



Published in final edited form as:

Ear Hear. 2015 ; 36(2): e14–e22. doi:10.1097/AUD.000000000000113.

Masked sentence recognition assessed at ascending target-to-masker ratios: Modest effects of repeating stimuli

Emily Buss¹, Lauren Calandruccio², and Joseph W. Hall III¹

¹Department of Otolaryngology/Head and Neck Surgery, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

²Division of Speech and Hearing, Department of Allied Health Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Abstract

Objectives—Masked sentence recognition is typically evaluated by presenting a novel stimulus on each trial. As a consequence, experiments calling for replicate estimates in multiple conditions require large corpora of stimuli. The present study evaluated the consequences of repeating sentence-plus-masker pairs at ascending target-to-masker ratios (TMRs). The hypothesis was that performance on each trial would be consistent with the cues available to the listener at the associated TMR, resulting in similar estimates of threshold and slope for procedures using novel vs. repeated sentences within an ascending-TMR block of trials.

Design—A group of 37 normal-hearing young adults participated. Each listener was tested in the presence of one of three maskers: a multi-talker babble, a speech-shaped noise, or an amplitude-modulated speech-shaped noise. There were two data collection procedures, both proceeding in blocks of trials with ascending TMRs. The novel-stimulus procedure used five lists of AzBio sentences, one presented at each of five TMRs, with a novel sentence and masker sample on each trial. The repeated-stimulus procedure used a single list of AzBio sentences, with each sentence presented at multiple TMRs, progressing from low to high; each sentence was paired with a single masker sample, such that only the TMR changed within blocks of repeated stimuli. Listeners completed one run with the novel-stimulus procedure and five runs with the repeated-stimulus procedure. The resulting values of percent correct at each TMR were fitted with a logit function to estimate threshold and psychometric function slope.

Results—The novel-stimulus and repeated-stimulus procedures resulted in generally similar data patterns. After controlling for effects related to the order in which listeners completed the six data collection runs, mean thresholds were slightly higher (<0.5 dB) for the repeated-stimulus procedure than the novel-stimulus procedure in all three maskers. Function slopes for the multi-talker babble and amplitude modulated noise maskers were slightly shallower using the repeated-stimulus than the novel-stimulus procedure, but slopes were comparable for the speech-shaped noise. The quality of psychometric function fits was significantly better for the repeated- than the

Address correspondence to Emily Buss, Department of Otolaryngology/Head and Neck Surgery, University of North Carolina at Chapel Hill, 170 Manning Dr. CB#7070, Chapel Hill, NC 27599, USA. Fax: (919) 966-7656, Phone: (919) 843-9163, ebuss@med.unc.edu.

novel-stimulus procedure, even when comparing a single run of the repeated-stimulus procedure (using one list) to a run of the novel-stimulus procedure (using five lists).

Conclusions—Repeating sentences at ascending TMRs is an efficient method for estimating thresholds and psychometric function slopes, both in terms of the number of sentences and the number of trials.

INTRODUCTION

The study of open-set masked sentence recognition is often limited by the number of unique items available in a particular corpus. Once a listener hears a sentence, it tends to be easier to identify on subsequent trials (Wilson et al. 2003; Yund and Woods 2010), so items are typically presented to the listener only once. Most efficient methods for characterizing speech recognition thresholds require a minimum of 20–30 trials, with more trials required to characterize psychometric function slope (Brand and Kollmeier 2002; Kontsevich and Tyler 1999; see also Plomp and Mimpen 1979). In rigorous experiments, where multiple conditions and replicate estimates are desirable, it can be challenging to identify sentence materials that are appropriate for both the listener group and the experimental question under study, while at the same time providing a sufficiently large number of comparable stimuli. There are a handful of speech corpora with 500 or more sentences, including AzBio (Spahr et al. 2012), IEEE (Rothausser et al. 1969), CUNY (Boothroyd et al. 1985), and BEL (Calandruccio and Smiljani 2012) sentences. These corpora vary on a number of dimensions (e.g., semantic context), and not all are appropriate for testing all listener groups, such as children or non-native speakers of English. Comparing performance across sentence lists within a corpus typically assumes that the items are balanced on a number of dimensions, including speaking rate, articulatory style, word frequency, phonetic content, prosody, syntactic structure, and semantic predictability (Calandruccio and Smiljani 2012; Kalikow et al. 1977; Nilsson et al. 1994). This balance is only approximate, however, and deviations in list equivalence are particularly evident when testing is performed under different conditions and with different types of listeners than used in the construction of the corpus (Bentler 2000; Bilger et al. 1984; Rimikis et al. 2013; Schafer et al. 2012). The present study examines an alternative to using a large speech corpus, namely using a relatively small number of sentences and repeating each target sentence and associated masker sample multiple times, with ascending target-to-masker ratios (TMRs). The rationale for repeated presentations with ascending TMRs is that each improvement in TMR improves the quality and quantity of speech cues available to the listener, which in turn supports better recognition. If performance on each trial reflects the cues available to the listener at the associated TMR, then recognition should be similar whether or not sentences are repeated with ascending TMRs. The purpose of the present study was to evaluate what effect, if any, prior exposure at a lower TMR has on masked sentence recognition for different types of maskers.

There are several experimental and clinical paradigms that call for repeating speech stimuli at ascending TMRs. This method is sometimes used at the beginning of an adaptive threshold estimation track, to determine the initial level (e.g., Nilsson et al. 1994; Plomp and Mimpen 1979). In this context, a sentence is repeated at increasing TMRs until the listener is

able to correctly recognize it, at which point the track begins, using novel stimuli on each subsequent trial. Many early studies of auditory and visual word recognition used the ascending method of limits, wherein a parameter affecting recognition (e.g., presentation level of a sound or duration of a visual presentation) is systematically increased, and threshold is defined as the point at which the word is correctly identified (reviewed by Goldiamond 1958). This general approach has been used to evaluate masked detection thresholds in modern studies of audiovisual integration for speech perception (Bernstein and Grant 2009; MacLeod and Summerfield 1987; Rosenblum et al. 1996), and less frequently in studies of auditory-alone speech perception (Summers and Leek 1998). For example, Summers and Leek (1998) estimated thresholds for masked sentence recognition by systematically increasing the TMR until the listener was able to accurately report three keywords. The final threshold was computed as the mean of the three-keyword-TMRs across 20 sentences. Repeating each stimulus over a block of trials with ascending TMRs is sometimes motivated in terms of the non-homogeneity of speech cues across tokens, particularly with respect to visual cues; when stimuli differ in terms of difficulty, using novel items at each TMR introduces variability and violates the assumptions of adaptive threshold estimation (MacLeod and Summerfield 1987).

While repeating sentence-plus-masker stimuli with ascending TMRs may produce results that are representative of those obtained using novel stimuli on each trial, it is also possible that stimulus exposure at a relatively low TMR could affect subsequent performance at higher TMRs. For example, repeated presentations could facilitate recognition through priming. Freyman et al. (2004) found that listeners were better able to recognize the final word of a masked nonsense sentence if they had just heard a prime consisting of the same sentence, but with the final word replaced by noise. That is, prior information about the initial part of the sentence helped the listener recognize the final word even in the absence of semantic cues. In the context of repeated stimulus presentation in an ascending-TMR block, those portions of a sentence that are recognized at a low TMR could likewise act as primes, helping the listener to more accurately identify the more difficult portions of the sentence. The introduction of semantic cues, absent for the nonsense sentences used by Freyman et al. (2004), could provide additional information to facilitate recognition.

Whereas repeating stimuli with ascending TMRs could improve performance, it is also theoretically possible that repetition could degrade performance. In the visual literature there is some indication that incorrect responses produced by the subject can act as primes, increasing the probability of future related errors in subsequent intervals, even as the task becomes easier (Morton 1964). In the auditory domain, listeners sometimes report clearly hearing words or phrases that are incorrect, a phenomenon referred to as a 'slip of the ear' (Bond 1999). When this occurs, the percept reported by the listener is often a high frequency word from a high density lexical neighborhood (Vitevitch 2002). These errors can be quite stable over time, such as in the context of misheard song lyrics (Beck et al. 2014). When stimuli are repeated with increasing TMRs, error priming or perseverations of 'slips of the ear' could reduce the likelihood of correct identification as the TMR is increased.

Relatively few data directly address the effect of repeating stimuli with ascending TMRs. In a study of the effect of duration on visual word recognition, Adis-Castro and Postman

(1957) compared results obtained with an ascending method of limits, where a single item was presented with increasing duration on sequential trials, and results obtained by random interleaving of stimuli. Performance was very similar with both methods, with slightly better thresholds when items were randomly interleaved. This finding suggests that repeating visual stimuli with increasing durations may slightly degrade visual word recognition. It is unclear whether this finding generalizes to auditory stimuli, particularly complex auditory stimuli like sentences. The audiovisual integration study of Rosenblum et al. (1996) suggests that it may not. That study included a control condition in which sentences were repeated at a fixed TMR of -27 dB. Recognition improved with the repetition of sentences in the control audio-alone condition, suggesting that repeated exposure may itself improve performance, quite apart from increasing TMR (see also Miller et al. 1951). Listeners in the study of Rosenblum et al. (1996) received feedback, indicating when they correctly identified a word, however, so trial-and-error may have played a role in listener improvement. Given the inconclusive nature of these results, one goal of the present study was to evaluate the effect of repeating stimuli at increasing TMRs on estimates of sentence recognition threshold.

The ascending method of limits, as it is applied to speech recognition, is typically used to estimate threshold, but responses across the range of TMRs tested can also be used to fit a psychometric function (e.g. Bernstein and Grant 2009). The additional level of detail obtained using this approach, such as psychometric function slope, can be highly informative. This is particularly of interest in light of the growing emphasis on characterizing performance across a range of TMRs as a means of better understanding the release from masking associated with masker envelope fluctuation in normal-hearing listeners (Bernstein and Grant 2009; Smits and Festen 2013). A second goal of the present study was to evaluate the effects of repeated stimulus presentation with increasing TMRs on estimates of psychometric function slope.

The final question of interest was whether repeated stimulus presentation has a different effect depending on the type of masker being used. Amplitude modulating a masker is thought to improve performance by providing the listener with ‘glimpses’ of the speech, coincident with epochs of improved TMR (Cooke 2006; Rhebergen et al. 2006; but see Stone et al. 2012). Due to the distribution of modulation minima, the quality of speech cues could vary substantially over the course of the stimulus. Repeated presentations of a sentence-plus-masker stimulus in which some cues are heard clearly and others are inaudible could provide the listener with additional opportunities to capitalize on high-level, cognitive cues. For example, a sentence in which approximately half of the phonemes are heard clearly may not result in a correct response initially, but repeated presentations of these glimpses could give the listener time to think about the alternative interpretations and arrive at the correct answer. These considerations motivated the use of speech-shaped noise and amplitude modulated (AM) speech-shaped noise in the present experiment.

In addition to noise maskers, the present study also evaluated performance in a speech-based masker. Noise maskers are traditionally thought to exert predominantly energetic masking, wherein peripheral encoding of the masker swamps out or suppresses response to the target (but see Stone et al. 2012). In contrast, more complex maskers such as background speech

are thought to exert primarily informational masking by reducing the listener's ability to selectively attend to the target in the context of the masker. Presenting a pre-trial prime has been shown to have a larger facilitative effect on masked speech recognition that is limited by informational than energetic masking (Freyman et al. 2004). If repeated stimulus presentations at ascending TMRs serve as primes, then repeating stimuli could have a greater beneficial effect in the context of informational than energetic maskers. Informational masking is generally attributed to stimulus uncertainty and/or similarity between the target and masker (Kidd et al. 2008). Repeating stimuli could reduce stimulus uncertainty, such that the listener can attribute changes in the stimulus following increases in TMR to the added target. Previous data on speech recognition in the presence of random vs. 'frozen' speech maskers indicates a small but significant benefit of repeated presentations (Brungart and Simpson 2004; Felty et al. 2009; Freyman et al. 2007). These observations support the prediction that repeated stimulus presentations may have more of a beneficial effect in the context of maskers dominated by informational masking (e.g., competing speech) than energetic masking (e.g., speech-shaped noise).

METHODS

In the present study masked sentence recognition was assessed with the ascending-TMR method using a fixed masker level, with either a single sentence repeated over a block of trials at increasing presentation levels, or novel sentences presented on each trial. The masker was a sample of multi-talker babble, speech-shaped noise, or speech-shaped noise that had been amplitude modulated at a rate of 10 Hz. All methods were approved by the Institutional Review Board associated with the University of North Carolina, School of Medicine.

Listeners

Listeners were 37 native English speakers, 18 to 60 years of age (mean 27 yrs). All had pure-tone thresholds of 20 dB HL or better at octave frequencies between 125 and 8000 Hz (ANSI 2010). Each listener was assigned to complete testing in one of the three background maskers, either multi-talker babble (n=12), speech-shaped noise (n=12), or AM noise (n=13). Listener age did not differ significantly among the three groups ($p>0.05$). None of these listeners had previously heard the target sentences. Testing was completed in two 1-hr visits, and listeners were paid for their participation.

Stimuli

Target sentences were a subset of 10 lists of the AzBio corpus (Spahr et al. 2012). These lists are approximately equivalent for normal-hearing listeners tested in noise (Schafer et al. 2012), although this does not imply list equivalence in other maskers. Each list in the AzBio corpus includes 20 unique sentences, with five recordings from each of four talkers, two male and two female. For the lists included in the present study, sentences ranged in duration from 1.3 to 3.9 sec, with a mean of 2.4 sec. There were between 3 and 11 words in each sentence (mean of 7.1 words), and between 133 and 154 words in each list. The multi-talker babble masker was based on four talkers (male and female) reading from books. These four speech streams were mixed, with levels adjusted to maintain equal perceptual

salience. Two sections of this mixture were summed to generate a multi-talker babble, with eight streams comprising four voices. The speech-shaped noise masker was generated by adjusting the long-term power spectrum of a Gaussian noise sample to match that of the multi-talker babble. The AM noise was speech-shaped noise that was gated on and off periodically at a 10-Hz rate, with transitions smoothed by 5-ms raised-cosine ramps to limit spectral splatter. This resulted in modulation with a 50% duty-cycle and 100% modulation depth. These three maskers had nearly identical power spectra, which were generally similar to the long-term power spectrum of the target sentences, as illustrated in Figure 1. Each of the three maskers was associated with a ~6-sec wav file.

Prior to the first presentation of each target sentence, a masker sample was randomly selected from the associated wav file, with a duration that was 800-ms longer than that target sentence. The target sentence was temporally centered in a masker sample, resulting in a 400-ms leading and lagging fringe. Masker gating was controlled with 100-ms raised-cosine ramps. Maskers were played at 70 dB SPL, and the target level was adjusted¹. Based on pilot data, five TMRs were selected for each masker, with the goal of sampling performance between 20 and 80% correct. For the multi-talker babble, those values were -6, -4, -2, 0, and 2 dB TMR. For the speech-shaped noise, those values were -12, -10, -8, -6 and -4 dB TMR. For the AM noise, those values were -28, -24, -20, -16, and -12 dB TMR.

Procedures

Experiments were carried out in a double-walled sound booth. A custom MATLAB script was used to construct the stimuli associated with each trial and to record data. Stimuli were played out of a real-time processor (TDT, RX6), passed through a headphone buffer (TDT, HB7), and presented diotically over a circumaural headset (Sennheiser, HD 265). The speech-shaped noise masker was calibrated using a 6-cc artificial ear (Brüel & Kjaer, 4153) and a sound-level meter (Larson Davis, 800B); calibration of other stimuli was inferred based on the RMS level relative to the speech-shaped noise masker. Listeners were instructed to report aloud their best guess at the sentence they heard following each trial. This response was picked up by a microphone mounted on the wall in the booth and routed to a headset worn by a research assistant sitting outside the booth. This research assistant was naïve with respect to the hypothesis of the study. After each listener response the assistant scored each word in the preceding sentence as correct or incorrect.

Listeners provided data in six runs, comprising up to 100 trials each: one run in which a novel sentence was presented on each trial and five runs in which each sentence was repeated up to five times. In both procedures, sequential trials cycled through the set of five TMRs, low-to-high, a total of 20 times. When sentences were novel on each trial, a run comprised sentences from five lists, one list associated with each of the five TMRs. A different randomly selected masker sample was paired with each sentence, with each sample starting at a different randomly selected point in the associated masker array. When sentences were repeated over trials, a run comprised sentences from a single list. Each sentence was paired with a single masker sample, such that the masker sample was ‘frozen’

¹On average, the level of the target sentences dropped by about 5 dB from the beginning (10% of the total duration) to the end (90% of total duration). The long-term average level of each target was used to determine the TMR.

across all TMRs for that sentence. Another difference between novel-stimulus and repeated-stimulus procedures is that once all words in a sentence were repeated back correctly in the repeated-stimulus procedure, further testing with that sentence was terminated. For both novel- and repeated-stimulus conditions, the presentation order of sentences within a list was randomized in software. As a result, the target talker was randomly selected (out of four) on sequential trials in the novel-stimulus procedure, whereas the target talker was fixed across a block of ascending-TMR trials in the repeated-stimulus condition.

The order in which each listener completed the six runs was quasi-random, at the discretion of the research assistant. In this way, each listener provided data for ten lists of 20 sentences each, five for the novel-stimulus procedure and five for the repeated-stimulus procedure, over the course of two 1-hr visits. The particular lists tested with each procedure were counterbalanced across listeners; for a particular masker, half the listeners heard one subset of five lists in the repeated-stimulus procedure, and half heard that same subset of five lists in the novel-stimulus procedure.

Analyses

Data in each condition were fitted with a logit, defined as: $y = 100/(1+\exp(4 \cdot b \cdot (a-x)))$, where y is the percent of correct responses, x is the TMR in dB, a is the threshold parameter of the function (50% correct word identification), and b is the slope parameter of the function, corresponding to the change in percent correct per unit change in TMR at the steepest point in the function (50%). Functions were fitted by minimizing Pearson's Chi-Square. The general approach was to compare recognition performance obtained when stimuli were repeated with those obtained when a novel stimulus was presented in each trial. The primary questions of interest were whether the data collected using the repeated-stimulus procedure converge on the same estimates of threshold and slope as the novel-stimulus procedure for each of the three maskers, and whether these two procedures differ in the number of sentences and/or the number of trials required to obtain reliable estimates of these parameters. A significance level of $\eta = 0.05$ was adopted for all statistical tests. Greenhouse-Geisser corrections were adopted as indicated.

RESULTS

Figure 2 shows the mean data in each of the three masker conditions, collected using either repeated or novel stimuli, presented in blocks of trials with ascending TMRs. These data also appear in the left portion of Table 1. Functions for both procedures are based on 12 or 13 listeners' responses to 100 sentences, with scores reflecting the percent of words identified correctly. In the novel-stimulus procedure, listeners heard each sentence once, so the number of sentences is the same as the number of trials. Recall, however, that in the repeated-stimulus procedure listeners heard each sentence at ascending TMRs until they repeated it back correctly or until the maximum TMR had been reached, whichever occurred first. With the repeated-stimulus procedure, the full set of five TMRs was presented for 39% of sentences in the multi-talker babble masker, 60% of sentences in the speech-shaped noise, and 59% of sentences in the AM noise masker. The smaller number of sentences at the highest TMR in the multi-talker babble is due to the higher percent correct associated with

the set of five TMRs tested in that masker. The mean number of trials in each run of the repeated-stimulus procedure was 70 for the multi-talker babble and 85 for both the speech-shaped noise and AM noise. Given that each listener completed five runs with this procedure², the functions shown in Figure 2 for the repeated-stimulus procedure are based on a factor of 3.5 to 4.3 more trials than the functions for the novel-stimulus procedure.

Several findings are evident based on the mean data shown in Figure 2. In the fits to mean data, there were marked differences in both threshold and slope *between maskers*. Averaged across the two procedures, thresholds were -19.1 dB for the AM noise, -7.4 dB for the speech-shaped noise, and -4.1 dB for the multi-talker babble. The associated slopes were $6.8\%/dB$ for the AM noise, $9.9\%/dB$ for the speech-shaped noise, and $10.7\%/dB$ for the multi-talker babble. Whereas the pattern of results obtained with repeated- and novel-stimulus procedures was broadly similar, there was a trend in the mean data for percent correct to be slightly lower with the repeated-stimulus procedure, particularly at the higher TMRs. This is reflected in slightly higher thresholds and shallower slopes using the repeated-stimulus procedure. For example, thresholds estimated based on data collected using repeated stimuli were elevated by 0.3 to 1.0 dB (multi-talker babble and AM noise maskers, respectively) relative to those based on novel stimuli.

Threshold and slope fitted to individual listener's data

The data from each of six runs for an individual listener were fitted, as were the combined data from the five runs using the repeated-stimulus procedure. These fits accounted for a median of 98.8% of the variance (90.3% to $>99.9\%$). The resulting estimates of threshold and slope for each listener in each condition were used to evaluate the significance of the results observed in the mean data (Fig 2).

A repeated-measures analysis of variance (rmANOVA) was performed on estimates of threshold, with three levels of the between-subjects factor masker type (multi-talker babble, speech-shaped noise, AM noise), and two levels of the within-subjects variable procedure (repeated, novel). There were significant main effects of masker type ($F_{2,34}=250.67$, $p < 0.001$, $\eta_p^2=0.94$) and procedure ($F_{1,34} = 10.76$, $p = 0.002$, $\eta_p^2=0.24$), but no interaction ($F_{2,34}=1.93$, $p = 0.161$, $\eta_p^2=0.10$). This analysis was repeated for estimates of psychometric function slope. In the analysis of slope, there were significant main effects of masker type ($F_{2,34}$, $p < 0.001$, $\eta_p^2=0.68$) and procedure ($F_{1,34} = 6.17$, $p = 0.018$, $\eta_p^2=0.15$), and a significant interaction ($F_{2,34}=4.31$, $p = 0.021$, $\eta_p^2=0.20$). Simple main effects testing revealed a significant difference in the multi-talker babble, with steeper slopes estimated for the novel- than the repeated-stimulus procedure ($11.3\%/dB$ vs. $9.7\%/dB$; $p=0.002$, $\eta_p^2=0.26$). While there was a similar trend in the mean data for the AM noise masker, this difference did not reach significance ($p=0.122$, $\eta_p^2=0.069$). There was no difference in slope for the steady noise ($p=0.481$, $\eta_p^2=0.015$). Recall that listeners heard different sets of sentence lists

²One listener was mistakenly presented with the same sentence list in two sequential repeated-stimulus runs with the speech-shaped noise masker. These replicate data were omitted, leaving only four repeated-stimulus runs for one listener in that masker.

in the repeated- and novel-stimulus procedures, with the list-to-method pairing counterbalanced across listeners. Including sentence set as a between-subjects factor did not change the pattern of significance, and neither main effects nor interactions with set approached significance ($p = 0.194$). Subsequent analyses therefore pooled data across sentence sets. Overall, analyses of threshold and slope based on psychometric function fits to each listener's data indicated a modest effect of repeating stimuli at ascending TMRs.

Figure 3 shows thresholds (left panel) and slopes (right panel) as a function of the order in which each listener completed the six runs, irrespective of whether stimuli were repeated or novel on each trial. Symbol shape indicates the masker type, as defined in the legend, and error bars indicate one standard error of the mean. This figure illustrates a trend for improved thresholds over the course of the experiment, but relatively consistent psychometric function slopes across runs. Comparing thresholds associated with the first and the sixth run, practice improved mean thresholds by 0.9 dB (speech-shaped noise), 2.0 dB (multi-talker babble), and 1.8 dB (AM noise). The effect of order on threshold was evaluated with an rmANOVA, with six levels of order and three of masker type. This analysis resulted in significant main effects of order ($F_{5,165} = 9.74$, $p < 0.001$, $\eta_p^2 = 0.23$) and masker ($F_{2,33} = 232.57$, $p < 0.001$, $\eta_p^2 = 0.93$), and no interaction ($F_{10,165} = 1.46$, $p = 0.160$, $\eta_p^2 = 0.08$). The linear contrast for order was significant ($F_{1,33} = 17.89$, $p < 0.001$, $\eta_p^2 = 0.35$), consistent with a gradual improvement in performance over the course of the experiment, on the order of -0.3 dB per run. No order effects were evident in a second analysis of psychometric function slope, a result that could reflect the relatively larger variance in estimates of slope.

The finding of a significant order effect across runs raises the possibility that the threshold differences between conditions incorporating repeated- and novel- stimulus procedures observed in Figure 2 could be affected by test order. In the quasi-random order of conditions, the novel stimulus procedure tended to be completed in the middle or towards the end of the experiment, so better performance in this condition could be due, in part, to a benefit conferred by opportunities for procedural learning (e.g., Yund and Woods 2010). This possibility was addressed by normalizing the data by run order. This was accomplished by adjusting the TMRs associated with each estimate of percent correct according to the ordinal position of the run and the mean effect of practice, and then fitting psychometric functions to the normalized data. The parameter fits to normalized data appear in the right columns of Table 1. With normalized data, the threshold difference between procedures was less than 0.5 dB in all masker conditions. A set of rmANOVAs was performed to assess the effect of procedure on estimates of psychometric function threshold and slope based on normalized data. In the analysis of threshold, there was a significant main effect of masker type ($F_{2,34} = 273.58$, $p < 0.001$, $\eta_p^2 = 0.94$), a non-significant trend for a main effect of procedure ($F_{1,34} = 3.63$, $p = 0.065$, $\eta_p^2 = 0.10$), and no interaction ($F_{2,34} = 0.73$, $p = 0.488$, $\eta_p^2 = 0.04$). In the analysis of slope, there was a significant main effect of masker type ($F_{2,34} = 42.32$, $p < 0.001$, $\eta_p^2 = 0.71$), a significant main effect of procedure ($F_{1,34} = 7.92$, $p = 0.008$, $\eta_p^2 = 0.19$), and a significant interaction ($F_{2,34} = 3.65$, $p = 0.037$, $\eta_p^2 = 0.18$). Whereas

slopes were nearly identical for the repeated- and novel-stimulus procedures in the speech-shaped noise (9.9%/dB and 9.6%/dB), estimates of slope were slightly shallower for the repeated-stimulus procedure in both the multi-talker babble (9.7%/dB and 11.3%/dB) and the AM noise maskers (5.8%/dB and 6.9%/dB). Simple main effects testing indicated a significant effect of procedure for both the multi-talker babble ($p=0.003$, $\eta_p^2=0.23$) and AM noise ($p=0.033$, $\eta_p^2=0.13$), but not for the speech-shaped noise masker ($p=0.612$, $\eta_p^2=0.01$). Compared to the original dataset, the main effect of procedure on threshold dropped below significance; in the analysis of slope, however, both the main effect of procedure and the interaction between procedure and masker remained significant.

Efficiency of novel- and repeated-stimulus procedures

While quality of the logit fits was relatively good in all conditions, the goodness of fit differed across stimulus conditions. This is illustrated in Figure 4, which shows the distribution of Chi-squared values for fits to the data of individual listeners. Fits to data collected using the novel-stimulus procedure tended to be poorer than those using the repeated-stimulus, compared either to fits to each of the five runs in this condition or to fits to the mean data for an individual listener. This observation was evaluated using a rmANOVA. The dependent measure was the Chi-square values characterizing the quality of the function fit, transformed by applying an exponent of 0.25 to approximate a normal distribution (Hawkins and Wixley 1986). The independent variables were dataset (novel-stimulus, mean of five repeated-stimulus runs, all repeated-stimulus runs pooled), and masker (multi-talker babble, speech-shaped noise, AM noise). This analysis resulted in a main effect of dataset ($F_{2,68} = 50.18$, $p < 0.001$, $\eta_p^2=0.60$), no main effect of masker ($F_{2,34} = 1.69$, $p = 0.200$, $\eta_p^2=0.09$), and no interaction ($F_{4,68} = 2.22$, $p = 0.076$, $\eta_p^2=0.12$). Pairwise comparisons indicated that fits differed significantly for all three datasets ($p \leq 0.002$).

One way to think about the quality of the psychometric function fits is in terms of the efficiency of the data collection procedure. Assuming that the logit is a good approximation of the function relating sensitivity to the TMR, then failure to achieve a good fit of this function can be attributed to noise in the data, due to either listener factors (e.g., variable strategy or attention) or to stimulus factors (e.g., variability among samples). When these sources of noise are evenly distributed across conditions, they tend to average out with larger numbers of trials. This is evident in Figure 4 in the better fits to data pooled across all the repeated-stimulus runs than data in each run fitted separately, despite the fact that sensitivity improved over the course of the experiment. In the case of fits to data from all the repeated-stimulus and the novel-stimulus runs, both are based on 100 sentences (5 lists of 20) from each listener, although the former is based on more trials (350–425) than the later (100). The larger number of trials in repeated-stimulus dataset could contribute to the better function fit. However, fits to data from each repeated-stimulus run fitted separately were also better than fits to data collected using the novel-stimulus procedure. In this comparison, the former is based on fewer trials (70–85 vs 100) and fewer sentences (20 vs 100) than the later.

Recall that testing of a particular sentence in the repeated-stimulus procedure was terminated if all the words were correctly identified before the maximum TMR was reached. The rationale for this approach is that any sentence that is correctly identified at one TMR would most likely continue to be correctly identified at higher TMRs. While the present data do not support a test of that assumption at the level of the sentence, it is possible to evaluate whether words correctly identified at one TMR continued to be correctly identified at higher TMRs. To that end, listener responses were evaluated on a word-by-word basis. Only 2.5% of words were correctly identified at one TMR and then missed at a subsequent (higher) TMR. Of these, the majority occurred when performance overall was poor (<50% correct overall). This provides support for the assumption that correctly identifying all words in a sentence at one TMR precludes the need for additional data collection with that sentence.

DISCUSSION

The purpose of the present study was to compare psychometric function fits for data collected using either novel stimuli on each trial or repeated presentations of each sentence-plus-masker pair at ascending TMRs. Of particular interest, we wanted to know whether similar estimates of threshold and slope were obtained with these two procedures, whether the type of masker affected data collected with the two procedures differently, and which method was more efficient in terms of the number of sentences and the number of trials required to obtain a good psychometric function fit.

Effect of repeating stimuli on estimates of slope and threshold

While estimates of both slope and threshold were similar using the novel-stimulus and repeated-stimulus procedures, there was some indication of consistent differences. Mean thresholds were higher for the repeated- than the novel-stimulus procedure, although this effect was less than 1 dB for all three maskers in the original data, and less than 0.5 dB when data were normalized for run order. Psychometric function slopes tended to be shallower for the repeated- than the novel-stimulus procedure for both the multi-talker babble and the AM noise maskers. There was no indication of shallower slopes with the repeated- than the novel-stimulus procedures in the speech-shaped noise data. Taken together, these results are consistent with the interpretation that repeated stimulus presentations at increasing TMRs modestly increases the probability of errors at high TMRs compared to the novel-stimulus procedure under some conditions.

While speech recognition is typically evaluated with novel stimuli on each trial in clinical evaluation and studies of speech perception, an argument could be made that repeating stimuli at increasing TMRs has some ecological validity. For example, when hearing-impaired listeners are unable to hear a conversation partner they often ask that the sentence be repeated (Tye-Murray 1991; Tye-Murray et al. 1992). In this situation the repeated utterance may not always be identical to the original, but the key content is likely to be similar. While the original focus of the present study was to evaluate the repeating-stimulus procedure as an alternative to presenting novel stimuli on each trial of a speech perception experiment, the slightly higher thresholds and shallower psychometric function slopes

obtained with the repeating-stimulus procedure could also provide insight into speech perception under these natural listening conditions.

One issue of interest was whether repeated stimulus presentations would have a different effect in the presence of different types of maskers, particularly those associated with envelope fluctuation and informational masking. Thresholds in the multi-talker babble were 3.3 dB higher than those in the speech-shaped noise, consistent with modest but significant informational masking. Freyman et al. (2004) evaluated speech maskers composed of different numbers of talkers by measuring the improvement in speech recognition associated with the introduction of binaural cues. By that metric, informational masking at 50% correct was approximately 3 dB for their four-talker masker and approximately 1–2 dB for their six- and ten-talker maskers. Those results are broadly consistent with those of the present study. Thresholds in the AM noise masker were 11.5 dB lower than those in the speech shaped noise, consistent with the introduction of ‘glimpses’ of speech. While the effect of masker AM is very sensitive to the speech materials and measurement procedures (Miller et al. 1951), the results of the present study can be broadly compared to those of Nelson et al. (2003). In that study sentence recognition was assessed at each of three TMRs for square-wave AM maskers and sentences presented at 65 dBA. Logits were fitted to the resulting estimates of percent correct (data in their Figure 1). As in the present data set, thresholds associated with 50% correct were higher for the steady (–3.5 dB TMR) than the AM noise (–20.1 and –16.6 dB TMR for 8- and 16-Hz AM, respectively); the psychometric function was also steeper for the steady noise (10%/dB) than the AM noise (3.5%/dB and 5.3%/dB for 8- and 16-Hz AM, respectively). These comparisons with published data support the idea that the multi-talker babble and AM noise maskers introduced informational masking and the opportunity to glimpse speech, respectively.

After correcting for order effects, repeating stimuli reduced the psychometric function slope for both the multi-talker babble and the AM noise maskers by about 15%. In contrast, the slope associated with the speech-shaped noise masker was not affected by repeating stimuli. Changes in slope could be related to an increased probability of perseverating on an incorrect answer in the context of sparse glimpses of the speech or informational masking, perhaps due to intrusions from the masker stream in the later case. While the present paradigm does not allow a test of this interpretation, future studies could evaluate the effect of repeating stimuli in the context of maskers introducing greater amounts of informational masking.

Efficiency of repeated-stimulus and novel-stimulus procedures

If repeating stimuli at increasing TMRs results in valid data for estimating threshold and psychometric function slope, then this method could reduce the number of target stimuli required to evaluate speech recognition. This would have the advantage of allowing greater flexibility in the selection of speech materials and the ability to complete more conditions with a limited set of materials. For example, the Revised SPIN sentences (Bilger 1984) consist of 200 low-predictability (LP) and 200 high-predictability (HP) sentences. These materials have been widely used in evaluating the effect of speech predictability, but the entire set of HP or LP stimuli may be required to estimate both threshold and psychometric

function slope (e.g., Pichora-Fuller et al. 1995; Wilson et al. 2012), precluding comparisons across listening conditions within listeners. With the procedures of the present study, the repeated-stimulus procedure could support five times the number of conditions possible using the novel-stimulus procedure. This would be particularly beneficial when working with hard-to-recruit populations, in that listeners could participate in multiple conditions or multiple experiments using stimuli from a single corpus.

In addition to the number of target sentences, the time required to obtain reliable data is an important consideration in speech testing. The present results indicate that for a fixed number of trials, repeating stimuli may result in more stable estimates of the psychometric function than presenting novel stimuli. Likewise, data of comparable stability would require fewer trials using the repeated-stimulus than the novel-stimulus procedure. This finding was not predicted. One possibility is that this outcome could be related to variability across stimuli. For example, if different samples of the multi-talker babble differ with respect to their ability to mask the target, then estimating percent correct at different TMRs with different samples could reduce the quality of the subsequent psychometric function fit. For small numbers of trials, presenting the same samples at all TMRs could reduce this source of error. This explanation is undermined somewhat by the lack of an effect of masker type in the quality of psychometric function fits, in that variability across masker samples would be expected to be smaller for speech-shaped noise than the multi-talker babble. Another factor to consider is variability across sentences within a list. Further research is needed to better understand the relatively good fits obtained with small numbers of repeated-stimulus trials.

Possible effects of ascending TMR

One question that arose during the review process was how the novel-stimulus procedure, using ascending TMRs, compares to the method of constant stimuli, wherein the TMR is randomized from one trial to the next. Supplemental data were collected to address this question. Twelve naïve listeners, ages 18 to 40 (mean 23 years), were recruited. All met the inclusion criteria of the main experiment. Data were collected in two conditions, both using speech-shaped noise. One condition was identical to the novel-stimulus condition of the main experiment, with RMR ascending (low to high) within sequential blocks of five trials. The other condition was based on the novel-stimulus condition, but the order of stimuli was randomly interleaved, such that the TMR was no longer predictable from trial to trial. The order of conditions and the subset of five sentences lists used in each condition were counterbalanced among the 12 listeners. Psychometric functions were fitted to each listener's data, as described above. Resulting estimates of threshold and slope were very similar for the two conditions. Across listeners, mean thresholds were -7.16 dB with ascending TMRs and -7.11 dB with randomized TMRs ($t_{11} = -0.22$, $p = 0.828$). The mean slopes in these conditions were $10.9\%/dB$ and $10.1\%/dB$, respectively ($t_{11} = 0.88$, $p = 0.398$). Both thresholds and slopes were comparable across procedures. The thresholds and slopes were also comparable to those observed in the speech-shaped noise conditions of main experiment, consistent with the conclusion that presenting stimuli with ascending TMRs on subsequent trials is functionally similar to randomizing TMRs. These estimates are also consistent with the results of Sinex (2013). In that study word-level percent correct was measured for AzBio sentences in speech-shaped noise, with blocks of trials utilizing a

single list and a fixed TMR. Based on mean data for five normal-hearing listeners, the threshold associated with 50% correct was -5.7 dB, and the slope was approximately 10%/dB.

Additional considerations

One way in which repeating stimuli could influence performance is by providing the listener with multiple opportunities to capitalize on sentence context. Semantic and syntactic context can improve performance (e.g., Kalikow et al. 1977; Miller et al. 1951), but the degree of context differs across corpora. It is therefore relevant to consider the quality of contextual information associated with the AzBio sentences. Contextual information can be characterized with the parameter j (Boothroyd and Nittrouer 1988), the exponent expressing the relationship between the probability of recognizing all the words in a sentence as a function of the probability of recognizing each word individually ($p_{\text{sent}} = p_{\text{word}}^j$). Based on data collected in the speech-shaped noise masker, values of j range from 1.5 (-12 dB TMR) to 2.3 (4 dB TMR)³. These values can be compared to estimates of what j would have been in the absence of context, which range from 5.0 (-12 dB TMR) to 6.5 (-4 dB TMR). These results indicate that the AzBio sentences provide the listener with relatively high quality context cues. If high context increases the likelihood of obtaining an effect of repeating stimuli, then the results observed here are likely to generalize to other corpora providing comparable or less context.

A second consideration relates to the content of the masker. As mentioned above, there is some indication that the repeated-stimulus procedure could have a slightly different affect on performance when the masker is dominated by informational masking as compared to energetic masking. The repeated-stimulus procedure depends critically on a monotonic effect of ascending TMR, and this is not always the case when informational maskers are used (Brungart 2001). If the psychometric function is non-monotonic, data with the ascending method would likely obscure reductions in percent correct with increasing TMR, due to prior stimulus exposure.

CONCLUSIONS

The data reported here support the following conclusions.

1. When masked sentence recognition is evaluated in blocks of trials with ascending TMRs, results are generally similar regardless of whether novel stimuli are presented on each trial or stimuli are repeated over a block of trials. Instead of facilitating performance, there is a significant (though small) trend for performance to be poorer at relatively high TMRs for the repeated-stimulus than the novel-stimulus procedure in some maskers.
2. While results were broadly similar for a multi-talker babble, speech-shaped noise, and AM noise, there was some indication that the difference between procedures

³In corpora for which all sentences contain equal numbers of words or keywords, the value of j associated with no benefit of context is equal to the number of keywords in each sentence. The number of keywords differs across AzBio sentences, however. Estimates of j in the absence of context were computed by estimating p_{sent} as the mean of p_{word}^n across all sentences, where n is the number of words in each of target sentence.

(repeated vs. novel stimuli on each trial) could have a larger effect in the context of a complex masker than a speech-shaped noise masker.

3. Repeating stimuli at increasing TMRs is an efficient method of characterizing psychometric function slope and threshold, both in terms of the number of unique sentences presented and the number of trials.

ACKNOWLEDGEMENTS

This work was supported by a grant from NIH NIDCD R01 DC007391. Huanping Dai, Richard Wilson and two anonymous reviewers provided helpful comments on an early draft.

Conflicts of Interest and Source of Funding: This work was funded by NIH NIDCD R01 DC007391.

References

- Adis-Castro G, Postman L. Psychophysical methods in the study of word recognition. *Science*. 1957; 125:193–194. [PubMed: 13390983]
- ANSI. ANSI S3.6-2010, American National Standard Specification for Audiometers. New York: American National Standards Institute; 2010.
- Beck C, Kardatzki B, Ethofer T. Mondegreens and soramimi as a method to induce misperceptions of speech content - influence of familiarity, wittiness, and language competence. *PLoS One*. 2014; 9:e84667. [PubMed: 24416261]
- Bentler RA. List equivalency and test-retest reliability of the Speech in Noise test. *Am J Audiol*. 2000; 9:84–100. [PubMed: 11200196]
- Bernstein JG, Grant KW. Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*. 2009; 125:3358–3372. [PubMed: 19425676]
- Bilger, RC. Speech recognition test development. In: Elkins, E., editor. *Speech Recognition by the Hearing Impaired*. ASHA Reports, 14. Rockville, MD: American Speech-Language-Hearing Association; 1984.
- Bilger RC, Nuetzel JM, Rabinowitz WM, et al. Standardization of a test of speech perception in noise. *J Speech Hear Res*. 1984; 27:32–48. [PubMed: 6717005]
- Bond, ZS. *Slips of the ear: Error in the perception of casual conversation*. New York: Academic Press; 1999.
- Boothroyd, A.; Hanin, L.; Hnath, T. Internal Report RC1. New York: City of University of New York; 1985. A sentence test of speech perception: Reliability, set equivalence, and short-term learning.
- Boothroyd A, Nittrouer S. Mathematical treatment of context effects in phoneme and word recognition. *J Acoust Soc Am*. 1988; 84:101–114. [PubMed: 3411038]
- Brand T, Kollmeier B. Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J Acoust Soc Am*. 2002; 111:2801–2810. [PubMed: 12083215]
- Brungart DS. Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am*. 2001; 109:1101–1109. [PubMed: 11303924]
- Brungart DS, Simpson BD. Within-ear and across-ear interference in a dichotic cocktail party listening task: effects of masker uncertainty. *J Acoust Soc Am*. 2004; 115:301–310. [PubMed: 14759023]
- Calandruccio L, Smiljani R. New sentence recognition materials developed using a basic non-native English lexicon. *J Speech Lang Hear Res*. 2012; 55:1342–1355. [PubMed: 22411279]
- Cooke M. A glimpsing model of speech perception in noise. *J Acoust Soc Am*. 2006; 119:1562–1573. [PubMed: 16583901]
- Felty RA, Buchwald A, Pisoni DB. Adaptation to frozen babble in spoken word recognition. *J Acoust Soc Am*. 2009; 125:EL93–EL97. [PubMed: 19275281]

- Freyman RL, Balakrishnan U, Helfer KS. Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J Acoust Soc Am*. 2004; 115:2246–2256. [PubMed: 15139635]
- Freyman RL, Helfer KS, Balakrishnan U. Variability and uncertainty in masking by competing speech. *J Acoust Soc Am*. 2007; 121:1040–1046. [PubMed: 17348526]
- Goldiamond I. Indicators of perception. I. Subliminal perception, subception, unconscious perception: an analysis in terms of psychophysical indicator methodology. *Psychol Bull*. 1958; 55:373–411. [PubMed: 13602019]
- Hawkins DM, Wixley RAJ. A Note on the Transformation of Chi-Squared Variables to Normality. *American Statistician*. 1986; 40:296–298.
- Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am*. 1977; 61:1337–1351. [PubMed: 881487]
- Kidd, G.; Mason, CR.; Richards, VM., et al. Auditory Perception of Sound Sources. New York: Springer; 2008. Informational Masking; p. 143-190.
- Kontsevich LL, Tyler CW. Bayesian adaptive estimation of psychometric slope and threshold. *Vision Res*. 1999; 39:2729–2737. [PubMed: 10492833]
- MacLeod A, Summerfield Q. Quantifying the contribution of vision to speech perception in noise. *Br J Audiol*. 1987; 21:131–141. [PubMed: 3594015]
- Miller GA, Heise GA, Lichten W. The intelligibility of speech as a function of the context of the test materials. *J Exp Psychol*. 1951; 41:329–335. [PubMed: 14861384]
- Morton J. The effects of context on the visual duration threshold for words. *Br J Psychol*. 1964; 55:165–180. [PubMed: 14168481]
- Nelson PB, Jin SH, Carney AE, et al. Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. *J Acoust Soc Am*. 2003; 113:961–968. [PubMed: 12597189]
- Nilsson M, Soli SD, Sullivan JA. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am*. 1994; 95:1085–1099. [PubMed: 8132902]
- Pichora-Fuller MK, Schneider BA, Daneman M. How young and old adults listen to and remember speech in noise. *J Acoust Soc Am*. 1995; 97:593–608. [PubMed: 7860836]
- Plomp R, Mimpen AM. Improving the reliability of testing the speech reception threshold for sentences. *Audiology*. 1979; 18:43–52. [PubMed: 760724]
- Rhebergen KS, Versfeld NJ, Dreschler WA. Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise. *J Acoust Soc Am*. 2006; 120:3988–3997. [PubMed: 17225425]
- Rimikis S, Smiljanic R, Calandruccio L. Normative data for the Basic English Lexicon sentences for non-native speakers of English. *J Speech Lang Hear Res*. 2013; 56:792–804. [PubMed: 23275396]
- Rosenblum LD, Johnson JA, Saldana HM. Point-light facial displays enhance comprehension of speech in noise. *J Speech Hear Res*. 1996; 39:1159–1170. [PubMed: 8959601]
- Rothauer EH, Chapman WD, Guttman N, et al. I.E.E.E. recommended practice for speech quality measurements. *IEEE Trans Aud Electroacoust*. 1969; 17:227–246.
- Schafer EC, Pogue J, Milrany T. List equivalency of the AzBio sentence test in noise for listeners with normal-hearing sensitivity or cochlear implants. *J Am Acad Audiol*. 2012; 23:501–509. [PubMed: 22992257]
- Sinex DG. Recognition of speech in noise after application of time-frequency masks: dependence on frequency and threshold parameters. *J Acoust Soc Am*. 2013; 133:2390–2396. [PubMed: 23556604]
- Smits C, Festen JM. The interpretation of speech reception threshold data in normal-hearing and hearing-impaired listeners: II. Fluctuating noise. *J Acoust Soc Am*. 2013; 133:3004–3015. [PubMed: 23654404]
- Spahr AJ, Dorman MF, Litvak LM, et al. Development and validation of the AzBio sentence lists. *Ear Hear*. 2012; 33:112–117. [PubMed: 21829134]

- Stone MA, Fullgrabe C, Moore BC. Notionally steady background noise acts primarily as a modulation masker of speech. *J Acoust Soc Am.* 2012; 132:317–326. [PubMed: 22779480]
- Studebaker GA. A "rationalized" arcsine transform. *J Speech Hear Res.* 1985; 28:455–462. [PubMed: 4046587]
- Summers V, Leek MR. Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners. *Hear Res.* 1998; 118:139–150. [PubMed: 9606069]
- Tye-Murray N. Repair strategy usage by hearing-impaired adults and changes following communication therapy. *J Speech Hear Res.* 1991; 34:921–928. [PubMed: 1956199]
- Tye-Murray N, Purdy SC, Woodworth GG. Reported use of communication strategies by SHHH members: client, talker, and situational variables. *J Speech Hear Res.* 1992; 35:708–717. [PubMed: 1608264]
- Vitevitch MS. Naturalistic and experimental analyses of word frequency and neighborhood density effects in slips of the ear. *Lang Speech.* 2002; 45:407–434. [PubMed: 12866911]
- Wilson RH, Bell TS, Koslowski JA. Learning effects associated with repeated word-recognition measures using sentence materials. *J Rehabil Res Dev.* 2003; 40:329–336. [PubMed: 15074444]
- Wilson RH, McArdle R, Watts KL, et al. The Revised Speech Perception in Noise Test (R-SPIN) in a multiple signal-to-noise ratio paradigm. *J Am Acad Audiol.* 2012; 23:590–605. [PubMed: 22967734]
- Yund EW, Woods DL. Content and procedural learning in repeated sentence tests of speech perception. *Ear Hear.* 2010; 31:769–778. [PubMed: 20562624]

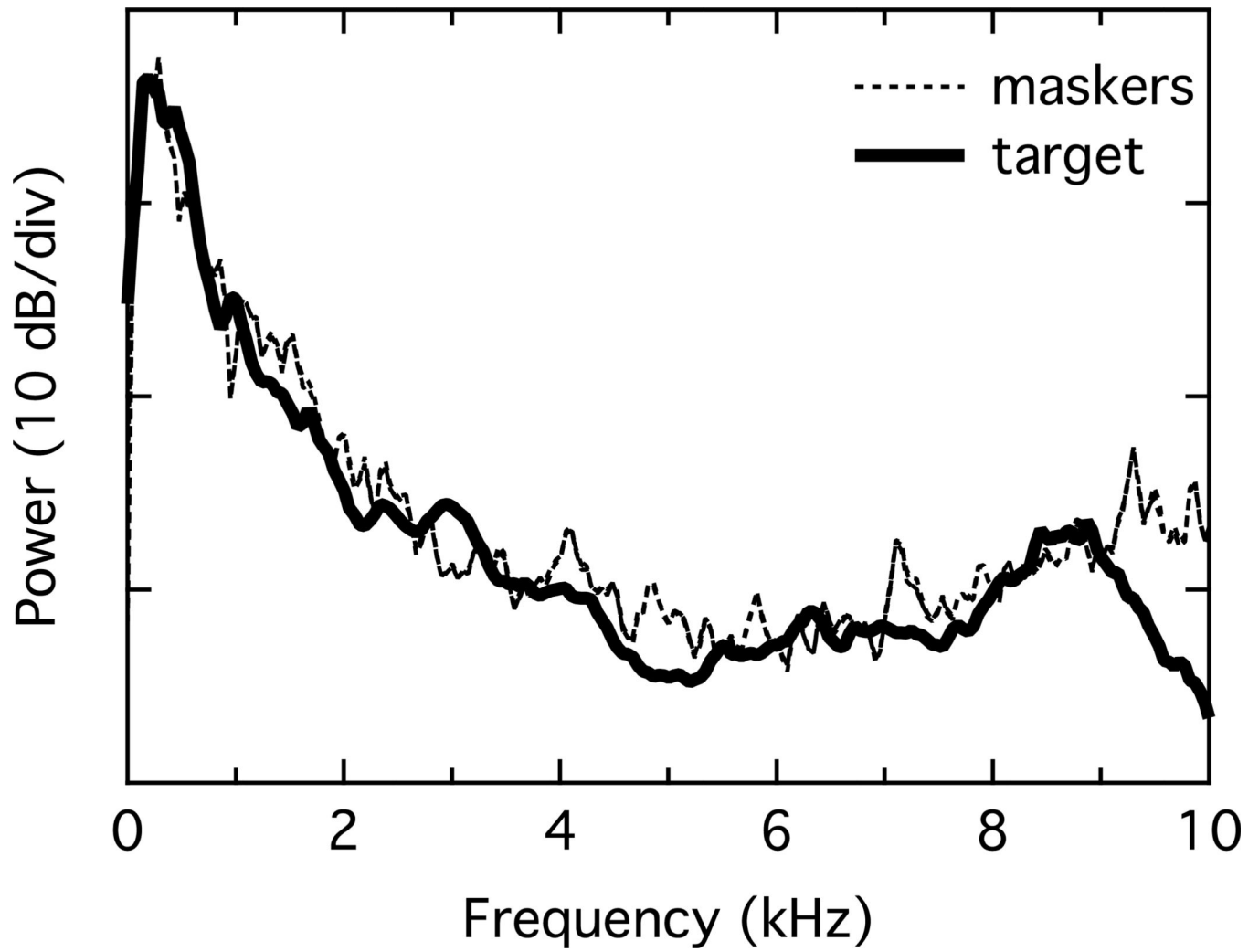


Figure 1. The power spectra of the three maskers (dotted lines) and target (thick line). Due to the similarity between masker spectra, the associated lines are largely overlapping.

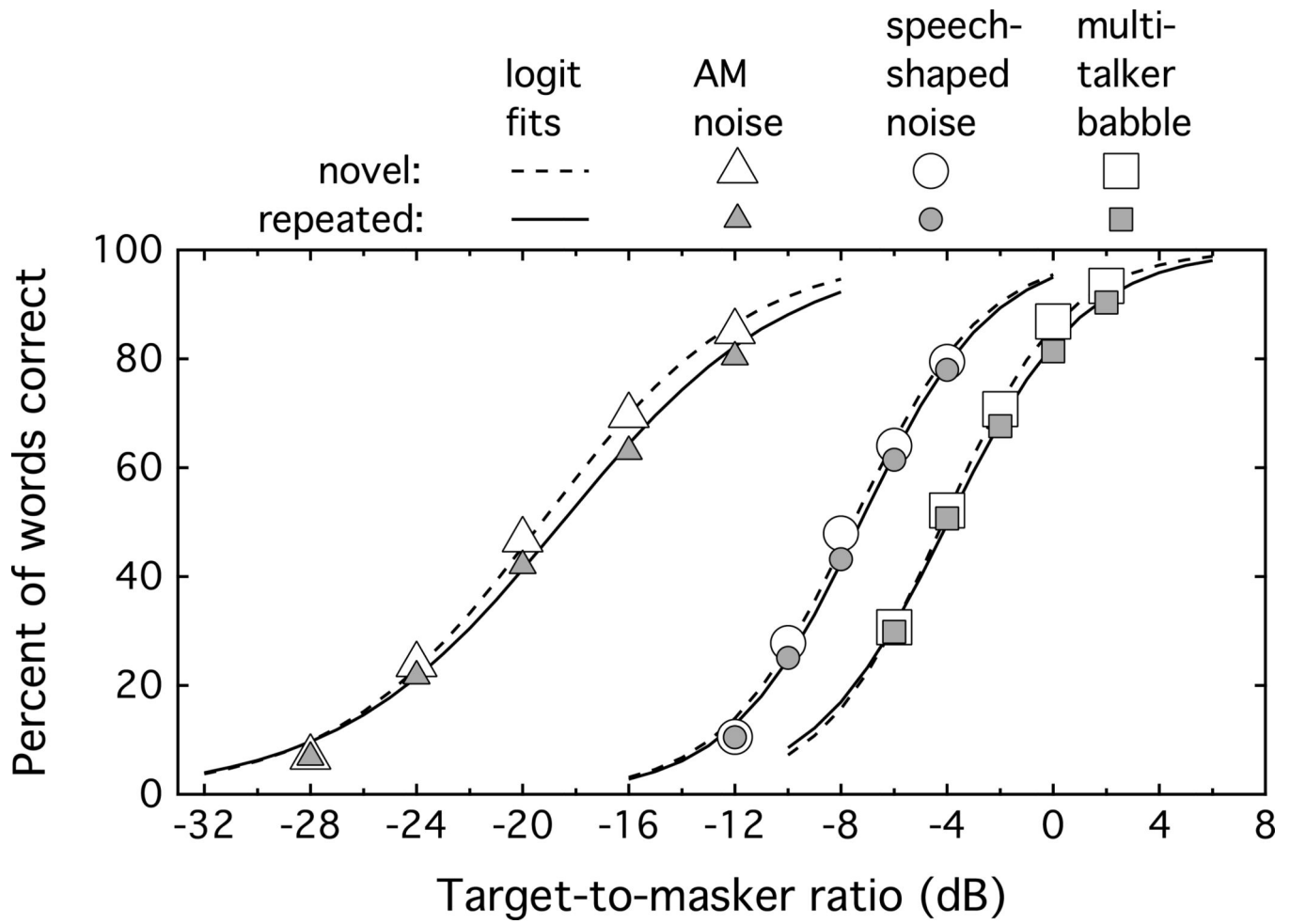


Figure 2. Percent correct for words, plotted as a function of the target-to-masker ratio. Symbol shape reflects the masker type, and symbol fill reflects the data collection method. Lines show psychometric function fits to the mean data.

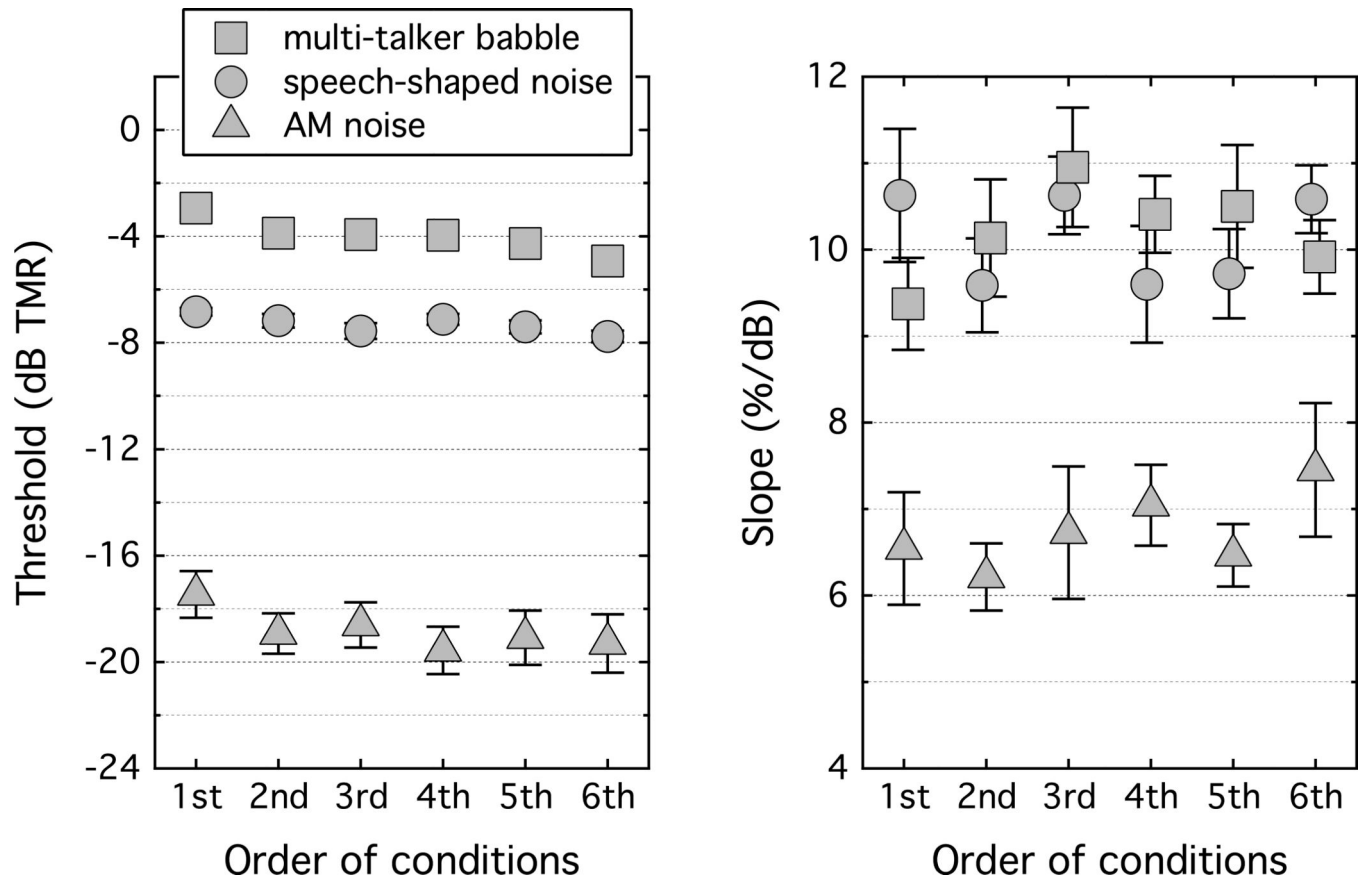


Figure 3.

Mean thresholds (left panel) and slopes (right panel) in each block of trials – five using the repeated-stimulus procedure and one using the novel-stimulus procedure – as a function of the order that each listener completed those blocks. Error bars indicate \pm one standard error of the mean.

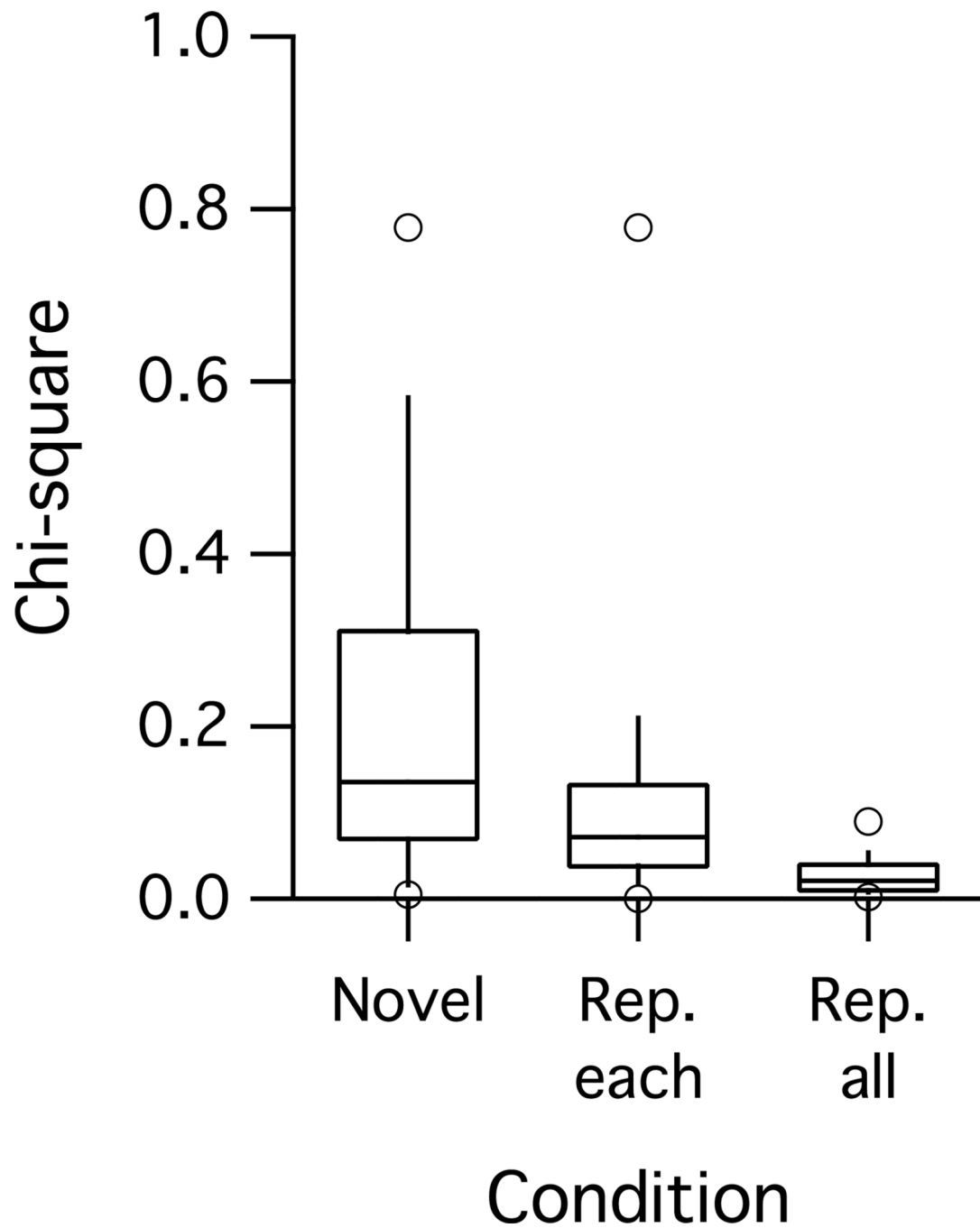


Figure 4.

Chi-squared values associated with the logit fits to individual listeners' data in each condition, indicated on the abscissa. The label 'Rep. each' refers to fits to each run using the repeated-stimulus procedure. The label 'Rep. all' refers to fits to the mean of data collected using the repeated-stimulus procedure. Horizontal lines indicate the median, boxes indicate the 25th-to-75th percentile range, and circles indicate the minimum and maximum values.

Summary statistics for individual data collected in each of the three maskers, using novel- and repeated-stimulus procedures. The left portion of the table shows the mean values of percent correct for the five TMRs associated with each masker. The right portion of the table shows the mean values of threshold (a, in dB TMR) and slope (b, in %/dB), computed based on psychometric function fits to individual listeners' normalized data. The standard error of the mean is shown in italics. The standard error should be treated with caution for percent correct values below 10% and above 90% (Studebaker 1985).

Table 1

multi-talker babble						
TMR	-6	-4	-2	0	2	a b
novel	30.7	52.2	70.7	86.9	93.4	-3.27 11.3
	<i>2.7</i>	<i>3.0</i>	<i>1.7</i>	<i>2.0</i>	<i>1.0</i>	<i>0.20 0.7</i>
repeated	29.8	50.7	67.6	81.4	90.4	-3.22 9.7
	<i>1.7</i>	<i>2.0</i>	<i>1.8</i>	<i>1.7</i>	<i>1.2</i>	<i>0.22 0.4</i>
speech-shaped noise						
TMR	-12	-10	-8	-6	-4	a b
novel	10.6	27.8	47.9	64.0	79.5	-7.29 9.7
	<i>1.5</i>	<i>2.6</i>	<i>2.4</i>	<i>2.6</i>	<i>1.8</i>	<i>0.23 0.5</i>
repeated	10.5	25.1	43.2	61.5	78.0	-6.95 9.9
	<i>1.2</i>	<i>1.3</i>	<i>1.8</i>	<i>1.5</i>	<i>1.4</i>	<i>0.18 0.3</i>
AM noise						
TMR	-28	-24	-20	-16	-12	a b
novel	6.8	23.8	46.7	69.6	85.0	-18.74 6.9
	<i>1.8</i>	<i>5.1</i>	<i>5.4</i>	<i>4.6</i>	<i>2.0</i>	<i>0.84 0.5</i>
repeated	6.8	21.7	42.0	63.0	80.4	-18.25 5.8
	<i>2.0</i>	<i>4.1</i>	<i>5.1</i>	<i>4.6</i>	<i>2.9</i>	<i>0.78 0.2</i>