Developmental Changes in Speech Frequency Weighting in Children

A. K. Bosen,[1] A. Frenette,[2] M. Spratford, D. Lewis,[1] E. A. Walker,[3] and R. W. McCreery[1]

Objectives: Estimates of speech intelligibility depend on the audibility of spectral bands in speech and the importance weight of each spectral band for speech recognition. In children, the transfer function that links estimates of speech intelligibility to recognition accuracy changes with age, indicating the need to incorporate developmental changes into clinical use of speech intelligibility metrics for children. Previous research has shown developmental shifts in how children use spectral cues to support speech recognition, but individual differences in band importance that form the basis for models of speech recognition have not been directly validated in children. The goal of the present study was to determine if children differ in band importance weights with age and, if such differences exist, determine how they relate to developmental changes in vocabulary and cognitive ability.

Design: Children (ages 6 to 13 years) with normal hearing (N = 42) completed a filtered word recognition task and a battery of vocabulary and cognitive assessments. Monosyllabic words and nonword auditory stimuli were filtered into six frequency bands that were centered on octave frequencies between 250 and 8000 Hz, each one octave wide. Each stimulus contained a pseudorandom combination of three out of six frequency bands, which children listened to and verbally repeated. Recognition accuracy was regressed against the presence or absence of each band across trials in a generalized linear mixed effects model to measure the contribution of each frequency band to speech recognition on an individual basis. Children also completed the Peabody Picture Vocabulary Test, the Wechsler Abbreviated Scale of Intelligence vocabulary task, the Psychology Experiment Building Language flanker and switcher tasks, and the Wechsler Abbreviated Scale of Intelligence matrix reasoning task.

Results: Changes in frequency importance weights with age were observed for bands centered on 2000 and 4000 Hz, with increasing weights for these bands in older participants. Frequency bands centered on 500, 1000, and 2000 Hz had higher weights for words than for nonwords. Interactions between age and stimulus type were not evident. None of the vocabulary and cognitive tasks predicted frequency importance weights after controlling for age. Flanker task performance was a significant predictor of recognition accuracy.

Conclusions: Frequency band importance weights undergo developmental changes in children. Models used to predict speech recognition in children may need to be updated to account for these developmental changes; otherwise, speech intelligibility estimates will likely overestimate speech recognition accuracy. Additional work is needed to determine whether the developmental trajectory of frequency importance weights is affected by hearing loss in children who are hard of hearing.

Key words: Band importance, Child development, Speech recognition.

(Ear & Hearing 2026;47;548–557)

This article has received OSF badges for Open Data and Open Material.

[1]Boys Town National Research Hospital, Omaha, Nebraska, USA; [2]University of Maryland, Baltimore, Maryland; and [3]University of Iowa, Iowa City, Iowa.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site (www.ear-hearing.com).

INTRODUCTION

The ability to understand speech in degraded conditions undergoes substantial developmental changes in children through adolescence. Children require higher levels of audibility to reach speech recognition accuracy comparable to that of adults (Stelmachowicz et al. 2000; Scollie 2008; McCreery & Stelmachowicz 2011). These developmental differences also manifest in how children weight different frequency bands during recognition tasks involving acoustically degraded speech (Nittrouer & Miller 1997; Eisenberg et al. 2000; Mlot et al. 2010; McCreery & Stelmachowicz 2011). Developmental differences in frequency weighting can be used to make predictions about speech recognition in noise (McCreery et al. 2017), hearing aid candidacy (McCreery et al. 2020), and hearing aid prescription for children with hearing loss (Scollie 2008). Current audibility-based models of speech perception used with children do not reflect age-based variability in frequency band importance weighting. In addition, the specific cognitive and linguistic mechanisms underlying frequency-specific changes in degraded speech recognition in children remain unclear. The goal of the present study was to determine if individual differences in the frequency band importance change with age, and if those age-related changes are related to individual differences in cognitive and linguistic abilities.

Much of the previous literature examining the development of degraded speech recognition has focused on age-related differences in recognition for the same level of audibility (Scollie 2008) or the same SNR (Leibold & Buss 2013; Corbin et al. 2016; Goldsworthy & Markle 2019). A more fundamental question is whether there are frequency-specific differences that underlie the maturation of speech recognition in children. During early childhood, infants attend to auditory signals across a broad frequency range in a nonselective manner that increases their susceptibility to masking compared with adults (Leibold & Werner 2006). As children develop and their experience with language increases, they learn to selectively attend to frequency regions that are informative for understanding speech. Nittrouer described the Developmental Weighting Shift (Nittrouer & Studdert-Kennedy 1987; Nittrouer 1996), where the acoustic cues that listeners use for speech recognition change throughout childhood based on increasing experience and proficiency with language. In these studies, the acoustic characteristics of specific speech cues are manipulated to examine how listeners of different ages use these cues to support speech recognition. Similarly, 5-year-old children show greater susceptibility to a masker signal during speech recognition that is spectrally remote from the target speech than 7-year-old children (Youngdahl et al. 2018), showing that the ability to selectively attend to frequency regions that are important for speech understanding continues to develop through school-age. Experiments such as these have been informative for understanding developmental changes in

0196/0202/2026/472-548/0 • Ear & Hearing • Copyright © 2025 The American Auditory Society

<zdoi; 10.1097/AUD.0000000000001749>

masked speech perception. However, these experiments have focused on manipulating specific acoustic cues to phonemic identity that were not constrained to a specific frequency range or masking signals that do not overlap with the target talker. Thus, observed differences in cue weighting between children and adults for specific speech sounds may not directly translate into more general audibility-based models for predicting speech recognition, such as the Speech Intelligibility Index (SII; ANSI S3.5 1997a). Rather than using specific speech contrasts, the SII predicts speech recognition based on the level of audibility and band importance for different frequency bands within the speech signal.

Early speech perception research showed that the distribution of acoustic cues needed for speech recognition is not uniform across the range of the speech spectrum (French & Steinberg 1947). Frequency band importance is most often estimated by measuring the change in speech recognition accuracy when a specific band has been removed from the stimulus using filtering (Studebaker & Sherbecoe 1991; McCreery & Stelmachowicz 2011; Healy et al. 2013) or by correlating speech recognition accuracy in noise with the signal to noise ratio (SNR) in each frequency band across a large number of trials where the SNR by band is varied (Calandruccio & Doherty 2007; Buss & Bosen 2021). Although differences in band importance have been observed across studies, frequencies around 2000 Hz typically have the highest importance for monosyllabic words or sentence stimuli. Band importance decreases for bands above or below 2000 Hz. Using the SII, band importance weights are used with measures of speech audibility to predict speech recognition. Speech audibility is calculated by measuring the intensity level of speech within each frequency band relative to the listener's threshold or hearing level to estimate band-specific audibility. Audibility in each frequency band is multiplied by each band's importance and then summed across bands to produce a metric of intelligibility (ANSI S3.5 1997a). Intelligibility is then transformed to a prediction of speech recognition accuracy via a nonlinear transfer function. Frequency importance weights and transfer functions must be experimentally measured, and predicting speech recognition depends on band importance and transfer functions that can be generalized across individuals. Such generalization can be problematic because frequency band importance and transfer functions vary across stimuli (Pavlovic 1994; Apoux & Healy 2012; Healy et al. 2013), listening condition (Buss & Bosen 2021; Shen & Langley 2023; Bosen et al. 2024), and between children and adults (Stelmachowicz et al. 2000; Scollie 2008; McCreery & Stelmachowicz 2011). These differences may be particularly important because the SII plays a fundamental role in audibility-based hearing aid candidacy (McCreery et al. 2020), pediatric hearing aid prescription (Scollie et al. 2005), and as a method of quantifying auditory experience (Tomblin et al. 2015; Wiseman et al. 2023) in children with hearing loss who use hearing aids. Current applications of the SII do not incorporate age-related differences in the frequency-band importance functions that form the basis of the mathematical calculations of audibility.

Quantifying developmental changes in frequency band importance could improve predictions of degraded speech recognition for children, but methodological limitations related to measuring band importance with children have introduced confounds that could obscure developmental differences in weights. For example, McCreery and Stelmachowicz (2011) measured band importance functions for a group of young adults and a group of 5 to 12-year-old children using nonword consonant-vowel-consonant stimuli. Nonword recognition was measured at four SNRs and seven filter conditions with all octave bands present or progressive amounts of high- and low-pass filtering to remove specific octave bands. Due to the large number of conditions, each child completed two SNRs for each of the seven filter conditions. The data were then combined with those of an age-matched child who completed the other two SNRs at the same filter conditions. Overall, nonword recognition increased as age increased, as predicted, but individual differences in band importance weighting were not observed between children and adults. However, children with higher expressive language abilities had better nonword recognition after controlling for age. It is possible that the methodological decision to combine data across children could have obscured individual differences in band importance weights across age groups.

Despite some preliminary evidence that band importance weights do not systematically vary by age, some evidence from the developmental speech perception literature supports the notion that frequency importance weights could shift with development. As noted earlier, the developmental changes in speech cue weighting (Nittrouer & Studdert-Kennedy 1987; Nittrouer 1996) suggest that children may change the cues that they use to recognize speech as they age. Thus, band importance weights should change with age as the acoustic cues that children use to identify phonemes change through adolescence (Nittrouer & Miller 1997). Specifically, the sharpness of categorical perception boundaries between phonemes (Nittrouer & Studdert-Kennedy 1987; Nittrouer 1992, 2002; McMurray et al. 2018) and the flexibility with which acoustic cues are integrated (Hazan & Barrett 2000) increase with age. These changes in phonetic cue use are driven by language experience and knowledge, specifically related to increasing phonological awareness and vocabulary (Mayo et al. 2003; Walley et al. 2003). Changes in acoustic cue weighting that reflect these changes in sensitivity to phonological and lexical cues in the stimuli should manifest in a change in band importance during childhood, where band importance shifts to become more adult-like as children get older. Such changes in frequency band importance should reflect the changes in acoustic cues weighting with age.

Recent methodological advancements make it possible to estimate individual differences in band importance weights in adults with normal hearing and with hearing loss in fewer stimulus trials and much less time than in previous studies (Bosen & Chatterjee 2016; Shen & Kern 2018; Yoho & Bosen 2019; Shen et al. 2020). However, these novel experimental methods for measuring band importance have not been tested with children. With this method, speech stimuli are divided into discrete frequency bands, and a random combination of bands is filtered out of each stimulus. Band combinations are selected to balance the frequency with which bands are presented and co-occur with one another. This approach accounts for synergistic interactions across frequency bands (Warren et al. 2005; Gustafson & Pittman 2011; Humes & Kidd 2016) in a manner that the conventional subtractive approach of progressive high- and low-pass filtering speech stimuli does not. Stimulus recognition accuracy is regressed against the presence or absence of each frequency band in each stimulus, which provides an estimate of the fixed effect of each frequency band on recognition accuracy averaged across its synergy with all other bands. The fixed effect of each frequency band is interpreted as the importance of that band for speech recognition. We predicted that this method would enable us to reliably estimate individual differences in frequency importance weights across children and across age.

Developmental changes in band importance weights likely reflect the maturation of linguistic and cognitive skills during childhood. As described previously, the ability to extract speech information from degraded stimuli improves with age in children with normal hearing. In the present study, we focused on measures of vocabulary, attentional control, and fluid intelligence as predictors of age-related changes in frequency band importance weights, accounting for the general effects of chronological age. Vocabulary was measured because vocabulary growth facilitates the development of phonological awareness (Walley et al. 2003), which in turn drives the development of awareness and flexible use of phonetic cues in acoustic signals (Hazan & Barrett 2000; Mayo et al. 2003). Because frequency importance weights have been shown to differ across words and nonwords in adult listeners (Pavlovic 1994; ANSI 1997b), it is possible that individual differences in lexical and phonological knowledge play a role in band importance weights in children. We anticipate that developmental changes in lexical knowledge (Storkel 2002) and phonological awareness (Nittrouer et al. 2014) will lead to higher accuracy for degraded words than nonwords. Language abilities also may differentially affect importance weights for words and nonwords, such that children with larger vocabularies are predicted to have more adult-like band importance functions than children with smaller vocabularies. Children also demonstrate protracted development of the ability to selectively attend to informative auditory cues in adverse listening conditions (Tarr & Nittrouer 2013; Buss et al. 2018). We measured attentional control (Eriksen & Ericksen 1974) to determine whether individual differences in attentional control govern a child's ability to flexibly attend to speech cues in available frequency bands. We predicted that children with more mature attentional control would have band importance functions that were more adult-like, reflecting greater ability to allocate attention selectively to informative frequency regions during a degraded speech recognition task. Nonverbal cognitive ability also affects performance on a variety of auditory tasks (Kovacs & Conway 2016). We predicted that changes in frequency band importance weight with age would be associated with improvements in vocabulary, selective attention, and nonverbal cognitive ability after controlling for age.

MATERIALS AND METHODS

Children ranging in age from 6 to 13 years completed filtered word and nonword recognition to measure changes in band importance weighting across stimuli and with age. They also completed a set of cognitive tasks selected to measure attentional control, fluid intelligence, and vocabulary to determine if these factors affected frequency importance. Data and analysis scripts are available as supplemental material through the Open Science Framework at https://osf.io/2fzh5/.

Participants

Forty-two typically developing children (24 male, 18 female) with normal hearing between the ages of 6 and 13 years participated in this study. Participant age was roughly uniformly distributed, with between 4 and 7 participants in each 1-year interval within this age range. Written or verbal assent and parental consent to participate were obtained for each child, and they were compensated monetarily for participation. The participant's hearing was screened using otoscopy, tympanometry, and audiometry. Audiometric thresholds were measured using a GSI-61 audiometer and ER-3A insert headphones in a sound-attenuated booth in both ears at frequencies of 250, 500, 1000, 2000, 4000, and 8000 Hz.

Frequency Importance Weighting

To estimate frequency importance weights, children listened to and repeated monosyllabic words and nonwords first described by McCreery and Stelmachowicz (2011) that were filtered to remove portions of the spectrum in each word. Consonant–vowel–consonant (CVC) nonword stimuli were created by taking all possible combinations of CVC using the consonants/b/,/ʧ/,/d/,/f/,/g/,/h/,/ʤ/,/k/,/m/,/n/,/p/,/s/,/ʃ/,/t/,/θ/,/ð/,/v/,/z/, and/ʒ/ and the vowels/a/,/i/,/I/,/ɛ/,/u/,/ʊ/, and/ʌ/. The resulting CVC combinations were entered into an online database based on the Child Mental Lexicon (Storkel & Hoover 2010) to identify all of the CVC stimuli that were real words likely to be within a child's lexicon and to calculate the phonotactic probability of each nonword using the biphone sum of the CV and VC segments. Any combination of phonemes that contained a biphone combination that was illegal in English (biphone sum phonotactic probability = 0) was eliminated. Review of the remaining CVCs was completed to remove slang words and proper nouns that were not identified by the calculator. To create a set of word and nonword stimuli with average phonotactic probabilities, the mean and SD of the biphone sum for the entire set were calculated. To limit the variability in speech understanding across age groups, the 735 CVC nonwords and 735 words with phonotactic probability within ±0.5 SD from the mean of the entire set were used for the present experiment. Stimuli were recorded for two female talkers at a rate of 44.1 kHz. Each stimulus was filtered as originally described by Bosen and Chatterjee (2016) to include three out of six one-octave-wide bands. The six one-octave bands were centered at audiometric frequencies, with filter edges placed one half-octave between band centers (edge frequencies are given in parentheses): 250 Hz (177 to 354 Hz), 500 Hz (354 to 707 Hz), 1000 Hz (707 to 1414 Hz), 2000 Hz (1414 to 2828 Hz), 4000 Hz (2828 to 5660 Hz), and 8000 Hz (5660 to 11314 Hz). Stimuli were filtered in MATLAB (Mathworks, Natick, MA) with a bandpass filter of order 2000 using the filtfilt function, then the output of the three filters was summed to produce the final stimulus. Stimuli were presented monaurally to an ear selected at random (17 left ears, 25 right ears) with an average presentation level of 60 dBA over headphones in a sound-attenuated booth.

Choosing three out of six bands for each stimulus yields a total of 20 possible combinations of bands. Speech intelligibility is affected by interactions between bands (Müsch & Buus 2001; Warren et al. 2005; Humes & Kidd 2016), so we chose to include all 20 band combinations to average estimates of band importance weights across those interactions. Each band combination was used 12 times each for words and nonwords, resulting in 240 filtered words and nonwords each. The assignment of band combinations to target words and nonwords was selected to distribute the frequency of each phoneme occurring with each band combination approximately evenly. The assignment of band combination to target word or nonword, and the order in which stimuli were presented, were fixed across participants. In addition to the filtered items, an unfiltered baseline condition was included that used 40 words and 40 nonwords that included all six bands in the stimulus. Words and nonwords were presented in separate blocks of 40 stimuli per block, for a total of one baseline and six filtered blocks each. Participants were given the opportunity to take a break between each block to manage fatigue.

Playback of each stimulus was initiated by participants via a mouse click on a computer, and their verbal responses were recorded for offline scoring. Responses were scored based on whole-item identification accuracy, with responses marked correct if every phoneme in the response matched the stimulus and incorrect otherwise.

Language and Cognitive Measures

Vocabulary was assessed via the Peabody Picture Vocabulary Test (PPVT-5; Dunn & Dunn 1997) and the Wechsler Abbreviated Scale of Intelligence vocabulary task. In the PPVT, participants are shown four pictures and asked to point to the picture that corresponds to a target word (e.g., "show me emoji"). Words were presented in a fixed order of increasing difficulty, starting from an age-appropriate point in the list (e.g., children between the ages of 6 and 7 years started with the word "full"; children between the ages of 7 and 8 years started with the word "package"). Testing was discontinued when the participant gave six incorrect responses in a row. Raw scores were calculated as the last item in the task they completed minus the number of incorrect responses, and raw scores were then used to calculate norm-referenced standard scores.

In the WASI vocabulary task, participants were asked to provide explanations of target words (e.g., "shirt"; McCrimmon & Smith 2013). Words were presented in a fixed order of increasing difficulty. Responses were scored on a 0- to 2-point scale, from 0 points for an incorrect response to 2 points for a complete definition. Testing was discontinued after participants gave three consecutive responses that received scores of zero, or after children aged 8 or younger reached a stopping point based on age as defined by the task instructions. Raw scores were calculated as the sum score across all responses the participant gave, and raw scores were then used to calculate norm-referenced *t*-scores.

Because the PPVT and WASI vocabulary tasks assess vocabulary, norm-referenced scores were averaged to produce a composite vocabulary score. Three participants did not complete the WASI vocabulary task due to time constraints, so for these participants, composite vocabulary scores were equal to the PPVT scores.

Attentional control was assessed via the Psychology Experiment Building Language (Mueller & Piper 2014) flanker and switcher tasks. In the flanker task, a row of arrows is displayed on a computer screen, and the participant is instructed to press the left shift or right shift key on a keyboard that corresponds to the direction that the arrow in the center of the row is facing. The center arrow is flanked by two additional arrows that either face the same direction as the center arrow (the "congruent" condition) or face the other direction (the "incongruent" condition). Response time and accuracy are automatically recorded by the Psychology Experiment Building Language software. Participants completed 12 practice trials and 40 experimental trials, with an equal number of trials that had a left and right target direction and were congruent or incongruent. The order of trials was randomized across participants, so sequential effects on response time (Davelaar 2013) were not controlled for in this task design. Accuracy was calculated as the proportion of trials in which the participant's response matched the center arrow direction. Response times and the slowdown in the incongruent condition were estimated as random effects in a generalized linear mixed effects model with a lognormal link function.

In the switcher task, which measures the costs of shifting attention, 10 symbols are shown on a computer screen that comprise 1 of 5 colors, letters, and shapes, with each unique color, letter, and shape appearing in two symbols. In each trial of the task, one symbol is selected as the starting symbol, and the participant clicks on successive symbols based on one of three matching rules ("color," "letter," or "shape") displayed at the top of the screen. Twelve rules were presented in each trial. Participants completed nine trials total, of which three alternated between two rules, three sequentially cycled through all three rules in a fixed order, and three switched between rules at random. Total errors were calculated as the total number of times the participant clicked on a symbol that did not match the prior symbol by the displayed rule. Total time was calculated as the cumulative time the participant spent completing all nine trials.

Nonverbal cognitive ability was assessed via the Wechsler Abbreviated Scale of Intelligence matrix reasoning task. In the matrix reasoning task, participants are shown an image of an incomplete pattern matrix and asked to choose from among five options the one that completes the pattern (McCrimmon & Smith 2013). Matrices were presented in a fixed order of increasing difficulty. Testing was discontinued after the participant gave three incorrect responses in a row, or until they reached a stopping point based on age as defined by the task instructions. Raw scores were calculated as the total number of correct responses the participant gave, and raw scores were then used to calculate norm-referenced *t*-scores.

Procedure

Experimental sessions began by administering informed consent, followed by audiometric screening. The cognitive tasks were administered first, in the order they are listed earlier, followed by the frequency band importance weighting task. Experimental sessions lasted approximately 3 h with breaks to minimize fatigue. Participants were compensated $15/h and were offered a prize and a book at the completion of the session. This study was approved by our institutional review board.

RESULTS

Frequency Importance Weights Change Across Stimuli and Age

Figure 1 shows the distribution of speech recognition accuracy across experimental conditions. As expected, baseline word recognition accuracy was high in most participants (≥95% in 38 of 42 participants), and baseline nonword recognition accuracy was lower (≥90% in 23 of 42 participants).

Filtering the words to only include three of the six octave frequency bands reduced recognition accuracy. Recognition accuracy for filtered nonwords and words increased with age (simple correlations, r = 0.62 for nonwords and r = 0.68 for words).

The weight of each frequency band was estimated on an individual basis by regressing recognition accuracy against the presence or absence of each band in a stimulus. A generalized linear fixed effects model with a logistic link function was fit to baseline and filtered trials for each participant for nonwords and words. To ensure results could be compared across participants, an intercept term was not included so that importance weights would be relative to a fixed value of 0. Figure 2 shows frequency importance weights for each participant for nonwords and words. Reliability of importance weights for each band was calculated using Monte Carlo resampling reliability (Pronk et al. 2022), which is described in detail in Supplemental Digital Content 1, https://links.lww.com/EANDH/B761. To summarize, individual differences in the importance weight of the 2000 and 4000 Hz bands had good reliability across individuals, and reliability was moderate for the remaining bands. Good reliability of individual differences in importance weights enables subsequent analysis of factors that affect importance weights across children.

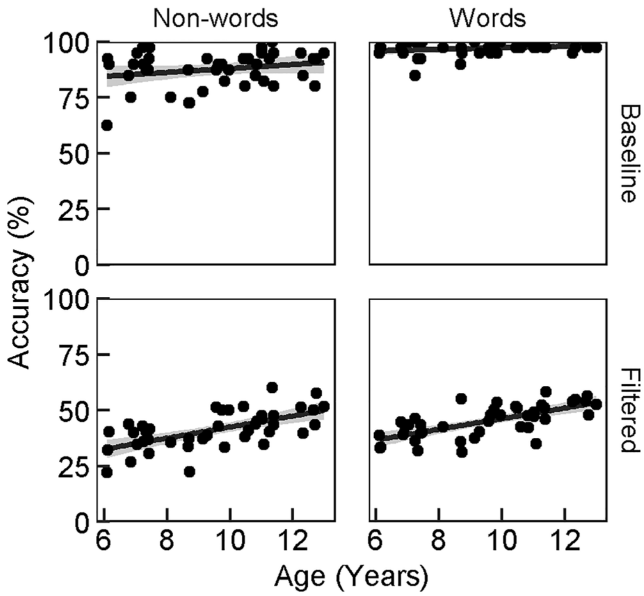

Fig. 1. Recognition accuracy for baseline and filtered nonwords and words as a function of age. Each point shows age and accuracy for one participant. Lines and shaded areas show linear regression best fits and 95% confidence intervals within each panel.

To determine if developmental differences and stimulus differences in frequency band importance weighting exist, standardized linear regressions were used to predict weights for each frequency as a function of stimulus type and participant age. These regressions showed mixed effects across frequency bands. The importance weights of the 250 and 8000 Hz bands were lower for words than for nonwords, while the importance of the 500, 1000, and 2000 Hz bands were higher ($p < 0.05$; regression coefficients for significant effects are shown in Fig. 2). For age, the importance weight of the 250 Hz band decreased with increasing age and the importance of the 2000 and 4000 Hz bands increased with increasing age ($p < 0.05$; regression coefficients for significant effects are shown in Fig. 3). Interactions of stimulus and age were not significant. Figure 3 shows frequency band importance weights as a function of age for each frequency band and stimulus type. A supplemental analysis, described in Supplemental Digital Content 2, https://links.lww.com/EANDH/B762, confirmed these differences across stimuli and age by fitting trial-level data with a generalized linear fixed effects model that included interacting fixed effects of band presence/absence, participant age, and stimulus type.

Age and Individual Differences Covary

Table 1 shows descriptive statistics for performance on each individual difference measure used in this study. For the WASI and PPVT tasks, mean group scores were above normative means (50 for WASI, 100 for PPVT), but the range of individual scores was fairly wide, with scores spanning a

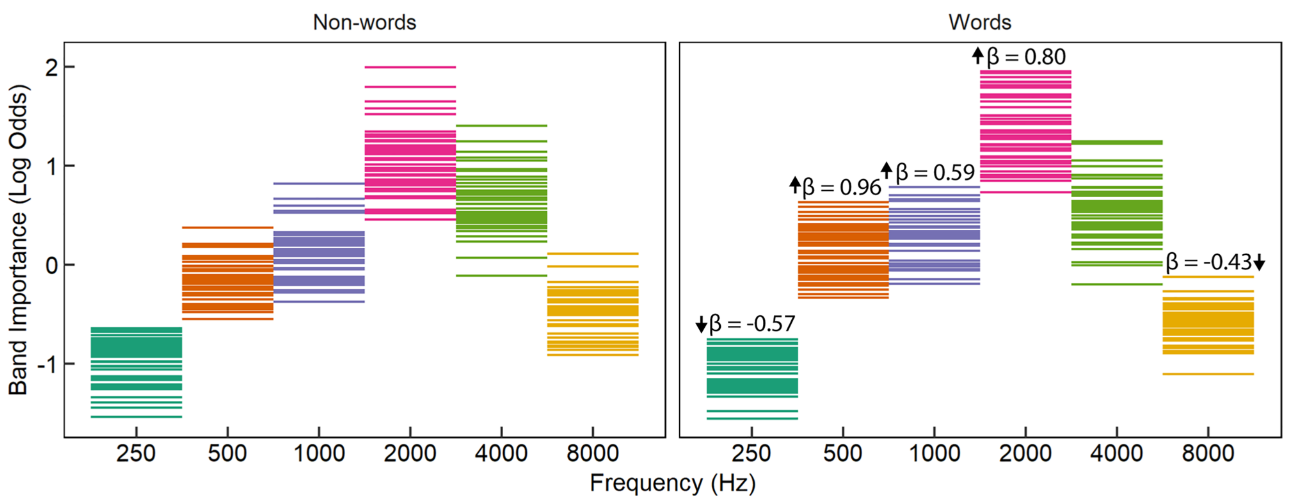

Fig. 2. Frequency importance weights for nonwords and words. Each horizontal line shows the importance weight of a frequency band for one participant. Arrows in the panel for words indicate frequency bands for which importance weights were greater (upward arrow) or lesser (downward arrow) than in the nonword condition, with associated effect sizes reported as standardized regression coefficients.

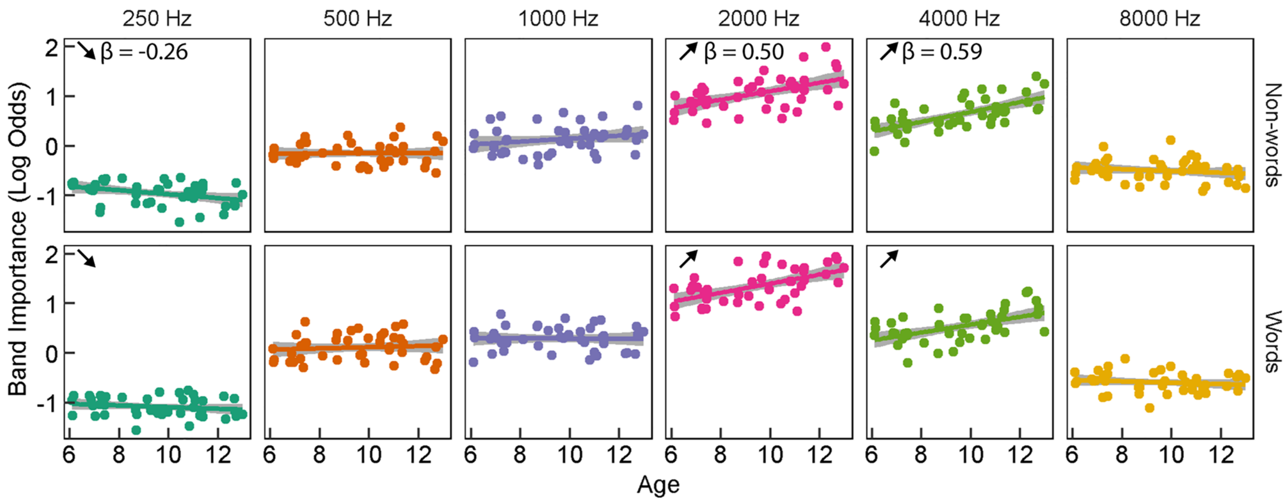

Fig. 3. Importance weights for each frequency band for nonwords and words as a function of participant age. Each point shows the age and importance weight for one participant. Lines and shaded areas show linear regression best fits and 95% confidence intervals within each panel. Arrows indicate frequency bands for which importance weights increased or decreased with increasing age (rising and falling arrows, respectively), with associated effect sizes reported as standardized regression coefficients.

TABLE 1. Range of performance on individual difference measures and Pearson correlation with participant age

Measure	Mean	Range	Age Correlation
WASI matrix reasoning (*t*-score)	53.3	38–76	0.04
WASI vocabulary (*t*-score)	56.3	36–73	0.23
PPVT (standard score)	114.9	84–143	−0.02
Flanker accuracy	89.8%	30–100%	0.43
Flanker response time	789 msec	424–1419 msec	−0.70
Flanker incongruent slowdown	91 msec	−18 to 287 msec	−0.18
Switcher total errors	25 errors	0–229 errors	−0.48
Switcher total time	548 sec	205–1761 sec	−0.80

WASI and PPVT measures were normalized for age, and the remaining measures were not. PPVT, Peabody Picture Vocabulary Test; WASI, Wechsler Abbreviated Scale of Intelligence.

range greater than one standard deviation above and below the mean.

Participants also varied in speed and accuracy in the flanker and switcher tasks. There were developmental changes in accuracy in both tasks. Some children below 10 years of age made a high number of errors, whereas children above this age made few errors. Thus, accuracy might be a useful measure of individual differences in younger children, but it becomes less useful in older children, who have near-perfect accuracy. For the flanker task, a separate analysis script was used to estimate individual differences in response time and slowdown in the incongruent stimulus condition to improve the reliability of estimated individual differences in reaction time differences across flanker congruence conditions (Rouder & Haaf 2019). Supplemental analyses in Supplemental Digital Content 3, https://links.lww.com/EANDH/B763, indicate that the task could reliably estimate individual differences in response speed but had poor ability to estimate individual differences in slowdown due to incongruent stimuli. Accuracy and speed were correlated with participant age in these tasks.

Comparison of Age, Language, and Cognitive Ability as Predictors of Speech Recognition Accuracy

We found developmental changes in frequency importance weights (Fig. 3), which we hypothesized would be associated with the maturation of linguistic and cognitive skills. To test this hypothesis, we compared a set of models that used scores from each task in addition to age to predict word and nonword recognition accuracy in a generalized linear mixed effects model. The goal of this analysis was to test whether including performance on any task provided a better explanation for the data than a model that only included effects of age on accuracy and band importance weights. The reference model, which included only age effects, had fixed effects of stimulus type, the presence or absence of each frequency band, participant age, and pairwise interactions between these terms (see Supplemental Digital Content 2, https://links.lww.com/EANDH/B762). Additional models each added the fixed effects of one task score and its pairwise interactions with stimulus type, band presence, and age. For these models, we normalized measures that were correlated with age (Table 1) by regressing the measure against participant age and used the residuals of the regression as normalized scores. Normalizing measures in this manner avoids introducing variance inflation into these models when adding additional predictor variables. In addition, scores for the WASI Vocabulary and PPVT tasks were combined into a single language composite as described in the methods section. Preliminary analysis also included third and fourth-order interactions between terms, but there was minimal effect of the higher-order interactions on model fits, and so they were subsequently dropped from the analysis reported here. Akaike's Information Criteria (AIC) was calculated for each model and compared across models to determine which individual difference measures improved model fit relative to a model that included only age, as shown in Table 2. Higher values of AIC indicate a worse fit to the data.

Table 2 indicates that adding flanker incongruent slowdown, flanker response time, and vocabulary composite scores to the model improved the model fit (i.e., decreased

TABLE 2. Model comparison using task scores in addition to age as a predictor of word and nonword recognition accuracy

Measure	Model AIC	AIC Difference
Flanker slowdown	25479	–15.5
Flanker response time	25481	–13.5
Vocabulary composite	25484	–10.8
Switcher total errors	25492	–2.8
Switcher total time	25493	–0.6
Age only	25494	–
Flanker accuracy	25496	1.2
WASI matrix reasoning	25500	5.1

Models are sorted by goodness of fit (AIC), with the best-fitting model at the top and the worst at the bottom. The model that did not include task scores (age only) is highlighted and was used as a comparison point for the goodness of fit of other models.
AIC, Akaike's Information Criteria; WASI, Wechsler Abbreviated Scale of Intelligence.

AIC) beyond using age alone. Adding switcher task performance also marginally improved model fits. To test whether these variables overlapped in their explanatory power, an additional model was fit to the data that included the flanker and vocabulary measures. Inspection of this model found that none of these variables had a significant interaction with importance for any band and that flanker slowdown and response time had significant effects, whereas the vocabulary composite did not. The improved model fit from adding flanker slowdown and response time primarily manifested as interactions with stimulus type, with smaller flanker slowdown effects leading to better nonword and word recognition accuracy and faster flanker response times leading to better word recognition accuracy. A summary of this model is provided in Supplemental Digital Content 4, https://links.lww.com/EANDH/B764.

DISCUSSION

In this study, we found distinct effects of stimulus content (words versus nonwords) and age on the band importance weights of octave-wide frequency bands for monosyllabic speech recognition. Flanker task performance further predicted speech recognition accuracy but did not affect the importance weight of specific frequency bands. Our results indicate that there are developmental changes in frequency importance weights that should be incorporated into estimates of speech audibility and audibility-based predictions of speech recognition for children.

Stimulus-Specific Effects on Frequency Importance Weights

Stimulus context had two effects on speech recognition accuracy. First, overall accuracy was greater for words than for nonwords. This effect was expected because lexical constraints limit the number of valid responses when target speech is degraded and thereby enable the listener to use lexical or phonological skills to resolve unclear portions of the target (Nittrouer & Boothroyd 1990; McCreery & Stelmachowicz 2011). Second, the importance weights of specific frequency bands differed across words and nonwords, with increased weight for bands centered on 500, 1000, and 2000 Hz, and decreased weight for bands centered on 250 and 8000 Hz for words relative to nonwords (Fig. 2). The net effect is that the frequency importance function has weights that are higher in the midfrequencies for words than nonwords. Differences in frequency importance function across stimuli indicate that lexical constraints facilitate the selective use of the most informative speech cues to identify target speech, in contrast to more diffuse attention to speech cues across the spectrum when such constraints are not present. The word and nonword stimuli used in this study were designed to have equal phonotactic probabilities, so it seems likely that these effects occur as part of word identification (or the lack thereof). There was also no interaction between stimulus content and age, which indicates that the effect of lexical constraints on band importance develops before 6 years of age, which is the lowest age of participants in this study.

Acoustically, vowel formants and dynamic transitions between vowels and consonants are most evident in the 500 to 2000 Hz bands. These cues are likely to be particularly important for the monosyllabic CVC stimuli used in the present study because in this structure, the static formant cues and the dynamic transition into and out of the vowel all facilitate identification of the vowel (Strange et al. 1983; Nittrouer & Studdert-Kennedy 1987; Nittrouer 1992; Nittrouer & Miller 1997), which in turn provides contextual information for identifying the adjacent consonants. This contextual information is present in both words and nonwords, but the addition of lexical constraints for words further enhances the role of the vowel in recognition; if the vowel is correctly identified, then it provides transition and lexical cues to constrain the identification of words. The use of lexical constraints would also explain the decrease in weights for the bands centered on 250 and 8000 Hz in words, as the consonant cues contained in these frequencies would be somewhat redundant with lexical constraints.

Developmental Changes in Frequency Importance Weights

All children exhibited the expected peaked shape in their frequency importance weighting function, but within that overall shape, the frequency importance weights changed from 6 to 13 years of age. Specifically, older children had greater weights for bands centered on 2000 and 4000 Hz than younger children, with an apparently linear increase in weight throughout this age range (Fig. 3). These changes in weights indicate that children learn to attend to speech cues in these frequency bands as they develop, which contributes to increases in overall speech recognition accuracy. Thus, our results extend previous work by demonstrating that developmental weighting shifts in speech perception (Nittrouer & Miller 1997) include changes in average band importance weights. Increases in the weighting of important bands are also consistent with the finding that there are developmental changes in the proficiency with which children use available speech cues (Stelmachowicz et al. 2000; Scollie 2008). These changes likely vary across specific phonemes (Mayo & Turk 2004), although our experiment was not designed to examine phoneme-specific weighting.

The process that children use to recognize speech changes with development. Early in language development, words are acquired as whole units, with awareness of phonetic distinctions arising as vocabulary size grows (Walley 1993). While initial awareness of phonetic categories occurs early in development, children grow more sensitive to fine-grained details in phonetic speech cues through 18 years of age (Rigler et al. 2015; McMurray et al. 2018). Similarly, phonemic awareness precedes changes in speech cue weighting (Mayo et al. 2003). Our work adds to this literature by showing that weighting changes are apparent in specific frequency bands, namely 2000 and 4000 Hz, across speech stimuli with diverse phonetic content. Phonemes contain multiple cues as to their identity, and awareness of secondary cues may develop slowly with experience. This explanation is consistent with weighting changes for specific phonetic cues during development, such as weighting of static spectral cues for fricatives (Nittrouer 1992) or F3 transitions (Nittrouer & Miller 1997), that occur in this frequency range.

Predictors of Speech Recognition Accuracy

In addition to age, individual differences in age-normed vocabulary and flanker task performance accounted for variability in speech recognition accuracy (Table 2). However, these measures were only associated with recognition accuracy and did not affect importance weights for any frequency band, which suggests that none of these tasks assess the mechanism underlying developmental changes in importance weights.

Although we found that age predicts importance weights for the bands centered around 2000 and 4000 Hz, it is likely that age is a proxy for an underlying developmental mechanism that causes these changes in weights in children with normal hearing. As described previously, children develop through experience-driven changes in the process by which they understand speech. Given that these weight changes occurred for both words and nonwords, it seems plausible that the mechanism underlying these changes occurs at the phonological perception level. Individual differences in phonological perception or awareness could drive both age-related changes in band importance weights and more general language development that would affect performance on the vocabulary tasks. However, the vocabulary composite measure did not account for individual differences in importance weights, which suggests that these tasks do not assess the mechanism underlying these differences. More proximal measures of phonological perception are likely required to account for these individual differences.

The association of flanker task performance with speech recognition accuracy suggests a role for individual differences in the inhibition of stimulus-driven conflict in both tasks. The flanker task is specifically designed to elicit such conflicts by presenting conflicting information on some trials that the participant needs to inhibit (Eriksen & Ericksen 1974). Although not intentionally designed to do so, contemporary theories of speech recognition posit that speech recognition tasks also produce competing activation of cohort competitors and thus require inhibition of those competitors (McClelland & Elman 1986; Marslen-Wilson 1987). Inhibition of cohort competitors develops throughout childhood, which could contribute to increases in speech recognition accuracy throughout development (Blomquist and McMurray 2023). Our results found that individual differences in slowdown due to conflict in the flanker task predicted overall recognition accuracy for both words and nonwords. This finding indicates that a modality-general stimulus-driven inhibition ability may determine performance on both tasks. More broadly, some theories of cognition emphasize the ubiquitous role of processing speed in the execution of many tasks (Salthouse 1996). Baseline response time in the flanker task was specifically related to word recognition but not nonword recognition, which suggests that lexical activation may depend in part on cognitive processing speed.

LIMITATIONS AND FUTURE DIRECTIONS

The developmental changes in importance weights that we observed are inconsistent with the finding of McCreery and Stelmachowicz (2011) that band importance weights did not differ between age groups, although several methodological differences could account for this difference. In the present study, the combination of target stimulus and frequency band audibility was fixed across participants, which may facilitate comparison of individual differences in importance weights across participants. In contrast, McCreery and Stelmachowicz pseudorandomly assigned participants to a subset of experimental conditions, so combinations of target stimulus and audibility manipulation varied across participants. It seems likely that the effects of manipulating band audibility depend on the acoustic content of the target speech. For example, words with fricative consonants would have greater high-frequency energy and thus greater high-frequency importance than words with nasal consonants. Thus, random effects of target interactions with audibility could confound the comparison of individual differences across participants if not controlled for. Because their participants only completed a subset of conditions, they could not estimate band importance functions on an individual basis, but rather combined data across participants and conditions to obtain an average importance function for each age group. Combining data in this manner could mask the presence of individual differences in importance weights within groups. In addition, McCreery and Stelmachowicz manipulated audibility by adding steady-state noise at a range of signal to noise ratios to each frequency band, whereas the present study filtered out frequency bands. Additional work is needed to determine the extent to which these methodological decisions affect estimates of importance weights.

Flanker task performance and vocabulary accounted for speech recognition accuracy in addition to participant age, but performance on these tasks did not account for changes in importance weights with age. This finding suggests that our battery of linguistic and cognitive tasks did not include an assessment of the developmental mechanism underlying developmental changes in importance weights. The inclusion of a measure of phonological awareness in future studies may help to more directly assess the role of phonological awareness in changes in frequency importance weights in children for words and nonwords.

Speech recognition and language outcomes tend to be more variable across children with hearing loss (Tomblin et al. 2015), so it is likely that increases in importance weights with age would be more variable and slower on average in children with hearing loss than we observed in children with normal hearing. A variety of developmental outcomes depend in part on auditory dosage in children with hearing loss (McCreery & Walker 2022). If awareness of speech cues in the 2000 and 4000 Hz bands drives developmental changes in importance weights of these bands, then band-specific auditory dosage should govern these developmental changes. Thus, even if a child has high broadband auditory dosage—that is, they use a device that provides high SII for the majority of the day—variability in the audibility of specific bands may still govern access to speech cues in those bands and impair the development of importance weights in bands with poor dosage. As a specific example, children with high-frequency hearing loss are often delayed in their perception (Stelmachowicz et al. 2004; Moeller et al. 2007) and production (Koehlinger et al. 2015) of fricative sounds, although it remains to be tested whether this relationship between audibility and cue awareness within frequency bands generalizes across different configurations of hearing loss. In children who use hearing aids, the auditory sensitivity to discriminate specific speech cues does not predict the weight those cues receive in speech recognition (Lowenstein & Nittrouer 2015; Nittrouer & Lowenstein 2015). Thus, it seems plausible that audibility of or sensitivity to cues in a frequency band puts an upper bound on the amount of information that a listener can extract from that band, while measures of importance reflect the extent to which listeners have learned to use that information in speech recognition.

It seems plausible that developmental changes in frequency importance weights would also occur in languages other than English, although to our knowledge, such changes have not been characterized. Importance weights vary across languages (Du et al. 2019), which indicates that weights reflect the spectral location of speech cues in a language (e.g., the increased importance of low frequencies in tonal languages) and that weights are learned based on the distribution of informative acoustic cues in that language. Future research should explore whether other languages show similar developmental changes in frequency importance from broadly distributed weights toward weighting at specific frequencies that are informative for speech understanding as were observed in English in this study.

CONCLUSIONS

Frequency importance weights undergo developmental changes in children with normal hearing. Metrics of speech audibility, such as the SII, that are based on adult speech perception overestimate the amount of speech information that children have access to and thus overestimate speech recognition accuracy. This problem has been addressed in previous studies through the use of age-based proficiency factors (Scollie 2008), but such factors do not capture individual variability in performance within groups of children who are the same age. Our results demonstrate that it is feasible to estimate importance weights in children using the method described here, which provides an opportunity to update metrics of speech audibility to accurately predict performance in children. Additional work is needed to determine how results obtained using this method relate to established metrics and how developmental changes in importance weights are affected by the configuration of hearing loss in children. Ensuring that clinical best practices for pediatric hearing aid fittings are sufficient to enable these changes in importance weights may facilitate language development in these children.

ACKNOWLEDGMENTS

The authors thank Jacob Oleson for advising preliminary statistical analyses and Johsette Witt for assisting with data collection.

This work was supported by NIH grants T35DC008757, awarded to Kristen Janky and used to support Alyssa Frenette's involvement in this project, and R01 DC013591, awarded to Ryan McCreery. Portions of this work were previously presented at the 2020 American Auditory Society Annual Scientific and Technology Meeting.

A.K.B. implemented the experiment, analyzed and interpreted the data, and led the writing of the manuscript. A.F. led data collection and wrote initial drafts of portions of the manuscript. M.S. supervised data collection and advised data interpretation and manuscript writing. D.L. supported data collection and advised on data interpretation and manuscript writing. E.A.W. advised experiment design, data analysis, interpretation of results, and manuscript writing. R.W.M. designed the experiment, analyzed and interpreted the data, and wrote and revised portions of the manuscript.

The data that support the findings of this study are openly available via Open Science Foundation at https://osf.io/2fzh5/.

The authors have no conflicts of interest to disclose.

Address for correspondence: R. W. McCreery, Boys Town National Research Hospital, 555 North 30th Street, Omaha, NE, USA. E-mail: ryan.mccreery@boystown.org

OPEN PRACTICES

This manuscript qualifies for an Open Data badge and an Open Materials badge. The materials and data have been made publically available at view_only=caedf8ea971642b29421d7dde9c1c5a2 and https://osf.io/kjytg/. More information about the Open Practices Badges can be found at https://journals.lww.com/ear-hearing/pages/default.aspx.

Received February 06, 2025; accepted August 26, 2025; published online ahead of print November 7, 2025

REFERENCES

ANSI (1997a). *S3.5-1997. Methods for Calculation of the Speech Intelligibility Index*. American National Standards Institute.

American National Standards Institute. (1997b). *Methods for Calculation of the Speech Intelligibility Index, S3.5*. Acoustical Society of America.

Apoux, F., & Healy, E. W. (2012). Use of a compound approach to derive auditory-filter-wide frequency-importance functions for vowels and consonants. *J Acoust Soc Am*, *132*, 1078–1087.

Blomquist, C., McMurray, B. (2023). The development of lexical inhibition in spoken word recognition. *Dev Psychol*, *59*, 186–206.

Bosen, A. K., & Chatterjee, M. (2016). Band importance functions of listeners with cochlear implants using clinical maps. *J Acoust Soc Am*, *140*, 3718–3727.

Bosen, A. K., Wasiuk, P. A., Calandruccio, L., Buss, E. (2024). Frequency importance for sentence recognition in co-located noise, co-located speech, and spatially separated speech. *J Acoust Soc Am*, *156*, 3275–3284.

Buss, E., & Bosen, A. (2021). Band importance for speech-in-speech recognition. *JASA Express Lett*, *1*, 084402.

Buss, E., Leibold, L. J., Lorenzi, C. (2018). Speech recognition for school-age children and adults tested in multi-tone vs multi-noise-band maskers. *J Acoust Soc Am*, *143*, 1458–1466.

Calandruccio, L., & Doherty, K. a. (2007). Spectral weighting strategies for sentences measured by a correlational method. *J Acoust Soc Am*, *121*, 3827–3836.

Corbin, N. E., Bonino, A. Y., Buss, E., Leibold, L. J. (2016). Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. *Ear Hear*, *37*, 55–63.

Davelaar, E. J. (2013). When the ignored gets bound: Sequential effects in the flanker task. *Front Psychol*, *3*, 1–13.

Du, Y., Shen, Y., Wu, X., Chen, J. (2019). The effect of speech material on the band importance function for Mandarin Chinese. *J Acoust Soc Am*, *146*, 445–457.

Dunn, D. M. (2019). Peabody Picture Vocabulary Test. *[Measurement instrument]*. (5th ed). Bloomington, MN: NCS Pearson.

Eisenberg, L. S., Shannon, R. V., Schaefer Martinez, A., Wygonski, J., Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *J Acoust Soc Am*, *107*, 2704–2710.

Eriksen, B. A., & Ericksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept Psychophys*, *16*, 143–149.

French, N. R., & Steinberg, J. C. (1947). Factors Governing the Intelligibility of Speech Sounds. *J Acoust Soc Am*, *19*, 90–119.

Goldsworthy, R. L., & Markle, K. L. (2019). Pediatric hearing loss and speech recognition in quiet and in different types of background noise. *J Speech Lang Hear Res*, *62*, 758–767.

Gustafson, S. J., & Pittman, A. L. (2011). Sentence perception in listening conditions having similar speech intelligibility indices. *Int J Audiol*, *50*, 34–40.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6-12. *J Phon*, *28*, 377–396.

Healy, E. W., Yoho, S. E., Apoux, F. (2013). Band importance for sentences and words reexamined. *J Acoust Soc Am*, *133*, 463–473.

Humes, L. E., & Kidd, G. R. (2016). Speech recognition for multiple bands: Implications for the Speech Intelligibility Index. *J Acoust Soc Am*, *140*, 2019–2026.
Koehlinger, K., Van Horne, A. O., Oleson, J., McCreery, R., Moeller, M. P. (2015). The role of sentence position, allomorph, and morpheme type on accurate use of s-related morphemes by children who are hard of hearing. *J Speech Lang Hear Res*, *58*, 396–409.
Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychol Inq*, *27*, 151–177.
Leibold, L. J., & Buss, E. (2013). Children's identification of consonants in a speech shaped noise and two talker masker. *J Speech Lang Hear Res*, *56*, 1144–1155.
Leibold, L. J., & Werner, L. A. (2006). Effect of masker-frequency variability on the detection performance of infants and adults. *J Acoust Soc Am*, *119*, 3960–3970.
Lowenstein, J. H., & Nittrouer, S. (2015). All cues are not created equal: The case for facilitating the acquisition of typical weighting strategies in children with hearing loss. *J Speech Lang Hear Res*, *58*, 466–480.
Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102.
Mayo, C., Scobbie, J. M., Hewlett, N., Waters, D. (2003). The influence of phonemic awareness development on acoustic cue weighting strategies in children's speech perception. *J Speech Lang Hear Res*, *46*, 1184–1196.
Mayo, C., & Turk, A. (2004). Adult–child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *J Acoust Soc Am*, *115*, 3184–3194.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cogn Psychol*, *18*, 1–86.
McCreery, R. W., Brennan, M., Walker, E. A., Spratford, M. (2017). Perceptual implications of level- and frequency-specific deviations from hearing aid prescription in children. *J Am Acad Audiol*, *28*, 861–875.
McCreery, R. W., & Stelmachowicz, P. G. (2011). Audibility-based predictions of speech recognition for children and adults with normal hearing. *J Acoust Soc Am*, *130*, 4070–4081.
McCreery, R. W., & Walker, E. A. (2022). Variation in auditory experience affects language and executive function skills in children who are hard of hearing. *Ear Hear*, *43*, 347–360.
McCreery, R. W., Walker, E. A., Stiles, D. J., Spratford, M., Oleson, J. J., Lewis, D. E. (2020). Audibility-based hearing aid fitting criteria for children with mild bilateral hearing loss. *Lang Speech Hear Serv Sch*, *51*, 55–67.
McCrimmon, A. W., & Smith, A. D. (2013). Review of the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II). *J Psychoeduc Assess*, *31*, 337–341.
McMurray, B., Danelz, A., Rigler, H., Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Dev Psychol*, *54*, 1472–1491.
Mlot, S., Buss, E., Hall, J. W. (2010). Spectral integration and bandwidth effects on speech recognition in school-aged children and adults. *Ear Hear*, *31*, 56–62.
Moeller, M. P., Hoover, B., Putman, C., Arbataitis, K., Bohnenkamp, G., Peterson, B., Wood, S., Lewis, D., Pittman, A., Stelmachowicz, P. (2007). Vocalizations of infants with hearing loss compared with infants with normal hearing: Part I—Phonetic development. *Ear Hear*, *28*, 605–627.
Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL test battery. *J Neurosci Methods*, *222*, 250–259.
Müsch, H., & Buus, S. (2001). Using statistical decision theory to predict speech intelligibility. II. Measurement and prediction of consonant-discrimination performance. *J Acoust Soc Am*, *109*, 2910–2920.
Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *J Phon*, *20*, 351–382.
Nittrouer, S. (1996). Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds. *J Speech Hear Res*, *39*, 278–297.
Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *J Acoust Soc Am*, *112*, 711–719.
Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *J Acoust Soc Am*, *87*, 2705–2715.
Nittrouer, S., & Lowenstein, J. H. (2015). Weighting of acoustic cues to a manner distinction by children with and without hearing loss. *J Speech Lang Hear Res*, *58*, 1077–1092.
Nittrouer, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *J Acoust Soc Am*, *101*, 2253–2266.
Nittrouer, S., Sansom, E., Low, K., Rice, C., Caldwell-Tarr, A. (2014). Language structures used by kindergartners with cochlear implants: Relationship to phonological awareness, lexical knowledge and hearing loss. *Ear Hear*, *35*, 506–518.
Nittrouer, S., & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *J Speech Hear Res*, *30*, 319–329.
Pavlovic, C. V. (1994). Band importance functions for audiological applications. *Ear Hear*, *15*, 100–104.
Pronk, T., Molenaar, D., Wiers, R. W., Murre, J. (2022). Methods to split cognitive task data for estimating split-half reliability: A comprehensive review and systematic assessment. *Psychon Bull Rev*, *29*, 44–54.
Rigler, H., Farris-trimble, A., Greiner, L., Walker, J., Mcmurray, B. (2015). The slow developmental timecourse of real-time spoken word recognition. *Dev Psychol*, *51*, 1690–1703.
Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychon Bull Rev*, *26*, 452–467.
Salthouse, T. A. (1996). The processing speed theory of adult age differences in cognition. *Psychol Rev*, *103*, 403–428.
Scollie, S. D. (2008). Children's speech recognition scores: The speech intelligibility index and proficiency factors for age and hearing level. *Ear Hear*, *29*, 543–556.
Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., Beaulac, S., Pumford, J. (2005). The desired sensation level multistage input/output algorithm. *Trends Amplif*, *9*, 159–197.
Shen, Y., & Kern, A. B. (2018). An analysis of individual differences in recognizing monosyllabic words under the speech intelligibility index framework. *Trends Hear*, *22*, 233121651876177.
Shen, Y., & Langley, L. (2023). Spectral weighting for sentence recognition in steady-state and amplitude-modulated noise. *JASA Express Lett*, *3*, 055202.
Shen, Y., Yun, D., Liu, Y. (2020). Individualized estimation of the Speech Intelligibility Index for short sentences: Test-retest reliability. *J Acoust Soc Am*, *148*, 1647–1661.
Stelmachowicz, P. G., Hoover, B. M., Lewis, D. E., Kortekaas, R. W. L., Pittman, A. L. (2000). The relation between stimulus context, speech audibility, and perception for normal-hearing and hearing-impaired children. *J Speech Lang Hear Res*, *43*, 902–914.
Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., Lewis, D. E., Moeller, M. P. (2004). The importance of high-frequency audibility in the speech and language development of children with hearing loss. *Arch Otolaryngol Head Neck Surg*, *130*, 556–562.
Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *J Child Lang*, *29*, 251–274.
Storkel, H. L., & Hoover, J. R. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behav Res Methods*, *42*, 497–506.
Strange, W., Jenkins, J. J., Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *J Acoust Soc Am*, *74*, 695–705.
Studebaker, G. a, & Sherbecoe, R. L. (1991). Frequency-importance and transfer functions for recorded CID W-22 word lists. *J Speech Hear Res*, *34*, 427–438.
Tarr, E., & Nittrouer, S. (2013). Explaining coherence in coherence masking protection for adults and children. *J Acoust Soc Am*, *133*, 4218–4231.
Tomblin, J. B., Harrison, M., Ambrose, S. E., Walker, E. A., Oleson, J. J., Moeller, M. P. (2015). Language outcomes in young children with mild to severe hearing loss. *Ear Hear*, *36*, 76S–91S.
Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Dev Rev*, *13*, 286–350.
Walley, A. C., Metsala, J. L., Garlock, V. M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Read Writ*, *16*, 5–20.
Warren, R. M., Bashford, J. A., Lenz, P. W. (2005). Intelligibilities of 1-octave rectangular bands spanning the speech spectrum when heard separately and paired. *J Acoust Soc Am*, *118*, 3261–3266.
Wiseman, K. B., McCreery, R. W., Walker, E. A. (2023). Hearing thresholds, speech recognition, and audibility as indicators for modifying intervention in children with hearing aids. *Ear Hear*, *44*, 787–802.
Yoho, S. E., & Bosen, A. K. (2019). Individualized frequency importance functions for listeners with sensorineural hearing loss. *J Acoust Soc Am*, *145*, 822–830.
Youngdahl, C. L., Healy, E. W., Yoho, S. E., Apoux, F., Holt, R. F. (2018). The effect of remote masking on the reception of speech by young school-age children. *J Speech Lang Hear Res*, *61*, 420–427.