Published in final edited form as:

Ear Hear. ; 42(2): 313–322. doi:10.1097/AUD.0000000000000951.

Contribution of Stimulus Variability to Word Recognition in Noise vs. Two-talker Speech for School-age Children and Adults

Emily Buss[1], **Lauren Calandruccio**[2], **Jacob Oleson**[3], **Lori J. Leibold**[4]

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

[2.] Case Western Reserve University, Department of Psychological Sciences, Cleveland, Ohio

[3.] University of Iowa, Department of Biostatistics, Iowa City, Iowa

[4.] Boys Town National Research Hospital, Center for Hearing Research, Omaha, Nebraska

Abstract

Background: Speech-in-speech recognition scores tend to be more variable than those for speech-in-noise recognition, both within and across listeners. This variability could be due to listener factors, such as individual differences in audibility or susceptibility to informational masking. It could also be due to stimulus variability, with some speech-in-speech samples posing more of a challenge than others. The purpose of this experiment was to test two hypotheses: 1) that stimulus variability affects adults' word recognition in a two-talker speech masker, and 2) that stimulus variability plays a smaller role in children's performance due to relatively greater contributions of listener factors.

Methods: Listeners were children (5–10 yrs) and adults (18 – 41 yrs) with normal hearing. Target speech was a corpus of 30 disyllabic words, each associated with an unambiguous illustration. Maskers were 30 samples of either two-talker speech or speech-shaped noise. The task was a four-alternative forced choice. Speech reception thresholds (SRTs) were measured adaptively, and those results were used to determine the signal-to-noise ratio associated with ~65% correct for each listener and masker. Two 30-word blocks of fixed-level testing were then completed in each of two conditions: 1) with the target-masker pairs randomly assigned prior to each block, and 2) with frozen target-masker pairs.

Results: The SRTs were lower for adults than children, particularly for the two-talker speech masker. Listener responses in fixed-level testing were evaluated for consistency across listeners. The target sample was the best predictor of performance in the speech-shaped noise masker for both the random and frozen conditions. In contrast, both the target and masker samples affected performance in the two-talker masker. Results were qualitatively similar for children and adults, and the pattern of performance across stimulus samples was consistent with differences in masked target audibility in both age groups.

Corresponding author: Emily Buss, 170 Manning Dr. CB#7070, Chapel Hill, NC 27599, ebuss@med.unc.edu, Phone: (919) 843-9163.

Conclusions: Whereas word recognition in speech-shaped noise differed consistently across target words, recognition in a two-talker speech masker depended on both the target and masker samples. These stimulus effects are broadly consistent with a simple model of masked target audibility. Although variability in speech-in-speech recognition is often thought to reflect differences in informational masking, the present results suggest that variability in energetic masking across stimuli can play an important role in performance.

Introduction

Speech recognition in the presence of background speech can be challenging, particularly for young children (e.g., Corbin et al. 2016; Wightman and Kistler 2005), older adults (Goossens et al. 2017; Helfer and Freyman 2014; Tun et al. 2002), and listeners with hearing loss (Festen and Plomp 1990). Speech-in-speech recognition is thought to be dominated by informational masking, which reflects a failure to segregate and selectively attend to target speech even when the cues necessary for recognition are well represented in the peripheral auditory system (Brungart et al. 2001; Kidd et al. 2008). The ability to recognize speech in a speech masker is important for everyday communication in both children and adults (Hillock-Dunn et al. 2015; Phatak et al. 2018), and a number of clinical tests evaluate speech recognition in either multi-talker babble (e.g., Spahr et al. 2012; Wilson 2003) or two-talker speech (e.g., Cameron et al. 2006; Jakien and Gallun 2018). Recent interest in two-talker speech maskers is motivated by the observation that informational masking is maximized with a small number of masker talkers (Freyman et al. 2004; Rosen et al. 2013). Estimates of speech-in-speech recognition can vary substantially across individual listeners, but they can also vary widely across estimates within a listener. The present study sought to characterize sources of variability for word recognition in a two-talker masker, with the goal of improving our ability to characterize performance in both children and adults.

In addition to the marked individual differences for speech-in-speech recognition as a function of listener age and hearing loss, there are reliable individual differences among young adults with normal hearing. For example, Carbonell (2017) recruited a group of 66 young adults with normal hearing and measured their single-syllable word recognition in three stimulus conditions: time compressed, noise vocoded, and in random samples of four-talker babble at −3 dB SNR (signal-to-noise ratio). Performance was evaluated in two sessions. In the first session, a unique 100-word list was used for each condition. The second session used the same three word lists, but with a different list-to-condition assignment. Scores were correlated across days and across tasks, indicating that individual differences were stable and somewhat consistent regardless of the method for degrading speech. However, the correlation across sessions was larger for the time-compressed and vocoded speech (r = .74 and .79, respectively) than for speech in a babble masker (r = .37). Carbonell (2017) suggested that the use of different masker samples across days could be responsible for the more modest reliability for words in a four-taker babble compared to the other two conditions.

Large individual differences for speech-in-speech recognition are often considered a hallmark of informational masking (Kidd et al. 2016). Individual differences in speech scores could reflect listener factors, such the ability to consistently segregate and selectively attend to the target, or the ability to recognize speech based on spectro-temporally sparse cues (Zekveld et al. 2007). Differences could also result from stimulus variability, with some speech-in-speech samples posing more of a challenge than others. Psychometric functions based on group data tend to be shallower for speech-in-speech than for speech-in-noise recognition (MacPherson and Akeroyd 2014), consistent with either stimulus variability, individual differences, or a combination of factors. However, functions fitted to individual listeners' data also tend to be shallower for speech in a two-talker masker than a speech-shaped-noise masker (Sobon et al. 2019). This result has been interpreted as reflecting variability in masked audibility across time and frequency. For a two-talker speech masker, a small number of cues are audible at low SNRs, and the number of audible cues increases gradually with increasing SNR; in contrast, for a speech-shaped noise masker audibility increases more rapidly with increasing target level. If the audibility interpretation is correct, then the ability to recognize a particular speech-masked target could depend critically on the particular stimulus sample, with some target and masker combinations resulting in greater target audibility than others. Understanding this source of variability is important because the goal of clinical assessment is to characterize an individual listener's abilities, while minimizing other sources of variability (e.g., stimulus variability).

One way to reduce effects of stimulus variability is to repeat stimuli, with those repeating stimuli sometimes described as "frozen." A potential drawback to this approach is that repeating stimuli tends to reduce stimulus uncertainty, a component of informational masking (Kidd et al. 2008). Repeating stimuli on sequential trials improves speech-in-speech recognition (Brungart and Simpson 2004; Felty et al. 2009). In one demonstration of this effect, Felty et al. (2009) tested word recognition in six-talker babble for two groups of listeners: one group heard a randomly selected masker sample on each trial, and the other group heard the same masker sample on all 300 trials. Listeners who heard the same masker sample improved more over the course of testing than listeners who heard random samples, with the largest gains made over the first 100 trials and a total effect on the order of six percentage points. These results suggest that repeating stimuli can improve performance, but that it takes multiple presentations for the full benefit to emerge. A similar result was observed by Langhans and Kohlrausch (1992) for tone detection in a narrowband noise masker; the benefit of repeating a frozen-noise masker sample emerged over 10–60 sequential trials. However, Richards and Neff (2004) observed improved ability to detect a tonal signal in a random-frequency multi-tone masker after a single presentation of the masker alone, indicating that extensive exposure is not always required to benefit from the reduced uncertainty associated with a frozen masker.

Another approach for reducing stimulus variability is to use a relatively large set of frozen stimuli that are interleaved over time, such that responses can be compared for identical stimuli, but the listener does not have the opportunity to learn from sequentially repeating exposures. This approach is widely used, including experiments designed to study informational masking with tonal stimuli (Durlach et al. 2005; Leibold et al. 2010) and tests of speech recognition where the target and masker are burned to CD. When Wilson et al. (2003) described development of the words-in-noise (WIN) test, they noted that the particular sample of multi-talker babble paired with the target word affected performance, so a fixed sample was used for each target. Using these methods, test-retest reliability for the

WIN is high, with r ≥ .80 for listeners with normal hearing (Wilson et al. 2003)[1] or hearing loss (Wilson and McArdle 2007). Recall that Carbonell (2017) reported reliability of r = .37 for word recognition when the masker was random (not frozen). While these results are broadly consistent with the idea that frozen speech-in-speech materials increase our ability to observe reliable individual differences compared to the use of random masker samples, a direct comparison of frozen vs. random stimuli is warranted, due to the other discrepancies in methods, stimuli, and listeners across these two studies.

The current study was designed to evaluate sources of variability in masked word-recognition thresholds for children and adults tested with a speech-shaped noise masker or a two-talker speech masker. The stimuli and methods used here are under development as part of the Children's English and Spanish Speech Recognition Test (ChEgSS), which uses stimuli and procedures introduced by Calandruccio et al. (2014). Our interest in characterizing sources of variability for this task was motivated by the need to optimize test efficiency for use in a clinical setting, particularly when testing children; under these conditions, the goal is to accurately characterize listener sensitivity with the smallest number of trials possible. At the outset there were two hypotheses: 1) that *stimulus variability is a dominant factor in adults' word recognition in a two-talker speech masker*, and 2) that *stimulus variability plays a smaller role in performance of children than adults.* The second hypothesis was based on the idea that children are often more variable and their responses are driven more by intrinsic listener factors compared to adults (McCreery et al. 2017; Miller et al. 2019).

Methods

Listeners

Listeners were 5- to 10-years-olds (n=28, mean 7.6 yrs, 16 F) and adults (n=23, 18–41 yrs, mean 24.5 yrs, 17 F). Two additional children provided incomplete data due to experimenter error; they were therefore omitted. All listeners had normal hearing, defined as thresholds no greater than 20 dB HL bilaterally at octave frequencies 250–8000 Hz (ANSI 2018), and all were native speakers of American English. Child listeners were also required to have a Type A tympanogram bilaterally to qualify for the study. Listeners were free from known cognitive or hearing related disorders, by self or parent report.

Stimuli

Targets were 30 disyllabic words .43 - .68 sec in duration, each associated with an unambiguous illustration, as described by Calandruccio et al. (2014). Maskers were 30 samples of either two-talker speech or speech-shaped noise, each 2.8 sec in duration. The two-talker speech masker was composed of two streams from a female talker reading passages from *Jack and the Beanstalk* (Walker 1999). These passages were edited to remove silent gaps greater than 300 ms, levels were equated, and the two streams were summed. The speech-shaped noise masker had the same long-term average power spectrum as the two-talker masker. Table 1 reports the target words, as well as the masker samples paired with

[1] Data used to estimate correlation between the first and second estimate were derived from Figure 2 of Wilson et al (2003).

each word in the *frozen* condition. Bold font represents the portion of the masker speech that temporally overlapped with the target word.

Procedures

Stimuli were presented diotically over circumaural headphones (Sennheiser, HD 280 PRO, Wedemark, Germany). The masker was gated on and off with 5-ms raised-cosine ramps, and the target was temporally centered in the masker sample. The overall level of the signal-plus-masker in the temporal center of the stimulus was fixed at 65 dB SPL. The SNR was manipulated by adjusting both the signal level and the masker level. Each listener was randomly assigned to begin testing with either the speech-shaped noise or the two-talker speech masker, and data were collected blocked by masker.

Before the experiment began, listeners were familiarized with the target stimuli. Each of the 30 illustrations were presented in random order, and listeners were asked to identify them. If listeners provided a response other than the associated target word (e.g., “kids” instead of “children”), they were asked if they could think of any other words to describe the picture. All listeners successfully produced the target word. Listeners then heard the stimulus recording of the target word presented in quiet.

Data were collected using a four-alternative forced-choice, with a touchscreen response. The four illustrations presented on each trial consisted of the target and three foils, randomly selected without replacement. The quadrant of the target illustration on the screen was randomly selected on each trial. The order of target words was determined by randomly scrambling the order of the 30 words (randperm, MATLAB). A new random sequence was generated for tracks containing more than 30 trials, with the caveat that no target word could be repeated in sequential trials.

For each masker, there were three stages of testing. The first stage was an adaptive procedure, estimating the SNR associated with 71% correct. The stepping rule was a 2-down, 1-up. The initial stepsize was 4 dB, and that was reduced to 2 dB after the second track reversal. Tracks continued for 8 reversals, and the final estimate of the Speech Reception Threshold (SRT) was the mean SNR at the last 6 reversals. Two such estimates were obtained. For this stage of testing, target and masker samples were randomly paired. This was achieved by randomly scrambling the order of the 30 target and masker samples independently at the beginning of each threshold estimation track and using the results to define stimuli presented on trials 1–30. A new pair of random target and masker sequences was generated for tracks containing more than 30 trials.

The second stage of testing measured percent correct for each of the 30 target words, randomly paired to masker stimuli, with randomization refreshed prior to each run. The selection of foils was also random on each trial. The initial SNR was selected based on the adaptive threshold estimate: 1–2 dB below the SRT for the speech-shaped noise masker and 4–8 dB below the SRT for the two-talker speech masker. The SNR was adjusted in subsequent runs until a value associated with approximately 65% correct was determined; this value was selected because it is approximately the midpoint of the psychometric function for a four-alternative forced choice. Two 30-word runs in the *random* condition were completed at the final SNR value.

The final stage of testing measured percent correct at the SNR determined in the previous stage, but with frozen stimuli. In the *frozen* condition, target and masker pairs were randomly selected once, prior to the beginning of the experiment, and used consistently across runs and across participants. The foils associated with each target and masker pair were likewise consistent. As for the *random* condition, two 30-word runs of fixed-SNR testing were completed in the *frozen* condition.

Analysis

Regression models were used to evaluate SRTs, and logistic regression models were used for trial-by-trial response data to evaluate percent correct. In all cases, these models accommodate repeated measures with a random intercept for each listener, as well correlations for within subject variables. The covariance structure for each model was selected to minimize the Akaike Information Criterion (AIC), and that choice was validated via the likelihood ratio test with a criterion of $\alpha = .05$.

Results and discussion

One of the younger children (5.5 yrs) provided data that were clearly outliers compared to other children. For example, this child's SRT in the two-talker speech masker was 21.6 dB SNR, which is more than 15 dB higher than the next-highest threshold. These data were omitted from further consideration, leaving 27 child listeners.

Speech reception thresholds

The top row of panels in Figure 1 shows mean SRTs obtained with random target and masker pairings, plotted as a function of child age on a log scale. Age is also represented by symbol shading. Results for the speech-shaped noise are shown in the left panel, and those for the two-talker speech masker are in the right panel. The distribution of SRTs for adult listeners is shown at the right of each panel.

Children's SRTs improved as a function of age, but the effect of age was more modest for the speech-shaped noise masker than the two-talker speech masker. This observation was confirmed with a regression model that accommodates correlations for within subject variables. The dependent variable was the mean SRT, averaged across the two threshold estimation tracks, and the independent variables were log-transformed age, sex[2], masker type, and the interaction between masker and age. Reference categories were female (sex) and speech-shaped noise (masker). The model was significantly improved by introducing different variances for each masker type (AIC = 217.4 vs. 219.5, p = .043), consistent with the observation of greater variance in the two-talker speech data than the speech-shaped noise data. This model indicates significant effects of child age ($\beta=-3.34$, $t=-2.74$, p = .009)

[2] The original motivation for including sex in this analysis was to be in compliance with the NIH policy on Sex as a Biological Variable (National Institutes of Health 2016).

and masker type (β=22.49, t=4.85, p < .001), and a significant interaction between age and masker type (β=−7.45, t=−3.28, p = .002). The older children's SRTs tended to fall within the range of adult data for both masker types. There was also a significant effect of sex (β=1.40, t=3.17, p = .003), reflecting higher thresholds for boys than girls. While an effect of sex was not predicted, there is some precedent in the literature for girls to have better speech-in-noise recognition than boys (Ross et al. 2015).

For adults, there were larger individual differences in the two-talker speech masker than the speech-shaped noise masker. This observation was confirmed using Bartlett's test for homogeneity of variances (χ^2 = 30.76, p < .001). The relative magnitude of individual differences and variability across replicate estimates was computed by fitting a linear model to the two SRT estimates for each listener, with a separate model for each group and masker type, and computing the percent of variance accounted for by the random effect of listener. Using this approach, individual differences in adults accounted for 37% of variance in the SSN and 56% of variance in the two-talker speech masker. For children, those values were 54% and 51%, respectively.

One interesting feature of these data is the trend for approximately constant performance in the two-talker speech masker between 5 and 9 years of age, followed by rapid improvement between 9 and 11 years of age. This trend is qualitatively similar to the results of Corbin et al. (2016), where monosyllabic open-set word recognition was measured in either a two-talker speech masker or speech-shaped noise masker, and listeners were 5- to 16-year-olds with normal hearing. As in the present study, performance improved steadily with age for the speech-shaped noise masker, but not in the two-talker speech masker. For the speech masker, thresholds were similar between 5 and 12 years of age, at which point they rapidly improved to adult levels. It is unclear how to explain this discontinuity, but one possibility is that young children are unable to segregate the target from the masker, such that they perform poorly unless the target is loud enough to dominate the target-plus-masker stimulus. By this view, the knee point associated with the onset of SRT improvement could reflect the age at which children begin to perceptually segregate the target from the speech masker.

The bottom row of panels in Figure 1 shows the two replicate estimates of SRT for each listener. Results for the speech-shaped noise are shown in the left panel, and those for the two-talker speech masker are shown in the right panel. Symbol shading reflects listener age, and lines connect the first and second estimate obtained for each listener. For the noise masker, the first and second estimate of SRT differed by an average of .3 dB for both children and adults (SD 2.0 dB and 1.5 dB, respectively). For the two-talker speech masker, the mean change in SRT across estimates was 1.4 dB (SD 3.8 dB) for children and .2 dB (SD 4.4 dB) for adults. This trend for improvement across estimates could reflect practice effects, but these changes are modest compared to the magnitude of group effects and individual differences. For example, mean SRTs for children and adults differ by 2.4 dB for speech-shaped noise and 10.7 dB for two-talker speech.

The stability of SRTs was evaluated using a regression model, with the difference in SRTs across estimates as the dependent measure, and independent variables of masker type (speech-shaped noise, two-talker speech), listener group (child, adult), and their interaction.

There was evidence of distinct variance by masker type (AIC = 484.5 vs. 515.9, p < .001), but no evidence of an added benefit of including random intercepts for listener group or the interaction of group and masker type. This is consistent with the observation of larger differences between sequential estimates for the two-talker speech masker than the speech-shaped noise masker, but comparable test-retest reliability for children and adults. This best-fitting model indicates no significant effect of age group (β = −.15, t = −.31, p = .755) or masker type (β = −.13, t = −.14, p = .887), and no interaction (β = 1.29, t = 1.05, p = .295). This result indicates that the mean change in SRT from the first to second estimate of SRT did not differ significantly across age groups or maskers.

These results confirm the premise of this experiment – that variability across sequential estimates of speech recognition within listeners is larger for the two-talker speech masker than the speech-shaped noise masker. However, the hypothesis of greater variability in children than adults was not supported. Behavioral testing depends on the listener understanding the task and maintaining attention throughout a measurement, either of which may be compromised in children. The ChEgSS protocol was designed to be easy to understand and engaging for young children, and failure to find evidence of greater variability in child data suggests that these goals were achieved.

Percent correct

The second stage of testing determined the SNR associated with approximately 65% correct. For children, those estimates were −10.6 dB (SD = 1.5) and −4.7 dB (SD = 4.6) for speech-shaped noise and two-talker speech maskers, respectively. For adults, those estimates were −12.8 dB (SD = 1.4) and −19.5 dB (SD = 4.6), respectively. The goal of achieving 65% correct was achieved in most cases. The mean percent correct observed in the second and third stages of testing fell between 61 and 72% across all combinations of age groups, maskers, and conditions.

If stimulus variability is a dominant factor determining listeners' responses, then listeners should tend to provide the same response when presented with the same stimulus. To evaluate the consistency of listener responses, the mean percent correct for each stimulus sample was computed in three ways: by masker segment for the *random* condition, by target word for the *random* condition, or by both masker sample and target word in the *frozen* condition. Figure 2 shows the group mean values of percent correct for each stimulus, sorted in ascending order (lowest to highest percent correct), plotted separately for each age group (rows) and masker type (columns). The sort order associated with each line is indicated on the abscissa using symbols associated with each target and masker sample, which are defined in Table 1. The grey shaded region in Figure 2 indicates the 95% confidence interval around chance performance, representing no systematic consistency across listeners' responses. This was computed using bootstrap procedures (n = 10,000), based on the mean percent correct values for each group and masker type. Deviation from the grey shaded region indicates more consistency in listener responses than expected by chance.

For the speech-shaped noise masker (A1 and A2, left panels of Fig 2), performance sorted by target deviates from chance for both the *frozen* condition and the *random* condition, but performance sorted by masker falls almost entirely within the shaded region for the *random* condition. These trends are observed for both child and adult listeners. This result suggests that some targets words are more likely to be recognized than others when presented in speech-shaped noise, but the masker sample has little or no effect on performance. For the speech-shaped noise masker, mean performance across target samples is correlated for children and adults for both the *random* condition (r = .93, p < .001) and the *frozen* condition (r = .89, p < .001), indicating that both age groups experience similar target effects.

For the two-talker speech masker (B1 and B2, Fig 2), performance clearly deviates from chance for both groups in the *frozen* condition. In contrast to the speech-shaped noise masker, in the *random* condition with the two-talker speech masker there is no evidence of consistent responses by target for either age group, but there is some indication of consistency by masker for the adults (as indicated by the dashed line falling outside the confidence interval around chance). For children, reliable responses in the *frozen* but not the *random* two-talker speech masker conditions support the idea that the target/masker combination is the most important stimulus feature contributing to performance. For adults, there may be additional effects related to masking exerted by particular two-talker speech masker samples. Mean performance across samples in the *frozen* condition is correlated for children and adults (r = .51, p = .002), but this correlation is significantly smaller than observed in the speech-shaped noise masker (z = 3.16, p = .002). While this result could indicate that the stimulus features associated with informational masking differ for children and adults, an alternative interpretation is offered below.

The data patterns illustrated in Figure 2 were evaluated with a series of mixed effects logistic regression models for each masker and condition, with fixed effects for age group and stimulus (target, masker). For the speech-shaped noise masker and the *random* condition, there was a significant effect of target (χ^2 = 162.0, p < .001), but no effect of masker segment (χ^2 = 27.6, p = .537), run number (χ^2 < .1, p = .985), or age group (χ^2 = 3.1, p = .080). There was no evidence of an interaction between group and either target or masker segment (χ^2 = 24.0, p = .729; χ^2 = 31.9, p = .324). For the speech-shaped noise masker and the *frozen* condition, there was a significant effect of stimulus (target and masker; χ^2 = 231.8, p < .001) and a significant interaction between stimulus and age group (χ^2 = 44.9, p = .030), but no effect of run number (χ^2 < .1, p = .828) or age group (χ^2 = .6, p = .452). Both of these models are consistent with the idea that performance in the speech-shaped noise masker varies across targets.

Parallel analyses were also performed for data collected with the two-talker speech masker. For the *random* condition, there were significant effects of target (χ^2 = 47.7, p = .016) and masker (χ^2 = 165.5, p < .001), as well as significant interactions with group for both factors (χ^2 = 53.8, p = .003; χ^2 = 72.6, p < .001). Effects of age group (χ^2 = 1.8, p = .179) and run number (χ^2 = 3.6, p = .058) were non-significant. Similarly, for the two-talker speech masker and the *frozen* condition, there was a significant effect of stimulus (target and masker; χ^2 = 246.8, p < .001) and age group (χ^2 = 4.8, p = .029), as well as an interaction between stimulus and group (χ^2 = 156.5, p < .001). There was not a significant effect of run number (χ^2 = 2.6, p = .108). These results provide support for the conclusion that both the target and masker samples affect performance in the two-talker speech masker conditions.

Interactions with group provide further evidence that children differ from adults with respect to the difficulty of particular target/masker pairs.

Effects of masked audibility

Stimuli were evaluated to determine whether the pattern of results in the *frozen* condition is consistent with differences in energetic masking. This analysis was modeled after the first stage of the ideal binary mask analysis (Brungart et al. 2006), which isolates spectro-temporal segments where the SNR exceeds some criterion (see also: Cooke 2006). In this implementation, stimuli were passed through a bank of 64 fourth-order gammatone filters, with center frequencies distributed from 125 to 12000 Hz, equally spaced on a log scale. Filter outputs were analyzed in sequential 20-ms Hann windows that overlapped at the half-rise point. The local SNR was computed for each temporal window in each frequency band, with a -6 dB criterion for audibility, based on results of Brungart et al. (2006). This analysis was completed for all stimuli in the *frozen* condition, with global SNRs set based on the mean values for either the child or adult data. Figure 3 shows percent correct as a function of the percent of epochs reaching the criterion for masked audibility, plotted separately for the two age groups and maskers.

For children in the present cohort, 60–70% correct performance in a 30-word list was associated with 21.7% masked audibility in speech-shaped noise and 63.7% masked audibility in the two-talker speech masker. For adults, those values were 15.9% and 25.8%, respectively. In other words, both groups required greater audibility in the two-talker speech masker than the speech-shaped noise masker, but this effect was more pronounced for children than adults (42 vs 10 percentage points). This finding is qualitatively similar to that reported by Sobon et al. (2019). In that study, the short-term speech intelligibility index (SII) was computed at the SRT for speech-in-noise for children, young adults, and older adults. This analysis was used to determine the criterion audibility required for recognition, which varies across listeners, and the individualized criterion was used to predict speech recognition in the two-talker speech masker. All listeners performed more poorly than predicted, but the discrepancy between predicted and observed SRTs in the two-talker speech masker was larger for children than young adults. This result was interpreted as indicating particularly large audibility requirements for children tested in the two-talker speech masker, as observed in the present dataset. This is also consistent with the general finding that the child/adult difference is larger for speech-in-speech recognition than speech-in-noise recognition (Sobon et al. 2019).

In addition to marked group differences in the audibility required to support recognition, differences in masked audibility were associated with performance across stimulus pairs in the *frozen* condition. For the speech-shaped noise masker, the correlation between percent correct and percent masked audibility was r = .67 (children) and r = .64 (adults). For the two-talker speech masker, those values were r = .55 (children) and r = .60 (adults); this trend for a lower correlation in child data was not significant (z = .275, p = .392). Recall that the correlation in mean percent correct for each of the 30 *frozen* stimuli was weaker across age groups for the two-talker speech masker than the speech-shaped noise masker (r = .51 vs. r = .89, respectively). One interpretation of this result is that different stimulus features limit performance in children vs. adults. For example, children are less efficient than adults at using differences in voice F0 to segregate the target and masker (Flaherty et al. 2018), and this cue could be more important for some stimulus samples than others. However, the acoustic analysis suggests an alternative interpretation – that differences in masked audibility across stimuli change with changes in SNR. In contrast to the speech-shaped noise maker, increasing SNR in the two-talker speech masker can have variable effects on the percent audibility of target cues, depending on the distribution of masker energy over time and frequency compared to the target. This raises the possibility that some or all of the discrepancy across age groups in the pattern of responses to frozen stimuli in the two-talker speech masker could be due to nonlinear changes in percent audibility in the speech masker. While results of the audibility analysis are suggestive, the very modest number of stimuli used in this experiment prevents a rigorous test of all the parameters and steps incorporated into this analysis. For example, it is not clear that the -6 dB criterion is appropriate for characterizing audibility in children, or that percent of audible glimpses is even the best metric to use.

Does the use of frozen samples increase reliability of SRT estimates?

If both target and masker samples affect performance, then theoretically test-retest reliability should be better in the *frozen* condition than the *random* condition. In other words, variability that is unrelated to reliable individual differences should be reduced. This benefit would be most pronounced for corpora composed of a small number of stimuli; as the stimulus set grows, effects related to stimulus variability would tend to average out. Test-retest reliability was evaluated by examining the mean difference between percent correct in the first vs. second fixed-SNR run for each listener and masker condition. For adults, the mean change in percent correct in the first vs. second run (absolute value of the difference percent correct at each time point) was 7.0 and 7.1 points for the *random* and *frozen* conditions, respectively. For children, those values were 11.7 and 6.9 points. These results indicate greater reliability with frozen stimuli compared to random stimuli for child data but not for adult data. While less variability in children's performance for the *frozen* condition is broadly consistent with reduced stimulus variability, it is also consistent with a practice effect, since performance was measured in the *random* condition prior to the *frozen* condition. In addition to potential practice effects, the staged data collection procedure could tend to reduce our ability to observe a benefit of using frozen stimuli. Recall that the SNR was manipulated in stage 2 until performance in the *random* condition was approximately 65% correct across two runs. This SNR was then used to evaluate performance in the *frozen* condition. This two-staged procedure would tend to replace outlier points in the *random* condition, in the search for an appropriate SNR, such that variance was reduced in the *random* condition relative to the *frozen* condition.

In order to assess effects of stimulus variability on percent correct performance, a second group of young adults (n = 15, 19 – 27 yrs, mean 22.4 yrs, 13 F) was recruited, meeting the same inclusion criteria as above. They completed four runs of fixed-SNR testing at −22 dB SNR with the two-talker speech masker, including two runs in the *random* condition and two in the *frozen* condition. Listeners who provided supplemental data had previously completed 10 adaptive tracks with the ChEgSS stimuli for an unrelated study prior to fixed-level testing, so they were fully practiced. They completed the fixed-level conditions in interleaved order, to limit bias related to fatigue. Median percent correct for the two stimulus conditions was 67% (*random*) and 72% (*frozen*), similar to results of the main study. In this dataset, the mean change in percent correct in the first vs. second run was 9.1 and 8.8 points for the *random* and *frozen* conditions, respectively. This result fails to demonstrate a meaningful improvement in reliability for the *frozen* condition in the current task, but we would expect a larger difference for a smaller stimulus set[3] or larger differences in performance across target and masker pairs.

Conclusions

- As observed previously, SRTs were lower for adults than children, particularly for the two-talker speech masker, and SRTs for the two-talker speech masker were more variable than those measured in the speech-shaped noise masker. Only ~50% of the variability observed for SRTs measured in the two-talker speech masker can be attributed to reliable individual differences.
- Group-level percent correct data indicate some reliable differences across stimuli. For the speech-shaped noise masker, differences across targets were reliable. For the two-talker speech masker, the combination of the target and the masker sample affected performance.
- Although children and adults were tested at different SNRs, the consistency of responses across individuals was similar for children and adults in the *frozen* condition. This outcome was unexpected in light of the observation that children’s performance often appears to suffer from variability in attention or listening strategy.
- The pattern of responses by stimulus sample in the *frozen* condition was broadly consistent with an analysis of masked audibility at the group mean SNR for each age group.

Acknowledgements

This work was supported by NIH NIDCD Grant No. R01 DC014460 (EB) and R01 DC015056 (EB & LJL). Preliminary data for this experiment were presented at the 42nd Midwinter Research Meeting of the Association for Research in Otolaryngology. Contributors to this project include Ryan McCreery, Dawna Lewis, Barbara Rodriguez, Diana Regalado, and Margaret Miller.

Conflicts of Interest and Source of Funding: The authors have no conflicts of interest to disclose. This work was supported by NIH NIDCD Grant No. R01 DC014460 (EB) and R01 DC015056 (EB & LJL).

References

ANSI. (2018). ANSI S3.6–2018, American National Standard Specification for Audiometers. New York: American National Standards Institute.

[3] Supplemental data were resampled 1000 times to estimate the change in percent correct between the first and second fixed-level runs for a subset of the 30 targets. For a randomly selected set of 5 targets, the mean magnitude of the change in percent correct was 22.4 and 16.0 points for the *random* and *frozen* conditions, respectively. These values dropped to 16.0 and 13.0 points for sets of 10 targets, and 11.4 and 10.2 points for sets of 20 targets. This analysis is consistent with the expectation that the use of *frozen* stimuli has larger beneficial effects on reliability with smaller stimulus sets.

Brungart DS, Chang PS, Simpson BD, et al. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. J Acoust Soc Am, 120, 4007–4018. [PubMed: 17225427]

Brungart DS, Simpson BD (2004). Within-ear and across-ear interference in a dichotic cocktail party listening task: effects of masker uncertainty. J Acoust Soc Am, 115, 301–310. [PubMed: 14759023]

Brungart DS, Simpson BD, Ericson MA, et al. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. J Acoust Soc Am, 110, 2527–2538. [PubMed: 11757942]

Calandruccio L, Gomez B, Buss E, et al. (2014). Development and preliminary evaluation of a pediatric Spanish-English speech perception task. Am J Audiol, 23, 158–172. [PubMed: 24686915]

Cameron S, Dillon H, Newall P. (2006). Development and evaluation of the listening in spatialized noise test. Ear Hear, 27, 30–42. [PubMed: 16446563]

Carbonell KM (2017). Reliability of individual differences in degraded speech perception. J Acoust Soc Am, 142, EL461.

Cooke M. (2006). A glimpsing model of speech perception in noise. J Acoust Soc Am, 119, 1562–1573. [PubMed: 16583901]

Corbin N, Bonino AY, Buss E, et al. (2016). Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. Ear Hear, 37, 55–63. [PubMed: 26226605]

Durlach NI, Mason CR, Gallun FJ, et al. (2005). Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers. J Acoust Soc Am, 118, 2482–2497. [PubMed: 16266169]

Felty RA, Buchwald A, Pisoni DB (2009). Adaptation to frozen babble in spoken word recognition. J Acoust Soc Am, 125, EL93–97.

Festen JM, 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. [PubMed: 2262629]

Flaherty MM, Buss E, Leibold LJ (2018). Developmental effects in children’s ability to benefit from F0 differences between target and masker speech. Ear Hear, 40, 927–937.

Freyman RL, Balakrishnan U, Helfer KS (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. J Acoust Soc Am, 115, 2246–2256. [PubMed: 15139635]

Goossens T, Vercammen C, Wouters J, et al. (2017). Masked speech perception across the adult lifespan: Impact of age and hearing impairment. Hear Res, 344, 109–124. [PubMed: 27845259]

Helfer KS, Freyman RL (2014). Stimulus and listener factors affecting age-related changes in competing speech perception. J Acoust Soc Am, 136, 748–759. [PubMed: 25096109]

Hillock-Dunn A, Taylor C, Buss E, et al. (2015). Assessing speech perception in children with hearing loss: What conventional clinical tools may miss. Ear Hear, 36, e57–60. [PubMed: 25329371]

Jakien KM, Gallun FJ (2018). Normative Data for a Rapid, Automated Test of Spatial Release From Masking. Am J Audiol, 27, 529–538. [PubMed: 30458523]

Kidd G Jr., Mason CR, Swaminathan J, et al. (2016). Determining the energetic and informational components of speech-on-speech masking. J Acoust Soc Am, 140, 132–144. [PubMed: 27475139]

Kidd G, Mason CR, Richards VM, et al. (2008). Informational Masking. In Auditory Perception of Sound Sources (pp. 143–190). New York: Springer.

Langhans A, Kohlrausch A. (1992). Differences in auditory performance between monaural and dichotic conditions. I: Masking thresholds in frozen noise. J Acoust Soc Am, 91, 3456–3470. [PubMed: 1619122]

Leibold LJ, Hitchens JJ, Buss E, et al. (2010). Excitation-based and informational masking of a tonal signal in a four-tone masker. J Acoust Soc Am, 127, 2441–2450. [PubMed: 20370027]

MacPherson A, Akeroyd MA (2014). Variations in the slope of the psychometric functions for speech intelligibility: a systematic survey. Trends Hear, 18.

McCreery RW, Spratford M, Kirby B, et al. (2017). Individual differences in language and working memory affect children’s speech recognition in noise. Int J Audiol, 56, 306–315. [PubMed: 27981855]

Miller MK, Calandruccio L, Buss E, et al. (2019). Masked English Speech Recognition Performance in Younger and Older Spanish-English Bilingual and English Monolingual Children. J Speech Lang Hear Res, 62, 4578–4591. [PubMed: 31830845]

National Institutes of Health. (2016). NIH Policy on Sex as a Biological Variable. 2020 from https://orwh.od.nih.gov/sex-gender/nih-policy-sex-biological-variable.

Phatak SA, Sheffield BM, Brungart DS, et al. (2018). Development of a test battery for evaluating speech perception in complex listening environments: Effects of sensorineural hearing loss. Ear Hear, 39, 449–456. [PubMed: 29570117]

Richards VM, Neff DL (2004). Cuing effects for informational masking. J Acoust Soc Am, 115, 289–300. [PubMed: 14759022]

Rosen S, Souza P, Ekelund C, et al. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. J Acoust Soc Am, 133, 2431–2443. [PubMed: 23556608]

Ross LA, Del Bene VA, Molholm S, et al. (2015). Sex differences in multisensory speech processing in both typically developing children and those on the autism spectrum. Front Neurosci, 9, 185. [PubMed: 26074757]

Sobon KA, Taleb NM, Buss E, et al. (2019). Psychometric function slope for speech-in-noise and speech-in-speech: Effects of development and aging. Journal of the Acoustical Society of America, 145, EL284.

Spahr AJ, Dorman MF, Litvak LM, et al. (2012). Development and validation of the AzBio sentence lists. Ear Hear, 33, 112–117. [PubMed: 21829134]

Tun PA, O'Kane G, Wingfield A. (2002). Distraction by competing speech in young and older adult listeners. Psychol Aging, 17, 453–467. [PubMed: 12243387]

Walker R. (1999). Jack and the Beanstalk. Cambridge, M.A.: Barefoot Books.

Wightman FL, Kistler DJ (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. J Acoust Soc Am, 118, 3164–3176. [PubMed: 16334898]

Wilson RH (2003). Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. J Am Acad Audiol, 14, 453–470. [PubMed: 14708835]

Wilson RH, Abrams HB, Pillion AL (2003). A word-recognition task in multitalker babble using a descending presentation mode from 24 dB to 0 dB signal to babble. J Rehabil Res Dev, 40, 321–327. [PubMed: 15074443]

Wilson RH, McArdle R. (2007). Intra- and inter-session test, retest reliability of the Words-in-Noise (WIN) test. J Am Acad Audiol, 18, 813–825. [PubMed: 18496992]

Zekveld AA, George EL, Kramer SE, et al. (2007). The development of the text reception threshold test: A visual analogue of the speech reception threshold test. J Speech Lang Hear Res, 50, 576–584. [PubMed: 17538101]

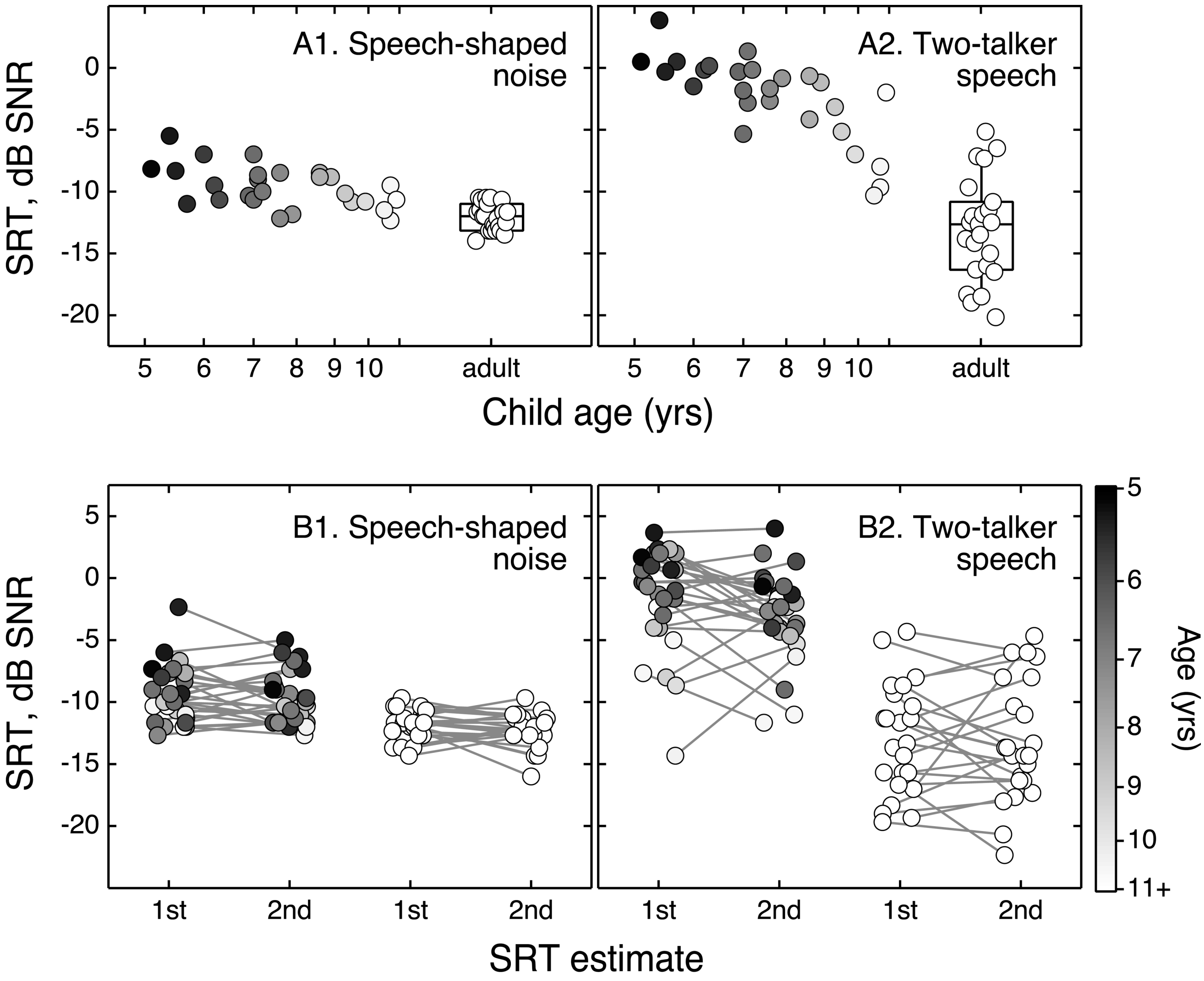

Figure 1.

Results for stage 1 of testing, adaptive threshold estimation with random target and masker pairings. A) The top row of panels shows mean SRTs as a function of child age, plotted separately for the speech-shaped noise masker (left) and the two-talker speech masker (right). Boxplots at the right of each panel indicate the distribution of SRTs for adults, and circles indicate individual data. B) The bottom row of panels shows the first and second SRT measured for each listener, plotted separately for the speech-shaped noise masker (left) and the two-talker speech masker (right). Connected circles indicate replicate estimates for an individual listener. Symbol shading reflects listener age, from dark fill (youngest children) to light (oldest children and adults).

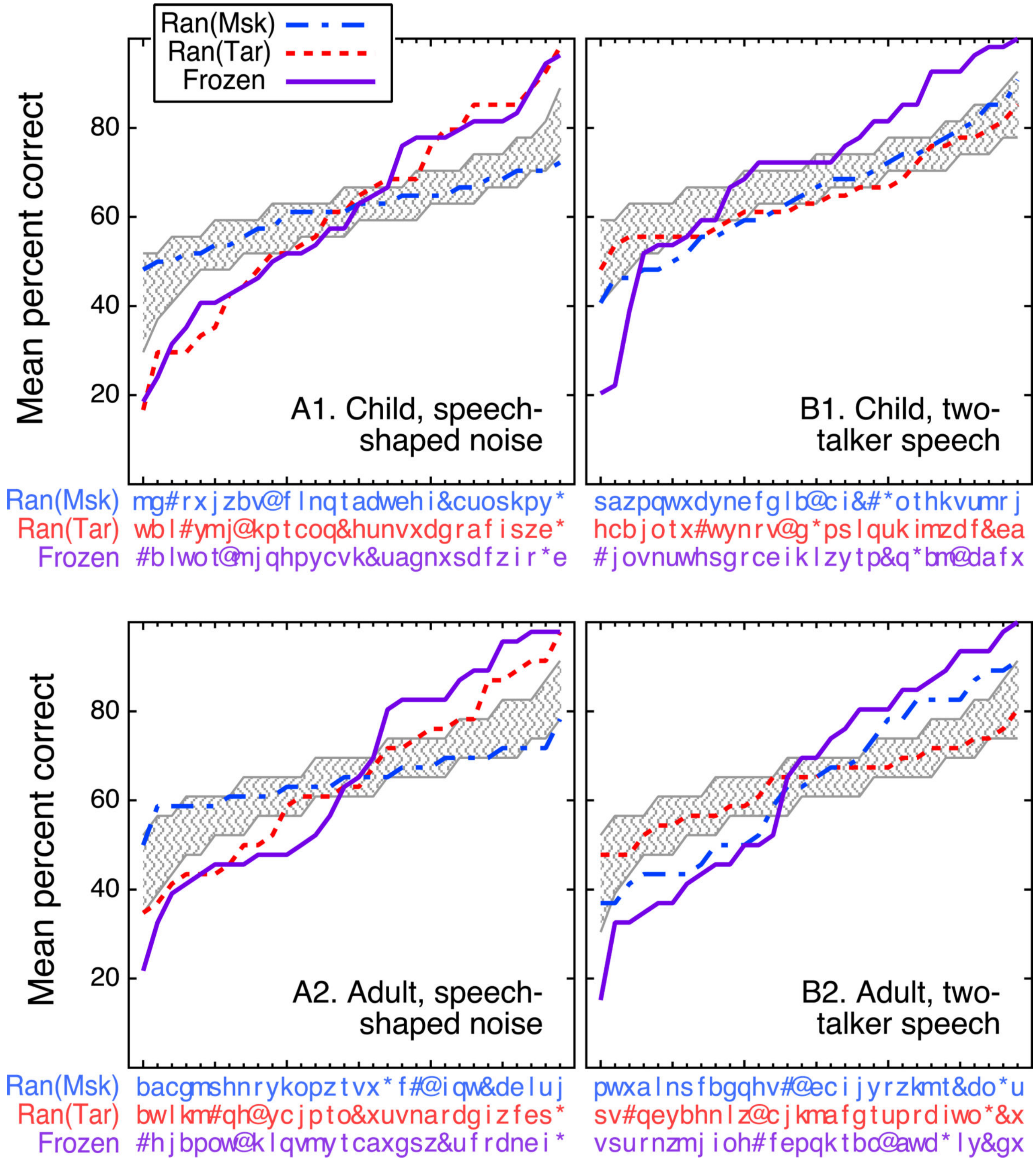

Figure 2.

Group mean percent correct by stimulus sample, sorted from low to high. Results for the random condition were evaluated twice: once relative to the masker sample [Ran(Msk)] and once relative to the target [Ran(Tar)]. Results for the frozen condition were evaluated once, by target and masker [Frozen]. The grey hatched region indicates the 95% CI around chance performance for each age group and masker type. Letters and symbols on the abscissa identify target samples in each ordinal position (see Table 1).

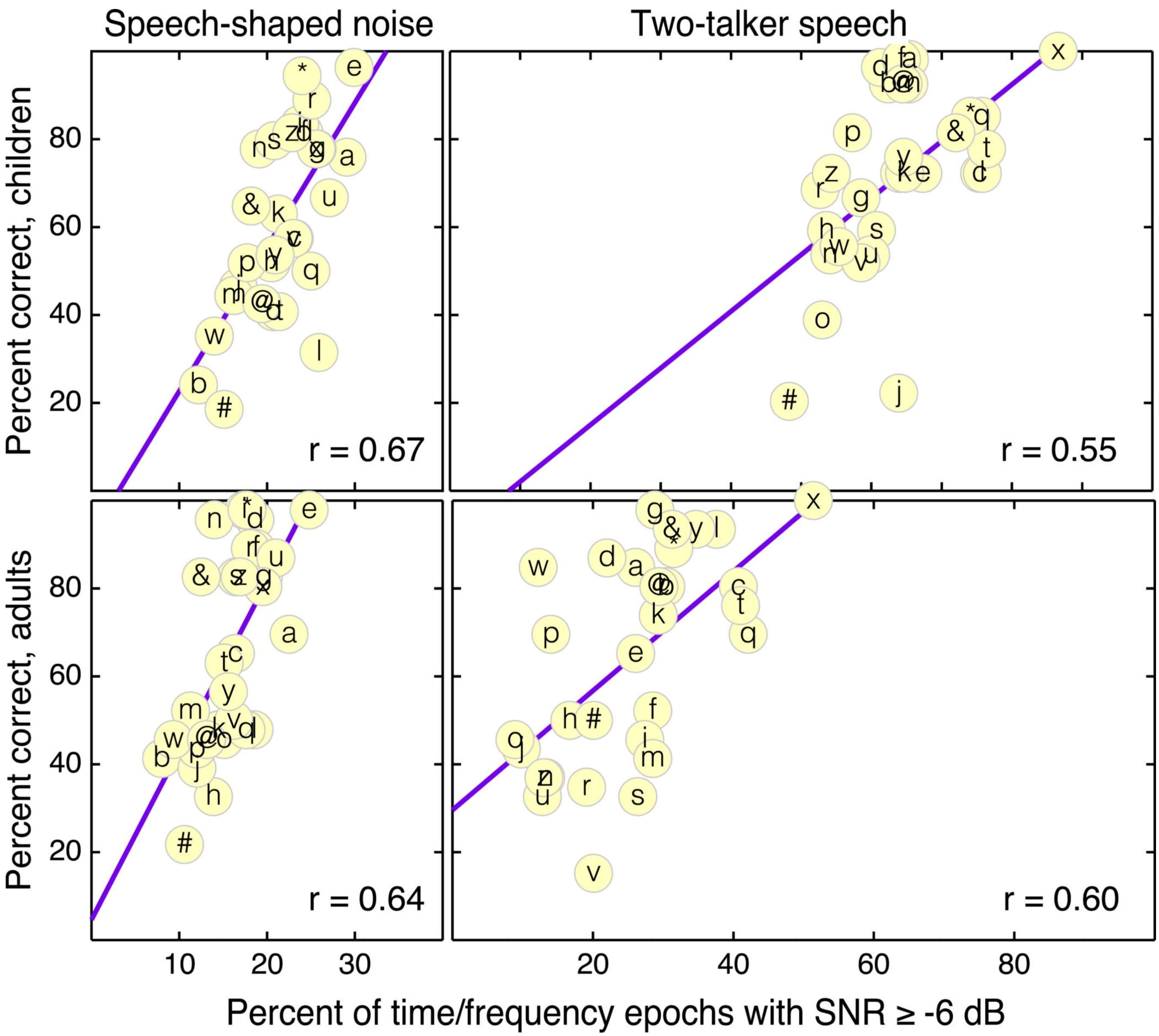

Figure 3.

Percent correct for each stimulus in the frozen condition, plotted as a function of percent audibility of the target. Results for children appear in the top row, and those for adults appear in the bottom row. The correlation between group mean percent correct and percent masked audibility is reported in the lower right of each panel. Letters and symbols identify target and masker samples (see Table 1). Lines indicate best fits to the data.

Table 1:

Symbols in the first column are used to identify stimuli in Figures 2 and 3. The correspondence of targets and maskers reflect stimulus pairings in the *frozen* condition. Bolded segments of each masker stream indicate the temporal overlap with the target word. Letters in parentheses indicate portions of words that were cut off at the beginning and end of the listening interval.

Symbol	Target	Masker
a	baby	...a little bit of this an**d a little bit** of that. Jack lived w(ith)... ...Daisy. At least we’ve got **something to** eat…well, we wi(ll)…
b	balloon	“...(Ja)ck,” his mom sai**d. “We’ll have** to sell poor ol(d)...” (wriggl)ed deep into the earth an**d shoots** pushed upward. They bur(st)...
c	button	...(ac)ross the room and drew b**ack his curt**ains. There, be(nding)... “Good morning to you!” said the ma**n. “Tha**t’s a nice-looking cow you ha(ve).”
d	candy	...castle. Jack wal**ked straight** up to it and knocked on... ...was nothing Jack lov**ed better than** magic, so he han(ded)...
e	chicken	...started to climb. So**on the house** was just a tiny... “Those?” asked **Jack**. “Yes,” said the funny little...
f	children	...was only one way to fi**nd out. With**out stopping to think... ...funny little ma**n and, plunging** a hand deep into o(ne)...
g	doctor	...(up)ward. Finally, he **reached the la**nd of the clouds... ...beans. Oh no! **These are** magic beans.
h	dolphin	...was lazy. When there w**as an advent**ure in the offing, he was... “What’s this?” ex**claimed his mo**m. “Oh dear,” though(t)...”
i	dragon	...not lazy at all. **But most of the** time, he just did a... ...(t)hought Jack. “They’re magic **beans, Mo**m.” I swapped them for Dai(sy).
j	elbow	...to do a little **bit of this** and a little bit of that Then she threw open **the window** and flung the bean(s)...
k	feather	She went white in the fa**ce and** shouted and stamped ...(f)armhouse, a little way **out of to**wn. Jack’s mom like(d)...
l	flower	...not even a crus**t of old br**ead, and no money le(ft)... ...garden, things began **to happen**. The beans slipped dow(n)...
m	garden	...ground. In the dis**tance, h**e could see a huge c(astle)... ... for them, so I’m not **sure what th**ey do.”
n	hanger	...her!” Jack knew b**etter than to ar**gue. Besides,... They kept on grow**ing and growing** until they reached the...

Symbol	Target	Masker
o	lemon	...(b)ending and swaying in the moo**nlight, w**as the most enormous... ...there. Do you fan**cy doing a** swap for her?" Jack
p	lion	...much are. Then o**ne day th**ere was nothing left to eat... ...bed feeling misera**ble and hun**gry. But in the gar(den)...
q	monkey	...his mom and Dais**y the cow**, in a tumbledown far(mhouse)... ...when they've grown." **Jack's** mom was furious.
r	monster	...as well. They didn't ha**ve very much** money, but they didn't m(uch)... ...outside. That n**ight, Jack a**nd his mom went to be(d)...
s	necklace	...twice, he clam**bered over the** windowsill and star(ted)... ...one of his pockets, h**e pulled out** six plump bean(s).
t	oven	...stepped off the beansta**lk on**to the fluffy gray... ...will have to be careful with the**m. I've** lost the instructions...
u	paper	...old Daisy. You had **better get up** early tomorrow morning... ...burst through the hard cru**st of the soil** and, twisting...
v	pencil	(t)ow. He had not gone **far** when he came around a cor(ner)... ....bedroom window. "**Who's that**?" Jack yawned.
w	ruler	...the door. I'm not g**oing to start b**y saying that Jack wa(s)... ...(hand)ed over Daisy, t**ook the bean**s and hurried home
x	sweater	...left either to buy **anything**. "It's no good, Ja(ck),"... ...down through the cracks in **the ground**. Their roots wriggle(d)...
y	table	...and take her to market. **Make su**re you get a good price for... ...and tangling togeth**er, they grew** high into the sky.
z	tiger	...sunrise and set **off down** the lane with Daisy in to(w.) ...tendril reached down **to the hou**se and tapped on Ja(ck's)...
*	turkey	"...(t)op of it goes to?" **Jack** said to himself. There w(as).... ..."What will you give me in exch**ange**?". "These!" decla(red)...
&	water	...beanstalk he had e**ver seen**. "I wonder where the to(p)... ...remembered what his mom **had told hi**m, so he asked...
#	woman	...(tin)y dot, far below. **Still** he made his way up(ward) ...man, "these! D**on't think th**ese are just ordinary be(ans).
@	zebra	(Beside)s, he was very hungry. **So th**e next day he got up at... ...land of the clouds. **Then** a long, wiry te(ndril)...