Neurophysiological, linguistic, and cognitive predictors of children's ability to perceive speech in noise

Elaine C. Thompson[a,b], Jennifer Krizman[a,b], Travis White-Schwoch[a,b], Trent Nicol[a,b], Ryne Estabrook[c], Nina Kraus[a,b,d,e,f,*]

[a] Auditory Neuroscience Laboratory, Northwestern University, Evanston, IL, USA
[b] Department of Communication Sciences, Northwestern University, Evanston, IL, USA
[c] Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
[d] Institute for Neuroscience, Northwestern University, Evanston, IL, USA
[e] Department of Neurobiology, Northwestern University, Evanston, IL, USA
[f] Department of Otolaryngology, Northwestern University, Chicago, IL, USA

ABSTRACT

Hearing in noisy environments is a complicated task that engages attention, memory, linguistic knowledge, and precise auditory-neurophysiological processing of sound. Accumulating evidence in school-aged children and adults suggests these mechanisms vary with the task's demands. For instance, co-located speech and noise demands a large cognitive load and recruits working memory, while spatially separating speech and noise diminishes this load and draws on alternative skills. Past research has focused on one or two mechanisms underlying speech-in-noise perception in isolation; few studies have considered multiple factors in tandem, or how they interact during critical developmental years. This project sought to test complementary hypotheses involving neurophysiological, cognitive, and linguistic processes supporting speech-in-noise perception in young children under different masking conditions (co-located, spatially separated). Structural equation modeling was used to identify latent constructs and examine their contributions as predictors. Results reveal cognitive and language skills operate as a single factor supporting speech-in-noise perception under different masking conditions. While neural coding of the F0 supports perception in both co-located and spatially separated conditions, neural timing predicts perception of spatially separated listening exclusively. Together, these results suggest co-located and spatially separated speech-in-noise perception draw on similar cognitive/linguistic skills, but distinct neural factors, in early childhood.

1. Introduction

Hearing in noise is an everyday challenge, affecting listeners of all ages. For children, hearing in noise is vital for academic success because the classroom environment seldom offers pristine listening conditions. School settings often exceed recommended noise levels (Bradley and Sato, 2008; Summers and Leek, 1998), and this interference can compromise a student's academic performance (Shield and Dockrell, 2008). As young children are more vulnerable to the detrimental effects of noise (Elliott, 1979; Fallon et al., 2002; Papso and Blood, 1989), those who struggle to perceive speech in noise are especially susceptible to academic challenges. Indeed, difficulty listening in noise is a hallmark symptom of many developmental disorders, including language impairments, attention-deficit hyperactivity disorder, and auditory

processing disorder (Bradlow et al., 2003; Brady et al., 1983; Moore et al., 2010; Ziegler et al., 2005, 2009).

Converging evidence suggests that speech-in-noise perception is facilitated by auditory-neurophysiological, cognitive, and linguistic processes (Pichora-Fuller et al., 1995). For instance, neural processing of the fundamental frequency, a cue thought to assist with auditory object formation (Shinn-Cunningham, 2008) and speaker identification (Andreou et al., 2011; Shamma et al., 2011), underlies a range of hearing in noise abilities (Anderson et al., 2010a; Song et al., 2011). Other cues, such as formant transitions in speech, provide information for consonant identification (Tallal and Stark, 1981), and are robustly encoded in individuals with superior speech-in-noise perception (Anderson et al., 2010b, 2013a). In addition, analysis of the auditory stream involves the separation of meaningful input from irrelevant

---

information, recruiting cognitive skills such as attention (Carlile and Corkhill, 2015; Jones et al., 2015; Strait and Kraus, 2011) to direct focus and suppress noise, and memory (Francis, 2010; Parbery-Clark et al., 2009; Rönnberg et al., 2010) to hold relevant details in short-term memory for lexical access and recognition. Lexical, semantic, and syntactic contexts facilitate recognition of words and phonemes (Boothroyd, 1970; Eisenberg et al., 2000, 2002; Nittrouer and Boothroyd, 1990). As speech-in-noise perception improves with age until adolescence (Elliott, 1979), children's hearing in noise is also constrained by developmental factors, such as vocabulary knowledge, phonemic categorization, and language competency (Boothroyd, 1970).

Everyday listening occurs within complex acoustic environments with multiple signals converging in location and spectrum. To segregate these signals into meaningful units, the nervous system must respond to the specific demands of the listening condition. For instance, the cognitive and linguistic mechanisms that support hearing in noise vary depending on the relative spatial locations of the target signal and its competing noise (Carlile and Corkhill, 2015; Ebata, 2003; Garadat and Litovsky, 2007); when speech and noise are co-located, this demands a large cognitive load and recruits working memory, whereas when speech and noise are spatially separated, the cognitive load is diminished, and alternative strategies are drawn upon for perception (Francis et al., 2011). Speech perception improves for listeners when the target signal is spatially separated from a masker (Bronkhorst and Plomp, 1988; Dirks and Wilson, 1969; Freyman et al., 1999), and this perceptual benefit, also known as spatial release from masking (SRM) (Hawley et al., 1999; Kidd et al., 1998), has been documented in children as young as 3 years old (Garadat and Litovsky, 2007; Litovsky, 2005). These results suggest the auditory physiological mechanisms supporting sound segregation and binaural processing are evident by young childhood, even though some auditory processing subskills are not mature until late childhood or even adolescence (e.g., gap detection (Wightman et al., 1989) and backward masking (Hartley et al., 2000)). It remains unknown, however, if auditory-neurophysiological mechanisms, in concert with cognitive and linguistic abilities, support hearing in noise in varying spatial configurations in young children.

While previous research demonstrates that neurophysiological, cognitive, and linguistic processes support hearing in noise in young childhood, these factors have been identified independently of one another. In the last decade, the hypothesis has emerged that speech in noise perception relies upon an integrated network of auditory-cognitive function (Wingfield et al., 2005), recognizing the need to consider the multiple supporting factors underlying hearing in noise within an integrated, dynamic system, rather than in isolation. This hypothesis has been tested in older adults (Anderson et al., 2013b) revealing the strongest predictors of hearing in noise are cognitive abilities and auditory neurophysiology, however, a comprehensive investigation into the relative contributions and interactions of these factors *in children* is needed. Some studies have crossed fields to explore these factors collectively, such as Hnath-Chisolm et al. (1998), who demonstrate young children's poor speech-in-noise perception may be due to cognitive factors coupled with immature phonological development (Hnath-Chisolm et al., 1998). Recent work from our laboratory revealed that in preschoolers (3–4 year olds), the development of word-in-noise perception is supported by both attention and neural processing of the fundamental frequency (Thompson et al., 2017). Still, there has yet to be a large-scale study examining multiple hypotheses about factors supporting children's speech-in-noise perception.

The first objective of this study is to understand how neurophysiological, cognitive, and linguistic factors come together to support speech-in-noise perception under different masking conditions in early childhood. We hypothesized that speech-in-noise perception in early childhood is supported by dissociable neurophysiological, cognitive, and linguistic processes, and that the relative contributions of these

processes depend on the degree of masking and the cognitive load of the task. We tested this hypothesis by evaluating 104 children, ages 3–7 years, on measures of speech-in-noise perception, neurophysiology, cognition, and linguistic knowledge. Structural equation modeling was used to determine the relative contributions of these measures in predicting speech-in-noise perception under two masking conditions: high masking (i.e., co-located), and low masking (i.e., spatial separation).

A second objective of this work is to examine predictive relationships over the course of childhood development. In pre-school children (3 year olds), the skills supporting functional listening are in flux, making it difficult to accurately measure performance on real-world measures of speech-in-noise perception (e.g., sentences in noise). Because of this gap, children are often not recognized as having difficulty perceiving speech in noise until classroom performance has been affected. To detect children at-risk for speech-in-noise deficits before they manifest, it is imperative to identify indices of a child's speech-in-noise ability early on (e.g., ˜age 3) that can predict performance when formal schooling begins (˜age 5).

Because all participants were enrolled in our lab's longitudinal study, we were poised to retroactively examine measures of neural processing, cognition, and language at a child's first visit to the lab (age ~ 3 years) and determine their predictive utility on sentence-in-noise perception two years later (age ~ 5 years). Following the identification of within-age predictors of sentence-in-noise perception using SEM, we used linear regression modeling to ascertain predictive relationships from data collected two years prior. Given established links between speech-in-noise perception and neurophysiological processing of acoustic cues (i.e., F0, transition timing) in children and adults (Anderson et al., 2010a; Thompson et al., 2017), *we hypothesize that processing of spectral and temporal cues are foundational neurophysiological processes underlying sentence perception in noise, and that these neural mechanisms when measured in preschoolers, can predict future (specifically, age ~ 5) sentence-in-noise perception.* Moreover, as speech perception in noise additionally relies upon a number of cognitive and linguistic factors including selective attention, short-term memory, and lexical knowledge (Lewis et al., 2010; Pichora-Fuller et al., 1995), *we also hypothesize that speech-in-noise perception is constrained by the development of cognitive and language skills in childhood, and that these skills, when measured in preschoolers, can predict future (age ~ 5) sentence-in-noise perception.*

2. Materials and methods

2.1. Participants

Participants included 104 children (46 F), ages 4 to 7 years old (Mean age = 5.99, Range: 4.13–7.71 years, SD = 0.85), recruited from the greater Chicago area; 51 children were 4–5 years old and 53 children were 6–7 years old. Ninety nine of these children were included in a retrospective analysis described below. Children were monolingual English speakers with non-verbal IQ > 85 (4–6 year olds: WPPSI, 7 year olds: WASI; see "Mental Ability/IQ"). Participants passed a peripheral hearing screener at the outset: normal otoscopy, Type A tympanometry, distortion product otoacoustic emissions ≥ 6 dB SPL above the noise floor from 0.5 to 4 kHz, and a click-evoked auditory brainstem response wave V latency within lab-internal normal limits (5.45 ms – 6.12 ms; Spitzer et al., 2015).

2.2. Analytical approach

Structural equation modeling (SEM) was used to delineate relationships among the contributors to speech-in-noise perception. SEM has two components: the measurement model and the structural model. The measurement model specifies the extent to which *latent variables,* or
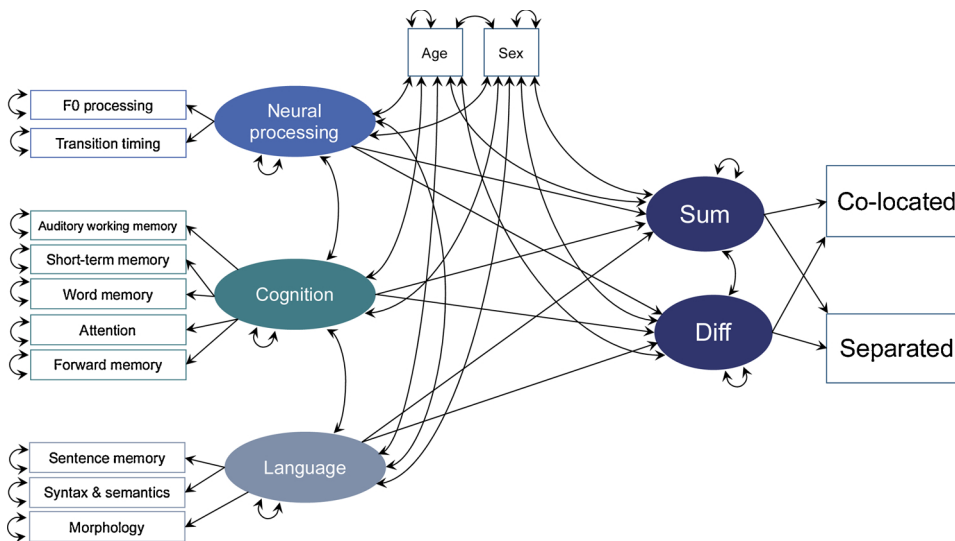
Fig. 1. The theoretically proposed structural equation model (M1) is comprised of latent constructs (ovals) and manifest variables (rectangles). Latent constructs include speech-in-noise perception (shared and differential), cognition, language, and neural processing. Manifest variables, or measured indicators, are listed for each latent construct.

unobserved constructs summarize a set of *manifest variables,* or observed indicators. The structural model evaluates relationships among the latent variables. Maximum likelihood estimation is used to approximate the set of model parameters simultaneously, and to estimate means, variances, and covariances of missing data (Loehlin, 2004).

In this study, *latent variables* included speech-in-noise perception, cognitive function, linguistic skills, and neurophysiology, all of which are composed of the *manifest variables* detailed below. Our initial hypothesized model can be observed in Fig. 1; we hypothesize that a constellation of factors predicts a significant amount of the variance in speech-in-noise perception, and their contributions vary with the masking condition and cognitive demands. While previous research has demonstrated unique contributions of neurophysiological, cognitive, and linguistic factors supporting speech-in-noise perception in adults (Anderson et al., 2013b), no study to date has used SEM to explore similar relationships in early childhood.

2.3. Speech-in-noise perception measures

Speech-in-noise perception was assessed using the Hearing in Noise Test (HINT, Bio-logic Systems Corp, (Soli and Wong, 2008)), an adaptive test of speech-in-noise processing that presents sentences in speech-shaped noise from a loud-speaker. During this test, the listener has access to acoustic, syntactic and semantic cues that increase the probability of selecting the correct target word from like-sounding competitors. Because of this, HINT performance does not solely rely on hearing thresholds but also depends on cognitive skills, such as auditory working memory (Parbery-Clark et al., 2009a) and attention (Strait and Kraus, 2011a), linguistic skills, such as syntax and semantics (Kalikow et al., 1977), and location of the speech signal (Hawley et al., 1999). Perception of speech in noise is typically superior in conditions where speech and noise are spatially separated; perception is also superior when speech is presented to the right ear. This phenomenon, often referred to as a "right-ear advantage", is thought to manifest due to processing of speech and language in the left hemisphere (Studdert-Kennedy and Shankweiler, 1970). The HINT determines signal to noise ratios (SNR) for individual subjects based on whether the noise and the target stimuli are presented from the same location (HINT$_{Front}$) or are spatially segregated (HINT$_{Right}$).

The HINT is composed of short semantically and syntactically simple English sentences (e.g., 'A boy fell from the window'; Bamford-Kowal-Bench sentences; Bench et al., 1979) spoken by a male speaker presented at adaptive levels in speech-shaped noise fixed at 65 dB SPL.

Participants are asked to repeat the target sentence out loud, and threshold SNR is defined as the dB SNR difference between the speech dB level and noise dB level in which the participant obtains 100% correct sentence repetition. Speech-in-noise performance was evaluated under two masking conditions: HINT$_{Front}$ (speech and masker emanate from the same speaker at 0° Azimuth, one meter directly ahead) and HINT$_{Right}$ (speech emanates from the front at 0° Azimuth, and the masker emanates from a speaker one meter away to the right at +90° Azimuth). Manifest variables of speech-in-noise perception include: *Co-located perception (*HINT$_{Front}$; high masking), and *Spatially separated perception* (HINT$_{Right}$; low masking).

To determine the shared and unique predictors of speech-in-noise perception in *co-located* and *spatially separated* conditions, "shared" and "difference" latent variables were created. "Shared" was created by constraining factor loadings of both manifest variables (HINT$_{Front}$ and HINT$_{Right}$) to 1, while "Difference" was generated by constraining factor loadings of HINT$_{Front}$ to 1 and HINT$_{Right}$ to -1.

2.4. Cognitive measures

Because of its reliance on contextual cues and sentence stimuli, sentence-in-noise perception engages cognitive processes, such as memory and attention (Francis, 2010a; Parbery-Clark et al., 2009; Strait and Kraus, 2011). To understand how these cognitive processes influence speech-in-noise performance in early childhood we tested participants on *Short-term Memory, Word Memory,* and *Auditory Working Memory* using the Woodcock Johnson-III (WJIII; Woodcock and Johnson, 1989) subtests Numbers Reversed (*Short-term Memory),* Memory for Words (*Word Memory*), and Auditory Working Memory, respectively. Additionally, *Attention* and *Forward Memory* were assessed using the Leiter-R (Roid and Miller, 1997), subtests Attention Sustained and Forward Memory, respectively. *Mental ability/Intelligence:* subtests of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI; < 6 year olds) and Wechsler Abbreviated Scale of Intelligence (WASI; ≥7 year olds) were administered to ensure normal cognitive function. Manifest variables of cognitive function included: *Short-term Memory, Auditory Working Memory, Word Memory, Attention, and Forward Memory*.

2.5. Language measures

Speech-in-noise perception in children is constrained by linguistic abilities (Eisenberg et al., 2002; Ren et al., 2015). The assessed

linguistic skills included: *sentence memory, semantics & syntax, and morphology*, evaluated using the Clinical Evaluation of Language Fundamentals (CELF-5; Wiig et al., 2013) subtests Recalling Sentences, Formulated Sentences, and Word Structure, respectively. Manifest variables of linguistic skills included: *Sentence memory, Semantics & syntax, and Morphology.*

2.6. Neurophysiological recording

2.6.1. Stimuli and presentation

Stimuli consisted of a click (for inclusionary criteria/hearing screening) and a 170 ms speech syllable [da], which is a voiced six-formant stop consonant constructed with a Klatt-based synthesizer at 20 kHz. The [da] has a fundamental frequency of 100 Hz, and during the consonant-vowel transition (0–50 ms) the lower three formants shift (f1: 400–720 Hz, f2: 1700-1240 Hz, f3: 2580-2500 Hz), while the fundamental frequency and upper three formants are steady (f0: 100 Hz, f4: 3300 Hz, f5: 3750 Hz, f6: 4900 Hz). During the vowel portion of the stimulus (50–120 ms) the six formants do not fluctuate.

The click stimulus was presented in rarefaction, and the [da] was presented in alternating polarities. Stimuli were presented monaurally to the right ear through electromagnetically-shielded insert earphones (ER-3A, Etymotic Research, Elk Grove Village, IL, USA) at 80 dB SPL.

2.6.2. Recording

Responses were collected using a BioSEMI Active2 recording system with an auditory brainstem response module. During the recording, the participant sat in a comfortable chair within an electrically- shielded and sound-attenuated booth (IAC Acoustics, Bronx, NY, USA), and watched a film of their choice; the left ear was unoccluded so the child could hear the soundtrack of the movie (~40 dB SPL; Skoe and Kraus, 2010). Electrodes were placed at Cz for active, right and left ear for reference/non-inverting, and +/- 1 cm on either side of Fpz for CMS/DRL, which serve as ground. All offsets are kept below 50 mV.

2.6.3. Data processing

Within the BioSEMI ActiABR module for LabView 2.0 (National Instruments, Austin, TX, USA), responses were online filtered from 100 to 3000 Hz (20 dB/decade roll-off), and digitized at 16.384 kHz. Using Matlab (The Mathworks, Inc., Natick, MA, USA), responses were bandpass filtered to the frequency region of interest (70–2000 Hz, Butterworth filter, 12 dB/octave roll-off, zero phase shift), epoched from -40-210 ms (stimulus onset at 0 ms), baselined, and artifact rejected (+/- 35 μV). The assessed neurophysiological factors include measures of frequency encoding and timing.

2.6.3.1. Measures of frequency encoding.
Fast Fourier transforms were applied to the de-meaned and windowed response period corresponding to the consonant-vowel transition (20–60 ms). The resulting frequency spectrum was analyzed with respect to the major stimulus frequency bands (including frequency bands corresponding to the speech formants) up to the limits of brainstem representation. Amplitudes (magnitudes) and phases of spectrum maxima (up to 2 kHz) were recorded. Because the frequency-following response imitates the signal that evokes it, larger amplitudes are interpreted to reflect a more robust representation of this acoustic cue.

2.6.3.2. Measures of timing.
Timing was examined by measuring latencies of the FFR, which occur at periodic intervals derived from the fundamental frequency (F0) and are thought to reflect phase locking of the auditory system (Skoe and Kraus, 2010). Latencies were identified in Neuroscan (Neuroscan Edit 4.5, Compumedics, Charlotte, NC) using a local maximum and minimum detection

algorithm followed by manual verification. Peaks and troughs were labeled based on the local maxima and minima within a latency range that corresponded to expected peak and trough latency values, respectively. Earlier latencies reflect a faster response.

Manifest variables of neurophysiology included *F0 processing,* which was calculated by taking the spectral amplitude of the fundamental frequency (F0) of the brain response, and *transition timing,* which refers to the absolute peak latencies within the formant-transition region of the brain response.

2.6.4. SEM approach

Structural equation modeling was used to estimate the pathways outlined in the initial model (M1, Fig. 1). SEM includes two main components: the measurement model, which defines latent unobserved constructs based on a set of measured (i.e., observed) variables, and the structural model, which measures the causal relationships between and among the constructs. To estimate the entire set of parameters in the model simultaneously, maximum likelihood estimation was used (Loehlin, 2004). All analyses were performed using Mplus (Muthén and Muthén, 2017). P values less than 0.05 were considered significant. STDXY standardized coefficients were reported.

To test overall fit of the structural equation model, we used several standardized fit statistics: the Model chi-square (X^2), Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and the Tucker-Lewis index (TLI). These indices assess how well the model's estimated population covariance matrix reproduces the sample covariance matrix. For RMSEA, values less than 0.04 are considered to be an "excellent" fit, less than 0.07 a "good" fit, and less than 0.1 a "fair" fit (Steiger, 2007); for CFI and TLI, models with an "excellent" fit are greater than 0.95 (Hu and Bentler, 1999). For details on these indices please see (Hooper et al., 2008).

2.6.5. Predicting speech-in-noise perception

The goal of this follow-up analysis was to determine whether the same factors at age 3 predict future speech-in-noise perception at age 5. Following the identification of neural, cognitive, and linguistic predictors of sentence-in-noise perception using SEM, we retroactively examined the predictive utility of these factors through linear regression modeling. Children were enrolled in a longitudinal study at age 3 or 4 and were tested annually for up to five years. Ninety-nine (n = 99) children were available for this retrospective analysis. Data was split into "visit 1" and "visit 2"; the average amount of time between these two tests was 1.98 years (SD = 0.45 years; range: 0.64–2.89 years). Two linear regressions were performed to predict 1) co-located and 2) spatially separated speech-in-noise perception. Included on the first step were sex and age at first visit; included on the second step were neural, cognitive, and linguistic predictors of sentence-in-noise perception identified using structural equation modeling but measured at first visit (age 3 or 4).

3. Results

Results are organized into three sections: 1) the measurement model, 2) the structural equation model, and 3) follow-up analyses.

3.1. Measurement model

To determine the reliability of the latent constructs, we used factor analysis to evaluate loadings of manifest variables on latent variables. Because F0 processing and transition timing did not equally comprise the neural latent construct (i.e. factor loadings were 0.073 and 0.573, respectively), we then refined the conceptual model by eliminating the multifactor neural latent variable and including these indicators in the
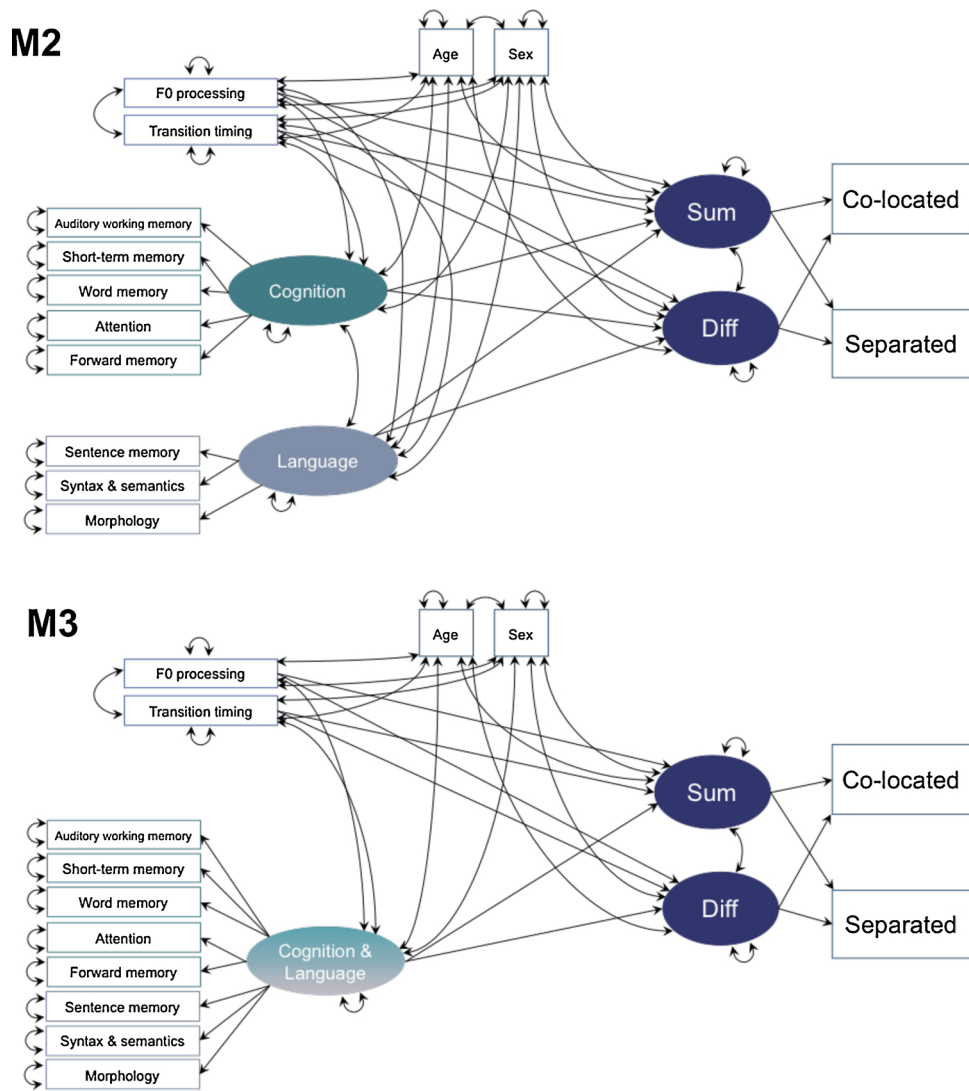
Fig. 2. Path diagrams for M2 and M3. These models reflect refined versions of the conceptual model (M1), where the multifactor latent variable is eliminated (M2), and the cognition and language latent constructs are combined (M3).

Table 1
Factor loadings of manifest variables on latent constructs of the final model (M2). HINT = Hearing in Noise Test; CELF = Clinical Evaluation of Language Fundamentals; WJIII = Woodcock Johnson Test of Cognitive Abilities-Third Edition; FFR = frequency following response.

Latent variable	Manifest variable	Test	Estimate	SE	p-value
Speech-in-noise perception SUM	Co-located	HINT Front (dB SNR)	1	—	—
	Spatially separated	HINT Right (dB SNR)	1	—	—
Speech-in-noise perception DIFF	Co-located	HINT Front (dB SNR)	1	0.088	< 0.001
	Spatially separated	HINT Right (dB SNR)	−1	0.024	< 0.001
Language	Syntax and semantics	CELF Formulated Sentences (scaled score)	0.599	0.088	< 0.001
	Word morphology	CELF Word Structure (scaled score)	0.689	0.077	< 0.001
	Sentence memory	CELF Recalling Sentences (scaled score)	0.792	0.063	< 0.001
Cognition	Auditory working memory	WJIII Auditory Working Memory (scaled score)	0.669	0.075	< 0.001
	Word memory	WJIII Memory for Words (scaled score)	0.809	0.055	< 0.001
	Short-term memory	WJIII Numbers Reversed (scaled score)	0.636	0.08	< 0.001
	Forward memory	Leiter Forward Memory (scaled score)	0.402	0.113	< 0.001
	Attention	Leiter Attention Sustained (scaled score)	0.36	0.109	0.001
Neural (F0)	F0 processing (µV)	FFR	1	—	—
Neural (Timing)	Transition timing (ms)	FFR	1	—	—

Table 2
Correlations among manifest variables. *p < 0.05; **p < 0.01.

	Co-located SPIN (dB SNR)	Spatially separated SPIN (dB SNR)	F0 amplitude (uV)	Transition timing (ms)	Auditory working memory (scaled score)	Memory for words (scaled score)	Numbers reversed (scaled score)	Attention (scaled score)	Forward memory (scaled score)	Formulated sentences (scaled score)	Word structure (scaled score)	Recalling sentences (scaled score)	Age (years)	Sex
Co-located SPIN (dB SNR)	1	.365**	-0.160	-0.124	-.245*	-0.081	0.042	-0.043	-.251*	-0.227	-.294*	-.286**	-.259**	0.024
Spatially separated SPIN (dB SNR)	.365**	1	-.268**	0.105	-0.121	-0.197	0.056	-0.061	-0.144	-0.128	-.273*	-0.161	-0.109	-0.053
F0 amplitude (uV)	-0.160	-.268**	1	0.067	0.025	0.011	0.003	-0.191	0.009	-0.030	0.113	0.084	0.122	.256*
Transition timing (ms)	-0.124	0.105	0.067	1	0.132	0.119	0.216	0.079	0.157	.260*	0.122	0.003	-0.041	.254*
Auditory working memory (scaled score)	-.245*	-0.121	0.025	0.132	1	.530**	.442**	.260*	0.183	.339**	.460**	.540**	-0.006	0.030
Memory for words (scaled score)	-0.081	-0.197	0.011	0.119	.530**	1	.519**	.256*	.263*	.400**	.424**	.552**	-.302**	0.108
Numbers reversed (scaled score)	0.042	0.056	0.003	0.216	.442**	.519**	1	0.143	.345**	.328**	.381**	.457**	-0.112	0.179
Attention (scaled score)	-0.043	-0.061	-0.191	0.079	.260*	.256*	0.143	1	.337**	.347**	0.192	.276*	-.235*	-0.171
Forward memory (scaled score)	-.251*	-0.144	0.009	0.157	0.183	.263*	.345**	.337**	1	.253*	0.220	0.228	-0.007	-0.093
Formulated sentences (scaled score)	-0.227	-0.128	-0.030	.260*	.339**	.400**	.328**	.347**	.253*	1	.543**	.469**	-0.027	-0.184
Word structure (scaled score)	-.294*	-.273*	0.113	0.122	.460**	.424**	.381**	0.192	0.220	.543**	1	.541**	0.061	-0.052
Recalling sentences (scaled score)	-.286**	-0.161	0.084	0.003	.540**	.552**	.457**	.276*	0.228	.469**	.541**	1	-0.043	0.066
Age (years)	-.259**	-0.109	0.122	-0.041	-0.006	-.302**	-0.112	-.235*	-0.007	-0.027	0.061	-0.043	1	0.090
Sex	0.024	-0.053	.256*	.254*	0.030	0.108	0.179	-0.171	-0.093	-0.184	-0.052	0.066	0.090	1

Table 3
Model comparisons show M2 had the best fit. When comparing models, a significant Chi Square test indicates the model with a greater number of parameters should be accepted, while non-significant Chi Square tests indicate the model with fewer parameters should be accepted. M1 refers to the theoretically proposed model with a neural latent construct, while M3 refers to the model in which cognition and language latent constructs were combined.

	M1	M2
Parameters	58	64
RMSEA	0.067	0.059
CFI	0.88	0.916
TLI	0.822	0.863
LL	−3498.704	−3263.385
Chi-Square Estimate	89.646	74.919
Chi-Square df	61	55
Model Comparison		*M1* vs *M2*
p-value		0.022

	M3	M2
Parameters	57	64
RMSEA	0.068	0.059
CFI	0.875	0.916
TLI	0.819	0.863
LL	−3271.792	−3263.385
Chi-Square Estimate	91.733	74.919
Chi-Square df	62	55
Model Comparison		*M2* vs *M3*
p-value		0.019

model as independent manifest variables (M2; Fig. 2). Table 1 presents latent constructs and their constituent manifest variables and factor loadings of the final model (M2). Table 2 reports correlations among all manifest variables.

3.2. Structural model

Although the cognition and language latent constructs were appropriately defined, they were also highly correlated (Table 4; R = 0.836, p < 0.001), and neither construct predicted speech-in-noise perception in the context of the other. This is not to say that cognition and language have no effect on speech-in-noise perception overall. Rather, there is no effect of cognition on speech-in-noise perception when controlling for language, and vice versa, indicating they should be thought of as a single latent construct that predicts speech-in-noise perception. To test this, we ran a third model (M3; Fig. 2) with cognition and language combined as one latent construct (Cognition/Language).

Model comparisons are reported in Table 3. The overall model fit improved with the elimination of the multi-factor neural latent variable (M1 vs M2; Chi square difference test: p = 0.022) but did not fit better when combining cognition and language into one latent construct (M2 vs M3; Chi square difference test: p = 0.019). Although M3 did not fit significantly better than M2, we report its estimates for interpretation. The final structural equation model (M2) showed moderate to good fit indices (RMSEA = 0.059, CFI = 0.916, TLI = 0.863). Standardized coefficients and their respective p-values for M2 and M3 are reported in Table 4.

Model 2 and, in the case of cognition/language, Model 3, demonstrate the shared predictors of speech-in-noise perception include F0 processing and a combined construct of cognition/language (M2: F0 processing, p-value = 0.012; M3: cognition/language, p-value = 0.002). This suggests both F0 processing and cognition/language have an overall effect on speech-in-noise perception, regardless

of listening condition.

In contrast, the differential predictors of co-located and spatially separated speech-in-noise perception were transition timing (i.e., neural timing) and F0 processing (M2: transition timing, p = 0.016; F0 processing, p = 0.04). However, because the effect of F0 processing was not strong (p = 0.04), this finding should be interpreted with caution as p-values between 0.04 and 0.06 may not be reliable (Loehlin, 2004). Nonetheless, given the differential effect of transition timing on speech-in-noise perception, these findings lend support to our hypothesis that the two listening conditions differ in their underlying neural mechanisms.

3.3. Predicting speech-in-noise perception, a retrospective analysis

To determine if the same factors at ˜age 3 predict speech-in-noise perception at ˜age 6, we used linear regression modeling with neural and cognitive/language predictors. To reduce the risk of over-parameterizing the linear regressions, we used one cognitive/language measure, "sentence memory", which was evaluated using the CELF subtest Recalling Sentences and is a task that engages both cognition and language. Included in the linear regressions were predictors measured at first test visit: age, F0 processing, transition timing, and sentence memory. Sex was also included as a predictor.

Over and above age and sex, neural and cognitive/language predictors measured at ˜age 3 did not significantly predict future co-located speech-in-noise perception (Total R^2 = 0.132; ΔR^2 = 0.078; p = 0.063). However, these measures did predict future spatially separated speech-in-noise perception (Total R^2 = 0.204; ΔR^2 = 0.138; p = 0.004). Unique predictors included F0 processing (β = -0.234; p = 0.020) and sentence memory (β = -0.226; p = 0.026), but not transition timing (β = 0.164; p = 0.108). Regression statistics for both co-located and spatially separated speech-in-noise perception are reported in Table 5.

Table 4

Standardized coefficients of latent constructs predicting the shared and differential effects of co-located and spatially separated speech-in-noise perception.

Regressions	M2			M3		
	Estimate	SE	p-value	Estimate	SE	p-value
SUM by F0 processing	−0.245	0.098	0.012	−0.251	0.092	0.007
Transition timing	0.074	0.098	0.453	0.083	0.096	0.384
Cognition	−0.114	0.419	0.785	—	—	—
Language	−0.195	0.4	0.626	—	—	—
Cognition/Language	—	—	—	−0.306	0.1	0.002
Sex	0.04	0.112	0.719	0.045	0.097	0.639
Age	−0.175	0.156	0.263	−0.192	0.093	0.04
DIFF by F0 processing	−0.222	0.108	0.04	−0.198	0.098	0.043
Transition timing	0.251	0.105	0.016	0.238	0.1	0.017
Cognition	−0.353	0.466	0.448	—	—	—
Language	0.247	0.448	0.582	—	—	—
Cognition/Language	—	—	—	−0.087	0.112	0.434
Sex	−0.038	0.121	0.752	−0.08	0.101	0.429
Age	−0.028	0.172	0.873	0.064	0.099	0.518
Cognition & Language	0.836	0.074	< 0.001			
F0 processing & Language	0.075	0.118	0.527			
F0 processing & Cognition	−0.008	0.121	0.946			
Transition tming & Language	0.126	0.122	0.302			
Transition tming & Cognition	0.213	0.116	0.066			
Transition tming & F0 processing	0.062	0.1	0.532			
DIFF & SUM	0.616	0.067	< 0.001			

Table 5

Over and above age and sex, neural and cognitive/language skills at ˜age 3 predict spatially separated speech-in-noise perception at ˜age 6. Because children participated in our longitudinal study, we were poised to retroactively examine these measures at a child's first visit (˜age 3–4) and their predictive utility on sentence-in-noise perception two years later (˜5-6). Unique predictors included F0 processing (β = -0.234; p = 0.020) and sentence memory (β = -0.226; p = 0.026), but not transition timing (β = 0.164; p = 0.108).

DV: Co-located

	ΔR^2	β	p-value
Step 1	0.054		0.091
Sex		0.072	0.490
Age		−0.225	0.034
Step 2	0.078		0.063
Sex		0.095	0.381
Age		−0.194	0.065
F0 processing		−0.100	0.334
Transition timing		−0.127	0.234
Sentence memory		−0.228	0.031
Total R^2	0.132		

DV: Spatially separated

	ΔR^2	β	p-value
Step 1	0.066		0.051
Sex		−0.027	0.792
Age		−0.253	0.017
Step 2	0.138		0.004
Sex		−0.064	0.532
Age		−0.199	0.049
F0 processing		−0.234	0.020
Transition timing		0.164	0.108
Sentence memory		−0.226	0.026
Total R^2	0.204		

4. Discussion

4.1. Results in review

This study examined the neurophysiological, cognitive, and linguistic factors supporting speech-in-noise perception under co-located and spatially-separated masking conditions in young children, ages 3–7 years. Our findings are twofold. First, we see that cognitive and language skills are tightly linked in this age range, and when combined into one latent construct, predict speech-in-noise perception across both masking conditions. Second, results show neural processing of sound contributes to speech-in-noise perception overall and differentially: while F0 processing is a shared predictor of speech-in-noise under both masking conditions, transition timing predicts spatially separated, but not co-located, perception. Taken together, these results suggest co-located and spatially separated speech-in-noise perception draw on similar cognitive/linguistic skills, but different neural mechanisms in early childhood.

4.2. Cognitive/language predictors of speech-in-noise perception

Here, we evaluated latent constructs of cognition and language using structural equation modeling and found that each appropriately summarized their constituent indices: cognition comprised tests of memory and attention, while language was indexed via measures of syntactic, semantic, and morphological knowledge. In addition, we saw a strong relationship between cognition and language, and that their predictive relationship with speech-in-noise perception emerged only upon combining the two into one latent predictor.

Throughout childhood development, there is a strong interplay between cognition and language such that both influence and inform the other. For example, language acquisition depends on the perception, rehearsal, and manipulation of phonemes, and is facilitated by cognitive skills such as working memory (Baddeley, 2003). As language

improves, the ability to synthesize linguistic complexities (e.g., new words, longer sentences, etc.) can challenge (and therefore, shape) cognitive function (Baddeley, 2003). This reciprocity is especially evident in the first decade of life as toddlers speaking simple sentences grow into school-aged children capable of reading lengthy chapter books. Given the ages of the children in this study, it may be the case that the cognitive-language interplay is still in flux, and that this reciprocal relationship can account for some of the overlap we observed between the two.

One important consideration for the link between cognition and language is the behavioral tests used to index these constructs. For example, while we specifically tested cognition using tasks known to relate with speech-in-noise perception (i.e., auditory working memory, word memory, short-term memory, and attention), in retrospect these tests also engage language, even when test instructions are non-verbal. For example, the sustained attention task specifically instructs the evaluator to use non-verbal cues when testing the participant, yet the task at hand—to circle target objects (e.g., flowers, snails, stars, etc.) amidst distractors—likely involved lexical access. In fact, performance on this task correlated with two of the three language measures (Table 2; Attention sustained with: Formulated sentences, R = 0.347, p < 0.05; Recalling sentences, R = .276, p < 0.05; Word structure, R = .192, p > 0.05).

4.3. Differential and shared task demands of various listening conditions

Varying degrees of masking are thought to influence the cognitive load of a speech perception task, and it is thought that the underlying mechanisms of hearing in noise vary depending on the listening condition. For instance, when speech and noise are co-located, this demands a large cognitive load and recruits working memory, while spatially separated speech and noise diminishes this load and draws on alternative strategies (Francis, 2010). We hypothesized that by co-locating a signal and noise, a high degree of masking degrades perceptually vulnerable acoustic cues of speech (e.g. formant transitions) (Hornickel et al., 2009) and linguistic contextual cues (Eisenberg et al., 2000), resulting in a task that is especially cognitively demanding. We also hypothesized that spatially separating a signal from noise reduces masking, allowing access to additional acoustic cues and the signal's content, and linguistic contextual cues to facilitate recognition, since cognitive demands have decreased.

Our findings partially support the proposed hypotheses. While cognition and language demonstrate an overall effect on speech-in-noise perception and do not differ based on the listening condition, the finding that neural processing of transition timing is a differential predictor of spatially separated perception suggests the listening conditions engage distinct neural mechanisms. The lack of a differential "cognitive load" does not align with previous research in older children and adults, yet this finding could be interpreted within the context of age. As discussed previously, it may be the case that in early childhood, cognition and language are inextricably linked, leading to the engagement of the two during speech-in-noise perception under different listening conditions. It is not unreasonable to suggest task demands under various listening conditions varies as a function of age.

4.4. Neural predictors of speech-in-noise perception

In addition to cognitive and language abilities, precise neurophysiological processing of acoustic cues augments perception of speech in noise. Our results show both neural processing of the fundamental frequency and transition timing are significant predictors of speech-in-noise perception in young children. Interestingly, these neurophysiological predictors uniquely contribute to speech-in-noise perception: while F0 processing is predictive of speech-in-noise perception overall, transition timing serves as a differential predictor of co-located and spatially-separated speech and noise perception. These findings suggest

that, regardless of the listening condition, enhanced neural processing of the F0 is linked with better perception of speech in noise, while neural timing supports spatially separated but not co-located speech-in-noise perception.

Enhanced processing of timing cues enables a listener to distinguish consonants (e.g., bad vs. dad) in listening conditions with greater access to acoustic cues. For example, individuals with faster neural timing have greater access to high frequency cues to differentiate speech syllables "ba" and "ga" (Strait et al., 2014), and temporal acoustic cues are more accessible when speech and noise are spatially separated than when they are co-located (Hornickel et al., 2009). It may be the case that a faster system more robustly encodes binaural cues in spatially separated conditions. After all, enhanced timing is associated with better binaural processing and sound localization (Grothe, 2003).

4.5. Structural equation modeling

This study is the first of its kind to employ structural equation modeling to delineate relationships between the multifaceted contributors to speech-in-noise perception in early childhood. SEM is a statistical approach that leverages matrix algebra and maximum likelihood estimation to evaluate predictive relationships among manifest (observed) variables and latent (unobserved) constructs. In SEM, the latent variables summarize the construct shared by a set of observed variables; latent variables in our initial model included speech-in-noise perception, cognitive function, linguistic skills, and neurophysiology. Through a series of model comparisons, we found that the best fitting model included latent constructs of cognition and language, as well as manifest variables of F0 processing and transition timing. This finding supports previous research in that neural processing of the fundamental frequency (F0) and neural timing are independent parameters, even though they come from the same electrophysiological recording (Skoe and Kraus, 2010).

4.6. Early predictors of speech-in-noise perception

One question that this work is able to address is whether early neurophysiological, cognitive, and language skills predict speech-in-noise perception two years later. By retroactively examining longitudinal data of our participants, we found that early neural and cognitive/language skills (age 3) predict future sentence-in-noise performance (age 5). This suggests both neural and cognitive/language skills are foundational for future functional listening success; these correlates could serve as useful tools for earlier identification of listening challenges of clinical populations.

Of note, though these models are useful for understanding how children perceive speech in noise, the total variance explained could be improved. Here, predictors were included based on evidence suggesting their relationship with speech-in-noise perception in older children and adults. Given the vast number of developmental changes that occur throughout young childhood, more accurate or precise age-related predictors likely exist. For example, while attention is thought to be engaged during speech-in-noise perception tasks, it may be the case that for children, certain *types* of attention, like tuning in to a speaker (i.e., selective attention) or tuning out the noise (i.e., cognitive inhibition) carry different weights at different developmental milestones. Improving the explained variance of the models requires additional research that can investigate these latent constructs throughout development with greater granularity and specificity.

5. Conclusions

For the first time, we provide evidence that children's speech-in-noise perception relies upon neural processing of acoustic cues and cognition/language skills. Yet, while cognition/language appears to support perception more generally (i.e., under various masking

conditions), neural processing of formant transitions is a differential predictor of spatially separated speech perception. In addition, by retrospectively examining the predictive utility of these factors in early development, results reveal both neural processing and cognition/language at age 3 predict future speech-in-noise perception at age 5. Taken together, these findings demonstrate that in early development, the mechanisms supporting functional listening are largely similar, although slightly different, from what is known to support adult-like perception. We interpret these results in the context of development, in that the maturation of the auditory system may provide a scaffolding for cognition and language to emerge as distinct processes.

Declaration of Competing Interest

The authors have no conflicts of interest to disclose.

Acknowledgement

References

Anderson, S., Skoe, E., Chandrasekaran, B., Zecker, S., Kraus, N., 2010a. Brainstem correlates of speech-in-noise perception in children. Hear. Res. 270, 151–157.

Anderson, S., Skoe, E., Chandrasekaran, B., Kraus, N., 2010b. Neural timing is linked to speech perception in noise. J. Neurosci. 30, 4922–4926.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., Drehobl, S., Kraus, N., 2013a. Effects of hearing loss on the subcortical representation of speech cues. J. Acoust. Soc. Am. 133, 3030–3038.

Anderson, S., White-Schwoch, T., Parbery-Clark, A., Kraus, N., 2013b. A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. Hear. Res. 300, 18–32.

Andreou, L., Kashino, M., Chait, M., 2011. The role of temporal regularity in auditory segregation. Hear. Res. 280, 228–235.

Baddeley, A., 2003. Working memory and language: an overview. J. Commun. Disord. 36, 189–208.

Bench, J., Kowal, A., Bamford, J., 1979. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. Br. J. Audiol. 13, 108–112.

Boothroyd, A., 1970. Developmental factors in speech recognition. Int. Audiol. 9, 30–38.

Bradley, J.S., Sato, H., 2008. The intelligibility of speech in elementary school classrooms. J. Acoust. Soc. Am. 123, 2078–2086.

Bradlow, A.R., Kraus, N., Hayes, E., 2003. Speaking clearly for children with learning disabilities: sentence perception in noise. J. Speech Lang. Hear. Res. JSLHR 46, 80–97.

Brady, S., Shankweiler, D., Mann, V., 1983. Speech perception and memory coding in relation to reading ability. J. Exp. Child Psychol. 35, 345–367.

Bronkhorst, A.W., Plomp, R., 1988. The effect of head-induced interaural time and level differences on speech intelligibility in noise. J. Acoust. Soc. Am. 83, 1508–1516.

Carlile, S., Corkhill, C., 2015. Selective spatial attention modulates bottom-up informational masking of speech. Sci. Rep. 5.

Dirks, D.D., Wilson, R.H., 1969. The effect of spatially separated sound sources on speech intelligibility. J. Speech Hear. Res. 12, 5–38.

Ebata, M., 2003. Spacial unmasking and attention related to the cocktail party problem. Accoustic Sci. Technol. 24, 208–219.

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

Eisenberg, L.S., Martinez, A.S., Holowecky, S.R., Pogorelsky, S., 2002. Recognition of lexically controlled words and sentences by children with normal hearing and children with cochlear implants. Ear Hear. 23, 450–462.

Elliott, L.L., 1979. Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. J. Acoust. Soc. Am. 66, 651–653.

Fallon, M., Trehub, S.E., Schneider, B.A., 2002. Children's use of semantic cues in degraded listening environments. J. Acoust. Soc. Am. 111, 2242–2249.

Francis, A.L., 2010. Improved segregation of simultaneous talkers differentially affects perceptual and cognitive capacity demands for recognizing speech in competing speech. Atten. Percept. Psychophys. 72, 501–516.

Francis, A.L., Connell, K., Anderson, L., 2011. Does spatial release from masking interact with working memory capacity in speech perception? J. Acoust. Soc. Am. 129, 2662.

Freyman, R.L., Helfer, K.S., McCall, D.D., Clifton, R.K., 1999. The role of perceived spatial separation in the unmasking of speech. J. Acoust. Soc. Am. 106, 3578–3588.

Garadat, S.N., Litovsky, R.Y., 2007. Speech intelligibility in free field: spatial unmasking in preschool children. J. Acoust. Soc. Am. 121, 1047–1055.

Grothe, B., 2003. Sensory systems: new roles for synaptic inhibition in sound localization. Nat. Rev. Neurosci. 4, 540–550.

Hartley, D.E., Wright, B.A., Hogan, S.C., Moore, D.R., 2000. Age-related improvements in auditory backward and simultaneous masking in 6- to 10-year-old children. J. Speech Lang. Hear. Res. JSLHR 43, 1402–1415.

Hawley, M.L., Litovsky, R.Y., Colburn, H.S., 1999. Speech intelligibility and localization in a multi-source environment. J. Acoust. Soc. Am. 105, 3436–3448.

Hnath-Chisolm, T.E., Laipply, E., Boothroyd, A., 1998. Age-related changes on a children's test of sensory-level speech perception capacity. J. Speech Lang. Hear. Res. JSLHR 41, 94–106.

Hooper, D., Coughlan, J., Mullen, M.R., 2008. Structural equation modelling: guidelines for determining model fit. Electron. J. Bus. Res. Methods 6, 53–60.

Hornickel, J., Skoe, E., Nicol, T., Zecker, S., Kraus, N., 2009. Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. Proc. Natl. Acad. Sci. 106, 13022–13027.

Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. Multidiscip. J. 6, 1–55.

Jones, P.R., Moore, D.R., Amitay, S., 2015. Development of auditory selective attention: why children struggle to hear in noisy environments. Dev. Psychol. 51, 353–369.

Kalikow, D.N., Stevens, K.N., Elliott, L.L., 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J. Acoust. Soc. Am. 61, 1337–1351.

Kidd, G., Mason, C.R., Rohtla, T.L., Deliwala, P.S., 1998. Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. J. Acoust. Soc. Am. 104, 422–431.

Lewis, D., Hoover, B., Choi, S., Stelmachowicz, P., 2010. Relationship between Speech Perception in Noise and Phonological Awareness Skills for Children With Normal Hearing. Ear Hear. 31, 761–768.

Litovsky, R.Y., 2005. Speech intelligibility and spatial release from masking in young children. J. Acoust. Soc. Am. 117, 3091–3099.

Loehlin, J.C., 2004. Latent Variable Models: An Introduction to Factor, Path, and Structural Equation Analysis. L. Erlbaum Associates, Mahwah, N.J.

Moore, D.R., Ferguson, M.A., Edmondson-Jones, A.M., Ratib, S., Riley, A., 2010. Nature of auditory processing disorder in children. Pediatrics 126, e382–390.

Muthén, L.K., Muthén, B.O., 2017. Mplus User's Guide, eighth edition. .

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.

Papso, C.F., Blood, I.M., 1989. Word recognition skills of children and adults in background noise. Ear Hear. 10, 235–236.

Parbery-Clark, A., Skoe, E., Kraus, N., 2009. Musical experience limits the degradative effects of background noise on the neural processing of sound. J. Neurosci. Off. J. Soc. Neurosci. 29, 14100–14107.

Pichora-Fuller, M.K., Schneider, B.A., Daneman, M., 1995. How young and old adults listen to and remember speech in noise. J. Acoust. Soc. Am. 97, 593–608.

Ren, C., Liu, S., Liu, H., Kong, Y., Liu, X., Li, S., 2015. Lexical and age effects on word recognition in noise in normal-hearing children. Int. J. Pediatr. Otorhinolaryngol. 79, 2023–2027.

Roid, G.H., Miller, L.J., 1997. Leiter International Performance Scale - Revised. Stoelting Co.

Rönnberg, J., Rudner, M., Lunner, T., Zekveld, A.A., 2010. When cognition kicks in: working memory and speech understanding in noise. Noise Health 12, 263–269.

Shamma, S.A., Elhilali, M., Micheyl, C., 2011. Temporal coherence and attention in auditory scene analysis. Trends Neurosci. 34, 114–123.

Shield, B.M., Dockrell, J.E., 2008. The effects of environmental and classroom noise on the academic attainments of primary school children. J. Acoust. Soc. Am. 123, 133–144.

Shinn-Cunningham, B.G., 2008. Object-based auditory and visual attention. Trends Cogn. Sci. 12 (5), 182–186.

Skoe, E., Kraus, N., 2010. Auditory brain stem response to complex sounds: a tutorial. Ear Hear. 31, 302–324.

Soli, S.D., Wong, L.L.N., 2008. Assessment of speech intelligibility in noise with the Hearing in Noise Test. Int. J. Audiol. 47, 356–361.

Song, J.H., Skoe, E., Banai, K., Kraus, N., 2011. Perception of speech in noise: neural correlates. J. Cogn. Neurosci. 23, 2268–2279.

Spitzer, E., White-Schwoch, T., Woodruff Carr, K., Skoe, E., Kraus, N., 2015. Continued maturation of the click-evoked auditory brainstem response in preschoolers. J. Am. Acad. Audiol. 26, 30–35.

Steiger, J.H., 2007. Understanding the limitations of global fit assessment in structural equation modeling. Personal. Individ. Differ. 42, 893–898.

Strait, D.L., Kraus, N., 2011. Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. Front. Psychol. 2.

Strait, D.L., O'Connell, S., Parbery-Clark, A., Kraus, N., 2014. Musicians' enhanced neural differentiation of speech sounds arises early in life: developmental evidence from ages 3 to 30. Cereb. Cortex 24, 2512–2521.

Studdert-Kennedy, M., Shankweiler, D., 1970. Hemispheric specialization for speech perception. J. Acoust. Soc. Am. 48, 579–594.

Summers, V., Leek, M.R., 1998. FO processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. J. Speech Lang. Hear. Res. JSLHR 41, 1294–1306.

Tallal, P., Stark, R.E., 1981. Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. J. Acoust. Soc. Am. 69, 568–574.

Thompson, E.C., Woodruff Carr, K., White-Schwoch, T., Otto-Meyer, S., Kraus, N., 2017. Individual differences in speech-in-noise perception parallel neural speech processing and attention in preschoolers. Hear. Res. 344, 148–157.

Wightman, F., Allen, P., Dolan, T., Kistler, D., Jamieson, D., 1989. Temporal resolution in children. Child Dev. 60, 611–624.

Wiig, E.H., Semel, E., Secord, W., 2013. Clinical Evaluation of Language Fundamentals.

Pearson.

Wingfield, A., Tun, P.A., McCoy, S.L., 2005. Hearing Loss in Older Adulthood. What It Is and How It Interacts With Cognitive Performance. Curr. Dir. Psychol. Sci. 14, 144–148.

Woodcock, R., Johnson, M., 1989. Woodcock-Johnson Psycho-Educational Battery-Revised. Tests of Cognitive Ability. DLM Teaching Resources, Allen, TX.

Ziegler, J.C., Pech-Georgel, C., George, F., Alario, F.-X., Lorenzi, C., 2005. Deficits in speech perception predict language learning impairment. Proc. Natl. Acad. Sci. 102, 14110–14115.

Ziegler, J.C., Pech-Georgel, C., George, F., Lorenzi, C., 2009. Speech-perception-in-noise deficits in dyslexia. Dev. Sci. 12, 732–745.