

Research Article

The Contribution of Individual Differences in Memory Span and Language Ability to Spatial Release From Masking in Young Children

Douglas MacCutcheon,^a Florian Pausch,^b Christian Füllgrabe,^c Renata Eccles,^d Jeannie van der Linde,^d Clorinda Panebianco,^e Janina Fels,^b and Robert Ljung^f

Purpose: Working memory capacity and language ability modulate speech reception; however, the respective roles of peripheral and cognitive processing are unclear. The contribution of individual differences in these abilities to utilization of spatial cues when separating speech from informational and energetic masking backgrounds in children has not yet been determined. Therefore, this study explored whether speech reception in children is modulated by environmental factors, such as the type of background noise and spatial configuration of target and noise sources, and individual differences in the cognitive and linguistic abilities of listeners.

Method: Speech reception thresholds were assessed in 39 children aged 5–7 years in simulated school listening environments. Speech reception thresholds of target sentences spoken by an adult male consisting of number and color combinations were measured using an adaptive

procedure, with speech-shaped white noise and single-talker backgrounds that were either collocated (target and background at 0°) or spatially separated (target at 0°, background noise at 90° to the right). Spatial release from masking was assessed alongside memory span and expressive language.

Results and Conclusion: Significant main effect results showed that speech reception thresholds were highest for informational maskers and collocated conditions. Significant interactions indicated that individual differences in memory span and language ability were related to spatial release from masking advantages. Specifically, individual differences in memory span and language were related to the utilization of spatial cues in separated conditions. Language differences were related to auditory stream segregation abilities in collocated conditions that lack helpful spatial cues, pointing to the utilization of language processes to make up for losses in spatial information.

Speech reception, namely, hearing, listening to, and comprehending spoken language, is a multifaceted process integrating basic auditory processing of speech acoustics, cognitive processes (such as attention and memory), and linguistic and knowledge-based processing (e.g., interpretations and expectations; Davis, 1964; Kiessling et al., 2003; B. C. J. Moore, 2012; Pichora-Fuller & Singh, 2006; Rönnerberg, Rudner, Foo, & Lunner, 2008; Tsui & Fullilove, 1998). Additionally, the auditory environment often contains competing talkers and background noises that affect speech reception as a function of their properties, such as level (B. C. J. Moore, 2012) or the number of talkers (Simpson & Cooke, 2005), the nature of the competing sounds (i.e., energetic vs. informational maskers; Brungart, 2001), and localization (i.e., identifying where the sounds are coming from in space; B. C. J. Moore, 2012). Signal properties, environmental factors, and personal factors, such as age, cognitive ability, language background, and the presence

^aDepartment of Building, Energy and Environmental Engineering, Hogskolan i Gävle, Sweden

^bMedical Acoustics Group, Institute of Technical Acoustics, Rheinisch Westfälische Technische Hochschule Aachen, Germany

^cSchool of Sport, Exercise and Health Sciences, Loughborough University, United Kingdom

^dDepartment of Speech-Language Pathology and Audiology, University of Pretoria, South Africa

^eDepartment of Music, University of Pretoria, South Africa

^fFaculty of Engineering and Sustainable Development–Environmental Psychology, Hogskolan i Gävle, Sweden

Correspondence to Douglas MacCutcheon:

Douglas.MacCutcheon@hig.se

Editor-in-Chief: Bharath Chandrasekaran

Editor: Julie Anderson

Received January 11, 2019

Revision received April 9, 2019

Accepted July 23, 2019

https://doi.org/10.1044/2019_JSLHR-S-19-0012

Disclosure: The authors have declared that no competing interests existed at the time of publication

of sensory impairments, are known to affect speech perception (Bronkhorst, 2000; Ching, van Wanrooy, Dillon, & Carter, 2011; Dole, Meunier, & Hoen, 2014; Füllgrabe, Moore, & Stone, 2015; Johnstone & Litovsky, 2006; Litovsky, 2005; Lotfi, Mehrkian, Moossavi, Zadeh, & Sadjedi, 2016; Marsh et al., 2018).

The ability to separate target speech from noisy backgrounds has been described as the “cocktail party effect” (Conway, Cowan, & Bunting, 2001; Moray, 1959) and is an essential auditory skill required daily by children. In the classroom, for example, children must separate out the voice of the teacher from noise sources (e.g., ventilation systems, traffic) and the competing voices of fellow pupils. To achieve successful intelligibility under such conditions, speech perception combines auditory processing of the signal with cognitive processing of speech and spatial listening skills (i.e., localization). Spatial listening skills are primarily due to interaural differences, but monaural localization cues provided by the pinna also contribute to a lesser extent (Wightman & Kistler, 1997). Interaural cues such as the head shadow effect (Shaw, 1974) are based on level differences between the two ears, as the level of the sound is highest for the ear that is closest to the sound source. Additionally, timing differences of signals between the two ears similarly occur due to the differing relative proximities of the ears to the sound source (Zurek, 1993). Therefore, interaural time and level differences are helpful for localizing sound sources (B. C. J. Moore, 2012) so that sounds co-occurring in space are successfully grouped together perceptually and perceived separately when they come from different positions (Bregman, 1994). However, this grouping comes at a cost, as sounds that co-occur are harder to differentiate from one another, particularly for children (Johnstone & Litovsky, 2006); the converse is also true, there is a significant benefit to spatially separating sounds, referred to as *spatial release from masking* (SRM; Freyman, Helfer, McCall, & Clifton, 1999).

The acoustical properties of background sounds also affect speech perception differently, depending on their characteristics, and are often categorized as energetic versus informational masking in the literature (Brungart, 2001; Lecumberri, Cooke, & Cutler, 2010). The masking effect of energetic maskers, such as steady-state wideband noise, is primarily produced as a result of overlapping energy representations of the target speech and masker signals on the basilar membrane, thereby impairing speech intelligibility (Brungart, 2001; for additional modulation masking produced by steady-state noises, see Stone, Füllgrabe, Mackinnon, & Moore, 2011; Stone, Füllgrabe, & Moore, 2012). Informational maskers (e.g., one or more competing talkers) also provide energetic masking but have an additional effect of speech intelligibility, which is attributable to the similarity of the acoustic information in target and masker, which leads to informational interference (Dole, Hoen, & Meunier, 2012; Stone et al., 2012). Therefore, informational masking produces poorer speech perception in children than energetic masking (Wightman & Kistler, 2005; Wightman, Kistler, & Brungart, 2006) as a result of the acoustic similarity of

grouped speech sources, which results in perceptual confusion (Brungart, 2001). Furthermore, similarities in language and semantic content lead to difficulties at the phonetic and semantic levels of processing (Brouwer, Van Engen, Calandruccio, & Bradlow, 2012; Schneider, Li, & Daneman, 2007).

Although children are worse than adults at processing speech in noise (Hall, Grose, Buss, & Dev, 2002), the benefits of spatial separation of sound sources in children are not consistently higher, despite the still-developing auditory system, and vary depending on the type of masker. Specifically, SRM is less pronounced when the target speech occurs in the presence of informational as opposed to energetic maskers, as has been shown in a number of studies (Ihlefeld & Shinn-Cunningham, 2008; Johnstone & Litovsky, 2006; Litovsky, 2005; Oh, Wightman, & Lotfi, 2001). A number of studies demonstrate that SRM in children is highly variable, and findings indicate that spatial cues are helpful, presumably for segregating auditory streams in the presence of informational maskers. For example, children aged 5–8 years and adults repeated back spondees presented with informational (spoken sentences) or energetic (speech-shaped noise [SSN]) maskers (Litovsky, 2005). The maskers were either collocated or spatially separated to the right from the centrally located target speech. SRM with the speech masker was 5.7 dB for the children and 0 dB for the adults. Children also showed much higher variability in SRM. Johnstone and Litovsky (2006) assessed the benefit derived from spatially separating auditory sources, that is, spatial release from informational and energetic masking, in adults and children aged 5–7 years by evaluating the perception of spondees in either white noise or speech that was either unprocessed or time-reversed (the latter has the same spectrotemporal complexity as unprocessed speech but lacks semantic information). Maskers were either collocated with the talker at 0° azimuth or spatially separated by 90° to the right or left. In children, all three maskers proved to be equally problematic in collocated conditions. However, when spatially separated, spatial cues significantly improved speech reception thresholds (SRTs) by 3.4 and 6.7 dB for unprocessed and time-reversed speech, while SRM in white noise was only 0.5 dB and nonsignificant. Cameron and Dillon (2007) evaluated SRM in children aged 5–11 years. SRTs for sentences presented collocated at 0° azimuth or spatially separated by 90° to the right or left were measured, presented with distracting talkers reading stories for children; the distractors had either the same or different voice to the target. Same-voice distractors produced SRM of 9.3 dB, and different-voice distractors produced SRM of 11.3 dB. A central hypothesis of this study is that some of this variability could be accounted for by differences in cognitive ability, although it is not yet known to what extent individual differences in cognitive ability in children (without diagnosed cognitive or sensory deficits) contribute to spatial listening. However, due to findings suggesting that children with suspected central auditory processing disorder (Cameron & Dillon, 2008) and hearing loss (Ching et al., 2011) benefit less from spatial separation,

the contribution of individual differences in cognitive ability seems plausible. Although no research has currently been provided regarding how individual differences in working memory (WM) contribute to spatial listening and SRM differences in children, why the advantages of better inhibition of nontarget sounds and better phonological processing for auditory stream segregation have been hypothesized in this study will now be explained.

It has been suggested that individual differences in WM capacity are linked to spatial listening because of higher inhibition of competing sounds in individuals with higher memory span (Conway et al., 2001). The function of WM is theorized to keep desired (as opposed to irrelevant) objects of perception in awareness long enough for cognitive processing to occur (Baddeley & Hitch, 1974). WM has been measured in children using, for example, backward digit span (BDS; St Clair-Thompson, 2010), and is considered to be synonymous with executive attention (Engle, 2002). Engle (2002) emphasizes that the capacity aspect of WM is more indicative of the ability to control attention to retrieve and actively maintain stored information as opposed to simply a measure of the limitations of storage. Conway et al. (2001) measured the WM capacity of a sample of adults using a memory span task and assessed participants on performance on a dichotic listening task. Irrelevant speech (to be ignored) was presented to the first ear, while the second was presented with speech, which they were required to attend to and “shadow” (repeat aloud what was heard). After a period, the participant’s name was included in the to-be-ignored speech, and the participant’s ability to inhibit distracting information was measured by whether or not they heard their name. Sixty-five percent of the low WM capacity group reported hearing their name as opposed to 20% in the high WM capacity group. The researchers concluded that lower WM capacity was linked with an inability to suppress irrelevant auditory information. In a study exploring developmental changes in the effects of irrelevant sounds (the “irrelevant sound effect”; Beaman & Jones, 1997) on WM, performance on a serial recall task in the presence of a variety of irrelevant sounds were compared between children and adults (Elliott, 2002). The ability to inhibit irrelevant speech was reported to be detrimental to serial recall in both age groups, but improved with age. Sounds that changed more, such as irrelevant speech sounds as opposed to irrelevant tones, were more detrimental for performance in children than adults, referred to as the “changing-state effect” (Elliott, 2002). The improvement was theoretically linked to the development of attentional control in children, which improves with age (Cowan, Nugent, Elliott, Ponomarev, & Saults, 1999). Another way in which WM is thought to be linked to speech reception is through phonological processing (Groeger, Field, & Hammond, 1999). Memory span tasks, such as forward digit span and BDS, present lists of numbers for immediate recall, either in the same serial order of presentation or the reverse. As the lists become successively longer, Baddeley (2000) claims that the temporarily active memory traces and subvocal repetition mechanism in phonological WM (the “phonological loop”) are put under

increasing strain. With the addition of challenging environmental factors such as noise or a loss of spatial cues, mismatches in implicit phonological processing result (Rönnberg et al., 2008).

Knowledge of the linguistic structure of language has been shown to shape speech perception, but how individual differences in language ability could be linked to benefits in SRM in children has not been explored. In the case of two competing talkers, successful language processing relies on the ability to attend to one talker over the other (i.e., successful auditory stream segregation), and although the boundary between language processing and speech perception is not entirely clear, auditory stream segregation is theorized to occur as a precursor to language processing (Cooke, Garcia Lecumberri, & Barker, 2008). This would indicate that language processing would not affect auditory stream segregation of speech. However, a native language benefit for speech reception masked by informational maskers has been shown in the area of second language listening and referred to as the “foreign language cocktail party problem” (Cooke et al., 2008), indicating that language ability and familiarity somehow assists with speech-in-noise perception. Indeed, the latter is more difficult for bilinguals than monolinguals, but results indicate that SRTs are lower (better) the earlier a language is acquired (Mayo, Florentine, & Buus, 1997). Johnson (2011) states that the exact manner in which knowledge-driven processes assist speech perception is not entirely agreed upon, but the interaction of language-related knowledge-driven processes with speech processing might explain why individual differences in language ability modulate speech perception under acoustically challenging conditions. Evidence indicates that linguistic experience begins to shape speech perception in infancy, at the time that speech sounds begin to acquire meaning from approximately 6 months of age (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). Findings by Ganong (1980) indicate that linguistic knowledge shapes speech perception in various ways, as listeners are more likely to hear acoustically and linguistically similar words in place of the non-words they are presented with (Ganong, 1980). Furthermore, when parts of words are replaced with noise, listeners tend to still hear the missing phoneme (called *phoneme restoration*; Samuel, 1991). Therefore, those with better knowledge of the language they are listening to are likely to fill in the gaps in perception more effectively.

Exploring how SRM might be modulated by individual differences in WM and language ability in children is the main focus of this experiment. The central hypothesis of this study is that differences could assist auditory stream segregation when sounds are collocated and more efficient use of spatial cues when sounds are separated (i.e., SRM). More specifically, higher WM could be indicative of more efficient matching of incoming phonological input with stored phonological representations in long-term memory (Rönnberg, Rudner, Lunner, & Zekveld, 2010) and better inhibition of distraction (Tun, O’kane, & Wingfield, 2002). Higher expressive language (EL) ability is indicative of the level of familiarity with the linguistic structure of language,

with hypothetical advantages for speech reception in challenging conditions where children need to fill in the perceptual gaps and utilize speech reception advantages provided by spatial cues. Participants were therefore assessed on memory span and language tasks suitable for their age group. Speech-in-noise scenarios were reproduced in a plausible virtual acoustic school classroom with simulated room acoustics. Speech-in-noise perception was measured adaptively under two spatial locations and masked by two different noise types, namely, spectrally matched speech-shaped white noise and a single talker. Based on an SRM paradigm used by Cameron and Dillon (2007), spatial locations of the target speech and noise masker were either collocated at 0° azimuth or spatially separated at 90° azimuth to the right. It was expected that individual differences in memory span and EL would interact with spatial listening as a function of ability and of the type of background masker used. More specifically, children with higher memory and language abilities were predicted to cope better with challenging acoustic conditions and to derive greater benefit from the presence of spatial cues.

Method

Participants

Thirty-nine male English-speaking children from a participating school for boys in South Africa took part in this study. The mean age was 6 years 3 months ($SD = 7$ months, range: 4 years 11 months to 7 years). Participants with a history of cognitive, sensory, or behavioral deficit, based on parental report, were excluded. All participants passed a hearing loss screening test using the smartphone Android OS application hearScreen that detects hearing losses of greater than 20 dB HL (at 1, 2, and 4 kHz) in 97.8% agreement with standard manual audiometry (Swanepoel, Myburgh, Mohamed, & Eikelboom, 2014). The application was run on Samsung Galaxy Pocket Plus S5301 phones connected to Sennheiser HD202 II headphones calibrated to prescribed standards (ANSI/ASA S3.6-2010; ISO 389-1:1998) for TDH 39 supra-aural headphones (described in Swanepoel et al., 2014). Assessments were conducted in a sound-isolated music room of the school. Ethical approval for the study was granted by the University of Pretoria Research Ethics Committee, Approval 25071999 (GW20171130HS), and parents provided written consent for their children to participate in the study.

Tasks

Expressive Language

The Renfrew Action Picture Test (Renfrew, 2011) was used to assess EL. The test requires participants to verbally describe 10 pictures (e.g., a girl hugging a teddy bear), and responses are scored according to information and grammar content. The Renfrew Action Picture Test yields information and grammar subscores (out of a maximum of 41 and 36, respectively), which were averaged to form a language ability score that was used in the analysis

as the EL variable. Therefore, to assess the effect of EL ability, two median-split EL groups were created, with scores below or equal to the median being assigned to the “low EL” group and those above the median being assigned to the “high EL” group.

Memory Span

The subtest Number Repetition–Backward from the Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel, Wiig, & Secord, 2003) was used to assess the memory span capacity. This version of a BDS test consists of 16 number sequences, ranging in length from two to nine digits, with each sequence length occurring twice. Participants are required to recall the sequence in reverse order immediately after hearing it. One point is awarded for every sequence that is correctly recalled. The maximum score is 16, and the test is discontinued after participants get two sequences of the same length wrong. To assess the effect of memory span capacity, two median-split groups (referred to as “low BDS” and “high BDS”) were created based on the same rule as used for the creation of the EL groups.

Speech-in-Noise Perception

The speech-in-noise paradigm is an adapted version of the Children’s Coordinate Response Measure (described in Vickers, Degun, Canas, Stainsby, & Vanpoucke, 2016). All target sentences followed the form “Show the dog where the <color><number> is!” where the call sign “color” could be “black,” “red,” “green,” “white,” “blue,” or “pink” and the call sign “number” was a number between one and nine, omitting the disyllabic number seven. Speech perception was assessed by measuring SRTs at 50% correct speech intelligibility. SRTs were measured in four experimental conditions obtained by combining two spatial combinations so that either the masking background noise was collocated with the target speech or spatially separated from the target speech at 90° azimuth to the right, with two background noise types—either SSN or speech that differed from the target speech in content and talker voice. The playback level of the background was fixed at 55 dB (A) for all experimental conditions, and the starting level of the target speech was 68 dB(A). SRTs were obtained using an adaptive up–down procedure with variable step sizes. The signal-to-noise ratio (SNR) in the next trial was either decreased or increased by changing the level of the target speech based on an up–down procedure (Levitt, 1971). The initial step size to change the SNR was 8 dB. After the first and second reversal, the step size decreased to 4 and 2 dB, respectively, converging at 50% positive responses. Thereafter, the participants needed an additional five reversals to finish the block. The SRT was then calculated based on averaged SNR values of the last four reversals (Halliday, Tuomainen, & Rosen, 2017). An average of the SNR values of the last four reversals were taken as the SRT, and the adaptive track ended automatically after four reversals had been achieved, terminating the test. Although the total number of potential trials within which to achieve the seven

required reversals was limited by the number of potential color/number combinations (i.e., 48), all participants achieved all seven reversals well within 48 trials.

Experimental Procedure and Stimuli

Speech-in-Noise Test

Each subject sat in front of a DELL Latitude E6430 laptop with a 14-in. display, and auditory stimuli were presented over a Focusrite Scarlett 2i2 audio interface using Sennheiser HD 650 headphones. All experimental conditions were tested in a four-block within-subject design, which was counterbalanced using a Latin square. Per condition, 48 different color/number combinations were randomly permuted for testing. A graphical user interface showed a picture of a dog next to six colored panels with numbered buttons representing all possible color/number combinations. Participants were instructed to repeat the correct color/number combination, and the investigator clicked on the corresponding button on the screen. The length of the target and the background was the same, and no two noise samples were the same (random sections of SSN and background speech were selected by the software). A maximum SNR of 20 dB limited the target speech level to 75 dB(A), so as to protect the participants' hearing.

Target and Noise Stimuli

All speech materials were recorded anechoically at a sampling rate of 44.1 kHz at 24-bit resolution using a Rode NT1A large-diaphragm condenser microphone and a Focusrite Scarlett 2i2 audio interface. The speech material for the background speech consisted of the 19 English news items in News Items 1 and 2 from the section "Focus: Listening" (Nilsson, 2016), which form part of the National Assessment Project. These news items were recorded with an adult male talker. After removing longer pauses between words and sentences that broke the natural flow of the masker, all news items were normalized to a common root-mean-square value. SSN was created by deriving 2^{11} linear predictive coding coefficients from the news items, which were subsequently used to filter zero-mean white Gaussian noise to achieve the same long-term average speech spectrum as the news items. The final SSN signal had the same length as the concatenated news items material.

Simulation of the Virtual Acoustic Environment

Room acoustics were simulated based on measurements taken in accordance with German standards (DIN3382-2, 2008) in a typical classroom with a mean midfrequency reverberation time T_{30} of 0.6 s (based on the arithmetic mean of the RTs between the 0.5 and 1 kHz octave bands). The software RAVEN (Room Acoustics for Virtual Environments; Schröder, 2011) was used for the simulation. Binaural room impulse responses were simulated based on a head-related transfer functions measured from a child dummy head (Fels, Buthmann, & Vorländer, 2004).

Results

Individual Difference Measures

Data were analyzed using IBM SPSS Statistics software, and plots were produced in R programming language and environment for statistical computing. When averaged across all spatial locations and background noise types, the overall mean SRT was -0.82 dB ($SD = 2.87$, 95% CI $[-1.75, 0.11]$). The mean SRT across collocated conditions was 0.65 dB ($SD = 2.72$, 95% CI $[-0.23, 1.53]$), the mean SRT across separated conditions was -2.29 dB ($SD = 3.93$, 95% CI $[-3.56, -1.01]$), the mean SRT across SSN conditions was -3.85 dB ($SD = 3.72$, 95% CI $[-5.06, -2.65]$), and the mean SRT across single-talker conditions was 2.21 dB ($SD = 3.17$, 95% CI $[1.18, 3.24]$). The mean BDS score was 4.21 ($SD = 1.81$, 95% CI $[3.62, 4.79]$) out of a maximum score of 8. Average EL was 24.17 ($SD = 5.16$, 95% CI $[22.50, 25.84]$) out of a maximum score of 37.5.

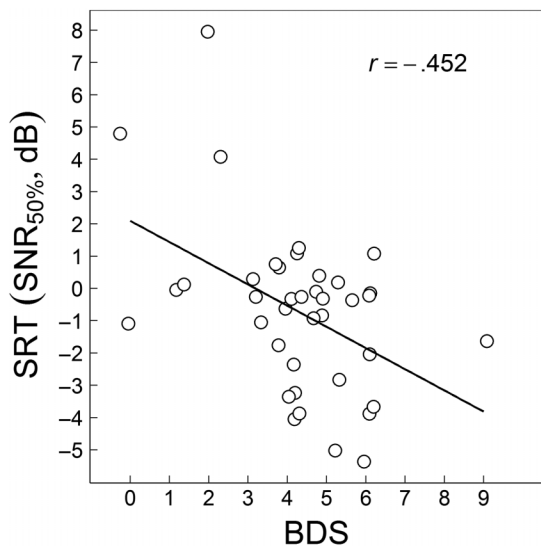
Multiple Regression

After testing for outliers (>3 SDs), a multiple linear regression was calculated to predict SRTs based on age, BDS, and EL. The part and partial correlations matrix provided in the regression output indicated that BDS was significantly negatively correlated with SRTs, $r = -.45$, $p = .002$, positively correlated with EL, $r = .36$, $p = .012$, and age, $r = .40$, $p = .006$. In the multiple regression, a significant relationship was found between the predictors and SRTs averaged across spatial locations and background types, $F(3, 35) = 7.648$, $p < .001$, with an R^2 of .40, therefore accounting for 40% of the variance in SRTs. However, only BDS was significant predictor of the variance in SRTs. Figure 1 illustrates that participants' predicted SRTs improved significantly ($p = .010$) by 0.96 dB for every 1-point improvement in BDS, improved nonsignificantly ($p = .729$) by 0.06 dB for every 1-point improvement in EL, and worsened nonsignificantly by 0.21 dB for every 1 point increase in age.

Analysis of Variance

A factorial four-way repeated-measures analysis of variance design was used with two within-subject factors spatial location (two levels: 0° , 90°) and background noise type (two levels: SSN, speech) and two between-subjects factors BDS (two levels: high BDS, low BDS) and EL (two levels, high EL, low EL). After BDS and EL were dichotomized into high and low groups, independent-samples t tests indicated that groups did not differ significantly in age: BDS groups, $t(37) = 0.390$, $p = .695$, $d = 0.14$ (low BDS $M_{\text{age}} = 75.9$ months, $SD = 6.9$ months; high BDS $M_{\text{age}} = 75$ months, $SD = 6.3$ months) and EL groups, $t(37) = -0.828$, $p = .601$, $d = 0.20$ (low EL $M_{\text{age}} = 74.7$ months, $SD = 6.7$ months; high EL $M_{\text{age}} = 76$ months, $SD = 6.4$ months). Analysis of variance results indicated significant main effects of spatial location, $F(1, 35) = 37.83$, $p < .001$, $\eta_p^2 = .52$, and background noise type, $F(1, 35) = 86.87$,

Figure 1. Scatter plot of backward digit span (BDS) scores and speech reception thresholds (SRTs). A significant regression indicated that BDS predicted variance in SRTs. Lower SRTs indicate better performance. For better visibility, overlapping data points have been slightly offset (i.e., jittered). SNR = signal-to-noise ratio.

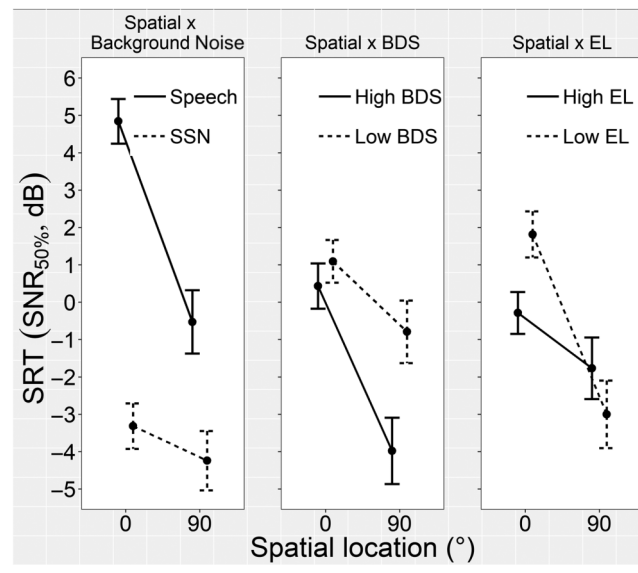


$p < .001$, $\eta_p^2 = .71$. SRM was 3.15 dB between the collocated conditions ($M = 0.76$ dB, $SE = 0.42$, 95% CI [-0.09, 1.61]) and separated conditions ($M = -2.39$ dB, $SE = 0.31$, 95% CI [-3.63, -1.15]). SRTs in the presence of background speech ($M = 2.16$ dB, $SE = 0.51$, 95% CI [1.12, 3.20]) were 5.94 dB higher (i.e., worse) than SRTs in the presence of SSN ($M = -3.79$ dB, $SE = 0.60$, 95% CI [-5.00, -2.57]). There was a significant between-subjects effect of BDS, $F(1, 35) = 4.45$, $p = .042$, $\eta_p^2 = .113$, with SRTs that were 1.93 dB lower (i.e., better) in the high BDS group ($M = -1.77$ dB, $SE = 0.67$, 95% CI [-3.12, -0.43]) than low BDS group SRTs ($M = 0.15$ dB, $SE = 0.63$, 95% CI [-1.12, 1.42]).

Simple effects analyses were conducted to investigate interaction effects, which are visualized in Figure 2. Estimated marginal means, standard deviations, and 95% confidence intervals for the interaction results presented in this section are given in Table 1. The two-way interaction between factors spatial location and background noise type was significant, $F(1, 35) = 8.47$, $p = .006$, $\eta_p^2 = .195$. When the background was speech, SRM was 5.37 dB, $F(1, 35) = 26.17$, $p < .001$, $\eta_p^2 = .428$, but SRM was not in evidence when the background noise was SSN, $F(1, 35) = 1.45$, $p = .237$, $\eta_p^2 = .040$. In the collocated condition, SRTs for speech backgrounds were significantly higher (worse) by 8.16 dB than those for SSN, $F(1, 35) = 87.19$, $p < .001$, $\eta_p^2 = .714$, and this difference was reduced to 3.71 dB in the spatially separated condition, $F(1, 35) = 11.35$, $p = .002$, $\eta_p^2 = .245$.

The spatial location by BDS group interaction was also significant, $F(1, 35) = 6.07$, $p = .019$, $\eta_p^2 = .148$. Simple effects analysis results indicated that speech-in-noise perception did not differ between the two BDS groups in the

Figure 2. Speech reception thresholds (SRTs) as a function of spatial location (0°, 90°) and background noise type (speech, speech-shaped noise, high/low backward digit span (BDS) group, and high/low expressive language (EL) group). Error bars indicate ± 1 SE. *Left:* The speech background noise (Speech) resulted in higher SRTs than speech-shaped noise (SSN) across spatial conditions, and significant spatial release from masking was observed only within the speech conditions. *Middle:* Low and high BDS performers significantly benefitted from the addition of spatial cues, but the low BDS group had greater spatial release from masking than the high BDS group. *Right:* Low and high EL performers significantly benefitted from the addition of spatial cues, but the low EL group benefitted more. SNR = signal-to-noise ratio.



collocated condition, $F(1, 35) = 0.633$, $p = .432$, $\eta_p^2 = .018$. Both high and low BDS groups benefitted significantly from additional spatial cues; the low BDS group's SRM was 1.87 dB, $F(1, 35) = 7.20$, $p = .011$, $\eta_p^2 = .171$, in relation to 4.41 dB in the high BDS group, $F(1, 35) = 35.12$,

Table 1. Estimated marginal means, standard errors, and 95% confidence intervals (CIs) of speech reception thresholds per experimental condition.

Conditions		M (dB)	SE	CI 95% [lower, upper]
0° Azimuth	SSN	-3.32	0.61	[-0.456, -2.08]
	Speech	4.84	0.60	[3.63, 6.05]
	Low BDS	1.09	0.57	[-0.07, 2.26]
	High BDS	0.43	0.61	[-0.80, 1.66]
	Low EL	1.81	0.62	[0.56, 3.06]
	High EL	-0.29	0.56	[-1.43, 0.85]
90° Azimuth	SSN	-4.24	0.80	[-5.86, -2.63]
	Speech	-0.53	0.85	[-2.25, 1.19]
	Low BDS	-0.79	0.84	[-2.50, 0.91]
	High BDS	-3.98	0.89	[-5.78, -2.18]
	Low EL	-3.00	0.90	[-4.84, -1.17]
	High EL	-1.76	0.82	[-3.44, -0.10]

Note. Lower speech reception thresholds indicate better performance. SSN = speech-shaped noise; BDS = backward digit span; EL = expressive language.

$p < .001$, $\eta_p^2 = .501$. However, the amount of SRM was significantly higher by 3.19 dB in the high BDS group, $F(1, 35) = 6.797$, $p = .013$, $\eta_p^2 = .163$.

Finally, the spatial location by EL interaction was also significant, $F(1, 35) = 10.63$, $p = .002$, $\eta_p^2 = .233$. The simple effects analysis indicated that both groups benefitted significantly from SRM, the high EL group benefitting by 1.48 dB, $F(1, 35) = 4.60$, $p = .039$, $\eta_p^2 = .116$, and the low EL group benefitting by 4.8 dB, $F(1, 35) = 40.55$, $p < .000$, $\eta_p^2 = .537$. Across separated conditions, high and low EL groups SRTs were not significantly different, $F(1, 35) = 1.02$, $p = .319$, $\eta_p^2 = .028$, but across collocated conditions, the low EL group's SRTs were significantly higher (worse) by 2.1 dB, $F(1, 35) = 6.34$, $p = .017$, $\eta_p^2 = .153$.

Discussion

A significant negative correlation of moderate strength between BDS scores and SRTs, with better speech-in-noise perception in the high BDS group, suggests a relationship between span-related aspects of WM capacity and speech-in-noise perception in young normal hearing children, which is further explained by interaction effects. This extends previous findings of significant correlations between BDS scores and speech-in-noise perception in normal hearing adults (Füllgrabe et al., 2015; Humes, Lee, & Coughlin, 2006) to the lower end of the life span. Although high and low BDS groups had almost equal average ages, the positive correlation between age and BDS indicated that developmental advantages were partly responsible for having better BDS in the sample. One way in which WM might be related to speech reception is explained by Rönnberg et al.'s (2008) ease of language understanding model in which WM is linked to improvements in matching of phonological input with stored phonological representations in long-term memory (Rönnberg et al., 2010). Francis and Nusbaum (2009) posit that perception of speech is dependent on the amount of WM resources available for the task, which are subject to capacity limitations. Therefore, when speech is masked, WM resources are spread more thinly and shared among other cognitive tasks, which might account for the reduction in intelligibility. Although a meta-analysis by Füllgrabe and Rosen (2016) suggest that, across a number of studies, WM has not been consistently associated with speech-in-noise perception in adults, they speculated that, even in normal hearing listeners, more WM resources may be required with age to compensate for the consequences of age-related deficits in suprathreshold auditory processing (Füllgrabe, 2013; Füllgrabe & Moore, 2018) on the ability to process speech in the presence of background sounds. Therefore, a similar hypothesis could be made for speech-in-noise perception in young children, as basic auditory processing abilities are still maturing in this population (e.g., D. R. Moore, Ferguson, Edmondson-Jones, Ratib, & Riley, 2010).

Perceptually grouping sounds with their respective sources (i.e., auditory stream segregation) is an essential constituent of speech perception in complex acoustic scenes (Bregman, 1994) and becomes more difficult when sound

sources are closer together. Therefore, SRTs were predicted to be better when the background noise was spatially separated from the target speech as a result of SRM, which has been found in a number of studies in both children and adults (Johnstone & Litovsky, 2006; Litovsky, 2005). A significant main effect of spatial location corroborated these findings in children, with collocated conditions being 3.15 dB higher than spatially separated ones. These results suggest that spatial cues (i.e., primarily interaural time and level differences) present in spatially separated conditions might be beneficial for speech perception by helping to define the directional properties of incoming sounds. They could therefore theoretically assist the process of auditory stream segregation in complex acoustic scenes such as classrooms.

It was also expected that the speech background would yield poorer SRTs than the SSN in children due to informational masking being highly effective in children (Wightman, Kistler, & Brungart, 2006). The effectiveness of this masker is possibly due to attentional capture produced by intelligible semantic content and the changing/fluctuating as opposed to constant/steady-state nature of the sound (i.e., changing-state effect; Elliott, 2002) in addition to the energetic masking, which is also present in an informational masker. The background types used in this study were selected for the purpose of providing an easy (energetic) versus difficult (informational) masking conditions that were expected to produce SRM differences that might be further modulated by cognitive abilities. Overall, SRTs were worse for speech backgrounds than SSN. A significant main effect indicated a 5.94-dB difference between energetic and informational masker conditions. This is consistent with Wightman and Kistler (2005), who found that informational maskers produce poorer speech reception and attribute this to low selective attention in children when the masker is informational. However, Litovsky (2005) found no significant SRT differences due to the nature of the masker. Johnstone and Litovsky (2006) also found that there was no significant main effect of energetic versus informational masker type, but post hoc tests revealed that the presence of spatial cues modulated this effect in children. Specifically, energetic masking SRTs (spatially separated to the right) were significantly higher than those for informational maskers, but this effect was not present when the target and masker were collocated.

In the spatial location by background noise type interaction, results are consistent with the literature. The speech background noise yielded significant SRM of 5.4 dB; however, with SSN that had the same spectral and virtual acoustic properties (i.e., the same long-term average speech spectrum), no significant SRM was observed. Litovsky (2005) found SRM of 5.7 dB when the masker was informational compared to a nonsignificant effect when the masker was energetic. Johnstone and Litovsky (2006) indicated a benefit of 3.4 dB for the informational background noise compared to a nonsignificant effect for the energetic background noise, and Cameron and Dillon's (2007) results were 9.3 dB (distractor has same voice as target) and 11.3 dB

(different-voice distractors). These findings indicate that the perception of speech in the presence of energetic maskers does not benefit from spatial cues (interaural time and level differences), but that spatial cues become beneficial when the masker is similar to the target (Litovsky, 2005). The effect of energetic masking on speech reception is at the peripheral rather than cognitive levels of speech processing; therefore, SRM advantages could be expected because spatial cues provide clues as to which sound source to attend to at the cognitive level. That is why greater SRM occurs when speech is masked by informational maskers, as such maskers interfere with the cognitive processes used to disentangle similar sounds (such as inhibiting irrelevant stimuli that co-occurs with the target, causing distraction or confusion at the semantic levels of processing) and are therefore more likely to benefit from spatial cues.

Additionally, Glyde et al. (2013) posited that one of the spatial cues the auditory system utilizes in order to assist with SRM is “better-ear glimpsing,” which is particularly effective for informational maskers. Better-ear glimpsing uses the head shadow effect on interaural-level differences to build up a representation of the signal by attending primarily to the ear with the best SNR (Glyde et al., 2013). In the case of the present experiment, the target was always presented at 0° azimuth, and the masker was presented 90° azimuth to the right in spatially separated conditions. Therefore, under separated conditions, the left ear would have benefitted from a higher (i.e., more favorable) SNR due to the head shadow effect, resulting in an interaural-level difference of the masker at the left ear, and a better-ear glimpsing strategy could therefore have been easily adopted to disentangle the similar target and masker. Furthermore, as the informational masker was a single talker with almost natural speech prosody, prosodic fluctuations allow the listener “glimpses” of spectrotemporal regions in which the masker was less/not present, however briefly, as described in Cooke’s glimpsing model of speech-in-noise perception (Cooke, 2006). Therefore, the advantages of better-ear glimpses could have been compounded with spectrotemporal glimpses.

Although both BDS groups showed significant SRM, the high BDS group benefitted more from the addition of spatial cues by over 3 dB. However, both groups were equally disadvantaged by the lack of spatial cues. As no other investigations of individual differences in WM capacity explaining SRM advantages for speech masked by informational maskers in young children exist, we looked into associated literature for comparative results. The link between cognitive ability and auditory stream segregation ability has been explored in children with a central auditory processing disorder (Lotfi et al., 2016). Results indicated that auditory stream segregation abilities began to covary with WM capacity as simultaneously presented 500- and 800-Hz tones became increasingly closer together (30° and 0°, respectively). These findings were extended to the speech domain in the current study, in which lower cognitive (specifically, WM capacity) performers’ auditory stream segregation abilities for competing talkers were also less benefitted by spatial cues. These results could indicate that higher

cognitive abilities related to WM capacity, such as executive attentional control (Engle, 2002) and inhibition of interfering sounds (Conway et al., 2001), provide benefits for perception of spatial cues that assist in auditory stream segregation of competing sounds (Bregman, 1994). Conway et al. (2001) showed that adults with lower WM capacity self-reported hearing their name in a to-be-ignored irrelevant message 45% more than a high WM capacity group, which suggests that WM capacity is also likely to be involved in the disentangling of co-occurring sounds in children.

Both EL groups showed significant SRM, producing similar SRTs when spatial cues were present. However, the low EL group was significantly disadvantaged by 3.3 dB in collocated conditions relative to the high EL group. This disadvantage could be attributable to the loss of spatial cues because the introduction of spatial cues in separated conditions led to a comparatively large and significant improvement in SRTs, of nearly 5 dB, but only in the low EL group. However, these results suggest that those with poorer language ability are more reliant on the presence of extralinguistic (e.g., spatial) cues to identify separate auditory streams. The usefulness of nonlinguistic cues, such as speech rate, has been shown for speech reception in the presence of a distractor in bilinguals (Cooke et al., 2008), and this could be analogous, provided this group is viewed as (in some sense) equivalent to low native language performers when performing second language tasks. However, this interpretation of these results is limited because it is unknown why language familiarity and knowledge might be linked to benefits in auditory stream segregation, which is considered to be a primarily signal-driven process, although this is debated (Cooke et al., 2008). The alternate and perhaps more likely interpretation is that the high EL group was better able to use cognitive/language processes to make up for speech perception deficits due to a loss of spatial information, resulting in 3.3-dB better speech reception than the low EL group in the collocated condition.

Limitations

A primary limitation of this study is that, because it is unclear how language ability is related to auditory stream segregation, the implications of why individual differences in language ability might be related to SRM were rather limited in scope. Further research is needed to ascertain the boundaries between signal-based processes, on the one hand, and cognitive and linguistic processes, on the other, and the way in which they interact with individual differences in ability. Particularly, the contribution of other nonlinguistic cues (e.g., speech rate, fundamental frequency, level differences) should be investigated to establish if these are collectively a greater source of benefit in speech-in-noise reception in low language performers. Furthermore, because a nonstandard assessment of hearing sensitivity was used for screening for hearing loss in participants, hearing thresholds were only measured up to 4 kHz and not 8 kHz or even 16 kHz as in standard manual audiometry. Therefore, a more accurate representation of the hearing ability of our sample would

have been desirable. In particular, losses in high-frequency sensitivity could have been investigated, which, despite not resulting in failing the hearing screening used in this study, are also thought to contribute to poorer speech reception in children (Stelmachowicz, Pittman, Hoover, Lewis, & Moeller, 2004). Finally, dichotomization into high and low BDS and EL groups was undertaken for the purposes of establishing the role of individual differences in cognitive and language ability. This could be a limitation as dichotomizations generally lead to reductions in statistical power, which could potentially increase the risk of Type II errors (Cohen, 1983) and, in extreme cases, spurious patterns of significance (Maxwell & Delaney, 1993). However, it should also be noted that the use of p values as indicators of significance alone are questionable as they have been identified as poor representations of the magnitude and importance of an effect (Sullivan & Feinn, 2012). Hence, effect sizes were also reported in this article as an aid for increased interpretability of the findings.

Conclusions

The aim of this study was to explore how cognitive and linguistic abilities modulate the benefits of spatial separation (i.e., SRM) between a target talker and noise sources of different types. Results suggest that informational maskers are more effective than energetic maskers in masking target speech in children. A role of WM capacity and EL for SRM is indicated in this age group, but only when the masker was informational. Namely, higher WM abilities could be linked to better inhibition of distraction, phonological processing, and executive control, which in turn assist the utilization of spatial cues, when present, in this group. Poorer EL seems to be related to greater problems in auditory stream segregation in collocated conditions, and children falling into this group derive a greater benefit from spatial cues than the group with higher EL abilities. This effect is possibly linked to benefits of non-linguistic cues for speech reception in lower language performers, but before more conclusive claims can be made, links between peripheral auditory processing, cognitive abilities, and language processing need to be better understood. Lines of causality between auditory and cognitive factors and the development of the skills tested require longitudinal studies to be confirmed. The potential of auditory training for improving speech-in-noise perception in those with poorer WM and EL should also be evaluated, particularly types of training that engage general cognitive skills in a variety of ways, and already shows a strong theoretical and empirical basis for supporting speech reception processes, such as music (Patel, 2011; Strait, Parbery-Clark, Hittner, & Kraus, 2012).

Acknowledgments

The authors report grant funding from the Swedish Foundation for International Cooperation in Research and Higher Education (IB2017-7004 awarded to Robert Ljung). The authors would

like to thank the children, their parents, and the staff at Waterkloof House Preparatory School for taking part in this study, with special thanks to Melanie Fouché.

References

- American National Standards Institute/Acoustical Society America.** (2010). *Specification for audiometers (ANSI/ASA S3.6-2010)*. Washington, DC: Author.
- Baddeley, A. D.** (2000). The phonological loop and the irrelevant speech effect: Some comments on Neath (2000). *Psychonomic Bulletin & Review*, 7(3), 544–549.
- Baddeley, A. D., & Hitch, G.** (1974). Working memory. *Psychology of Learning and Motivation*, 8, 47–89.
- Beaman, C. P., & Jones, D. M.** (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(2), 459–471.
- Bregman, A. S.** (1994). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Bronkhorst, A. W.** (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica United with Acustica*, 86(1), 117–128.
- Brouwer, S., Van Engen, K. J., Calandruccio, L., & Bradlow, A. R.** (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *The Journal of the Acoustical Society of America*, 131(2), 1449–1464.
- Brungart, D. S.** (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, 109(3), 1101–1109.
- Cameron, S., & Dillon, H.** (2007). Development of the Listening in Spatialized Noise-Sentences Test (LISN-S). *Ear and Hearing*, 28(2), 196–211.
- Cameron, S., & Dillon, H.** (2008). The Listening in Spatialized Noise-Sentences Test (LISN-S): Comparison to the prototype list and results from children with either a suspected (central) auditory processing disorder or a confirmed language disorder. *Journal of the American Academy of Audiology*, 19(5), 377–391.
- Ching, T. Y., van Wanrooy, E., Dillon, H., & Carter, L.** (2011). Spatial release from masking in normal-hearing children and children who use hearing aids. *The Journal of the Acoustical Society of America*, 129(1), 368–375.
- Cohen, J.** (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7(3), 249–253.
- Conway, A. R., Cowan, N., & Bunting, M. F.** (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, 8(2), 331–335.
- Cooke, M.** (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, 119(3), 1562–1573.
- Cooke, M., Garcia Lecumberri, M. L., & Barker, J.** (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, 123(1), 414–427.
- Cowan, N., Nugent, L., Elliott, E., Ponomarev, I., & Saults, J.** (1999). The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development*, 70(5), 1082–1097.
- Davis, H.** (1964). Physiological and psychological functions in relation to anatomy and physiology. *International Audiology*, 3(2), 209–215.

- DIN3382-2, German Standard.** (2008). *Acoustics—Measurement of room acoustic parameters. Part 2: Reverberation time in ordinary rooms (No. 3382-2:2008)*. Berlin, Germany: German Institute for Standardization.
- Dole, M., Hoen, M., & Meunier, F.** (2012). Speech-in-noise perception deficit in adults with dyslexia: Effects of background type and listening configuration. *Neuropsychologia*, *50*(7), 1543–1552.
- Dole, M., Meunier, F., & Hoen, M.** (2014). Functional correlates of the speech-in-noise perception impairment in dyslexia: An MRI study. *Neuropsychologia*, *60*, 103–114.
- Elliott, E. M.** (2002). The irrelevant-speech effect and children: Theoretical implications of developmental change. *Memory & Cognition*, *30*(3), 478–487.
- Engle, R. W.** (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23.
- Fels, J., Buthmann, P., & Vorländer, M.** (2004). Head-related transfer functions of children. *Acta Acustica United With Acustica*, *90*(5), 918–927.
- Francis, A. L., & Nusbaum, H. C.** (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics*, *71*(6), 1360–1374.
- Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K.** (1999). The role of perceived spatial separation in the unmasking of speech. *The Journal of the Acoustical Society of America*, *106*(6), 3578–3588.
- Füllgrabe, C.** (2013). Age-dependent changes in temporal-fine-structure processing in the absence of peripheral hearing loss. *American Journal of Audiology*, *22*, 313–315.
- Füllgrabe, C., & Moore, B. C. J.** (2018). The association between the processing of binaural temporal-fine-structure information and audiometric threshold and age: A meta-analysis. *Trends in Hearing*, *22*, 1–14.
- Füllgrabe, C., Moore, B. C. J., & Stone, M. A.** (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. *Frontiers in Aging Neuroscience*, *6*, 347.
- Füllgrabe, C., & Rosen, S.** (2016). Investigating the role of working memory in speech-in-noise identification for listeners with normal hearing. In P. van Dijk, D. Başkent, E. Gaudrain, E. de Kleine, A. Wagner, & C. Lanting (Eds.), *Physiology, psychoacoustics, and cognition in normal and impaired hearing* (pp. 29–36). Cham, Switzerland: Springer.
- Ganong, W. F., III.** (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125.
- Glyde, H., Buchholz, J., Dillon, H., Best, V., Hickson, L., & Cameron, S.** (2013). The effect of better-ear glimpsing on spatial release from masking. *The Journal of the Acoustical Society of America*, *134*(4), 2937–2945.
- Groeger, J. A., Field, D., & Hammond, S. M.** (1999). Measuring memory span. *International Journal of Psychology*, *34*(5–6), 359–363.
- Hall, J. W., III., Grose, J. H., Buss, E., & Dev, M. B.** (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear and Hearing*, *23*(2), 159–165.
- Halliday, L. F., Tuomainen, O., & Rosen, S.** (2017). Language development and impairment in children with mild to moderate sensorineural hearing loss. *Journal of Speech, Language, and Hearing Research*, *60*(6), 1551–1567.
- Humes, L. E., Lee, J. H., & Coughlin, M. P.** (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *The Journal of the Acoustical Society of America*, *120*(5), 2926–2937.
- Ihlefeld, A., & Shinn-Cunningham, B.** (2008). Spatial release from energetic and informational masking in a selective speech identification task. *The Journal of the Acoustical Society of America*, *123*(6), 4369–4379.
- International Organization for Standardization (ISO).** (1998). ISO 389-1: 1998. *Acoustics—Reference zero for the calibration of audiometric equipment—Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones*. Geneva, Switzerland: Author.
- Johnson, K.** (2011). *Acoustic and auditory phonetics* (3rd ed.). Hoboken, NJ: Wiley.
- Johnstone, P. M., & Litovsky, R. Y.** (2006). Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults. *The Journal of the Acoustical Society of America*, *120*(4), 2177–2189.
- Kiessling, J., Pichora-Fuller, M. K., Gatehouse, S., Stephens, D., Arlinger, S., Chisolm, T., . . . von Wedel, H.** (2003). Candidature for and delivery of audiological services: Special needs of older people. *International Journal of Audiology*, *42*(Suppl. 2), 92–101.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B.** (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.
- Lecumberri, M. L. G., Cooke, M., & Cutler, A.** (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11–12), 864–886.
- Levitt, H.** (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467–477.
- Litovsky, R. Y.** (2005). Speech intelligibility and spatial release from masking in young children. *The Journal of the Acoustical Society of America*, *117*(5), 3091–3099.
- Lotfi, Y., Mehrkian, S., Moossavi, A., Zadeh, S. F., & Sadjedi, H.** (2016). Relation between working memory capacity and auditory stream segregation in children with auditory processing disorder. *Iranian Journal of Medical Sciences*, *41*(2), 110–117.
- Marsh, J. E., Ljung, R., Jahncke, H., MacCutcheon, D., Pausch, F., Ball, L. J., & Vachon, F.** (2018). Why are background telephone conversations distracting? *Journal of Experimental Psychology: Applied*, *24*(2), 222–235.
- Maxwell, S. E., & Delaney, H. D.** (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, *113*(1), 181.
- Mayo, L. H., Florentine, M., & Buus, S.** (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, *40*(3), 686–693.
- Moore, B. C. J.** (2012). *An introduction to the psychology of hearing* (6th ed.). Leiden, the Netherlands: Brill.
- Moore, D. R., Ferguson, M. A., Edmondson-Jones, A. M., Ratib, S., & Riley, A.** (2010). Nature of auditory processing disorder in children. *Pediatrics*, *126*(2), e382–e390.
- Moray, N.** (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60.
- Nilsson, S.** (2016). *Engelska 6—Exempel på uppgiftstyper [English 6—Examples of task types]*. Gothenburg, Sweden: Göteborgs universitet. Retrieved from http://nafs.gu.se/prov_engelska/exempel_provuppgifter/engelska_6_exempelupp
- Oh, E. L., Wightman, F., & Lutfi, R. A.** (2001). Children's detection of pure-tone signals with random multitone maskers. *The Journal of the Acoustical Society of America*, *109*(6), 2888–2895.
- Patel, A. D.** (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, 142.

- Pichora-Fuller, M. K., & Singh, G.** (2006). Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation. *Trends in Amplification, 10*(1), 29–59.
- Renfrew, C.** (2011). *Action Picture Test—Revised Edition*. Milton Keynes, England: Speechmark.
- Rönnerberg, J., Rudner, M., Foo, C., & Lunner, T.** (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology, 47*(Suppl. 2), S99–S105.
- Rönnerberg, J., Rudner, M., Lunner, T., & Zekveld, A. A.** (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise and Health, 12*(49), 263–269.
- Samuel, A. G.** (1991). A further examination of attentional effects in the phonemic restoration illusion. *The Quarterly Journal of Experimental Psychology Section A, 43*(3), 679–699.
- Schneider, B. A., Li, L., & Daneman, M.** (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology, 18*(7), 559–572.
- Schröder, D.** (2011). *Physically based real-time auralization of interactive virtual environments* (Vol. 11). Berlin, Germany: Logos Verlag Berlin.
- Semel, E., Wiig, E. H., & Secord, W. A.** (2003). *Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF-4)*. San Antonio, TX: The Psychological Corporation.
- Shaw, E. A. G.** (1974). Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *The Journal of the Acoustical Society of America, 56*, 1848–1861.
- Simpson, S. A., & Cooke, M.** (2005). Consonant identification in *N*-talker babble is a nonmonotonic function of *N*. *The Journal of the Acoustical Society of America, 118*(5), 2775–2778.
- St Clair-Thompson, H. L.** (2010). Backwards digit recall: A measure of short-term memory or working memory? *European Journal of Cognitive Psychology, 22*(2), 286–296.
- 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. *Archives of Otolaryngology—Head & Neck Surgery, 130*(5), 556–562.
- Stone, M. A., Füllgrabe, C., Mackinnon, R. C., & Moore, B. C.** (2011). The importance for speech intelligibility of random fluctuations in “steady” background noise. *The Journal of the Acoustical Society of America, 130*(5), 2874–2881.
- Stone, M. A., Füllgrabe, C., & Moore, B. C.** (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *The Journal of the Acoustical Society of America, 132*(1), 317–326.
- Strait, D. L., Parbery-Clark, A., Hittner, E., & Kraus, N.** (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language, 123*(3), 191–201.
- Sullivan, G. M., & Feinn, R.** (2012). Using effect size—Or why the *p* value is not enough. *Journal of Graduate Medical Education, 4*(3), 279–282.
- Swanepoel, D. W., Myburgh, H. C., Howe, D. M., Mahomed, F., & Eikelboom, R. H.** (2014). Smartphone hearing screening with integrated quality control and data management. *International Journal of Audiology, 53*(12), 841–849.
- Tsui, A. B., & Fullilove, J.** (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied Linguistics, 19*(4), 432–451.
- Tun, P. A., O’kane, G., & Wingfield, A.** (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging, 17*(3), 453.
- Vickers, D., Degun, A., Canas, A., Stainsby, T., & Vanpoucke, F.** (2016). Deactivating cochlear implant electrodes based on pitch information for users of the ACE strategy. In P. van Dijk, D. Başkent, E. Gaudrain, E. de Kleine, A. Wagner, & C. Lanting (Eds.), *Physiology, psychoacoustics and cognition in normal and impaired hearing* (pp. 115–123). Cham, Switzerland: Springer.
- Wightman, F. L., & Kistler, D. J.** (1997). Monaural sound localization revisited. *The Journal of the Acoustical Society of America, 101*(2), 1050–1063.
- Wightman, F. L., & Kistler, D. J.** (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America, 118*(5), 3164–3176.
- Wightman, F. L., Kistler, D. J., & Brungart, D.** (2006). Informational masking of speech in children: Auditory-visual integration. *The Journal of the Acoustical Society of America, 119*(6), 3940–3949.
- Zurek, P. M.** (1993). Binaural advantages and directional effects in speech intelligibility. In G. A. Studebaker & I. Hochberg (Eds.), *Acoustical factors affecting hearing aid performance* (2nd ed.). Boston, MA: Allyn & Bacon.