

On Detectable and Meaningful Speech-Intelligibility Benefits

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

Abstract The most important parameter that affects the ability to hear and understand speech in the presence of background noise is the signal-to-noise ratio (SNR). Despite decades of research in speech intelligibility, it is not currently known how much improvement in SNR is needed to provide a meaningful benefit to someone. We propose that the underlying psychophysical basis to a meaningful benefit should be the just noticeable difference (JND) for SNR. The SNR JND was measured in a series of experiments using both adaptive and fixed-level procedures across participants of varying hearing ability. The results showed an average SNR JND of approximately 3 dB for sentences in same-spectrum noise. The role of the stimulus and link to intelligibility was examined by measuring speech-intelligibility psychometric functions and comparing the intelligibility JND estimated from those functions with measured SNR JNDs. Several experiments were then conducted to establish a just meaningful difference (JMD) for SNR. SNR changes that could induce intervention-seeking behaviour for an individual were measured with subjective scaling and report, using the same stimuli as the SNR JND experiment as pre- and post-benefit examples. The results across different rating and willingness-to-change tasks showed that the mean ratings increased near linearly with a change in SNR, but a change of at least 6 dB was necessary to reliably motivate participants to seek intervention. The magnitude of the JNDs and JMDs for speech-intelligibility benefits measured here suggest a gap between what is achievable and what is meaningful.

W. M. Whitmer (✉) · D. McShefferty
MRC/CSO Institute of Hearing Research—Scottish Section, Glasgow, UK
e-mail: bill@ihr.gla.ac.uk

D. McShefferty
e-mail: david@ihr.gla.ac.uk

M. A. Akeroyd
MRC Institute of Hearing Research, Nottingham, UK
e-mail: maa@ihr.mrc.ac.uk

© The Author(s) 2016

P. van Dijk et al. (eds.), *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*, Advances in Experimental Medicine and Biology 894, DOI 10.1007/978-3-319-25474-6_47

Keywords Signal-to-noise ratio · Just-noticeable difference · Speech intelligibility · Hearing impairment

1 Introduction

The inability to provide appreciable speech-in-noise benefits can lead to the non-use of hearing aids (e.g., McCormack and Fortnum 2013). Non-use can come from unmet expectations (i.e., benefits without satisfaction; Demorest 1984), which may be at least partly a result of whatever changes the hearing aids provided being either undetectable to the user or not important enough to the user. The current study looked at both of these: how large a change in signal-to-noise ratio (SNR) has to be for it to psychophysically discriminable or clinically meaningful.

This is a curiously under-studied area: previously there has been only one small study related to measuring the just-noticeable difference (JND) in SNR by Killion (2004), in which unspecified participants were at chance discriminating a 2 dB SNR difference for speech in noise. A JND, though, is just a psychophysical benchmark; it is not clear *a priori* that a change of one JND is necessary or sufficient for the change to have any subjective importance to an individual. What is meaningful and clinically significant is key to service-wide treatments, and could be vital for determining provision criteria as well as tempering expectations for the patient. The current study looks at what could induce intervention-seeking behaviour for an individual; that is, how much subjective value do patients ascribe to discriminable speech intelligibility benefits? Here, we use psychophysical methods to rigorously measure (a) the JND for SNR in decibels, (b) the JND for intelligibility in %, and (c) the JND for meaningful benefit, using the same stimuli as examples of pre- and post-benefit situations.

2 Detectable Benefits

2.1 *The JND for Changes in SNR*

The JND for SNR was measured for adults of varying hearing ability at different reference SNRs using adaptive and fixed-level procedures based on the classic level discrimination paradigm (cf. McShefferty et al. 2015). The stimuli were equalised IEEE sentences uttered by a native British English speaker (Smith and Faulkner 2006) in same-spectrum noise, presented over headphones in a sound-attenuated booth. Noises began and ended simultaneous with the speech signal, and there was a 500-ms gap between intervals. Both intervals contained the same randomly chosen sentence, and after both intervals, participants were prompted to respond to which one was clearer. In each interval the levels of both speech and noise were adjusted to maintain a presentation level of 75 dB A; to minimize other cues, the level of each interval was roved independently by a maximum of ± 2 dB. In a two-interval/

Table 1 Mean (μ) SNR JND for reference SNRs of -6 , 0 and $+6$ dB pooled across experiments, showing standard deviation (σ) and participant number (n) tested at each reference SNR. Median BE4FAs and ages are given with ranges in parentheses

	-6 dB	0 dB	$+6$ dB
μ (dB SNR)	2.8	2.9	3.7
σ	1.0	1.0	1.4
n	35	99	72
BE4FA (dB HL)	28 (3–56)	29 (–1–85)	32 (–1–71)
Age (yrs)	63 (22–72)	65 (23–76)	64 (22–76)

alternative forced-choice task, participants were presented, in randomized order, with a reference interval at a fixed reference SNR (either -6 , 0 , or $+6$ dB), and a target interval at the reference SNR plus an increment (Δ SNR) that was either varied adaptively or presented at fixed, randomly interleaved Δ SNRs. The adaptive procedure estimated 79% correct from the geometric means of the best two of three three-up/one-down tracks. The fixed-level procedure estimated 79% correct from fitting a maximum-log-likelihood logistic function to the data. The procedures produced equivalent results to 0.1 dB, so results were combined across procedures for each reference SNR.

The results are shown in Table 1. The SNR JNDs at the three SNR references of -6 , 0 and $+6$ dB SNR were 2.8, 2.9, and 3.7 dB respectively; the first two were statistically no different and both less than the third [$t_{(91.3)}=4.07$; $p=0.0001$ and $t_{(121.4)}=4.11$; $p=0.00007$, respectively].

2.2 The JND for Changes in Intelligibility

The JND for intelligibility was estimated by first measuring psychometric functions for IEEE sentences based on keywords at SNRs of $-16/-12/-8/-4/0$ dB or $-8/-4/0/+4/+8$ dB. Twenty-four adult participants, median better-ear four frequency average (BE4FA) of 16 dB HL (range $-3-49$ dB HL) and median age of 57 years (range 27–74), were first presented with sentences at -16 and 0 dB SNR. Fourteen participants could not respond with half of the keywords (5 keywords/sentence) at 0 dB; those participants were tested at the $-8-8$ dB SNR range. The individual results, averaged across 50 sentences (250 keywords) at each SNR, were fit with a maximum-log-likelihood logistic function. The SNR corresponding to 50% correct on each individual’s psychometric function (SNR50) was calculated and then used as the reference for an adaptive-track JND experiment; the mean SNR50 was 0.20 dB. This gave a mean SNR JND of 3.0 dB ($\sigma=1.0$ dB), which, converted to intelligibility using the slope of the psychometric function, corresponded to a mean Intelligibility JND of 26% ($\sigma=7.5\%$).

This raises the question as to whether the JNDs initially measured (in Sect. 2.1 above) were intelligibility rather than SNR JNDs (i.e., were the listeners basing their responses on changes in word intelligibility or in changes in signal-to-noise ratio?) We reasoned that by changing the stimuli to ones with a far steeper psycho-

metric function (MacPherson and Akeroyd 2014), namely triple digits, the JND in the appropriate domain would be constant, but the JND in the other domain would change. On testing this hypothesis, we found that the mean SNR JND for digits was 2.5 dB ($\sigma=0.8$ dB) and the mean Intelligibility JND estimate for digits was 17.5% ($\sigma=7.2\%$). SNR and Intelligibility JNDs were significantly less using digits in unfiltered random (white) noise than sentences in same-spectrum noise [$t_{(23)}=2.98$; $p=0.007$ and $t_{(23)}=5.00$; $p=0.00005$]. Across stimuli, both the SNR JND differences (mean 0.5 dB) and the Intelligibility JND differences (mean 8.92 dB) were significantly non-zero ($z_{(23)}=2.98$; $p=0.003$, and $z_{(23)}=5.00$; $p\approx 0$). That *both* differences were non-zero was unexpected; it indicates that one cannot be certain whether the 3 dB SNR JND is indeed the JND for SNR or instead the JND for Intelligibility.

3 Meaningful Benefits

To ascertain what is a meaningful benefit to someone requires more than detectability; it requires the subjective input of the participant. The previous method for deriving the SNR JND was hence changed to subjective-comparison tasks using SNR and SNR + Δ SNR pairs as examples of, respectively, pre-benefit and post-benefit situations to measure the “just meaningful difference” (JMD) for SNR. It was not clear in advance to us what query, however, best represents *meaningfulness*. Four tasks are discussed below: a conventional better/worse rating task, a novel conversation-tolerance rating task, a swap-paradigm task and a clinical significance task. For the latter two tasks (Sect. 3.3 and 3.4), the JMD was considered as the Δ SNR where the proportion of affirmative responses (i.e., willingness to swap devices or attend the clinic) were significantly greater than chance (50%).

3.1 Rateable Benefits

The simplest, most direct rating of preference is a better/worse scaling procedure. Thirty six participants (median age: 62 years; range: 31–74) of varying hearing ability (median BE4FA: 21 dB HL; range 3–85 dB HL) heard two intervals of IEEE sentences in same-spectrum noise, one at 0 dB SNR, and the other at 0 dB plus a Δ SNR of 1, 2, 4, 6 or 8 dB. The overall level of each interval was 63 dB A (73 for those with more severe losses); unlike the JND experiments, the level was not roved across intervals. Participants were asked how much better/worse the second interval was on a discrete, signed 11-point scale with -5 marked “much worse” and $+5$ marked “much better.” Each of the ten conditions (5Δ SNRs $\times 2$ orders) was repeated 12 times in randomized order, resulting in 120 trials completed in three blocks of 40.

The results, including 95% within-subject confidence intervals, are shown in the left panel of Fig. 1. Ratings increase near linearly as a function of Δ SNR, and are asymmetric: when the $+\Delta$ SNR interval preceded the reference, ratings were less than when the $+\Delta$ SNR interval followed the reference (e.g., when the second

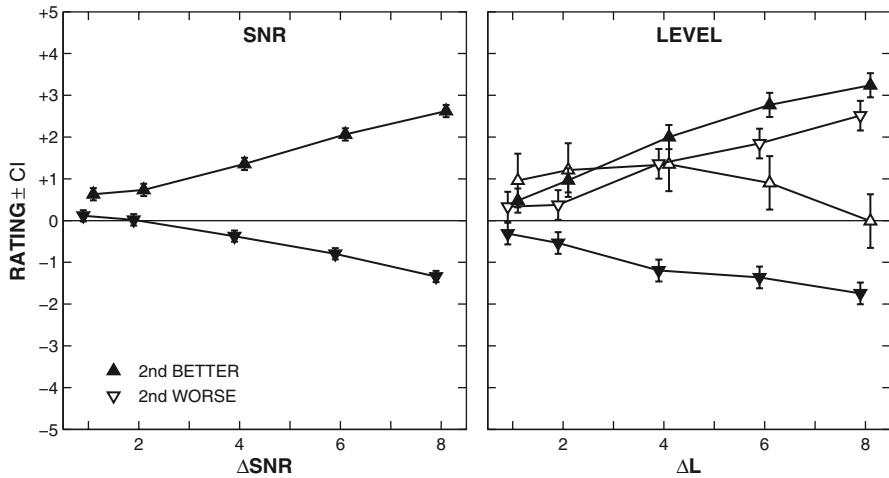


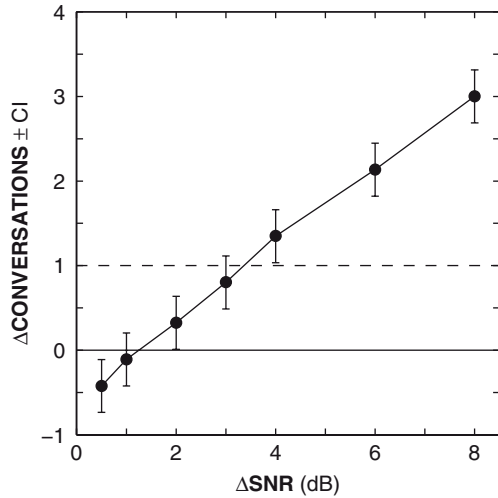
Fig. 1 Better (*upward triangles*) and worse (*downward triangles*) ratings as function of changes in SNR (*left panel*) and overall level (*right panel*). The reference SNR was 0 dB; the reference level was 70 dB A. Open *symbols/lines* (*right panel*) indicate mean ratings for six participants whose max./min. ratings did not occur at the extremities. *Error bars* show 95% within-subject confidence intervals (based on the analysis of variance subject \times condition interaction term)

interval was the better SNR, ratings were greater than one increment at a Δ SNR of 4 dB, but when the second interval was the worse SNR, ratings were greater than one increment only when Δ SNR was 8 dB). This asymmetry was most likely due to an order effect; the second iteration of a sentence has been shown to be more intelligible (e.g., Thwing 1956), hence second-interval deficits would be deflated, and second-interval benefits inflated. Regarding the JMD, the ratings significantly increased only at 4 dB, regardless of order, lending support to the notion of detectability being a requisite for meaningfulness. SNR JNDs for these participants, however, were not well correlated with their individual mean responses at any Δ SNR.

An analogous procedure was used to examine what ratings would apply to a change in level per se as opposed to a change in relative level due to a change in SNR. Thirty-six participants (median age: 67 years; range: 27–77) of varying hearing ability (median BE4FA: 37 dB HL; range: 4–84) heard stimuli at a fixed SNR of 0 dB. The reference level (L) was 70 dB A, and Δ L on any trial was 1, 2, 4, 6 or 8 dB. The level JND was also measured for each participant; this mean was 1.4 dB ($\sigma=0.4$ dB), similar to previous level JNDs for older hearing-impaired adults (Whitmer and Akeroyd 2011).

The results are shown in the right panel of Fig. 1. Six participants exhibited varying negative reactions to the greatest levels, finding a particular non-maximal level rated highest (e.g., 66 and 74 dB A). For the remaining 30 participants, the results are very similar to those for SNR, though Δ Ls of 4–8 dB were rated modestly higher than Δ SNRs of 4–8 dB. That is, a change in level without a change in SNR was considered as good as—indeed if not better than—a change in SNR. While this finding does not coincide with their role in intelligibility, it is coincident with detectability of level and SNR changes (JNDs of 1.5 and 3.0 dB, respectively).

Fig. 2 Mean change in conversations tolerated as a function of the change in SNR. Error bars show 95% within-subject confidence intervals



3.2 Tolerable Benefits

A second, novel method for establishing a meaningful difference was developed based around complaints of not being able to endure noisy conversation. For the conversation tolerance test, listeners heard the same sentence at two SNRs, one at an estimate of their speech-in-noise reception threshold (SNR₅₀) and one at SNR₅₀ plus an increment of 0.5–8 dB. In each of a pair of intervals, the participant was asked, “How many conversations in this situation would you tolerate?” (i.e., a paired single-interval task). Twenty-one adults (median age: 66 years; range: 52–73) with varying hearing ability (median BE4FA: 24 dB HL; range: 4–61) completed the task, repeating each increment in randomized order ten times.

The results are shown in Fig. 2. On average, participants were not prepared to tolerate an additional conversation until the SNR had increased by at least 3 dB, and not prepared to tolerate *more* than one extra conversation until a 4 dB change in SNR, similar to the results in Sect. 3.1. Though a change of one conversation is an arbitrary change, it is arguably a less arbitrary unit of meaningfulness than one point on a better/worse rating scale.

3.3 Swappable Benefits

If another hearing aid offers more speech-in-noise benefit, would someone be willing to trade for it? Using the same stimuli as the earlier rating experiment, 35 participants (median age: 62 years; range: 38–74) of varying hearing ability (median BE4FA: 34 dB HL; range: 4–80) were asked to consider the reference SNR (either –6 or +6 dB) interval as an example of their current device, and the reference SNR + ΔSNR (2, 4, 6 or 8 dB) interval as a different device. They were asked on each

trial if they would swap their current for the different device in order to get that change (yes/no). Each condition (reference SNR \times Δ SNR) was repeated 30 times in randomized order for a total of 240 trials.

The results are shown in the left panel of Fig. 3. When the initial SNR was -6 dB, the proportion of times that participants, on average, were willing to swap exceeded chance only when the Δ SNR was greater than 4 dB (i.e., 70% at 6 dB SNR, and 87% at 8 dB SNR). When the initial SNR was $+6$ dB, willingness to swap just exceeded chance (56% willing to swap) at a Δ SNR of 8 dB. The JMD based on the swap paradigm is therefore dependent on the difficulty of the situation (i.e., the reference SNR), but appears to be at least 6 dB SNR.

3.4 Clinically Significant Benefits

While there are statistical bases for a minimal clinically important difference (cf. Jaeschke et al., 1989), the goal here was to develop a benchmark for what speech intelligibility benefit is necessary to motivate an individual to seek intervention. Hence, “clinical significance” was applied in a literal sense: participants were asked whether a positive change in SNR, as an example of what a visit to the clinic would provide, was worth attending the clinic to get that change. Thirty-six participants (median age: 63 years; range: 22–72) of varying hearing ability (median BE4FA: 28 dB HL; range: 2–56) were presented stimuli as in Sect. 3.1 and 0 with 12 trials of each reference SNR (-6 and $+6$ dB) and each Δ SNR (1, 2, 4, 6 and 8 dB) for a total of 120 trials.

The results are shown in the right panel of Fig. 3. They were similar to the swap experiment (left panel of same figure), though there was a greater tendency towards

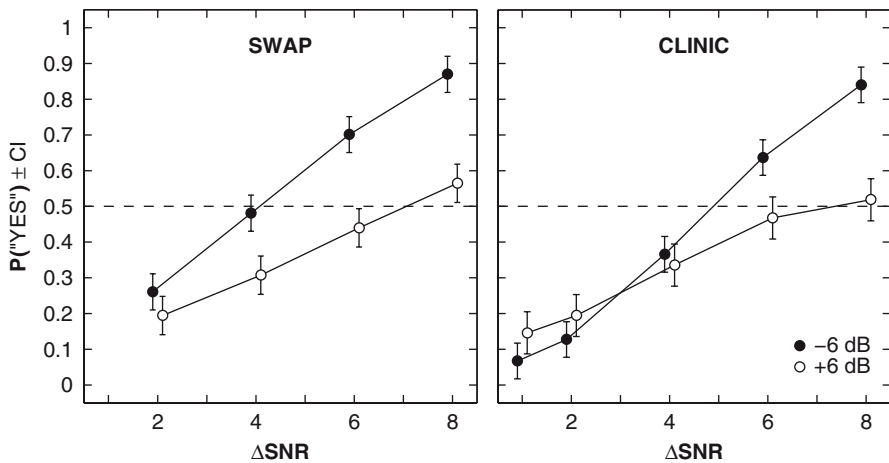


Fig. 3 Proportion of yes responses for willingness to swap devices (*left panel*) and willingness to attend the clinic (*right panel*) as a function of positive change in SNR (Δ SNR) for two reference SNRs: -6 dB (filled circles) and $+6$ dB SNR (open circles). Error bars show 95% within-subject confidence intervals

not attending the clinic compared to not swapping devices at the lowest, sub-JND Δ SNRs. For the -6 dB SNR reference condition, the mean proportion of yes responses only exceeded chance at 6 and 8 dB SNR. For the $+6$ dB SNR reference condition, mean responses never significantly exceeded chance; furthermore, the function appears asymptotic, so the responses may not have exceeded chance regardless of the change in SNR. If the SNR is already relatively high enough to be clear, then making it clearer still does not appear to induce intervention-seeking behaviour. The JMD derived from clinical significance is the same as the JMD derived from the swap task: at least 6 dB SNR.

4 Conclusions

To determine both detectable and meaningful speech-intelligibility benefits, participants were presented paired examples of speech in noise, one at a reference SNR and the other at a variably higher SNR. The threshold (JND) for a discriminable change was roughly 3 dB SNR. In more advantageous conditions, the SNR JND increased, and decreased for simpler speech (digits vs. sentences; cf. MacPherson and Akeroyd 2014). The threshold (JMD) for a meaningful benefit was at least 6 dB. Curiously, age and hearing loss were not factors well correlated with either individuals' just noticeable or just meaningful differences.

Given the difficulty in achieving SNR benefits greater than 6 dB in realistic environments with current hearing aid processing outwith wireless transmission of the signal (e.g., Whitmer et al. 2011), the evidence here indicates that there is no demonstrable “wow” in speech-intelligibility benefits. Rather, assessing other potential avenues of benefit—such as long-term improvements, attentional/cognitive ease, and psycho-social engagement—could show the advantages of hearing aids despite the benefits they give in SNR being less than the JND or JMD.

Acknowledgments We would like to thank Dr. Michael Stone for supplying some of the stimuli used. 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.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Demorest ME (1984). Techniques for measuring hearing aid benefit through self-report. In: Pickett J (ed) Symposium on hearing aid technology: its present and future. Gallaudet, Washington, DC
- Jaeschke R, Singer J, Guyatt GH (1989) Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 10:407–415
- Killion MC (2004) Myths about hearing in noise and directional microphones. *Hear Rev* 11(14–19):72–73
- MacPherson A, Akeroyd MA (2014) Variations in the slope of the psychometric functions for speech intelligibility: a systematic survey. *Trends Hear* 18:1–26
- McCormack A, Fortnum H (2013) Why do people with hearing aids not wear them? *Int J Audiol* 52:360–368
- McShefferty D, Whitmer WM, Akeroyd MA (2015) The just-noticeable difference in speech-to-noise ratio. *Trends Hear* 19:1–9. doi:10.1177/2331216515572316
- Smith MW, Faulkner A (2006) Perceptual adaptation by normally hearing listeners to a simulated ‘hole’ in hearing. *J Acoust Soc Am* 120:4019–4030
- Thwing EJ (1956) Effect of repetition on articulation scores for PB words. *J Acoust Soc Am* 28:302–303
- Whitmer WM, Akeroyd MA (2011) Level discrimination of speech sounds by hearing-impaired individuals with and without hearing amplification. *Ear Hear* 32:391–398
- Whitmer WM, Brennan-Jones CG, Akeroyd MA (2011) The speech intelligibility benefit of a unilateral wireless system for hearing-impaired adults. *Int J Audiol* 50:905–911