hearing-aid users' auditory ecology improve their rehabilitation? In A. N. Rasmussen, T. Poulsen, T. Andersen & C. B. Larsen (Eds.), *Hearing Aid Fitting* (pp. 235-260). Kolding, Denmark: The Danavox Jubilee Foundation.

Keidser, G. (2016). Introduction to Special Issue: Towards Ecologically Valid Protocols for the Assessment of Hearing and Hearing Devices. *J Am Acad Audiol, 27*, 502-503.

Keidser, G., Dillon, H., Flax, M., et al. (2011). The NAL-NL2 prescription procedure. *Audiology Research, 1*, 88-90.

Larson, V. D., Williams, D. W., et al. (2000). Efficacy of 3 commonly used hearing aid circuits: A crossover trial. *JAMA*, 284, 1806-1813.

Lin, F. R., Thorpe, R., Gordon-Salant, S., et al. (2011). Hearing loss prevalence and risk factors among older adults in the United States. *J Gerontol A Biol Sci Med Sci, 66*, 582-590.

Naylor, G. (2016). Theoretical Issues of Validity in the Measurement of Aided Speech Reception Threshold in Noise for Comparing Nonlinear Hearing Aid Systems. *J Am Acad Audiol, 27*, 504-514.

Pearsons, K. S., Bennett, R. L., & Fidell, S. (1977). Speech levels in various noise environments (Report No. EPA-600/1-77-025). Washington, DC: U.S. Environmental Protection Agency.

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *J Speech Lang Hear Res, 29*, 146-154.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Ricketts, T. A. (2000). Impact of noise source configuration on directional hearing aid benefit and performance. *Ear Hear, 21*, 194-205.

Ricketts, T. A., & Galster, J. (2008). Head angle and elevation in classroom environments: Implications for amplification. *J Speech Lang Hear Res*, *51*, 516-525.

Sarampalis, A., Kalluri, S., Edwards, B., et al. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *J Speech Lang Hear Res, 52*, 1230-1240.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annu Rev Clin Psycho, 4*, 1-32.

Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *J Am Acad Audiol, 26*, 183-196.

Stone, A. A., Broderick, J. E., Schwartz, J. E., et al. (2003). Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. *Pain*, *104*, 343-351.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J Acoust Soc Am*, *26*, 212-215.

Takahashi, G., Martinez, C. D., Beamer, S., et al. (2007). Subjective measures of hearing aid benefit and satisfaction in the NIDCD/VA follow-up study. *J Am Acad Audiol, 18*, 323-349.

VanDam, M., Ambrose, S. E., & Moeller, M. P. (2012). Quantity of parental language in the home environments of hard-of-hearing 2-year-olds. *J Deaf Stud Deaf Educ, 17*, 402-420.

Vestergaard, M. D. (2006). Self-report outcome in new hearing-aid users: Longitudinal trends and relationships between subjective measures of benefit and satisfaction. *Int J Audiol, 45*, 382-392.

Wagener, K. C., Hansen, M., & Ludvigsen, C. (2008). Recording and classification of the acoustic environment of hearing aid users. *J Am Acad Audiol, 19*, 348-370.

Walden, B. E. (1997). Toward a model clinical-trials protocol for substantiating hearing aid user-benefit claims. *Am J Audiol, 6*, 13-24.

Walden, B. E., Demorest, M. E., & Hepler, E. L. (1984). Self-report approach to assessing benefit derived from amplification. *J Speech Lang Hear Res, 27*, 49-56.

Walden, B. E., Surr, R. K., Cord, M. T., et al. (2004). Predicting hearing aid microphone preference in everyday listening. *J Am Acad Audiol, 15*, 365-396.

Walden, B. E., Surr, R. K., Grant, K. W., et al. (2005). Effect of signal-to-noise ratio on directional microphone benefit and preference. *J Am Acad Audiol, 16*, 662-676.

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear Hear, 36*, e153-165.

Wolters, F., Smeds, K., Schmidt, E., et al. (2016). Common sound scenarios: A context-driven categorization of everyday sound environments for application in hearing-device research. *J Am Acad Audiol*, *27*, 527-540.

Woods, W. S., Merks, I., Zhang, T., et al. (2010). Assessing the benefit of adaptive null-steering using real-world signals. *Int J Audiol, 49*, 434-443.

Wu, Y. H., & Bentler, R. A. (2010a). Impact of visual cues on directional benefit and preference: Part I--laboratory tests. *Ear Hear, 31*, 22-34.

Wu, Y. H., & Bentler, R. A. (2010b). Impact of visual cues on directional benefit and preference: Part II--field tests. *Ear Hear, 31,* 35-46.

Wu, Y. H., & Bentler, R. A. (2012). Do older adults have social lifestyles that place fewer demands on hearing? *J Am Acad Audiol, 23,* 697-711.

Wu, Y. H., Stangl, E., & Bentler, R. (2013). Hearing-aid users' voices: A factor that could affect directional benefit. *Int J Audiol, 52*, 789-794.

Wu, Y. H., Stangl, E., Zhang, X., et al. (2015). Construct Validity of the Ecological Momentary Assessment in Audiology Research. *J Am Acad Audiol, 26*, 872-884.

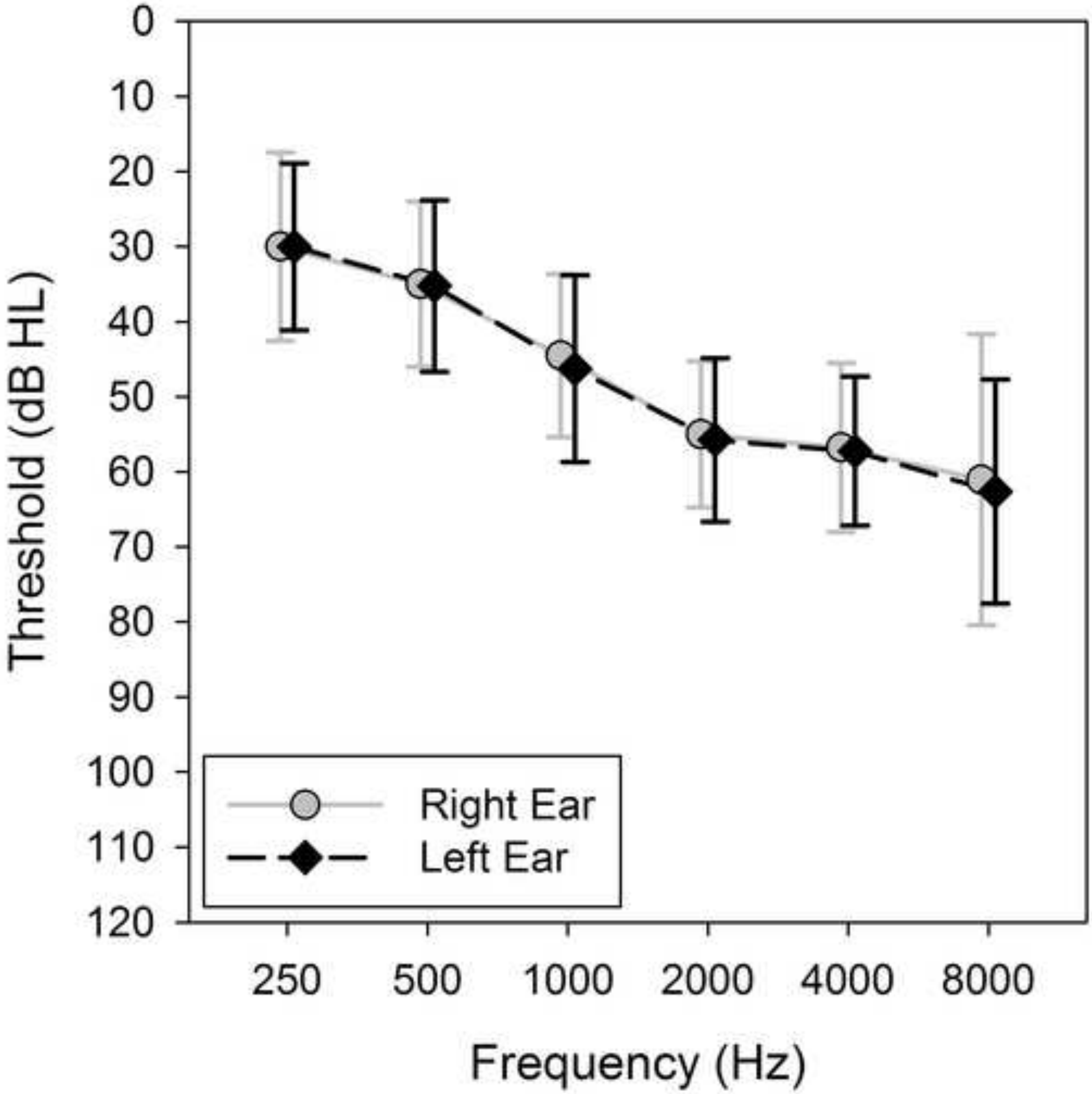
Wu, Y. H., Stangl, E., Zhang, X., et al. (2016). Psychometric Functions of Dual-Task Paradigms for Measuring Listening Effort. *Ear Hear, 37*, 660-670.

FIGURE LEGENDS

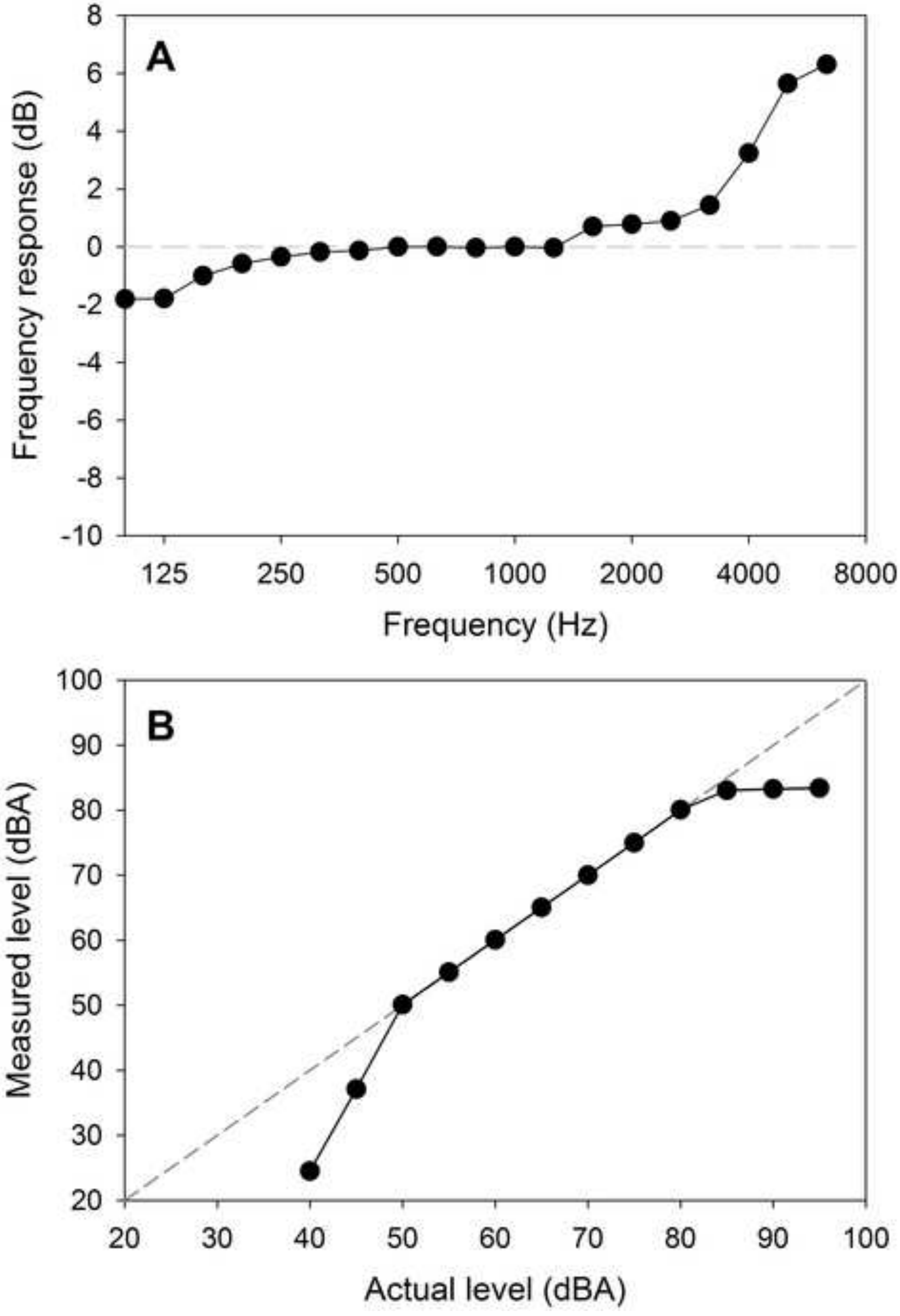
Figure 1	Average audiograms for left and right ears of twenty study participants. Error bars = 1 SD.
Figure 2	Frequency response (2A) and the relationship between the measured and actual level (2B) of the digital audio recorder.
Figure 3	3A. Speech level as a function of noise level reported in the current study (circles and thick black solid curve), Smeds et al. (2015), and Pearsons et al. (1977). Chest-level microphones were used in the current study while ear-level microphones were used in Smeds et al., and Pearson et al. Diagonal light gray line represents where the speech level is equal to the noise level. 3B. Signal-to-noise ratio as a function of noise level reported in the current study.
Figure 4	Distribution of signal-to-noise ratios (SNRs) measured using chest-level microphone. Gray bars represent a histogram (refer to the left y-axis). Dashed curve (refer to the left y-axis) represents an asymmetric peak function that fits the histogram data of occurrence frequency and bin center value. Open circles represent the frequency of occurrence of the SNRs that are lower than a given SNR (i.e., the cumulative frequency; refer to the right y-axis).
Figure 5	Boxplots of speech level, noise level (refer to the left y-axis), and signal-to-noise ratio (SNR; refer to the right y-axis) as a function of self-reported listening environment. The boundaries of the boxes represent the 25th and 75th percentile and the line within the boxes marks the median. Error bars indicate the 10th and 90th percentiles.

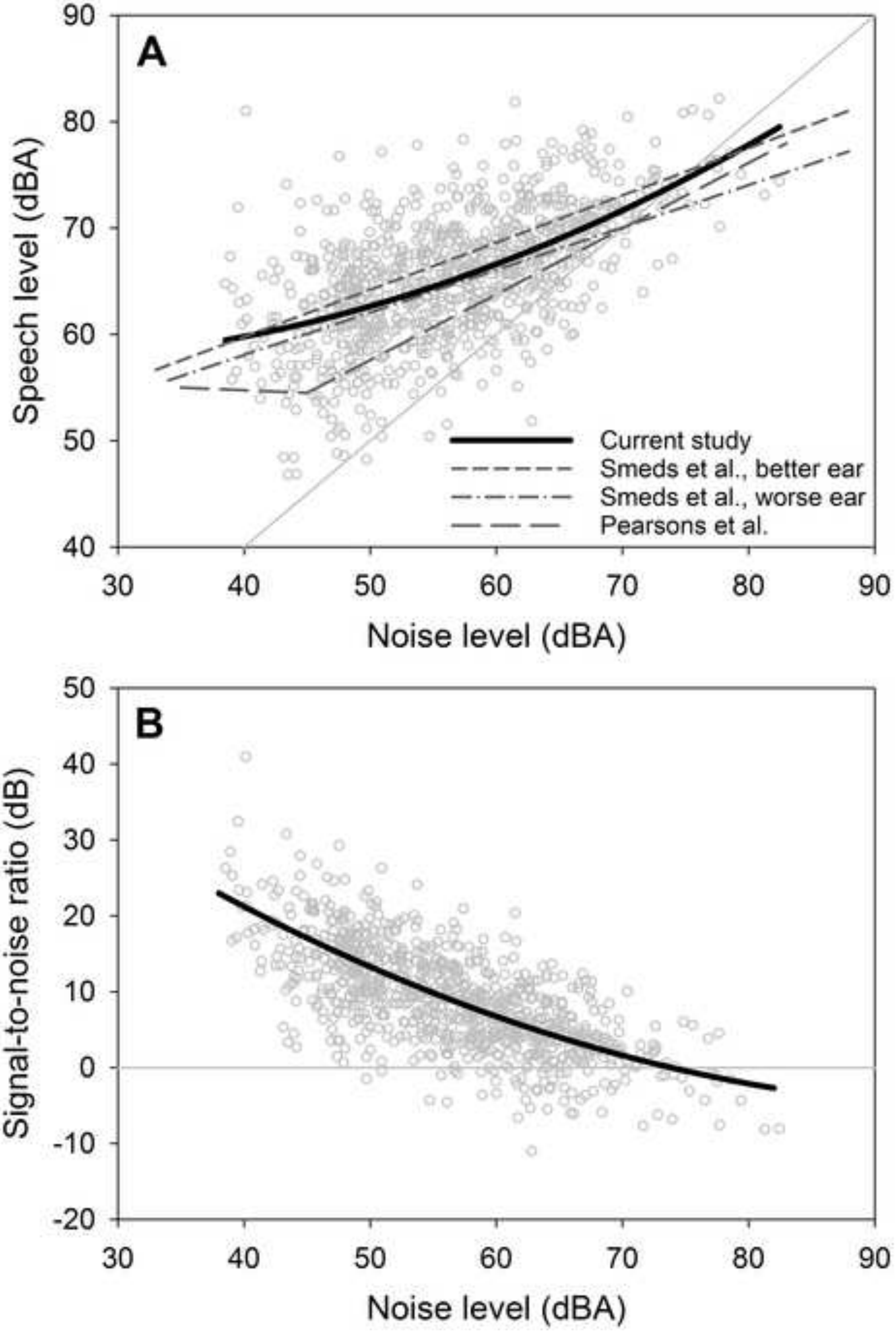
Figure 6 Boxplot of signal-to-noise ratio as a function of self-reported noisiness. The boundaries of the box represent the 25th and 75th percentile and the line within the box marks the median. Error bars indicate the 10th and 90th percentiles.

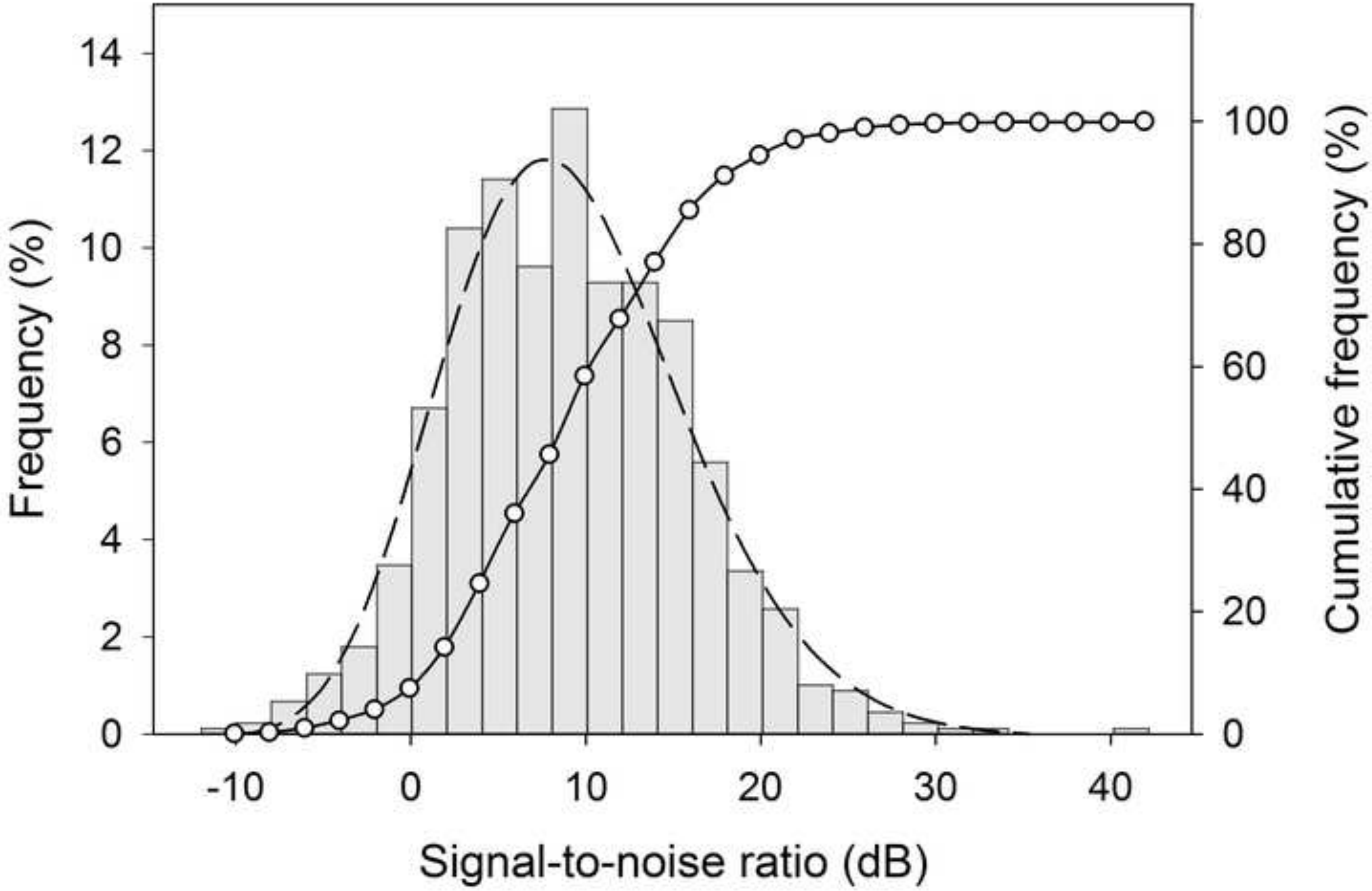
Figure 7 Signal-to-noise ratio (SNR) distribution curve of the current study and histograms of SNRs reported by Pearsons et al. (1977) (7A) and Smeds et al. (2015) (7B). The light gray shade and dark gray shade in Figure 7B represent the histograms of the better SNR ear and worse SNR ear, respectively. Chest-level microphones were used in the current study while ear-level microphones were used in Smeds et al., and Pearson et al.

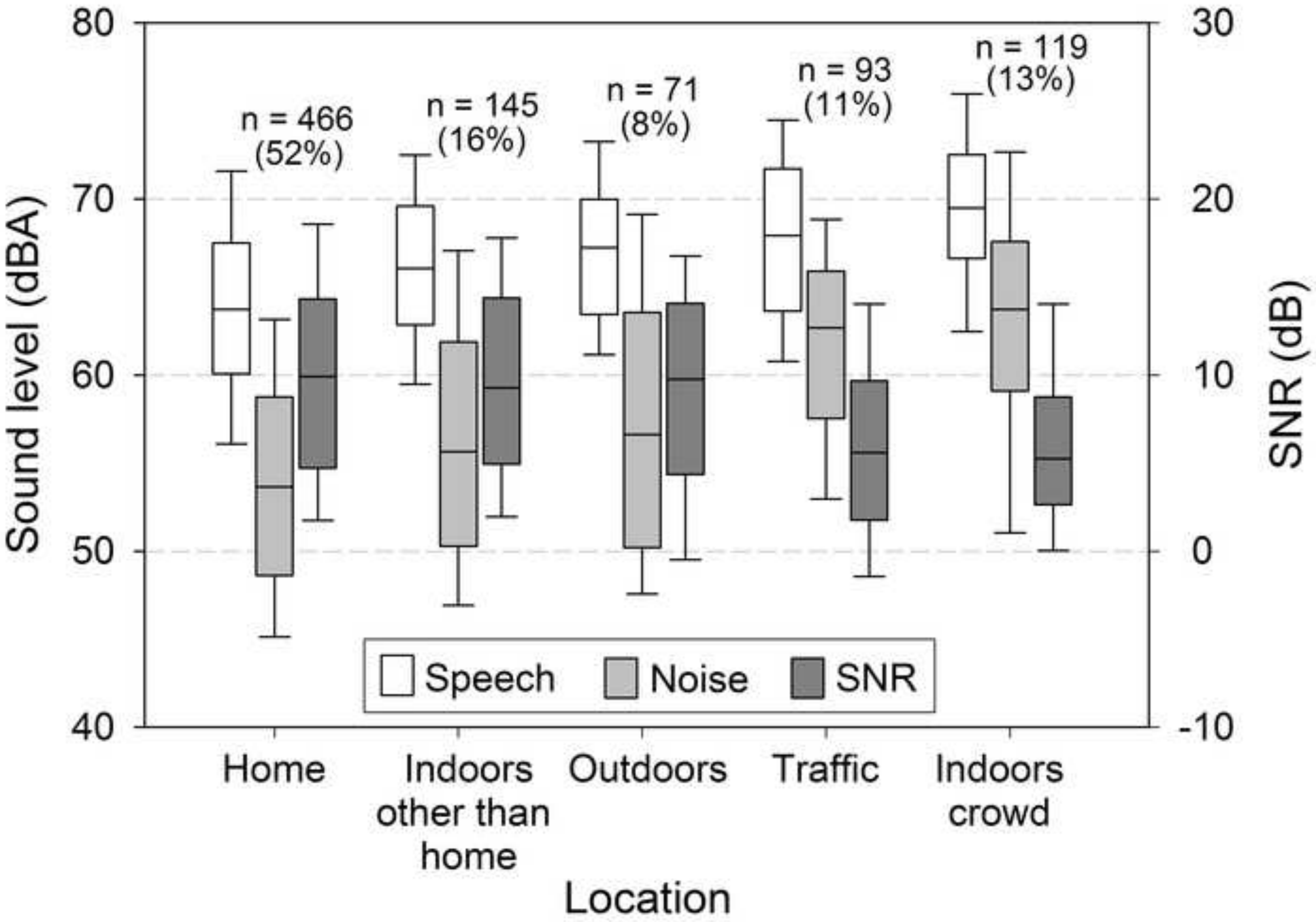


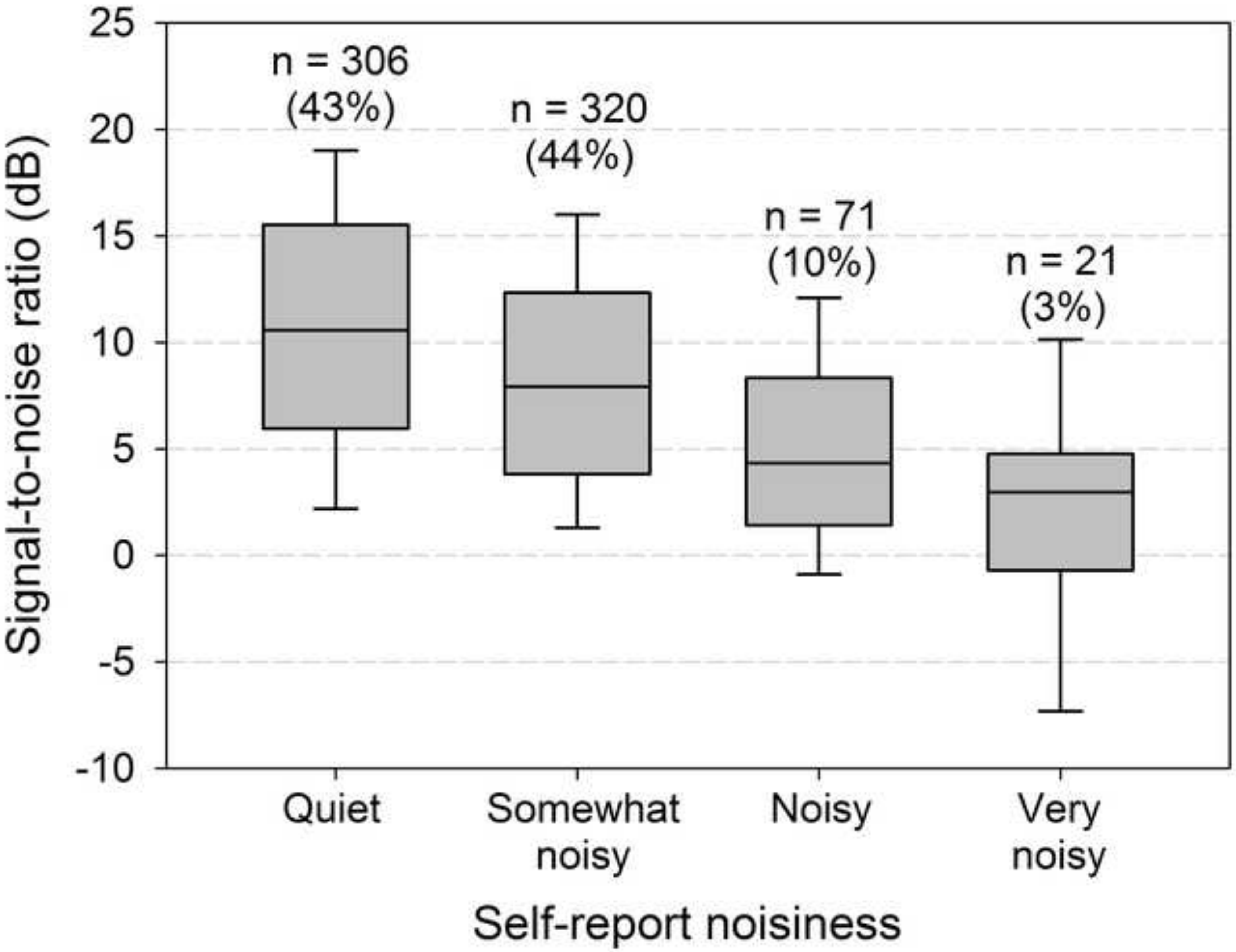
Threshold (dB HL)
0
10
20
30
40
50
60
70
80
90
100
110
120
250
500
1000
2000
4000
8000
Frequency (Hz)
● Right Ear
◆ Left Ear


A
Frequency response (dB)
8
6
4
2
0
-2
-4
-6
-8
-10
125
250
500
1000
2000
4000
8000
Frequency (Hz)
B
Measured level (dBA)
100
90
80
70
60
50
40
30
20
20
30
40
50
60
70
80
90
100
Actual level (dBA)


A
Speech level (dBA)
90
80
70
60
50
40
30
40
50
60
70
80
90
Noise level (dBA)
Current study
Smeds et al., better ear
Smeds et al., worse ear
Pearsons et al.
B
Signal-to-noise ratio (dB)
50
40
30
20
10
0
-10
-20
30
40
50
60
70
80
90
Noise level (dBA)


Frequency (%)
14
12
10
8
6
4
2
0
Cumulative frequency (%)
100
80
60
40
20
0
-10
0
10
20
30
40
Signal-to-noise ratio (dB)


n = 466
(52%)
n = 145
(16%)
n = 71
(8%)
n = 93
(11%)
n = 119
(13%)
80
70
60
50
40
30
20
10
0
-10
Sound level (dBA)
SNR (dB)
Speech
Noise
SNR
Home
Indoors other than home
Outdoors
Traffic
Indoors crowd
Location


Signal-to-noise ratio (dB)
25
20
15
10
5
0
-5
-10
n = 306
(43%)
n = 320
(44%)
n = 71
(10%)
n = 21
(3%)
Quiet
Somewhat noisy
Noisy
Very noisy
Self-report noisiness

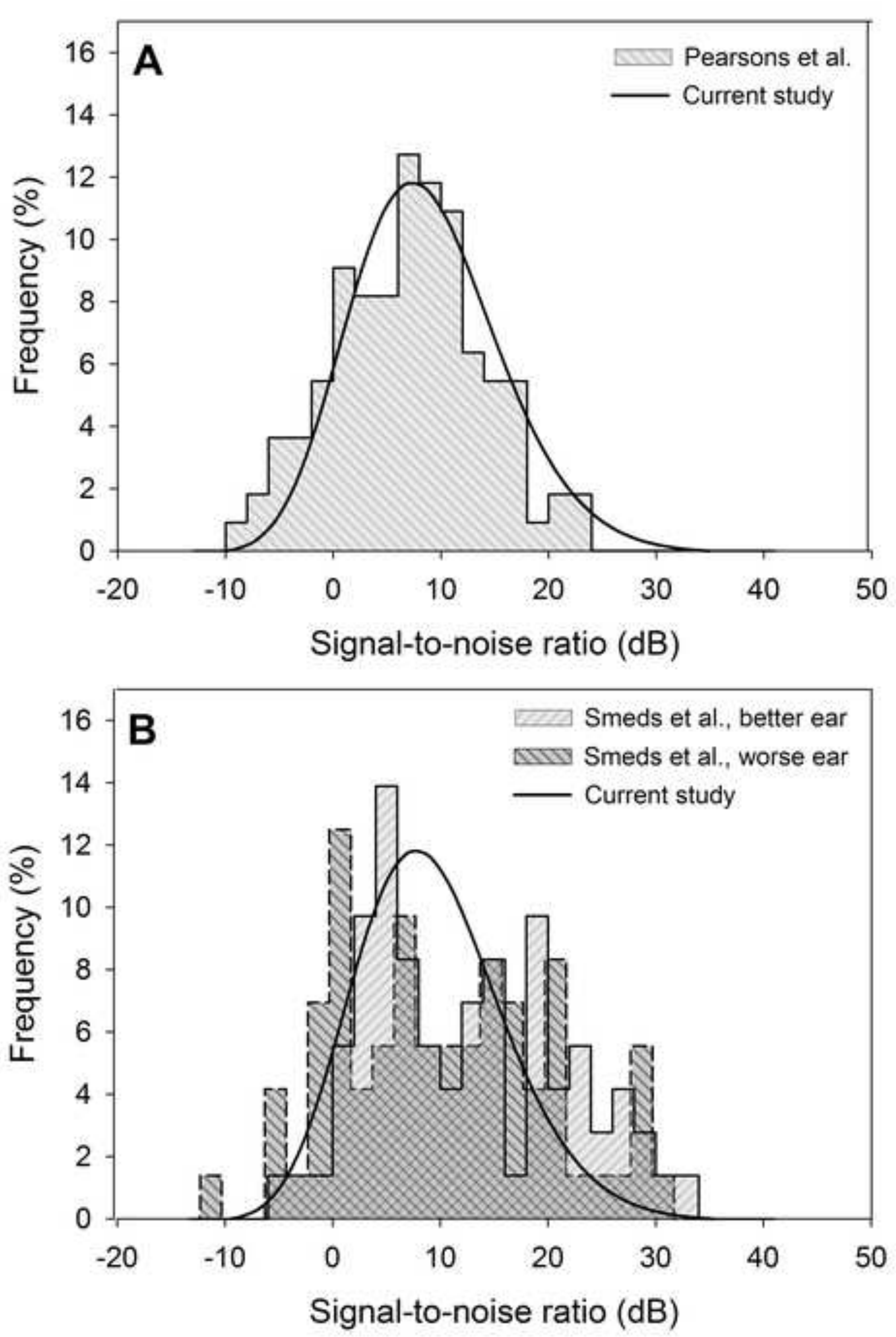

A
Pearsons et al.
Current study
Frequency (%)
16
14
12
10
8
6
4
2
0
-20
-10
0
10
20
30
40
50
Signal-to-noise ratio (dB)
B
Smeds et al., better ear
Smeds et al., worse ear
Current study
Frequency (%)
Signal-to-noise ratio (dB)

Table 1. General Prototype Listening Situations (gPLS)

Subgroup	Numbering	Cluster size	Speech level (dBA)	Noise level (dBA)	SNR (dB)	Visual cues	Talker location	Noise location
Quiet	1	115 (16%)	63.9	50.5	13.4	Always	Front	N/A (quiet)
	2	96 (13%)	61.5	50.6	10.9	Sometimes	Side	N/A (quiet)
	3	45 (6%)	60.4	50.4	10.0	Sometimes	Front	N/A (quiet)
	4	37 (5%)	65.4	51.0	14.4	Always	Side	N/A (quiet)
	5	20 (3%)	62.6	50.7	11.9	Sometimes	Back	N/A (quiet)
Diffuse Noise	6	93 (13%)	68.5	59.9	8.6	Always	Front	All around
	7	87 (12%)	67.3	60.9	6.4	Sometimes	Side	All around
	8	74 (10%)	68.8	64.0	4.8	Sometimes	Front	All around
	9	53 (7%)	68.7	59.4	9.2	Always	Side	All around
	10	20 (3%)	67.4	60.6	6.7	Sometimes	Back	All around
Non-Diffuse Noise	11	42 (6%)	64.4	54.9	9.5	Always	Front	Front
	12	36 (5%)	69.5	61.9	7.6	Sometimes	Side	Side

Table 2. Noisy Prototype Listening Situations (nPLS)

Numbering	Cluster size	Speech level (dBA)	Noise level (dBA)	SNR (dB)	Visual cues	Talker location	Noise location
1	153 (55%)	67.4	63.7	3.8	Always	Front	All around
2	127 (45%)	67.6	62.8	4.8	Sometimes	Side	All around