The just meaningful difference in speech-to-noise ratio

David McShefferty, William M. Whitmer and Michael A. Akeroyd

MRC/CSO Institute of Hearing Research - Scottish Section

Authors and Affiliations

David McShefferty, MRC/CSO Institute of Hearing Research - Scottish Section;

William M. Whitmer, MRC/CSO Institute of Hearing Research - Scottish Section;

Michael A. Akeroyd, MRC Institute of Hearing Research

Correspondence concerning this article should be addressed to David McShefferty, MRC/CSO Institute of Hearing Research - Scottish Section, Glasgow Royal Infirmary, 10-16 Alexandra Parade, Glasgow, G31 2ER, United Kingdom. E-mail: david@ihr.gla.ac.uk

Abstract

The speech-to-noise ratio (SNR) in an environment plays a vital role in speech communication for both normal-hearing (NH) and hearing-impaired (HI) listeners. While hearing-assistance devices attempt to deliver as favorable an SNR as possible, there may be discrepancies between noticeable and meaningful improvements in SNR. Furthermore, it is not clear how much of an SNR improvement is necessary to induce intervention-seeking behavior. Here we report on a series of experiments examining the just-meaningful difference (JMD) in SNR. All experiments used sentences in same-spectrum noise, with two intervals on each trial mimicking examples of pre- and post-benefit situations. Different groups of NH and HI adults were asked (a) to rate how much better or worse the change in SNR was in a number of paired examples, (b) if they would swap the worse for the better SNR (e.g., their current device for another) or (c) if they would be willing to go to the clinic for the given increase in SNR. The mean SNR JMD based on better/worse ratings (one arbitrary unit) was similar to the just-noticeable difference, approximately 3 dB. However, the mean SNR JMD for the more clinically relevant tasks -- willingness (at least 50% of the time) to swap devices or attend the clinic for a change in SNR -- was 6-8 dB regardless of hearing ability. This SNR JMD of the order of 6 dB provides a new benchmark, indicating the SNR improvement necessary to immediately motivate participants to seek intervention.

The just meaningful difference in speech-to-noise ratio

INTRODUCTION

The ability to hear and understand speech in the presence of background noise is highly dependent on the speech-to-noise ratio (SNR), i.e., the level of the speech relative to the level of the background noise. Generally, hearing-impaired (HI) listeners require a higher SNR than normal-hearing (NH) listeners to achieve equivalent scores in speech intelligibility tests (e.g., Summerfield, 1987; Grant & Walden, 2013). For most forms of hearing impairment, the standard medical intervention is provision of a hearing aid, and in some circumstances hearing aids can increase SNRs, for example by incorporating directional microphones (e.g., Picou et al., 2014), although these increases in SNR are small in realistic environments (e.g., Ricketts & Hornsby, 2003; Dittberner & Bentler, 2003). Such increases in SNR should provide increases in intelligibility, though the amount can vary, as it depends on the slope of the psychometric function (e.g. MacPherson and Akeroyd 2014), but it may not always be the case that the increases are noticeable, meaningful, or important to users.

We argue that noticeability, meaningfulness, and importance need be carefully distinguished. Our previous work has shown the *just-noticeable difference* (JND) for a change in SNR, using sentences in same-spectrum noise, to be approximately 3 dB regardless of hearing loss (McShefferty et al., 2015). An SNR change of 3 dB is necessary, then, for an immediately and reliably noticeable change. However, this does not indicate how large a change in SNR needs to be for it to be meaningful. Given that a hearing aid is a medical intervention that someone wears to improve their hearing, we define this change, the *just meaningful difference* (JMD), as the minimum increase in SNR necessary for someone to seek an intervention, such as by the uptake or renewal of a hearing device.

The JMD bears a strong resemblance to the *clinically important difference* (CID), as the CID is regarded as a change in outcome that would be considered meaningful to a patient after some form of intervention. Various terms have been used in prior work to describe such changes, including the minimal clinically important change (e.g., van den Roer et al., 1976), the minimal important change (e.g., Juniper et al., 1994) and the minimum clinically important difference (Jaeschke et al., 1989). The latter is a threshold value that has been defined as “the smallest difference in score in the domain of interest which patients perceive as beneficial” (ibid., p. 408) or alternatively “the smallest change that is important to patients” (Stratford et al., 1998, p. 1188). What is beneficial or important to an individual, though, is often neither a decrease in disease prevalence (e.g., “clinically impressive”) nor determined solely by statistical inference, such as confidence intervals (Newman et al., 1991) or critical differences (e.g., Cox et al., 2001) for normative data. What is unclear from these statistical definitions of CID is whether any of these statistically relevant benefits are *perceptually* relevant to patients; this perceptual relevance is the crucial distinction between the JMD here and the various previous forms of the CID.

The JND can be measured using laboratory psychophysical techniques and as such can be regarded as objective. Its measurement scale, decibels, is easily appreciable to the scientist or clinician but can be of uncertain meaning to the patient. In contrast the JMD is subjective, as it fundamentally relies on a person’s opinion. Subjective patient-reported outcomes are commonly used to establish improvements (or lack of) after clinical intervention, and they often have abstract and ordinal units of measurement. In the case of hearing aid benefit, outcomes are important since improvement in an objective measure, such as a speech recognition in noise test (e.g., Bilger et al., 1984; Nilsson et al., 1992), does not always correspond to a patient’s subjective evaluation of benefit after intervention (Saunders & Forsline, 2006; McClymont Browning & Gatehouse, 1991). Analysis of hearing ability and hearing-aid benefit typically combines both subjective and objective measures, but rarely bridges the gap between the subjective and the objective.

In an attempt to reconcile differences between subjective and objective ratings of hearing ability and hearing aid benefit, Saunders et al. (2004) developed the Performance-Perceptual Test. It was based on measuring both the SNR for 50% correct identification of speech (the HINT sentences; Nilsson et al., 1992) and the SNR at which participants self-reported that they could *just* understand all of the speech (cf. NH estimates of consonant recognition; Rankovic & Levy, 1997). The difference in SNRs was termed the Performance-Perceptual Discrepancy (PPDIS) and was used to quantify how much a listener under- or over-estimates their hearing ability. The same test materials, testing format and unit of measurement (SNR in decibels) were used to measure both thresholds. Listeners were tested unaided. Results showed that while NH listeners had significantly better thresholds than HI listeners, PPDIS values did not differ between NH and HI groups and were not related to age. Reported hearing handicap (using the Hearing Handicap Inventory for the Elderly/Adults; Newman et al., 1990; Ventry & Weinstein, 1982) was affected just as much by listeners' perception of their hearing ability (their PPDIS) as by their speech-recognition ability. That is, the PPDIS indicates an aspect of handicap at a given SNR not revealed by speech-recognition ability at that SNR. These results indicate that the PPDIS can be important for clinical practice as it probes handicap and expectations (Saunders & Forsline, 2006), but it does not measure either the just noticeable or just meaningful change.

There are two previous instances of measuring a "just meaningful difference" from two disparate fields: economics and birdsongs. Zedeck and Smith (1968) appear to have first coined the term JMD as the standard deviation for salaries based on subjective responses to different values (namely categories of fair pay, more than fair pay or less than fair pay). The authors suggested that the JMD for salary indicates the range within which different levels of experience can be rewarded while still deemed equitable. Nelson and Marler (1990) separately developed a JMD for birdsongs, being the minimal change in a signal feature (e.g., pitch, duration) that elicited a measurable difference in behavior (e.g., wings flapping).

Both of these previous instances of a JMD used a change of at least x units of standard deviation as the underpinning definition of importance or measurability (e.g., for Nelson and Marler it was 2.5 units). They are arbitrary in the amount of change required – the value of x – but also standard deviation is, by definition, derived from a population of responses. As it is not *a priori* obvious to us that a particular *individual* should regard as meaningful to her or him an arbitrary change calculated from a population, our definition of the speech-to-noise JMD deliberately avoids standard deviation in its definition. However, it maintains two aspects of these previous uses of the term: we measure subjective responses to achieve an objective benchmark of meaningful change (cf. Zedeck & Smith, 1968) and we aim to measure the smallest difference in SNR that would elicit a change in behavior (cf. Nelson & Marler, 1990).

The four experiments of the current study were designed to examine what is a meaningful increase in SNR using both objective and subjective methods. Items from a corpus of short sentences partially masked by a speech-shaped noise were presented in a two-interval fixed-level procedure. Participants compared the SNR of a reference interval (SNR_R) with the SNR of a test interval ($SNR_T = SNR_R + \Delta SNR$), with the value of the change (ΔSNR) chosen from predefined sets of values. The tasks required of the listeners varied across the four experiments, though all used similar stimuli as examples of pre- and post-benefit situations. In Experiment 1 participants performed a paired-comparison better/worse rating task. Paired examples of reference and target intervals were presented, and participants were asked to rate the second presentation compared to the first. In Experiment 2 participants performed a derivative of the willingness-to-pay paradigm (cf. Chisolm & Abrams, 2001), probing whether participants were willing to swap devices. The yes/no task asked participants if they would swap the reference SNR (which they were told represented their current device) for the improved SNR example (representing a new or different device). In Experiment 3 participants performed a novel subjective-comparison task that took clinical significance literally: they were asked if they would be willing (yes/no) to attend the clinic for a given SNR increase (benefit) or decrease (deficit). In Experiment 4 the same clinical significance task was re-examined using a different, larger set of participants and a reduced set of conditions. In Experiments 1 and 4 participants also performed an SNR JND task to corroborate previous results (McShefferty et al., 2015) and to examine how the JND compared to the JMD. The JMD was calculated from the ΔSNR condition where responses were statistically greater than a particular limen (one unit in Experiment 1, 50% in Experiments 2-4).

METHODS

Participants

In all four experiments, participants were recruited from local hearing clinics. This study was approved by the West of Scotland research ethics service (WoS REC(4) 09/S0704/12) and informed written consent was obtained from all participants prior to commencing experimentation. Pure-tone thresholds were measured using the modified Hughson-Westlake method (British Society of Audiology, 1981). Participants were classified as NH if their better-ear four-frequency pure-tone average hearing loss (BE4FA; average of 0.5, 1, 2 and 4 kHz) was less than 25 dB HL (hearing level) (cf. Clark, 1981). The loss type of HI participants was based on air-bone threshold differences (British Society of Audiology and British Academy of Audiology Guidelines, 2007). Table 1 gives the number of participants, the range of BE4FAs and ages for each experiment.

Table 1. General demographics of participants in each experiment, showing the number (*N*) of participants including gender distribution, and medians and ranges in parentheses for better-ear four-frequency average hearing thresholds (BE4FA) and age.

Experiment	**N / N female**	**BE4FA** (dB HL)	**Age** (years)
1	32 / 18	21 (3 - 58)	64 (31 - 74)
2	31 / 19	33 (4 - 48)	62 (38 - 74)
3	21 / 13	24 (-1 - 56)	63 (41 - 76)
4	36 / 15	28 (3 - 56)	63 (22 - 72)

For Experiment 1, 35 participants (21 female) were recruited. One of the participants was unresponsive, failing to understand the task despite demonstration. Two others were excluded as the severity of their hearing loss meant the stimuli were presented at a sensation level (SL) of < 15dB. Fourteen of the remaining 32 participants were classified as HI; all had a sensorineural hearing loss. In Experiment 2, 39 participants (22 female) were recruited. One participant was unable to complete the task due to time constraints, three were unresponsive, and four were excluded due to presentation levels < 15 dB SL based on BE4FA. Twenty of the remaining 31 participants were classified as HI. Three had a conductive hearing loss, and 17 had a sensorineural hearing loss. Participants for Experiment 2 were also queried about their use of hearing aids. Nineteen participants responded that they had at least tried a hearing aid (median BE4FA = 35 dB HL; median age = 65 years); the remaining 12 participants had not (median BE4FA = 19 dB HL; median age = 60 years). In Experiment 3, 27 participants (15 female) were recruited. One participant was unable to complete the task due to time constraints, four were unresponsive, and one was excluded due to presentation levels < 15dB SL. Ten of the remaining 21 participants were classified as HI, all with a sensorineural hearing loss. In Experiment 4, 46 participants (20 female) were recruited. Ten were unresponsive. Nineteen of the remaining 36 participants were classified as HI; one had a conductive hearing loss and 18 had a sensorineural loss.

Stimuli

The stimuli for Experiments 1 through 4 were male-talker IEEE sentences (Rothauser et al., 1969) embedded in a speech-shaped noise. These were chosen to allow a direct comparison with our previous JND work (McShefferty et al., 2015). The corpus consisted of 720 individual sentences with durations ranging from 1360 to 2997 ms. The sentences were originally recorded at University College London with a native speaker of British English at a sampling rate of 48 kHz (Smith and Faulkner, 2006). Sentences were then filtered to match the SII standard speech spectrum (ANSI, 1997) for normal vocal effort (i.e. a constant spectrum level for frequencies up to 500 Hz then a slope of -9 dB/octave). White noise of the same duration as each chosen sentence was generated in Matlab (R2013b version 8.2.0.701, The Mathworks Inc.) and filtered using coefficients obtained from the average spectrum of the entire equalised male-talker sentence set. Both the speech and the noise were resampled to 44.1 kHz for playback to participants. In each single trial, the duration of the noise was set to equal that of the randomly chosen sentence. Speech and noise were added together for simultaneous presentation and raised-cosine ramps of 20-ms were applied to the onset and offset of the composite speech-and-noise stimulus.

In each trial of every experiment, a sentence was chosen at random and presented in noise in two intervals: a reference interval with one value of speech-to-noise ratio (SNR_R) and a target interval (SNR_T) at the reference SNR plus an increment (ΔSNR) chosen from a predefined set of values. Differences in SNR_R and ΔSNR used in each of the experiments are given in the Procedures section below. Note that the same sentence was used in both intervals but the samples of noise differed across the intervals. The interstimulus interval on each trial was 500 ms.

The actual presentation levels of the speech and the noise were obtained from the SNRs using a three-step algorithm (McShefferty et al., 2015). First, in the reference interval, the speech was presented at an A-weighted level of 63 dB SPL plus ½ of SNR_R and the noise was presented at an A-weighted level of 63 dB SPL minus ½ of SNR_R. In the target interval, the speech was presented at 63 dB (A) plus ½ of SNR_R plus ½ of ΔSNR and the noise at 63 dB (A) minus ½ of SNR_R minus ½ of ΔSNR. Second, both of the two combined speech-plus-noise mixtures were adjusted to give an overall level of 63 dB (A) SPL. Third, if the participants' BE4FA was < 65 dB HL the reference A-weighted presentation level was 63 dB SPL but otherwise the stimuli were presented at 73 dB SPL, ensuring at least 15 dB SL based on BE4FA for all participants. For the SNR discrimination (JND) task in Experiments 1 and 4, the overall levels of the combined stimuli in each interval were then roved independently by a maximum of ±2 dB in randomized (rectangular distribution) increments of 0.1 dB to partially reduce the possibility that participants would use the level of either the noise or the speech as a cue (McShefferty et al., 2015).

Apparatus

During all four experiments, participants were seated in a sound-proof audiometric booth. Stimuli were presented diotically via a PC and USB external sound card (High Resolution Technologies microStreamer) to circumaural headphones (AKG K702). Participants' responses were recorded via a touch screen monitor.

Procedures

Experiment 1

In Experiment 1, participants undertook both an SNR discrimination task and a rating task. The order of the tasks was alternated across participants. SNR discrimination thresholds were obtained using a 2AFC fixed-level procedure. The SNR_R was 0 dB, and ΔSNR was 1, 2, 4, 6 or 8 dB. Participants were instructed to select the interval that was clearest to them and informed that it may not necessarily be the loudest interval. After a short practice (ten trials, two at each value of ΔSNR) to introduce the task, participants were asked if the sounds were too loud or too quiet and if necessary the presentation level was changed by ±10 dB (i.e., 63 to 73 if too quiet, 73 to 63 dB if too loud). Following the practice, six blocks of 20 trials were run, resulting in 12 repeats of each of the five ΔSNR values where SNR_T was presented in the first interval and 12 repeats where SNR_T was presented in the second interval.

Prior to commencing the rating task in Experiment 1, participants were given the following on-screen instructions: "*In each trial of this experiment you will hear a sentence presented in noise twice. We will ask you to judge if the second example is better, the same, or worse than the first.*" If the participant asked for clarification, "better" was further defined as being clearer or easier to listen to. After each trial participants were asked "*How was the second example compared to the first?*" and responded by pressing one of eleven buttons (marked -5 to +5) to indicate their rating. Text anchors with the words "*Much Worse*", "*Same*" and "*Much Better*" were placed below buttons -5, 0 and +5 respectively. Of the 14 HI participants, 13 completed the experiment at an A-weighted presentation level of 63 dB SPL and one did so at 73 dB SPL.

Experiment 2

In Experiment 2, two SNR_R values (-6 and +6 dB) were tested in a subjective "willing to swap" comparison task to estimate the JMD for SNR. The ΔSNR values tested were 2, 4, 6 and 8 dB. Participants completed three blocks for each reference condition in random order. During the reference interval the touchscreen displayed the phrase "*Your device sounds like this.*" During the target interval, the phrase "*A different device sounds like this*" was displayed. After both intervals, participants were asked "*Would you swap your device for the different device*?" and responded by choosing the appropriate button marked "*Yes*" or "*No*" on the touchscreen. After eight practice trials (one for each reference SNR at all ΔSNRs), participants completed 240 trials: three blocks of 40 trials at each SNR_R with 10 repeats of each SNR increment per block. Level roving was not applied to any of the stimuli in Experiment 2. All NH and HI participants in Experiment 2 completed the experiment at an A-weighted presentation level of 63 dB SPL.

Experiment 3

In Experiment 3, three SNR_R conditions (-6, 0 and +6 dB) were used in a subjective "clinical significance" comparison task to estimate the JMD for SNR. In half of the blocks of trials a positive SNR change was used, and in the other half a negative SNR change was used. Participants completed all of one block type before commencing the other with the starting type alternated across participants (this was done to avoid confusion). Prior to the positive-change blocks, participants were given the following instructions verbally and written: "*Consider the first presentation as an example of a conversation you are having. Consider the second as an example of the benefit (compared to the first) you would get if you attended a clinic (e.g. getting a new/adjusted hearing aid). After both presentations we will ask you if the improvement is worth going to a clinic (and the time and effort involved in doing so).*" Prior to the negative-change blocks, the following instructions were given: "*Consider the first presentation as an example of a conversation you were having. Consider the second as an example of the increased deficits/difficulties you are now having in that conversation. After both presentations, we will ask you if it is worth going to the clinic (and the time and effort involved) if it made the second presentation as clear as the first.*" On each trial, participants were prompted with "*Would you go to the clinic if it made the first sound as clear as the second*?" in the positive SNR change conditions and "*Would you go to the clinic if it made the second sound as clear as the first*?" in the negative SNR change conditions. In both cases participants responded by choosing the appropriate button marked "*Yes*" or "*No*" on the touchscreen. Twenty-one practice trials (one at each SNR_R and ΔSNR) of the appropriate type were completed before both negative and positive condition blocks. After practice, each participant completed 420 trials: ten repeats with ΔSNR values of 0.5, 1, 2, 3, 4, 6 and 8 dB and ten repeats with ΔSNR values of -0.5, -1, -2, -3, -4, -6 and -8 dB at three SNR_R values of -6, 0, and +6 dB. Level roving was not applied to any of the stimuli in Experiment 3. Of the 10 HI participants in Experiment 3, eight completed the experiment at an A-weighted presentation level of 63 dB SPL and two did so at 73 dB SPL.

Experiment 4

In Experiment 4, participants undertook both an SNR discrimination task and a truncated version of the clinical significance task (Experiment 3). The task order was alternated across participants. SNR discrimination thresholds were obtained using the same procedure as in Experiment 1 except that two conditions were tested, with SNR_R = -6 dB and +6 dB. The practice comprised 10 trials, one at each value of ΔSNR for each SNR_R. Following ten practice trials, each participant completed a total of 120 trials: six repeats of each of five ΔSNR values at 1, 2, 4, 6 and 8 dB where SNR_T was presented in the first interval and six repeats of the same ΔSNR values where SNR_T was presented in the second interval, for both the -6 and +6 dB SNR_R conditions.

The instructions for the clinical significance task of Experiment 4 were identical to those for Experiment 3 (for positive-SNR changes). After each trial, participants were asked "*Would you go to the clinic if it made the first sound as clear as the second?*" and responded by pressing one of two buttons marked "*Yes*" or "*No.*" As in the SNR discrimination task, two SNR_R conditions were tested: -6 and + 6 dB SNR. The same five ΔSNR values (1, 2, 4, 6 and 8 dB) were used and the same number of practice trials were completed. After those ten practice trials, each participant completed three blocks of 20 trials for each SNR_R condition, resulting in 12 repeats of each ΔSNR. One of each SNR_R type was run in random order, followed by a further two more of each in random order. Twelve of the 19 HI participants in Experiment 4 completed the experiment at an A-weighted presentation level of 63 dB SPL and seven did so at a presentation level of 73 dB SPL.

Data Analysis

The value of the SNR JMD was calculated as the change in SNR which gave a significant (based on within-subject confidence intervals; $p = 0.05$) increase compared to 1 response unit (Experiment 1) or to 50% affirmative (Experiments 2-4). While any criteria could be chosen, we chose one unit as the criterion for the rating experiment as responses were given in discrete one-unit steps, and chose 50% for the other, proportional-response experiments as we wanted to know what SNR change would induce intervention-seeking behaviour at least half of the time (i.e., when participants were more likely than not to seek such an SNR change). The JNDs in Experiments 1 and 4 were measured using a fixed-level procedure, estimating 79% correct using a log-likelihood logistic fit to the data. To counteract the problem of multiple comparisons, the Holm-Bonferroni method was used to adjust the rejection criteria of the individual comparisons where necessary (Holm, 1979).

RESULTS

Experiment 1

In Experiment 1, across all 32 participants, the JND for a change in SNR was 2.8 dB, 95% CI [2.34, 3.34]. NH participants ($n = 18$) gave a JND of 2.7 dB, 95% CI [2.06, 3.35]. HI participants ($n = 14$) gave a JND of 3.0 dB, 95% CI [2.24, 3.8]. From an independent-samples t test, no significant difference was found between NH and HI groups. There was no significant correlation between age and hearing loss, as measured by BE4FA (Pearson product-moment correlation coefficient $r = 0.07$, $p = 0.70$). Nor was there a significant correlation between age and JND ($r = 0.25$, $p = 0.16$), or between hearing loss and JND ($r = 0.09$, $p = 0.61$).

Figure 1 shows the rating results for Experiment 1. The ratings increased almost linearly as ΔSNR increased. Ratings for benefit (increased SNR) were significantly higher than those for deficit at all ΔSNR values tested. However, this may represent an order effect, as the interval with the increased benefit was always the second interval of the trial. The difference ranged from 0.53 at a ΔSNR value of 1 dB to a difference of 1.27 at a ΔSNR value of 8 dB. For Experiment 1, we defined the JMD as the SNR increase rated significantly better or worse than one discrete unit on the scale. A Wilcoxon signed-rank test showed that ratings for benefit were not significantly greater than one unit (+1) until a ΔSNR of 4 dB ($z = -3.00$; $p = 0.003$). Ratings for deficit were not significantly less than one unit (-1) at the maximum ΔSNR tested ($z = -1.96$; $p = 0.05$).

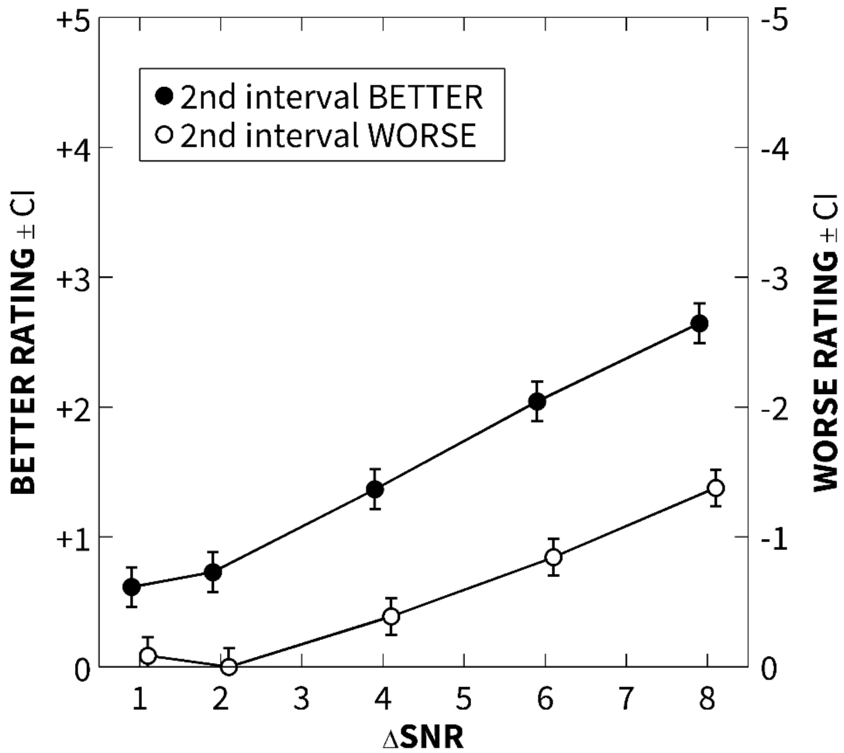

Figure 1. Mean rating results for all 32 (normal-hearing and hearing-impaired) participants in Experiment 1 as a function of ΔSNR (dB). Black circles show ratings for benefit (i.e.,

where the second interval was judged to be better than the first), white circles show ratings for deficits (i.e., where the second interval was judged to be worse than the first); error bars show 95% confidence intervals.

Experiment 2

For Experiment 2 we defined the JMD as the threshold for willingness to swap devices. Separate analyses were conducted for those participants who had at least tried hearing aids, and those who had never tried them (see Figure 2). For the -6 dB SNR_R condition, the JMDs for participants who had and had not tried hearing aids were 6 and 4 dB, respectively. For the +6 dB SNR_R condition, the JMDs for both those who had and had not tried hearing aids was greater than 8 dB (the highest ΔSNR tested). Responses at the lowest ΔSNR tested (2 dB) were well below 50% for all conditions except for participants who had not tried hearing aids at -6 dB SNR_R, indicating a bias towards responding “no.”

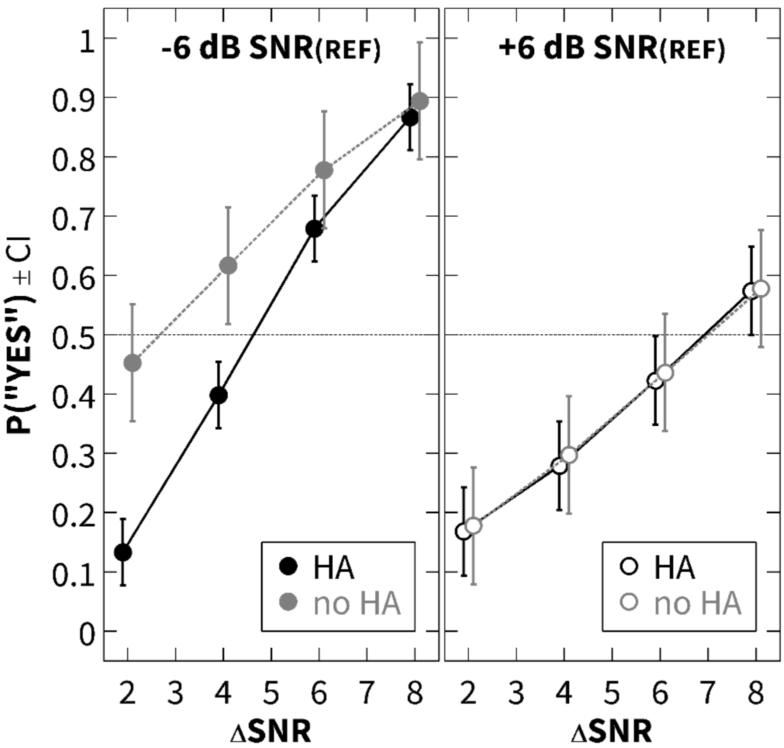

Figure 2. Mean proportion of “Yes” responses for all 31 (normal-hearing and hearing-impaired) participants in Experiment 2 as a function of ΔSNR (dB). Left panel shows

responses for the -6 dB reference SNR condition. Right panel shows responses for the +6 dB reference SNR condition. In both panels, black line and black circles show responses for those participants who had at least tried a hearing aid (n = 19), grey line and grey circles show responses for those who had never tried a hearing aid (n = 12). Error bars in both panels show 95% confidence intervals.

Experiment 3

For Experiment 3 we defined the JMD as the threshold for willingness to seek intervention (i.e., to go to the clinic) based on a change in SNR; results are shown in Figure 3. When ΔSNR was positive, the JMDs were 6, 6 and 8 dB for SNR_R of -6, 0 and +6 dB, respectively. When ΔSNR was negative, the JMDs were 8 dB for all SNR_R. While independent samples t tests revealed significant differences in willingness to attend a clinic at various ΔSNR values when SNR_R was -6 dB, the two participants who had the higher presentation level could be regarded as outliers in this condition. That is, when ΔSNR was negative, one of the two showed almost 100% willingness at all ΔSNR values tested and when ΔSNR was positive both responded at approximately 50% across all values tested. Hence, *p* values are not reported here.

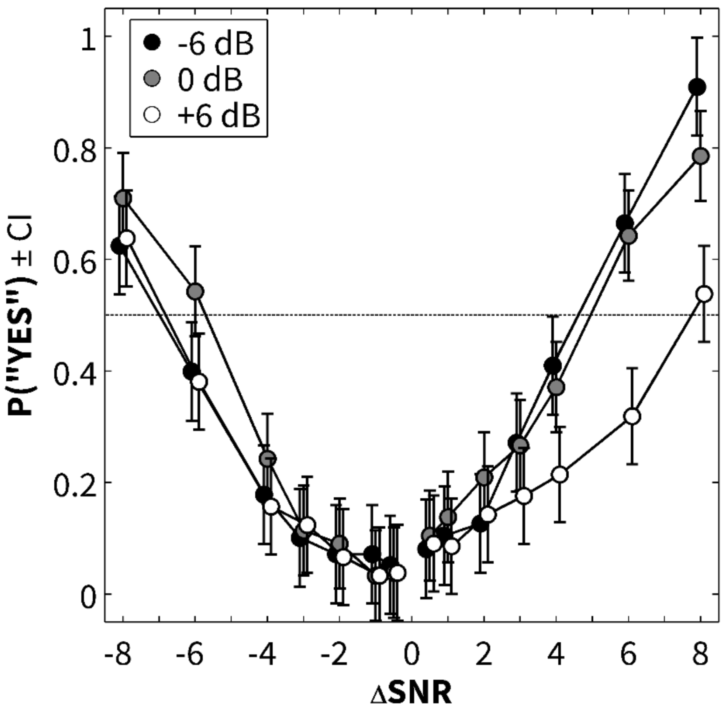

Figure 3. Mean proportion of "Yes" responses for all 21 (normal-hearing and hearing-impaired) participants in Experiment 3 as a function of ΔSNR (dB). Black filled circles show responses for the -6 dB reference SNR condition. Grey filled circles show responses for the 0 dB reference SNR condition and white filled circles show responses for the +6 dB reference SNR condition. Error bars show 95% confidence intervals.

Experiment 4

The mean SNR JNDs are shown in Table 2. When SNR_R was +6 dB, eight participants had unusually high JNDs (μ = 10.2 dB, 95% CI [9.0, 11.5]), due to the fact that they did not achieve > 79% correct at the highest ΔSNR value tested (8 dB) and the logistic fits to their data were of poor quality. Hence, for the remainder of the analysis we consider these 8 as a separate group (termed Group H, for High) from the remaining 28 participants (termed Group L). One participant in the -6 dB SNR_R condition had a JND over 3 standard deviations from the group mean (7.5 dB). Hence this result was not included in the group averages (and comparisons for that condition).

Table 2. Summary of SNR JND results for Experiment 4, showing paired comparisons between groups. Student's *t* statistic is shown for each comparison; *p* values for significantly different means are shown in parentheses. For the NH/HI distinction see text. Asterisk denotes comparison rejected by Holm-Bonferroni method for adjusting for multiple comparisons (.048 → 0.143).

Group	***n***	**-6 dB SNR$_R$**	← *t* (*p*) →	**+6 dB SNR$_R$**
All	36(28)	**2.8 dB**	2.97 (0.0043)	**3.7 dB**
Group L	28	**2.5 dB**	4.47 (0.00053)	**3.7 dB**
↑ *t* (*p*) ↓		2.84 (0.0077)		
Group H	8	**3.6 dB**		
Group L-NH	15	**2.4 dB**	2.17 *	**3.3 dB**
↑ *t* (*p*) ↓		0.70		1.82
Group L-HI	13	**2.7 dB**	4.95 (0.0017)	**4.3 dB**
Group H-NH	2	**3.75 dB**		
↑ *t* (*p*) ↓		-0.22		
Group H-HI	6	**3.52 dB**		

As shown in Table 2, across all participants there was a significant difference between mean JNDs in the -6 and +6 dB SNR$_R$ conditions. Examining only the 28 participants in group L, there was still a significant difference between these two conditions (post hoc comparisons shown between means in Table 2). When group L was divided into NH and HI sub-groups, there was a significant difference between the -6 and +6 dB SNR$_R$ conditions for the L-HI group only. For the -6 dB SNR$_R$ condition, there was a significant difference between the L and H groups. There were no significant correlations between age, hearing loss and JND for either participant group.

The JMD results (clinical significance) are shown in Figure 4. The JMD in the -6 dB SNR_R condition was 6 dB for both JND groups (L and H). For the +6 dB SNR_R condition, the JMD was greater than 8 dB for both groups.

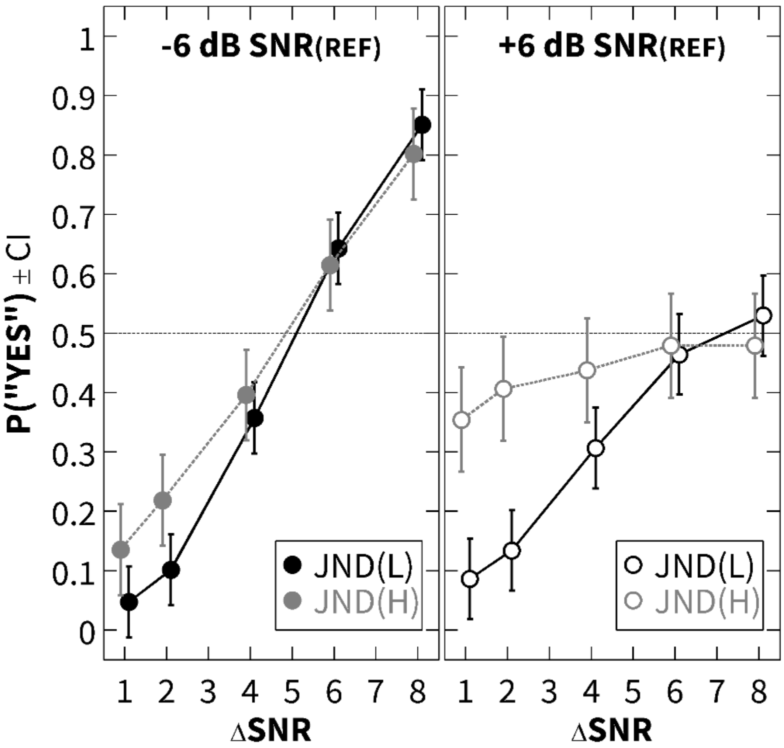

Figure 4. Mean proportion of “Yes” responses for all 36 (normal-hearing and hearing-impaired) participants in Experiment 4 as a function of ΔSNR (dB). Left panel shows responses for the -6 dB reference SNR condition. Black line and black filled circles show responses for participants who had low SNR JNDs (n = 28), grey line and grey filled circles show responses for those who had high SNR JNDs (n = 8). Right panel shows responses for the +6 dB reference SNR condition. Black line and white filled circles show responses for participants who had low SNR JNDs, grey line and white filled circles show responses for those who had high SNR JNDs. Error bars in both panels show 95% confidence intervals.

DISCUSSION

Table 3. Summary of JND and JMD results across experiments, showing mean limens in dB SNR. JND results are collated from Experiments 1 and 4 and show mean limens ± one standard deviation. Rating JMDs (Experiment 1) are shown for when the better-SNR interval was second. Swap JMDs (Experiment 2) are shown for those who had at least tried a hearing aid in the past (*n* = 19). Clinical significance JMDs (CS I & II; Experiments 3 & 4, respectively) results are shown for all participants.

		Reference SNR		
		-6 dB	0 dB	+6 dB
	JND	**2.8 ± 1.0**	**2.8 ± 1.4**	**3.7 ± 1.5**
JMD	Rating		**4**	
	Swap	**6**		**>8**
	CS I	**6**	**6**	**8**
	CS II	**6**		**>8**

The JND in SNR

The SNR JND was measured in Experiments 1 and 4 of the current study. The SNR JNDs for SNR_Rs of -6, 0 and +6 dB were 2.8, 2.8 and 3.7 dB SNR, respectively (see Table 3 above). The latter two JNDs are similar to the 2.9 and 3.5 dB SNR JNDs measured in our previous study for 0 and +6 dB SNR_R (McShefferty et al., 2015), despite overall presentation levels being lower in the current study. This suggests that overall presentation level did not affect SNR JND, at least within the range used across both studies. Further work should be undertaken to establish if this holds across a full range of presentation levels. Similar to our previous study, across both current experiments, NH participants gave on average slightly lower SNR JNDs than their HI counterparts, and SNR JNDs increased slightly in the conditions where SNR_R was more favorable. In both our previous and current studies, the JNDs were lower (better) when SNR_R was less favorable. This may be due to the less favorable SNRs, on average, being on a steeper point of the psychometric function. From a higher performance point along the function, a greater change in SNR would be necessary to elicit the same change in performance. This explanation, though, assumes both that the less favorable SNRs were indeed along the steeper slope of the function and that the JND represents a fixed change in intelligibility. Neither assumption was tested in the current study.

The JMD in SNR

When participants were asked to rate the second of a pair of stimuli in relation to the first in Experiment 1, ratings for both benefit and deficit trials were not significantly different from that for the minimum ΔSNR tested until ΔSNR was 4 dB. Benefits were rated on average as better by one unit at a ΔSNR of 4 dB, whereas deficits were rated worse by one unit only at 8 dB. However, the primary issue with using better/worse ratings is the interpretability of responses; not only is it difficult to interpret "one unit" better on a ±5-point scale, but it is also unclear what "one unit" better means clinically. There was also a clear order effect in Experiment 1. Other studies have shown order effects in speech intelligibility (e.g., Thwing, 1956), and it is possible that our results could have overestimated benefit based on increased intelligibility in the second presentation.

To measure the just-meaningful difference (JMD) in SNR with more clinical relevance, two methods were used across three experiments. When asked if they would swap their current device for a different one in Experiment 2, participants did not respond "Yes" more than 50% of the time until ΔSNR was 4 - 6 dB in the least favorable SNR_R condition. Participants who had never tried hearing aids were more likely to swap at each ΔSNR value but the difference between groups was reduced as ΔSNR increased. In the more favorable reference condition "Yes" responses from both groups did not exceed 50% even at the highest ΔSNR tested, and there were no significant differences between groups at any of the ΔSNR values tested. It seems likely that when the speech was 6 dB greater in level than the noise in the SNR_R interval, and therefore more audible, for both participant groups there was less advantage to be gained by swapping devices and the proportion of “Yes” responses fell accordingly. This pattern also occurred in Experiments 3 and 4. When asked if they would attend the clinic for a given increase in SNR in Experiment 3, participants did not respond affirmatively more than 50% on average until ΔSNR was -8 dB (when ΔSNR was negative) in all three reference SNR conditions. When ΔSNR was positive, “Yes” responses did not exceed 50% until ΔSNR was 6 dB (and 8 dB for the most favorable SNR_R). The mean proportions of “Yes” responses were consistently higher when ΔSNR was positive than when it was negative, except for the most favorable SNR_R condition. When asked the same question in Experiment 4, the mean proportion of “Yes” responses for participants in both L and H groups (based on their JND thresholds) did not exceed 50% until ΔSNR was 6 dB when SNR_R was least favorable (-6 dB), and responses for neither group significantly exceeded 50% even at the highest ΔSNR value tested when SNR_R was most favorable (+6 dB). These findings across Experiments 2-4 correspond to a 50% JMD estimate of approximately 6 dB for -6 and 0 dB SNR conditions, and 8 dB for +6 dB SNR (see Table 3). As these are JMDs for changes in SNR, a JMD of 6 dB means that a change of 6 dB of SNR needs be supplied for someone, on average, to consider it worth seeking intervention, whether by swapping their device(s) or attending the clinic.

The current study also highlights the difference between what is a *noticeable* and what is a *meaningful* difference in SNR (there was a lack of JND to JMD correlations). While participants were able to detect differences in SNR of 3 dB, those differences were not deemed to be clinically important (i.e., participants were unwilling to swap devices or to attend the clinic for differences of that magnitude). Only when differences in SNR reached at least 6 dB did participants find them meaningful enough to consider intervention. The varying gap between JND and JMD for each individual could stem from the additional variance in the subjective decision-making process of measuring the JMD. That is, the varying gap between JMD and JND could be due to the varying complexity of the tasks used to measure them. When asked to detect a difference, subjects were often consistently accurate without too much effort. Being asked to swap devices or attend a clinic involves a much more complex thought process.

Another distinction is that the JMD was calculated in Experiments 2-4 as a change in SNR equivalent to 50% “Yes”, while the JND was calculated as the 79% point on the psychometric function. That is, the SNR JMDs reported here only represent a participant being willing to swap or attend the clinic more than 50% of the time.

Limitations

Several of the experiments in the current study had a relatively high number of participants who were excluded from the reported results. A small number of these were due to time constraints, some were due to an apparent failure to understand the task and in some cases participants were unresponsive (i.e., they gave the same response to all stimuli in all conditions). It is unclear why some participants had these difficulties, but not others, since all were given the same written instructions. The reduced condition set in Experiment 4 was an attempt to eradicate these difficulties but in fact Experiment 4 had the highest proportion of exclusions of all the experiments. The lowest number of exclusions was for better/worse ratings, which conversely were the least interpretable. Despite attempts to make a clinically significant JMD task that was simple enough to be fathomable to all, further refinement may be required. Across Experiments 1-3, several participants were also excluded from the reported results due to poor audibility of the stimuli (i.e., the stimuli were presented at < 15 dB SL). It is possible that for some of the remaining participants, the outcomes of these experiments may not be representative of what would be obtained under conditions of greater audibility. With hindsight, frequency-selective amplification could have been used to partially compensate for the hearing losses of some participants.

In the current experiments, the SNR was adjusted without regard to signal spectrum. The noise reduction schemes of current digital hearing aids, whether single microphone (e.g., spectral subtraction) or multiple microphone (e.g., directionality), are frequency specific. It is unclear how frequency-dependent changes would affect either the JND or JMD.

The noise masker used in this series of experiments was a speech-shaped unmodulated noise, based on the average spectrum of the entire male-talker IEEE corpus. It is possible that both the JND and JMD could change using other potential maskers (e.g., a single competing talker or multi-talker babble) or in a more realistic scenario with spatial separation between speech and masker. Measuring the SNR JMD differently, such as with ratings of listening effort or fatigue, may also affect the value as well as the definition, although noise reduction has not been recently shown to affect effort (Wu et al., 2014) or fatigue (Hornsby, 2013).

Finally, we note that our experiments used two-interval methods in which one stimulus quickly followed another. They therefore essentially measure what is meaningful instantaneously – here over 2-3 seconds. It is possible that what becomes meaningful over hours, days and weeks may differ greatly. The scale of the JMDs measured here indicates that when fitting a hearing aid with noise-reduction features, those features may not be wholly convincing right away, but they may be appreciated over time.

CONCLUSIONS

The data of the current study confirm earlier results which showed the JND in SNR to be approximately 3 dB for sentence-in-noise stimuli. The JMD for the same stimuli, when measured as a change of 1 unit on a 11-point rating scale was also approximately 3 dB, but when the JMD was measured as a participant's willingness – 50% of the time – to swap devices or attend clinics for a change in SNR, it was approximately 6 dB for more difficult (lower SNR) situations, and 8 dB for less difficult situations (see Table 3). These latter, less arbitrary JMD values exceed what is currently possible with conventional hearing-aid technology.

ACKNOWLEDGEMENTS

We thank Prof. Andrew Oxenham, Prof. Brian Moore and an anonymous reviewer for comments on this manuscript. Portions of this research were presented at the 2014 International Hearing Aid Research Conference, Lake Tahoe, California, and the 2015 International Symposium on Hearing, Groningen, Netherlands. The Scottish Section of IHR is supported by intramural funding from the Medical Research Council (grant number U135097131) and the Chief Scientist Office of the Scottish Government.

REFERENCES

American National Standards Institute. (1997). *Methods for calculation of the speech intelligibility index (ANSI 3.5-1997).* New York: Acoustical Society of America.

Bilger, R.C., Nuetzel, J.M., Rabinowitz, W.M., Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27,* 32-48.

British Society of Audiology. (1981). Recommended procedures for pure tone audiometry using a manually operated instrument. *British Journal of Audiology, 15*, 213-216.

British Society of Audiology, & British Academy of Audiology. (2007). Guidance on the use of real ear measurements to verify the fitting of digital signal processing hearing aids. Retrieved from http://www.thebsa.org.uk/wp-content/uploads/2014/04/REM.pdf

Chisolm, T.H., Abrams, H.B. (2001). Measuring hearing aid benefit using a willingness-to-pay approach. *Journal of the American Academy of Audiology*, *12*, 383-389.

Clark, J.G. (1981). Uses and abuses of hearing loss classification. *ASHA*, *23*, 493-500.

Cox, R.M., Gray, G.A., Alexander, G.C. (2001). Evaluation of a revised speech in noise (RSIN) test. *Journal of the American Academy of Audiology*, *12*, 423-432.

Dittberner, A.B., Bentler, R.A. (2003). Interpreting the Directivity Index (DI). *Hearing Review, 10*, 16-19.

Grant, K., & Walden, B. (2013). Understanding excessive SNR loss in hearing-impaired listeners. *Journal of the American Academy of Audiology*, *24*, 258-273.

Holm. S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6(2)*, 65-70.

Hornsby, B.W. (2013) The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear and Hearing*, *34*, 523-534.

Jaeschke, R., Singer, J., Guyatt, G.H. (1989). Measurement of Health Status: Ascertaining the Minimum Clinically Important Difference. *Controlled Clinical Trials, 10(4),* 407-415.

Juniper, E.F., Guyatt, G.H., Willan, A., Griffith, L.E. (1994). Determining a minimal important change in a disease-specific Quality of Life Questionnaire. *Journal of Clinical Epidemiology*, *47*, 81-87.

McClymont, L.G., Browning, G.G., Gatehouse, S. (1991). Reliability of patient choice between hearing aid systems. *British Journal of Audiology, 25,* 35-39.

McShefferty, D., Whitmer, W.M., Akeroyd, M.A. (2015). The just-noticeable difference in speech-to-noise ratio. *Trends in Hearing, 19,* 1-9.

Nelson, D.A., & Marler, P. (1990). The perception of bird song and an ecological concept of signal space. In *Comparative perception: complex signals*, vol. 2 (eds W.C. Stebbins & M.A. Berkley), pp. 443–478. New York, NY: Wiley.

Newman, C.W., Weinstein, B.E., Jacobson, G.P., & Hug, G.A. (1990). The Hearing Handicap Inventory for Adults: Psychometric adequacy and audiometric correlates. *Ear and Hearing, 11,* 430-433.

Newman, C.W., Jacobson, G.P., Hug, G.A., Weinstein, B.E., & Malinoff, R.L. (1991). Practical method for quantifying hearing aid benefit in older adults. *Journal of the American Academy of Audiology*, *2*, 70-75.

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. *Journal of the Acoustical Society of America, 95,* 1085-1099.

Picou, E.M., Aspell, E., Ricketts, T.A. (2014). Potential benefits and limitations of three types of directional processing in hearing aids. *Ear and Hearing, 35,* 339-352.

Rankovic, C.M., Levy, R.M. (1997). Estimating articulation scores. *Journal of the Acoustical Society of America, 102,* 3754-3761.

Ricketts, T.A., Hornsby, B.W. (2003). Distance and reverberation effects on directional benefit. *Ear and Hearing, 24,* 472-484.

Rothauser, E., Chapman, W., Guttman, N., Hecker, M., Nordby, K., Silbiger, H., Urbanek, G., & Weinstock, M. (1969). IEEE Recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*, 225-246.

Saunders, G.H., Forsline, A. (2006). The performance-perceptual test (PPT) and its relationship to aided reported handicap and hearing aid satisfaction. *Ear and Hearing, 27,* 229-242.

Saunders, G.H., Forsline, A., Fausti, S.A. (2004). The performance-perceptual test and its relationship to unaided reported handicap. *Ear and Hearing, 25,* 117-126.

Smith, M.W., & Faulkner, A. (2006). Perceptual adaptation by normally hearing listeners to a simulated 'hole' in hearing. *Journal of the Acoustical Society of America*, *120*, 4019-4030.

Stratford, P.W., Binkley, J.M., Riddle, D.L., Guyatt, G.H. (1998). Sensitivity to change of the Roland-Morris back pain questionnaire: part 1. *Physical Therapy, 78,* 1186-1196.

Summerfield, Q. (1987). Speech perception in normal and impaired hearing. *British Medical Bulletin*, *43*, 909-925.

Thwing, E. J. (1956). Effect of repetition on articulation scores for PB words. *Journal of the Acoustical Society of America*, *28*, 302-303.

Wu, Y.H., Aksan, N., Rizzo, M., Stangl, E., Zhang, X., Bentler, R. (2014). Measuring listening effort: Driving simulator versus simple dual-task paradigm. *Ear and Hearing*, *35*, 623-632.

van der Roer, N., Ostelo, R.W., Bekkering, G.E., van Tulder, M.W., & de Vet, H.C. (1976). Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. *Spine*, *31*, 578-582.

Ventry, I.M., & Weinstein, B.E. (1982). The Hearing Handicap Inventory for the Elderly: A new tool. *Ear and Hearing, 3,* 40-46.

Zedeck, S., & Smith, P. C. (1968). A psychophysical determination of equitable payment: A methodological study. *Journal of Applied Psychology*, *52*, 343-347.