

- Nelson, P. B., and Jin, S. H. (2002). "Understanding speech in single-talker interference: Normal-hearing listeners and cochlear implant users," *J. Acoust. Soc. Am.* **111**, 2429.
- Nelson, P. B., and Jin, S. H. (2004). "Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **115**, 2286-2294.
- Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (2003). "Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **113**, 961-968.
- Peters, R. W., and Moore, B. C. J. (1992). "Auditory filters and aging: filters when auditory thresholds are normal," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 179-185.
- Peters, R. W., Moore, B. C. J., and Baer, T. (1998). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," *J. Acoust. Soc. Am.* **103**, 577-587.
- Pichora-Fuller, M. K., and Schneider, B. A. (1991). "Masking-level differences in the elderly: a comparison of antiphase and time-delay dichotic conditions," *J. Speech Hear. Res.* **34**, 1410-1422.
- Pichora-Fuller, M. K., and Schneider, B. A. (1992). "The effect of interaural delay of the masker on masking-level differences in young and old subjects," *J. Acoust. Soc. Am.* **91**, 2129-2135.
- Schooneveldt, G. P., and Moore, B. C. J. (1987). "Comodulation masking release (CMR): effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," *J. Acoust. Soc. Am.* **82**, 1944-1956.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303-304.
- Stone, M. A., Füllgrabe, C., and Moore, B. C. J. (2010). "Relative contribution to speech intelligibility of different envelope modulation rates within the speech dynamic range," *J. Acoust. Soc. Am.* **128**, 2127-2137.
- Stone, M. A., Moore, B. C. J., and Füllgrabe, C. (2011). "The dynamic range of useful temporal fine structure cues for speech in the presence of a competing talker," *J. Acoust. Soc. Am.* (in press).
- Strelcyk, O., and Dau, T. (2009). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing," *J. Acoust. Soc. Am.* **125**, 3328-3345.
- Studebaker, G. A., and Sherbecoe, R. L. (2002). "Intensity-importance functions for bandlimited monosyllabic words," *J. Acoust. Soc. Am.* **111**, 1422-1436.
- Wang, D., Kjems, U., Pedersen, M. S., Boldt, J. B., and Lunner, T. (2009). "Speech intelligibility in background noise with ideal binary time-frequency masking," *J. Acoust. Soc. Am.* **125**, 2336-2347.
- Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., Cao, K. (2005). "Speech recognition with amplitude and frequency modulations," *Proc. Natl. Acad. Sci. USA* **102**, 2293-2298.

Controlling signal-to-noise ratio effects in the measurement of speech intelligibility in fluctuating maskers

JOSHUA G. W. BERNSTEIN

Audiology and Speech Center, Walter Reed National Military Medical Center, Bethesda, MD 20889, USA

The measurement of speech intelligibility in noise is often complicated by floor and ceiling effects. Because of this, adaptive methods are often used to determine the signal-to-noise ratio (SNR) required for a fixed performance level. Unfortunately, such methods relinquish control of the test SNR, confounding data interpretation when the effect of interest is SNR-dependent. For example, the intelligibility improvement afforded by glimpsing the target speech during brief dips in the level of a fluctuating masker is highly SNR-dependent. Thus, comparisons of performance in stationary and fluctuating maskers are susceptible to SNR confounds. Various methods of controlling SNR differences in the measurement of speech intelligibility are discussed, including the development and validation of a standardized intelligibility testing procedure that uses a variable response set size to control SNR differences. The application of these techniques to studies of hearing loss or simulated hearing loss demonstrate that impaired listeners may retain the ability to listen in the dips of a fluctuating masker to a much greater extent than previously thought.

INTRODUCTION

Normal-hearing (NH) listeners typically demonstrate better speech recognition when the target is presented in a fluctuating background (e.g., competing speech or modulated noise) than when it is presented at the same signal-to-noise ratio (SNR) in stationary noise (e.g., Miller and Licklider, 1950). This phenomenon, referred to as the fluctuating-masker benefit (FMB) or masking release, is thought to reflect dip listening – the ability to excise speech information during momentary dips in the masker level. Hearing-impaired (HI) listeners show little or no fluctuating-masker speech-reception advantage (e.g., Festen and Plomp, 1990), suggesting reduced dip-listening ability.

Signal processing methods have been used to simulate individual aspects of hearing loss in attempts to identify aspects of hearing loss that underlie the reduced FMB often observed for HI listeners. In particular, several studies have focused on the possible role of reduced frequency selectivity or an inability to use TFS information (i.e., fast timing information carried by phase locking in the auditory nerve) in limiting the FMB. These studies used spectral smearing algorithms to simulate reduced frequency selectivity (ter Keurs *et al.*, 1993; Baer and Moore, 1994; Gnansia *et al.*, 2009) or vocoding to remove TFS and present only the envelope (Qin and Oxenham, 2003; Gnansia *et al.*, 2009; Hopkins and Moore, 2009). In each case,

FMB was reduced by the stimulus processing, suggesting that reduced frequency selectivity or TFS processing ability could reduce the FMB for HI listeners.

Bernstein and Grant (2009) and Bernstein and Brungart (2011) identified an important confound that questions conclusions regarding the effects of hearing loss or stimulus processing on dip listening. They pointed out that the magnitude of the FMB depends on the SNR at which the test is performed (Oxenham and Simonson, 2009). At very low SNRs (i.e., negative dB values), the FMB for NH listeners is large. The FMB then decreases with increasing SNR, becoming negative (i.e. a fluctuating-masker deficit rather than a benefit) for SNRs greater than about 0 dB. Most studies of FMB for HI listeners (or for NH listeners presented with processed stimuli) have used adaptive tests to estimate the SNR needed to achieve a given level of performance, usually 50% correct (SNR_{50}). Adaptive measures allow SNR to vary according to listener performance, leaving the results susceptible an SNR confound. Because HI listeners (or NH listeners presented with processed stimuli) have a general speech-processing deficit, their SNR_{50} will be higher than for NH listeners presented with unprocessed stimuli. Reduced FMB would be expected at these higher SNRs, even for NH listeners who, by definition, have normal dip-listening ability. Thus, the reduced FMB measured for HI and simulated HI listeners might not be due to reduced dip-listening ability, but instead might result from the higher SNR_{50} values tracked by the measurement algorithm. Several alternative techniques are proposed here to avoid such SNR confounds in the estimate of FMB.

MODEL: SPEECH INTELLIGIBILITY IN FLUCTUATING MASKERS

Figure 1 presents a working model of speech intelligibility in fluctuating backgrounds based on the Extended Speech Intelligibility Index (ESII; Rhebergen *et al.*, 2006). This framework is presented as a basis to understand how SNR confounds can affect the FMB measurement, and how alternative speech intelligibility measures might avoid such confounds. The model consists of three stages. Stage 1 takes speech and masker signals as inputs, and calculates the audible speech information available to the listener [the ESII; Fig. 1(a-c)], taking into account the masker statistics and the listener’s audiogram and dip-listening ability. This is the only model component where stationary- and fluctuating-masker conditions might be differently affected by hearing loss. For example, if a HI listener had excess forward masking, resulting in reduced ability to listen in the dips, this deficit would manifest as a reduction (relative to the NH listener) in the fluctuating-masker ESII. There would be no concurrent reduction for stationary noise, yielding a reduced FMB. When an experiment aims to measure the effect of impairment on dip listening, the goal is to identify deficits at this stage.

Stage 2 [Fig. 1(d-f)] incorporates suprathreshold distortions, or the psychoacoustic effects of hearing loss that reduce the intelligibility of audible speech information. This has been referred to as a “distortion factor” (D; Plomp, 1986) or “desensitization” (e.g., Ching *et al.*, 2001). Such factors might include reduced frequency selectivity or reduced TFS processing ability. This stage multiplies the ESII (Stage 1) by factor D, yielding an effective ESII. D applies equally to all masking conditions, with any signal that is audible to the listener treated the same

way. Individual differences in D do not impact the FMB in Stage 1. Any distortion that reduces the ability to listen in the gaps (e.g., a TFS processing deficit making it difficult to stream simultaneous talkers) is incorporated in Stage 1, not Stage 2.

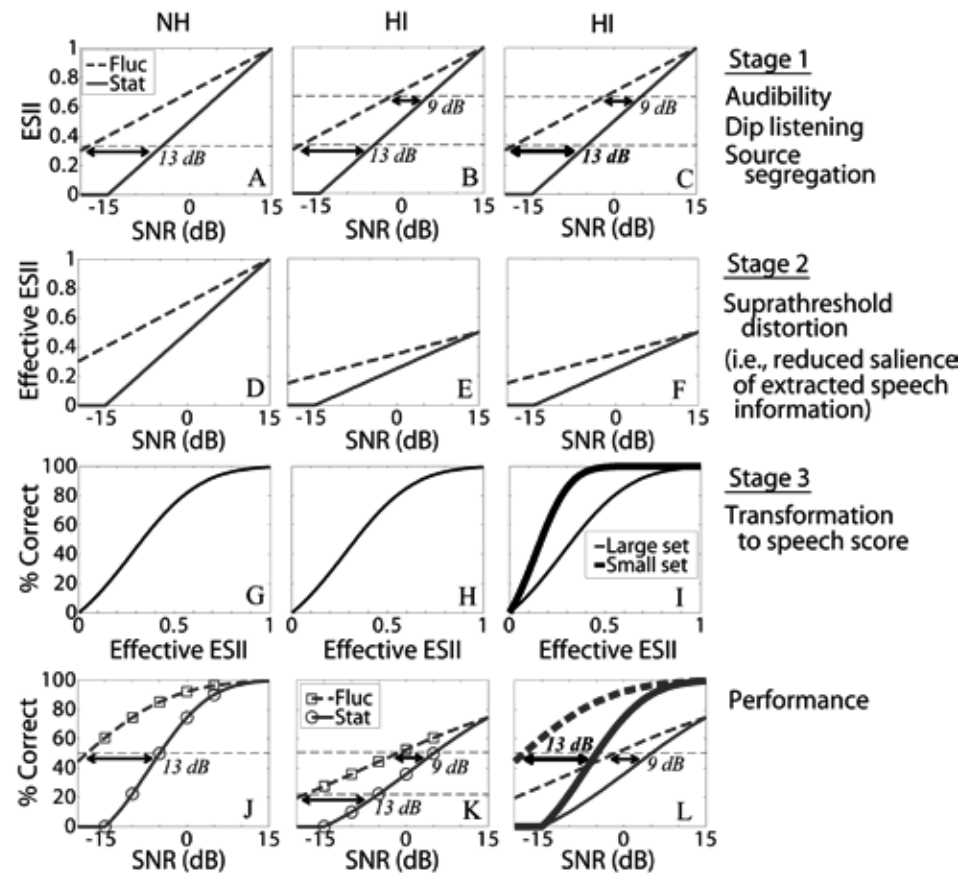


Fig. 1: Working model of speech intelligibility in stationary and fluctuating-maskers. Column 1 traces the stages of the model for a NH listener. Columns 2 and 3 trace the stages of the model for a hypothetical HI listener with normal audibility and dip listening ability (Stage 1) but suprathreshold distortion (Stage 2) affecting speech intelligibility for all maskers.

Stage 3 [Fig. 1(g-i)] is the task-specific transformation from effective ESII to performance (e.g., ANSI, 1969). As the task becomes more difficult with decreasing availability of context, the score for a given effective ESII decreases. The output of Stage 3 is the speech score [Fig. 1(j-l)]. The goal of a dip-listening experiment is to infer effects occurring in Stage 1 based on measurements at the output of Stage 3.

To illustrate the potential effects of an SNR confound on the FMB estimate, consider a hypothetical HI listener with normal audibility and dip-listening ability,

and therefore normal ESII [Fig. 1(b)]. Suppose this listener had a psychoacoustic distortion (e.g., reduced frequency selectivity) that reduced the intelligibility of the audible speech signal (Stage 2). The distortion would be modelled with a lower value of D , say $D=0.5$, where $D=1$ represents a NH listener with no distortion. This would lower the effective ESII [Fig. 1(e)] and in turn the speech score [Fig. 1(k)] for both stationary and fluctuating maskers. The FMB is usually estimated by calculating the difference between stationary- and fluctuating-masker SNR_{50} estimates [Fig. 1(j-k), horizontal dashed lines]. Here, the FMB would be 13 dB for the NH listener [Fig. 1(j), arrow], but only 9 dB for the HI listener [Fig. 1(k), upper arrow], a result that might imply reduced dip-listening ability for the HI listener. However, the FMB difference is actually due to a difference in the stationary-noise SNR_{50} , which is higher for the HI listener due to suprathreshold distortion (Stage 2). Reflecting the SNR_{50} -based FMB estimate back to the underlying ESII, the FMB is calculated at a higher stationary-noise SNR and higher underlying ESII value for the HI listener [Fig. 1(b), vertical dashed line, $\text{SNR} = +5$ dB, $\text{ESII} = 0.67$] than for the NH listener [Fig. 1(a), vertical dashed line, $\text{SNR} = -5$ dB, $\text{ESII} = 0.33$]. Such an experiment would erroneously conclude that the HI listener has a dip-listening deficit. In fact, this listener has normal dip-listening ability (Stage 1), but suprathreshold distortion (Stage 2) affecting all masking conditions. Measuring percentage-correct differences between maskers at a fixed SNR could lead to a similar erroneous conclusion, because changes in D could have different effects at different percentage-correct values.

ALTERNATIVE FMB MEASURES

This discussion highlights the need for an alternative measure of the FMB that is not susceptible to an SNR confound and is therefore less likely to yield erroneous conclusions regarding the source of the speech intelligibility deficits faced by an impaired listener. In the context of the working model presented in Fig. 1, the ideal measure would identify dip-listening deficit that manifests in a reduced ESII (Stage 1) for the fluctuating-masker condition, but would not interpret a global suprathreshold deficit (Stage 2) as a dip-listening problem.

The FMB (in dB) at the same baseline stationary-noise SNR

For this FMB estimate, the FMB is compared between NH and HI listeners starting at the same baseline SNR, but different percentage-correct points on the psychometric function. The idea is that the different percentage-correct points for NH and HI listeners for a particular SNR [Fig. 1(j-k)] represent the same underlying ESII at the output of Stage 1 [Fig. 1(a-b)]. First, an estimate of the percentage-correct at a fixed SNR in the stationary-noise condition is estimated for both listener groups. In this example, an SNR of -5 dB yields 50% correct for the NH listener [Fig. 1(j), vertical dashed line] and 22% correct for the HI listener [Fig. 1(k), vertical dashed line]. HI performance is lower due to suprathreshold distortion (Stage 2). Second, the SNR required in fluctuating-masker condition to yield the same percentage-correct as the stationary-noise condition is estimated for each listener (in this example, an SNR of -18 dB for both listeners). Thus, both example listeners have the same 13-dB FMB when computed using the -5-dB stationary-noise SNR as

a baseline. Reflecting back to the underlying ESII [Fig. 1(a-b)], the FMB is calculated for both listeners at the same SNR and ESII, yielding the same horizontal (dB) distance between the curves. It is therefore concluded that the HI listener does not have a (Stage 1) dip-listening deficit, but rather a general deficit applying equally to all masking conditions. If this method were to identify a reduced FMB for the HI listener, this would suggest a true (Stage 1) dip-listening deficit.

Bernstein and Grant (2009) used this approach to investigate the FMB for HI listeners. They measured psychometric functions for sentences spoken by a female talker in the presence of a speech-shaped stationary noise, an interfering male talker or a speech-shaped noise modulated in two bands by the envelope of the interfering talker signal. To reduce audibility limitations for the HI group, they applied a filter with high-frequency emphasis and presented target signals at a higher overall level for the HI (87 dB) than for the NH listeners (57 dB SPL). Results were generally consistent with previous results. The FMB, calculated as the difference between the stationary-noise and fluctuating-masker SNR_{50} , was larger for the NH listeners [7 to 10 dB; Fig. 2, large filled circles] than for the HI listeners [-2 to -1 dB; large filled squares]. However, the stationary-noise SNR_{50} was higher for the HI (0 dB) than for the NH listeners (-6 dB), which might have contributed to the reduced FMB.

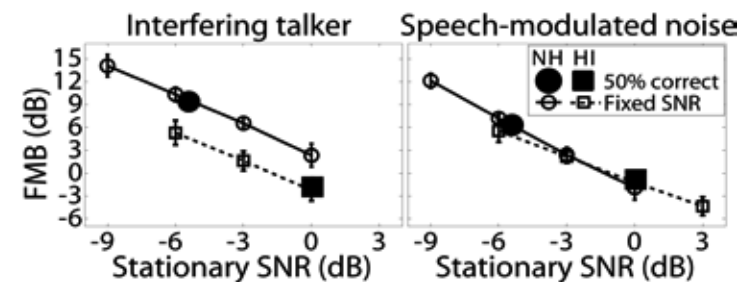


Fig. 2: FMB plotted against the stationary-noise SNR for NH and HI listeners in two fluctuating-masker conditions. Data from Bernstein and Grant (2009).

To avoid this SNR confound, the FMB was estimated for a range of stationary-noise SNRs by calculating horizontal distances between the psychometric functions for the stationary-noise and each fluctuating-masker condition. Fig. 2 plots these FMB estimates as a function of stationary-noise SNR in 3-dB steps (curves and open symbols). For speech-modulated noise, the NH and HI curves overlap, suggesting no dip-listening deficit for the HI listeners. The apparent FMB deficit observed using the traditional measure is likely attributable to an SNR confound. In contrast, the interfering-talker FMB was still reduced for HI listeners using this method, albeit less so [about 5 dB less than the NH FMB; Fig. 2(b), vertical distance between curves] than indicated by the traditional method (a 12-dB FMB difference). This suggests a real (Stage 1) dip-listening deficit for the HI listeners in this condition.

In summary, when steps were taken to avoid SNR confounds, there was less reduction in the FMB for HI listeners than estimated using traditional methods. HI listeners may retain more dip listening ability than previously thought, and for some conditions may benefit from masker fluctuations as much as NH listeners.

Plotting NH against HI performance for the same stimulus condition

The alternative FMB metric described above requires the fitting of a psychometric curve to the data, which can lead to substantial uncertainty in the FMB estimate for regions where the psychometric function is fairly flat. Another method of avoiding SNR confounds in the analysis of fluctuating-masker intelligibility data is plot the raw percentage-correct data for the NH and HI listeners against each one another for each SNR point and masker condition. This technique factors out differences in D between the two listener groups that apply equally to each masker type. If there is a real dip-listening deficit specifically affecting fluctuating masker conditions, the curves describing the relationship between performance for the two listener groups should differ across masker type.

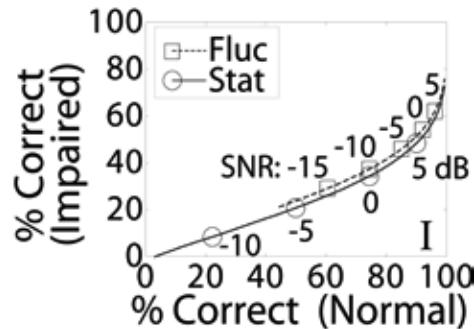


Fig. 3: An illustration of the method of plotting percentage-correct data for NH and HI listeners [Figs. 1(d) and (h)] against one another.

As an example, recall the hypothetical situation of the HI listener with normal dip-listening ability and normal ESII at the output of Stage 1, but suprathreshold distortion reduced effective ESII at the output of Stage 2 [Fig. 1, middle column]. Points along each psychometric function in Fig. 1(j-k) represent performance for SNRs from -15 to +5 dB in 5-dB steps. Fig. 3 plots HI performance [Fig. 1(k)] against NH performance [Fig. 1(j)] for each SNR and masker condition, with one curve for each masker type (dashed = fluctuating; solid = stationary). Points associated with the same position along the x-axis represent the same percentage-correct score across masker conditions, but different SNRs. In this example, the stationary- and fluctuating-masker curves lie on top of one another, reflecting normal dip listening (Stage 1). The suprathreshold distortion (Stage 2) had the same effect on performance for both masking conditions, reducing the percentage-correct score in each case according to the (coincident) curves shown in Fig. 3.

This method is applied to data from Bernstein and Brungart (2011), who investigated the effect of noise vocoding and spectral smearing on the FMB for NH listeners. Noise vocoding extracts envelopes from speech passed through a bank of bandpass filters, then uses these envelopes to modulate noise passed through the same filterbank (Shannon *et al.*, 1995). The processing removes TFS, leaving mainly envelope information. Thus, vocoding can be thought of as a simulation of a TFS-processing deficit. Spectral smearing removes spectral detail, simulating the reduction in frequency selectivity that accompanies hearing loss (ter Keurs *et al.*, 1993; Baer and Moore, 1994). Both manipulations have been shown to reduce the FMB for NH listeners (ter Keurs *et al.*, 1993; Baer and Moore, 1994; Qin and Oxenham, 2003; Gnansia *et al.*, 2009; Hopkins and Moore, 2009), suggesting that deficits in TFS or spectral processing impair dip listening and that such deficits might contribute to the reduced FMB for HI listeners. However, these FMB estimates were subject to the SNR confound described above. Because the signal processing reduced performance overall, the stationary-noise SNR₅₀ was higher for the processed stimuli, likely contributing to the reduced FMB.

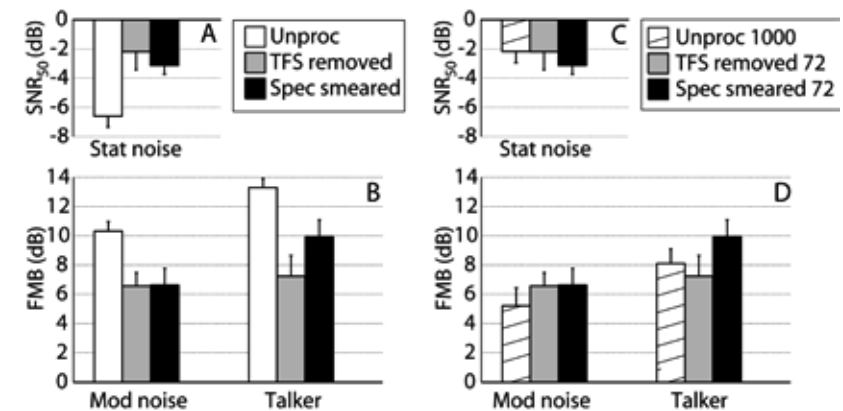


Fig. 4: (a) Mean stationary-noise SNR₅₀ and (b) mean FMB, calculated as the difference between the SNR₅₀ for stationary noise and each fluctuating masker. (c) Mean stationary-noise SNR₅₀, and (d) mean FMB, with set size increased to 1000 for the unprocessed conditions to increase the stationary-noise SNR₅₀ to equal that for the 72-word processed conditions. Data from Bernstein and Brungart (2011).

Bernstein and Brungart (2011) measured psychometric functions for words spoken by a male talker presented in backgrounds of stationary noise, speech-modulated noise or an female interfering talker. Signal and masker were combined, then processed by a 32-channel noise vocoder (Hopkins *et al.*, 2008), spectrally smeared to simulate auditory-filter bandwidths five times normal (Baer and Moore, 1993), or

left unprocessed. In Fig. 4(b), the FMB derived by calculating the difference between the stationary- and fluctuating-masker SNR_{50} was reduced for both processing algorithms (grey and black bars) relative to the unprocessed case (white bars). However, the SNR_{50} was higher for the processed [Fig. 4(a), grey and black bars] than for the unprocessed conditions (white bar), due to distortion introduced by stimulus processing. Fig. 5 plots the percent-correct data for each processed condition against that for the unprocessed condition. The fluctuating-masker and stationary-noise curves overlap, with no interaction between stimulus processing and masker type. This led to the conclusion that processing did not impair dip listening (i.e., Stage 1), but instead affected intelligibility only generally by introducing distortion (i.e., Stage 2). These results suggest that the apparent reduction in FMB with the traditional SNR_{50} -based FMB measurement was due to an SNR confound.

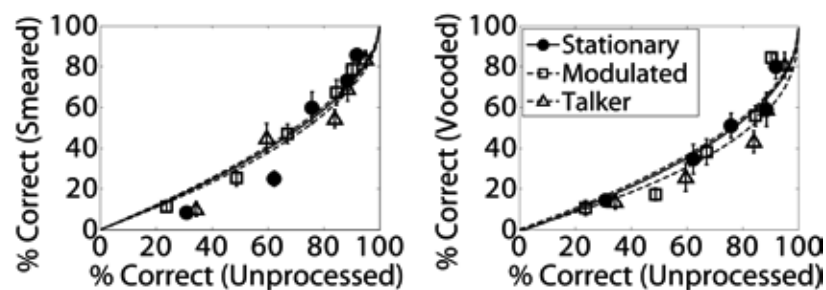


Fig. 5: NH performance for processed stimuli plotted against performance for unprocessed stimuli. Data from Bernstein and Brungart (2011).

Adjusting task difficulty to equalize NH and HI performance

The two methods described above require NH and HI performance to be compared at the same SNRs. Because of the steep performance-intensity functions associated with speech, there are often situations where no SNR exists for which HI listeners perform substantially above chance level and NH listeners performs substantially below ceiling level. A third alternative method of FMB estimation addresses this issue by adjusting the response set size to control task difficulty. The idea is to simultaneously equate both SNR and speech score for NH and HI listeners. Recall the hypothetical HI listener with normal dip listening [Fig. 1(b-c)] but general suprathreshold distortion [Fig. 1(e)] that reduces speech intelligibility [Fig. 1(k-l)]. Decreasing the response set size for this listener will make the task easier, altering the percentage-correct transform [Fig. 1(i)] to yield a higher level of performance for a given effective ESII. If the set size for the HI listener is manipulated to offset the reduced effective ESII caused by the suprathreshold distortion, stationary-noise performance for the HI listener with the smaller response set [Fig. 1(l), thick solid curve] would be equal that for the NH listener with the large set [Fig. 1(j), solid curve]. Because the listeners have the same stationary-noise performance, an SNR_{50} -based FMB estimate can be computed without an SNR confound, yielding a 13-dB

FMB for each listener. This result correctly reflects the normal dip listening ability of the HI listener. Any FMB deficit estimated for a HI listener using this method would represent a true (Stage 1) dip-listening deficit.

Bernstein and Brungart (2011) used this method to estimate the FMB for the stimuli processed to remove TFS information or smear the stimulus spectrum. Recall that the traditional SNR_{50} measurement yielded a reduced FMB for the processed conditions [Fig. 4(b)], but that this result was likely influenced by an SNR confound due to the difference in stationary-noise SNR_{50} for the processed and unprocessed condition [Fig. 4(a)]. These data were collected using the same 72-word response set for each of the processing conditions. The intelligibility measurements were repeated for the unprocessed stimuli using an open-set response paradigm. Stimuli were selected from a 1000-word list, and listeners typed their responses without the benefit of a printed list of possible responses. This increased the stationary-noise SNR_{50} for the unprocessed condition [Fig. 4(c), striped bar] to be equal to that for the 72-word processed conditions [Fig. 4(c), grey and black bars]. When the fluctuating-masker conditions for the unprocessed stimuli were also tested with the open set, the FMB [Fig. 4(d), striped bars] was equal to that for the 72-word processed conditions [Fig. 4(d), grey and black bars]. This analysis led to the same conclusion: that TFS removal and spectral smearing did not impair dip listening.

SET SIZE PROCEDURE: SYSTEMATIC EVALUATION

Of the three methods for controlling SNR in the estimate of FMB, the set-size adjustment is potentially the most versatile. The other two methods require both NH and HI listeners to perform above chance and below ceiling at a common SNR, a situation not always possible with the steep performance functions for speech. Furthermore, these other methods require measurement of the psychometric function. The set-size approach allows for a less time-intensive SNR_{50} measure.

An important assumption underlying the set-size method that it affects only the transformation from effective ESII to percent correct (Stage 3), and not the underlying ESII (Stage 1). This validity of this assumption was tested for NH listeners across a range of maskers and set sizes. Furthermore, Bernstein and Brungart's (2011) set-size procedure had several drawbacks that might limit its use in general practice. First, it required extensive training for each randomly-selected subgroup of words. Second, because listeners were required to learn a large list of words, memorization ability could substantially impact performance. Finally, the word subgroups were selected randomly, possibly resulting in substantial variation in task difficulty depending on the particular subset. Because a given subset was fixed for a period of time for training purposes, this could lead to considerable measurement variability. A revised methodology addressed these drawbacks.

Stimuli consisted of 160 consonant-vowel (CV) or vowel-consonant (VC) tokens, similar to the set described by Vestergaard *et al.* (2009). The set included VC and CV tokens for combinations of five vowels and 16 consonants spoken by seven male talkers from the Linguistic Data Consortium LDC-2005S22 corpus (Fousek *et al.*, 2004). Response alternatives were arranged in a grid with 16 columns (consonants)

and two sets of five rows (vowels; VC and CV context). Training consisted of one trial for each of the 160 tokens presented in quiet.

For a 160-response set, all of the response buttons were available as possible choices (blue background). For smaller sets, some response buttons were marked “inactive” (grey background). Response subsets were selected pseudo-randomly from the 160 tokens, while ensuring the minimum number of active choices with the same vowel or the same consonant as the target. Response sets were re-selected on each trial, and the target token selected from this subset. Four masker conditions were tested [speech-shaped stationary noise, a female interfering talker, and 4- and 32-Hz sinusoidally amplitude-modulated (SAM) noise], each over a range of SNRs.

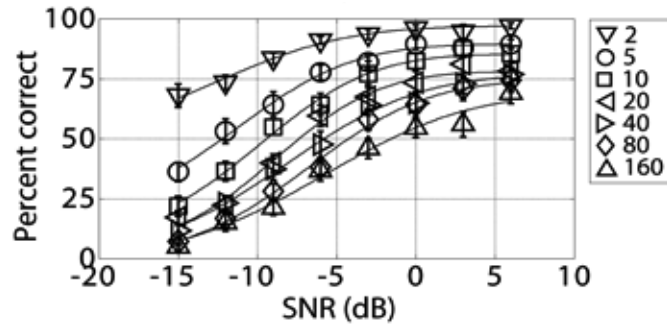


Fig. 6: Psychometric functions for VC/CV identification in stationary noise with parameter set size.

Fig. 6 plots psychometric functions for the stationary noise. Consistent with Miller *et al.* (1951), performance improved with decreasing set size. Estimating the SNR_{50} for each set size gives an idea of the degree to which the set size manipulation can adjust performance. The SNR_{50} ranged from -12 dB (five words) to -2 dB (160 words), indicating that performance can be adjusted by up to 10 dB using this method. For example, if a HI listener has an SNR_{50} 10 dB poorer than a NH listener for a five-word set, then NH and HI performance could be equalized by testing the NH listener with a 160-word set. The results also indicate that the set-size adjustment can increase performance above chance or decrease performance below ceiling. For the -15-dB SNR, performance was near zero for a set size of 160, but increased to 35% (above chance) for a set size of five. For the +6-dB SNR, performance was at ceiling for a set size of two, but below ceiling for large sets.

The question of whether set-size adjustment affects dip listening was tested by plotting percent-correct for two set sizes against one another (the same method used to test the effect hearing loss on dip listening). If set size affects only the percent-correct transformation (Stage 3) and not dip listening, the curves for each masker should overlap. In each panel of Fig. 7, the curves for the four maskers overlap, suggesting that set size can be adjusted between 2 and 160 without affecting FMB.

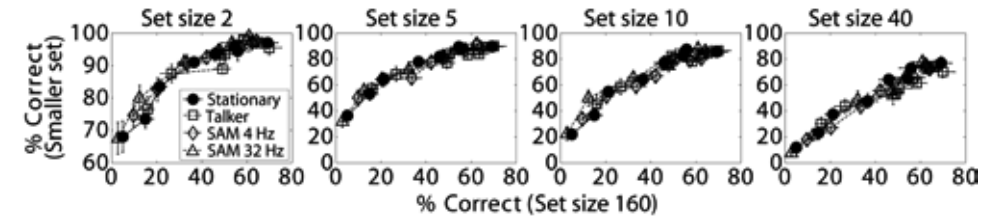


Fig. 7: CV/VC token identification performance for smaller set sizes plotted against performance for a set size of 160. The curves for the four maskers overlap, indicating no interaction between set size and masker type.

CONCLUSIONS

Because the FMB is SNR dependent, traditional FMB measures can lead to erroneous conclusions regarding dip listening. Three methods are proposed to control SNR confounds in the measurement of FMB: (1) estimating the FMB at the same SNR but different percentage-correct levels for two listener groups, (2) plotting speech scores for each group against one other for each masker type, and (3) equalizing stationary-noise performance by testing the two groups with different response set sizes. Using these methods, real and simulated HI listeners were found to retain the ability to dip listen to a greater extent than previously thought, in some cases as well as NH listeners.

ACKNOWLEDGEMENTS

Ken Grant, Douglas Brungart, Van Summers and Nandini Iyer contributed to the studies described here. Martin Vestergaard provided the CV/VC stimulus matrix. Brian Moore and Kathryn Hopkins provided vocoding and spectral-smearing software. The Oticon Foundation and NIH/NIDCD (1-R03-DC010264) provided financial support. The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Department of the Army, Department of the Navy, Department of Defense, nor the U.S. Government.

REFERENCES

- ANSI S3.5 - 1969. *Methods for the calculation of the articulation index*. American National Standards Institute, New York.
- Baer, T., and Moore, B. C. J. (1994). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech" *J. Acoust. Soc. Am.* **95**, 2277-2280.
- Bernstein, J. G. W., and Grant, K. W. (2009). "Auditory and auditory-visual speech intelligibility in fluctuating maskers for normal-hearing and hearing-impaired listeners" *J. Acoust. Soc. Am.* **125**, 3358-3372.

- Bernstein, J. G. W., and Brungart, D. S. (2011). "Effects of spectral smearing and temporal fine-structure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio" *J. Acoust. Soc. Am.* **130**, 473-488.
- Ching, T. Y. C., Dillon, H., Katsch, R., and Byrne, D. (2001). "Maximizing effective audibility in hearing aid fitting" *Ear Hear.* **22**, 212-224.
- Festen, J. M., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing" *J. Acoust. Soc. Am.* **88**, 1725-1736.
- Fousek, P., Svojanovsky, P., Grezl, F., and Hermansky, H. (2004). "New nonsense syllables database — analyses and preliminary ASR experiments" *Proceedings ICSLP 2004, Jeju, South Korea.*
- Gnansia, D., Pean, V., Meyer, B., and Lorenzi, C. (2009). "Effects of spectral smearing and temporal fine structure degradation on speech masking release" *J. Acoust. Soc. Am.* **125**, 4023-4033.
- Hopkins, K., and Moore, B. C. J. (2009). "The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise" *J. Acoust. Soc. Am.* **125**, 442-446.
- Hopkins, K., Moore, B. C. J., and Stone, M. A. (2008). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech" *J. Acoust. Soc. Am.* **123**, 1140-1153.
- Miller, G. A., and Licklider, J. C. R. (1950). "The intelligibility of interrupted speech," *J. Acoust. Soc. Am.* **22**, 167-173.
- Miller, G. A., Heise, G. A., and Lichten, W. (1951). "The intelligibility of speech as a function of the context of the test materials" *J. Exp. Psych.* **41**, 329-335.
- Oxenham, A. J., and Simonson, A. M. (2009). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference" *J. Acoust. Soc. Am.* **125**, 457-468.
- Plomp, R. (1986). "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired" *J. Speech Hear. Res.* **29**, 146-154.
- Qin, M. K., and Oxenham, A. J. (2003). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers" *J. Acoust. Soc. Am.* **114**, 446-454.
- Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (2006). "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise" *J. Acoust. Soc. Am.* **120**, 3988-3997.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues" *Science* **270**, 303-304.
- ter Keurs, M., Festen, J. M., and Plomp, R. (1993). "Effect of spectral envelope smearing on speech reception. II" *J. Acoust. Soc. Am.* **93**, 1547-1552.
- Vestergaard, M. D., Fyson, N. R. C., and Patterson, R. D. (2009). "The interaction of vocal characteristics and audibility in the recognition of concurrent syllables" *J. Acoust. Soc. Am.* **125**, 1114-1124.

Recognition rates and linguistic processing: Do we need new measures of speech perception?

THOMAS BRAND, VERENA USLAR, DOROTHEA WENDT, BIRGER KOLLMEIER

Medizinische Physik, Universität Oldenburg, D-26111 Oldenburg, Germany

Speech perception goes far beyond the recognition of phonemes, words, and sentences. The Oldenburg Linguistically and Audiologically Controlled Sentences (OLACS) were developed for investigating the interactions between the listener's linguistic and auditory capabilities in speech perception. Using these sentences with normal-hearing and hearing-impaired listeners in different listening conditions, a small but significant influence of the sentences linguistic complexity was detected. To some degree this influence was related to other cognitive measures of the listeners. In an eye-tracking experiment delayed eye movements for more complex sentences indicated a higher cognitive load during the speech recognition process. These delayed eye movements were sensitive even in conditions where classical recognition rate and speech reception measures were not sensitive because the recognition rate was near 100%.

INTRODUCTION

When speech perception is assessed in audiology in many cases recognition rates are measured. For this purpose different speech intelligibility tests are available in different languages, such as phoneme or word recognition tests. However, a principal problem of these tests is that they are not very efficient, because for a given number of tested speech items the confidence interval of the recognition rate estimate is relatively wide, due to the binomial distributed data. When the speech reception threshold (SRT) is measured using a test with a steep level-intelligibility curve, such as a sentence test, relatively precise measurements with a standard error of about one dB are possible within only a few minutes per listener. Such SRT tests work very well in many conditions and are a working horse in many applications. However, the SRT is usually located in a level range which is too low for some applications. For instance, some noise reduction algorithms used in hearing aids do only work properly at positive SNRs. The use of threshold values related to recognition rates higher than 50% (for example 80%) can sometimes help here because the signal-to-noise ratio is a few dB higher.

In this article we address the question if there might be further measures besides speech recognition rates and SRTs which might be useful in speech audiometry and which work even in conditions with a recognition rate near 100%. For this purpose we investigated how the linguistic complexity of test sentences influences recognition rates and SRTs as well as eye movements. One of our hypotheses is that an increased linguistic complexity of the used sentence material causes an increased cognitive load of the listener which results in increased SRTs if an interfering noise