J Am Acad Audiol 18:718–719 (2007)

Letters to the Editor

To the Editor:

In reply to Wilson R, Carnell CS, Cleghorn AL. (2007) The Words-in-Noise (WIN) Test with multitalker babble and speech-spectrum noise maskers. *J Am Acad Audiol* 18:522–529.

There are several issues in this article that should be addressed before one can come to a conclusion that the "MTB was an easier listening condition" or "the MTB has more face validity in that listeners with hearing loss complain of difficulty understanding speech in noisy backgrounds" (p. 528).

We would take issue with some of the methods and conclusions of this paper. To wit:

1. It would seem intuitively obvious to most observers that the spectra of the competing noises, shown in figure 1 of the article, are inherently different. From this observation, it should follow that using an averaging method to measure the equivalency of these masking signals might prove difficult. Certainly, comparing their rms values may not be the best method to equate them for comparison. This difference in calibration may have been the reason for the 2 dB difference in the effective masking between the two maskers (p. 526). In fact, experience with equalization of test materials has shown that equating rms levels for individual words leads to widely differing loudness and intelligibility perceptions.

2. If "release from masking" contributes to the scores obtained with MTB, does this not imply that, if the ear is released from masking, the task no longer is masking dependent? That is, the MTB masker confounds the SNR measurement as it, according to the "release" argument, ceases to mask the signal. While the MTB may be more "real world" in concept, a clinical test would seem to require, first of all, exact and complete control of variables. This "release" argument may also explain the differences in normal ranges between the two maskers as noted in the conclusion section on page 528.

3. The correlation between pure-tone hearing loss and test scores as an indication of test sensitivity is not a useful comparison. If a test provides the same information as the pure tone, then why develop the test? This is not proof that speech intelligibility in noise is more accurately being measured with the new materials. Rather, it may be just a demonstration of the poor correlation of the task to pure-tone audibility.

The authors comment on the slope of the functions produced by the two maskers in the two test populations—normal and hearing impaired. We would argue that the slope of the "abnormal" populations likely represents nothing more than the average of the various scores from the hearing loss subjects. If one notes the variance of scores from this group, one is impressed by the large differences in range. Likely this reflects something other than pure-tone thresholds that were used to define this hearing loss group. If the SNR tasks were sampling something other than the influence of pure-tone loss (presumably outer hair cell damage or loss), then one would argue that a sample of subjects chosen by pure-tone loss would show a wide range of responses. Indeed, past investigations have shown that there is a limited relationship between pure-tone loss and SNR score (Nilsson et al, 1994; Killion et al, 2004).

Face validity is a valuable concept in research design and clinical application. However, there are likely many situations where face validity is not a good defining criterion for a clinical test. Pure-tone audiograms come to mind.

In short, the conclusions drawn from the data in this study are, in large part, difficult to defend. Certainly the reader should not conclude that a multitalker babble is more precise in measuring what has come to be known as "SNR hearing loss." In fact, if one were interested in precise measures of this task, the data would lead to the use of a predictable, controllable, and constant masker such as the speech spectrum noise.

REFERENCES

Nilsson M, Soli SD, Sullivan JA. (1994) Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am* 92:1086–1099.

Killion MC, Niquette PA, Gudmundsen G, Revit RJ, Banerjee S. (2004) Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 116:2395–2405.

Michael J. Metz
Audiology Associates, P.C., Irvine, CA

Michael J. Nilsson
Sonic Innovations, Inc., Salt Lake City, UT

Dr. Wilson Responds:

Drs. Metz and Nilsson bring up several points about our manuscript that are timely and deserve mention. The results of our study clearly indicate that recognition performances were better (i.e., at lower signal-to-noise ratios, SNRs) when listening in multitalker babble (MTB) than when listening in speech-spectrum noise (SSN). The advantage was about 2 dB for the listeners with normal hearing and reduced to 0.6 dB for the listeners with hearing loss. This relation was evident in figure 5 in the manuscript.

We agree that calibration is always an issue when dealing with speech stimuli, either as signals or as maskers. In fact, any signal with any degree of amplitude modulation presents calibration issues in that the amplitude is constantly changing as a function of time. We used effective root-mean-square (rms) sound pressure because it provided the only method we had available that both quantified the energy in the signal and provided a method that could be replicated by other investigators. Again, we think everyone would agree that a better way to characterize the energy in an amplitude modulated signal is needed but, to our knowledge, lacking. True, the difference in calibration

between the two maskers could account for the 2 dB difference observed between the two maskers for the listeners with normal hearing. What about the 0.6 dB difference that was observed between the two maskers with the listeners with hearing loss? The major point of the study was not necessarily a focus on the difference between maskers but, rather, a focus on the between group differences between the maskers. The listeners with hearing loss exhibited a different degree of recognition performance on the two maskers compared to the listeners with normal hearing.

We are keenly aware that equating rms of speech signals does not provide equal intelligibility. Likewise, equating intelligibility of a set of words or sentences psychometrically on one population (e.g., listeners with normal hearing) does not ensure equal intelligibility of the same set of materials on another population (e.g., listeners with hearing loss).

"Release from masking" simply means that for one reason or another, the effective masking achieved with one masker is not as great as the effective masking achieved with a second or reference masker. In our study, the MTB and SSN were the respective maskers. Masking dependency is a matter of degree. We agree that a clinical test should require "exact and complete control of variables." That concept can be obtained with a pure-tone signal. Speech signals, however, by their very nature defy that requirement.

Again, we agree that "the correlation between pure-tone (threshold) hearing loss and test scores as an indicator of test sensitivity is not necessarily a useful comparison." Rightfully or wrongfully, the pure-tone audiogram has become the default proverbial "gold standard" for hearing loss. Likewise the pure-tone average at 500, 1000, and 2000 Hz has become in most circles the accepted descriptor of hearing loss, which completely ignores the higher frequencies. Numerous papers over the past 30 to 40 years amply demonstrate that "hearing loss" is composed of many components, each of which is reflected in a different psychometric or electrophysiologic measure. Pure-tone thresholds, which are elicited by the simplest auditory signals, are but one aspect of the functional status of an auditory system; these are the "acuity" and "attenuation" component of hearing loss put forth by Carhart (1951) and Plomp (1978). Measures of speech recognition in quiet assess another aspect of auditory function using a signal that is more complex than pure tones. Measures of speech-in-noise further elevate the complexity of the listening task and examine yet other aspects of auditory function. These more complex measures assess aspects of the "clarity" and "distortion" component of hearing loss described by Carhart and Plomp, respectively. We believe that these three measures of auditory function should form the nucleus of an audiological evaluation.

As described by Wilson and Margolis (1983), the slopes of psychometric functions are dependent on how the slopes are calculated, and the variability of the test items, and the variability of the performances by the listeners. The smaller the variability of the data, the steeper the slope of a function. In our study, we too were "impressed by the large differences in range." In studying listeners with hearing loss we have come to expect substantial variability that we believe reflects a multitude (and probable combination) of auditory abnormalities that range from metabolic changes in the cochlea to various types of neural degeneration at various locations throughout the auditory system. We hope that everyone understands that "there is a limited relationship between pure-tone loss and SNR score." Efforts to predict "SNR hearing loss" (typically viewed as the 50% correct point of a speech recognition task in background noise; Killion, 2002) from pure-tone thresholds have not met with success. To paraphrase Killion, if you want to know how a listener understands speech in background noise, then you must measure that performance.

We certainly did not mean to indicate that multitalker babble is more precise in measuring "SNR hearing loss." Although different types of materials provide varying degrees of information regarding recognition performance in background noise (Wilson et al, 2007), for the moment, we are not sure what alternatives there are to speech-recognition measures in noise (words or sentences) as a measure of SNR hearing loss. Regarding the type of noise, our data importantly indicate that with listeners with hearing loss the type of noise makes little or no difference, at least between MTB and SSN. Further, the subtle differences in masking effectiveness reflected by different maskers in listeners with normal hearing are not realized by many listeners with hearing loss. That is, the auditory systems of many listeners with hearing loss are not differentially affected by different maskers. Although face validity may not always be a “good defining criterion for a clinical test,” the majority of the older adult patients report difficulty understanding speech in a multitalker environment. To date, I have not encountered a patient complaining of difficulty understand speech in SSN. Given that MTB and SSN provide similar measures of SNR hearing loss for listeners with hearing loss, it seems appropriate from the perspective of the patient to utilize MTB for clinical testing.

REFERENCES

Carhart R. (1951) Basic principles of speech audiometry. *Acta Otolaryngol* 40:62–71.

Killion MC. (2002) New thinking on hearing in noise: a generalized Articulation Index. *Semin Hear* 23:57--75.

Plomp R. (1978) Auditory handicap of hearing impairment and the limited benefit of hearing aids. *J Acoust Soc Am* 63:533–549.

Wilson RH, Margolis RH. (1983) Measurement of the auditory thresholds for speech stimuli. In: Konkle D, Rintelmann W, eds. *Principles of Speech Audiometry*. Baltimore: University Park Press, 79–126.

Wilson RH, McArdle R, Smith SL. (2007) An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *J Speech Lang Hear Res* 50:844–856.

Richard H. Wilson
VAMC, Mountain Home, TN