Remote-Frequency Masking and Speech Perception in Adults

Taylor L. Arbogast

A dissertation s~~S~~ubmitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

in

Partial ~~fulfillment F~~ fulfillment of the Requirements

for the degree of

Doctor of Audiology

Department of Communication Sciences and Disorders

May 2022

FACULTY COMMITTEE:

Committee Members: Yingjiu Nie, Ph.D.

Lincoln Gray, Ph.D.

Ayasakanta Rout, Ph.D. ~~and Lincoln Gray, Ph.D.~~

Style Definition: Heading 1: Line spacing: Double

Style Definition: Heading 2: Keep with next

Style Definition: TOC 1: Font: (Default) Times New Roman, Do not check spelling or grammar, Line spacing: Double, Tab stops: 6.24", Right,Leader: ...

Style Definition: TOC 2: Indent: Left: 0.15", Line spacing: Double, Tab stops: 6.24", Right,Leader: ...

Style Definition: TOC 3: Line spacing: Double, Tab stops: 6.24", Right,Leader: ...

Acknowledgments

This project would not be possible without the unwavering support and guidance of Dr. Yingjiu Nie, who has graciously leant me her knowledge and shared her passion for research with me over the past five years. Dr. Nie, you have been an instrumental part of my academic endeavors for the past six years and I am so thankful to have had you as a mentor. To Drs. Ayasakanta Rout and Lincoln Gray, my graduate experience would certainly not have been as enriched without your advice and experience in both my clinical and professional endeavors. To Husna Firdose, the completion of this project would not be possible without the countless hours you volunteered to help with figure creations, poster presentation, and statistical analysis. The value of your enthusiasm for research and willingness to assist with this project is immeasurable. To my fellow classmates and colleagues, none of this would be possible without the willingness of your gracious and giving spirits to donate your time and efforts for data collection amid an unpredictable and relentless pandemic. Lastly, to my classmates, the entire faculty of the Department of Communication Sciences and Disorders at James Madison University, and my family, thank you for your emotional, academic, and professional support throughout every phase of life the past four years.

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Table of Contents

List of Tables

List of Figures

Formatted: Font: 12 pt
Formatted: Font: Bold, Not Italic
Formatted: Font: Not Italic, No underline
Formatted: Font: 12 pt
Formatted: Font: 12 pt, Not Italic
Formatted: Font: 12 pt
Formatted: Font: Bold
Formatted: Font: Not Italic, No underline
Formatted: Font: 12 pt
Formatted: Font: 12 pt, Not Italic
Formatted: Font: 12 pt
Formatted: Font: 12 pt

Abstract

The primary purpose of this study is threefold: to use SRT measurements to examine the effect of various remote-frequency, narrowband maskers on adult's perception of narrowband speech, to compare the performance between low and high band speech stimuli, and to evaluate the combination of these approaches by examining the correlation between the masking effect observed with speech and pure tone stimuli. Twelve subjects aged 22-34, with hearing thresholds no worse than 15 dB HL for frequencies 500-8000 Hz, participated in two listening tasks. In the speech perception task, coordinate response measure (CRM) sentences and their maskers were separately filtered into two ½-octave wide frequency bands with respective center frequencies of 500 Hz (low-band) and 2500 Hz (high-band). Three types of maskers were utilized: Gaussian noise, CRM sentences spoken by a talker different from the talker of the target sentences (speech-masked conditions), and time-reversed CRM sentences. Speech reception thresholds (SRTs) of either low- or high-band sentences were assessed in quiet and in the presence of a high- or low-band masker. Speech recognition scores (SRSs), or the percentages of keywords correctly identified, were measured in the same conditions. In the informational masking task, detection thresholds of a 1 kHz tone were measured in quiet and in the presence of a muli-tonal masker. SRTs in quiet were found to be significantly higher than in GNB and reverse speech maskers. SRTs were also found to be lower for high band target speech. In the SRS task, only the forward speech masker produced significantly worse recognition scores. Using pure tone stimuli, an average masking effect of approximately 18 dB was observed across participants. The pure tone masking effect was not found to correlate

with the SRT masking effect, however, a trend of correlation appears to exist that may potentially reach significance with a larger sample size.

Formatted: Not Different first page header

Formatted: Heading 1, Indent: First line: 0", Line spacing: single

I. ~~Background~~Introduction

Formatted: Heading 1, Left, Line spacing: single

Existing research documents that children have more difficulty understanding speech in noise, thereby requiring a higher signal-to-noise ratio for adequate speech recognition (Hall & Grose, 1991; Litovsky, 2005; Leibold & Neff, 2011; Werner & Bargones, 1991; Youngdahl et al., 2018). More specifically, these studies have established that, when compared to adults, infants and young children demonstrate a susceptibility to masking when the masker is remote in frequency from the target signal. That is, the signal of interest and distractor do not overlap in the frequency domain. This type of masking is referred to as informational masking. The exact mechanisms to explain this discrepancy are unknown but given that the peripheral auditory filters reach maturity as early as 5 months of age, it is believed to be due to the maturation and development of more central structures (Hall & Grose, 1991). This difference poses a significant disadvantage for children, considering that one of their most important and frequented environments (i.e. the classroom) is characterized by a notoriously poor ~~signal to noise~~signal-to-noise ratio.

In addition to the maturation and of central structures, studies have also documented that adults appear to adopt specific listening strategies to assist with understanding speech in noisy environments. That is, it has been demonstrated that adults can selectively attend to a frequency region of interest while ignoring or paying less attention to frequency regions are deemed less important to speech recognition. Typically, speech is comprised of mid-high frequencies while background noise is

Formatted: Right

comprised of more low frequencies. Research suggests that the adult population can selectively ~~chose~~ choose to pay more attention to the mid-high frequency region and thereby experience less of a masking effect from the background noise (Dai et al., 1991; Scharf et al., 1987). While this appears to be a successful strategy for adults, in 1994, Bargones and Werner found that infants and children tend to adopt a broader listening strategy and do not focus ~~in~~ on a specific frequency region of interest.

Current research on the topic has utilized a variety of stimuli to evaluate performance in children and adults. Werner and Bargones (1991) and Leibold and Neff (2011) utilized a 500-msec 1kHz pure tone as the signal and 50 different samples of a band-passed noise with cutoff frequencies of 4-10 kHz as the masker. In both cases, performance was determined by calculating the 1kHz pure tone threshold in quiet and noise. In contrast, Youngdahl et al. (2018) utilized a different approach in which the target stimuli consisted of 120 sentences, each with 3-5 ~~key words~~keywords. Sentences were presented in the presence of a MATLAB-generated ~~speech shaped~~speech-shaped noise masker for a control condition and in in the presence of a spectrally remote noise for five additional conditions. Subjects were asked to repeat the sentences and performance was determined by the calculating percentage of ~~key words~~keywords correct.

Despite differences in stimuli, these studies demonstrate similar findings and appear to identify an important developmental period regarding central auditory maturation. That is, child participants show adult-like performance in ~~in~~ their adopted listening tasks at approximately five to seven years of age. Even still, other studies have revealed that, compared with adults, children are still more negatively affected by speech maskers in recognizing target speech until adolescence (Wightman & Kistler, 200~~1~~5).

Formatted: Indent: First line: 0"

Formatted: Not Different first page header

Formatted: Right

Several factors may account for the discrepancy in the age of maturity for speech recognition in the presence of speech maskers. The type of stimuli used may be one example. To our knowledge, no study has assessed performance in an informational masking task using speech reception thresholds (SRT) to evaluate performance. SRTs are widely used in clinical settings to determine the lowest intensity of speech required for a listener to correctly understand and repeat 50% of the words presented. Using SRT as a measure of performance may prove to be a more realistic measure than pure tone detection, as users must recognize and distinguish the specific stimuli, as opposed to simply detecting them.

Youngdahl et al. 2018 measured performance by calculating the percentage of keywords correctly identified when the stimuli were presented at fixed levels that yielded average scores of 60% or higher, which may have not created adequately adverse conditions for the 7-year-old group to show the deficit when compared with adults. Regarding tone detection tasks as demonstrated by Werner and Bargones (1991) and Leibold and Neff (2011), the distracting effect of background noise may be more robust if an unpredictable, multi-tonal remote-frequency masker is utilized.

The masking effect of distractors that are spectrally separate from target stimuli is referred to as informational masking. The detection of a tone in informational masking is considered an approach to assessing listeners' frequency-focused auditory attention. By far, the literature is limited on assessing the relationship between this auditory attention and speech recognition in noise for both adults and children.

In this study, we are recruiting adult participants to begin collecting data to determine the effects of remote masking using SRTs to analyze performance and measure frequency-focused auditory attention using the method of measuring informational masking. Collecting

Formatted: Right

this data is a critical component to be able to later extend this research to children and further our understanding of why this difficulty exists in the younger population. The long-term goal of this line of research is to further existing literature on the differences between children and adults regarding the effect of speech recognition in the presence of a frequency-remote masker. In contrast to existing studies that measured performance in the percentage of words identified correctly, we will utilize speech reception thresholds to analyze performance. Speech reception threshold, or SRT, refers to the lowest level at which an individual can correctly identify speech. Specifically, when using the SRT method, we aim to determine if we can still see the same trend of vulnerability to frequency-remote masking in children as compared to adults

Formatted: Right

II. Materials and Methods

Participants

Thirteen normal-hearing adults were recruited via word of mouth through the James Madison University Communication Sciences and Disorders department. Participants were unpaid volunteers with ages ranging from 22 to 34, with a mean age of 24.15 (SD = 3.16). Of the thirteen participants, twelve were female, and one was male. Inclusion criteria for the present study necessitates that all participants have audiometric thresholds less than or equal to 15 dB HL for octave frequencies 250 – 8000 Hz.

Stimuli

Two sets of stimuli were used to facilitate the two tasks: the speech perception task and the pure tone informational masking task. For the speech perception task, target speech stimuli were randomly chosen from 28 sentences taken from the speech corpus for multitalker communication research database established by Bolia et al. (2000). Target stimuli consisted of the coordinate response measure (CRM) sentence "Ready (call sign) to go (color) (number) now" wherein the call sign (Charlie) remained constant, but the color and number listed changed with each presentation. Participants were selected from seven numbers (one through eight, excluding seven) and four colors (blue, green, red, and white). An example would be "Ready Charlie go to red two now." This study utilized two talkers from the Bolia et al. speech corpus: Talker 0 and Talker 2. The talker employed was alternated between participants, such that Talker 0 was used as the target speech for 7 participants and Talker 2 was used as the target speech for 6 participants.

Formatted: Right

Stimuli for the speech perception task were generated through a custom MATLAB program, downloaded onto a Dell Optiplex 9010 PC and subsequently routed through a Hi-Definition Lynx22 Soundcard, a DAC1 D/A converter, and the analog amplifier/attenuator of a Tucker-Davis Technologies (TDT) RZ6 signal processor and its headphone buffer. Stimuli were ultimately presented through the right ear cup of Sennheiser HDA 200 transducers. Both speech stimuli and masking stimuli (if present) were passed through either a low bandpass (low-band) or high a bandpass (high-band) filter, both of which were half octave wide. To ensure separation in the frequency domain, if speech stimuli were passed through a low bandpass filter, masking stimuli were passed through a high bandpass filter, and vice versa. The low and high frequency bandpass filters were respectively centered around 500 Hz (ranging from 420 – 595) Hz and 2500 Hz (ranging from 2101 – 2973 Hz).

Speech reception thresholds (SRTs) were calculated using low-band target speech in quiet and in the presence of high-band Gaussian noise band and high-band time-reversed CRM sentences uttered by a female talker randomly selected from the corpus. SRTs using high-band target speech were also calculated in quiet and low-band Gaussian noise band conditions. Speech recognition scores (SRSs) were determined using only low-band speech in quiet, high-band Gaussian noise, high-band forward speech masker uttered by a female talker randomly selected from the corpus (but the masker talker stayed consistent within a given block of trials), and high-band time-reversed speech masker uttered by the same female talker(s) of the forward speech masker. A complete list of conditions can be seen *in Table 1.*

Formatted: Font: Italic

Formatted: Right

Target	**Masker**	**SRT or SRS?**
Low-band speech	None	SRT
Low-band speech	None	SRS
Low-band speech	High-band Gaussian Noise	SRT
Low-band speech	High-band Gaussian Noise	SRS
Low-band speech	High-band forward speech (female speaker)	SRS
Low-band speech	High-band time-reversed speech (female speaker)	SRT
Low-band speech	High-band time-reversed speech (female speaker)	SRS
High-band speech	None	SRT
High-band speech	Low-band Gaussian Noise Band	SRT

Table 1. List of conditions for low and high band target speech used for Speech Reception Threshold (SRT) and Speech Recognition Score (SRS) tasks

For the pure tone detection task, the target stimulus was a 512-ms 1000 Hz tone burst. The stimulus was presented both in quiet and in the presence of a variable-frequency, multi-component masker. The masker consisted of 10 randomly-selected tones falling outside of a 1-octave protective band (707 – 1414 Hz) centered around 1000 Hz. The level of each masker component was set to 60 dB SPL. Five masking components were randomly selected from below 707 Hz and 5 masking components were randomly selected from above 1414 Hz. To be included, the frequency of any two masking components must differ by more than 0.9% from any other components. The ten-frequency components of the masker were randomized from trial-to-trial but never changed within a single trial. Each trial consisted of three observational intervals each with the presentation of a sound burst: a 1000 Hz pure tone followed by two bursts of noise with the 1000 Hz pure tone either present (real trial) or absent (catch trial). Bursts were 512ms in duration with a 10ms rise and fall time and were separated by 200ms inter-burst intervals. The initial level for the 1000 Hz target tone burst was 70 dB SPL.

Procedure

Participation in this study was divided into two sessions, with each session lasting approximately two hours. All sessions took place in a double-walled, sound-treated booth in the Laboratory for Auditory Perception in Children and Adults, located in the Health and Behavioral Sciences building at James Madison University. The two sessions were never conducted in the same day but were not separated by more than three weeks. For each participant, the first session assessed speech perception while the second session assessed detection of non-speech, pure tone stimuli.

For the first session, listeners were seated in a sound-attenuated booth in front of a monitor. Participants were instructed to listen for a male voice uttering a sentence similar to “Ready Charlie go to red two now” and were informed that the color and number voiced by the speaker would change with each presentation. Instructions were given to select the color and number on the screen that corresponds with what the speaker said. Prior to beginning real trials, participants completed one practice trial of 25 sentences. Practice trials were conducted using low-band speech presented at 55 dB SPL in quiet. For practice trials only, the correct answer was displayed on the screen after the listener made his or her selection. After the completion of the first set of practice trials, the first SRT condition was run. The order of conditions was randomized for each participant, but SRT for a given condition was always established before calculating SRS for the same condition.

Formatted: Right

For all conditions, the starting level of speech was presented at 30 dB SPL, and the level of the masker, if present, was held constant at 70 dB SPL. Speech intensity increased in 4 dB steps until the first correct response for both color and number, upon which reversals were decreased to 2 dB steps. A total of 27 trials were conducted for each block. The threshold was determined by averaging the dB SPL after the first correct response. After SRT was calculated for a given condition (e.g. quiet, Gaussian noise band, time-reversed speech), it was immediately replicated. If the two SRTs did not fall within 3 dB of each other, a third replication was conducted. The two closest SRTs were averaged together and recorded as the threshold. The average SRT was also used to determine the presentation level for the corresponding speech recognition score task. That is, target speech for SRSs was presented at 7 dB SL (re: averaged SRT for that condition). Masker level, if present, was held constant at 70 dB SPL. SRSs were also immediately replicated prior to beginning the next condition. The average of the two scores was recorded as the SRS for each condition.

For the second session, listeners were seated in the same sound-treated booth in front of a different Dell monitor. Sennheiser HDA 200 transducers were used. The second session consisted of two tasks: the determination of eleven pure tone thresholds in quiet and the determination of a 1000 Hz threshold in the presence of a frequency-remote multi-tonal masker. In the first task, thresholds in quiet were recorded for 11 frequencies (250, 315, 397, 500, 630, 1000, 1587, 2000, 2519, 3174, and 4000 Hz) utilizing the maximum likelihood algorithm described by Gray et. al (2002). Of note, this data was collected but will not be analyzed for the current study.

Using a single trial paradigm, thresholds were obtained in sequential order, beginning with 250 Hz and extending to 4000 Hz. Prior to every trial, the word "Trial"

Formatted: Right

prompted participants to listen for the pure tone stimulus. After the signal presentation, the words “Please decide” were displayed on the monitor to prompt listeners to indicate whether or not they heard a tone. To do so, participants were instructed to press either a button labeled “tone” or “no tone.” Immediate visual feedback was displayed on the screen as either “yes” or “no” corresponding to correct and incorrect answers, respectively. The initial presentation level of the stimuli was presented at an audible level, specifically 30 dB SPL. When present, the pure tone was presented in two 200ms bursts with a 10ms inter-burst interval. Each run consisted of 15 real trials and 5 catch trials. At the end of each run, the threshold and false alarm rate for each frequency were displayed on the screen. Participants were instructed to record these values on a sheet of paper provided at the beginning of the session before moving on to the next frequency.

After thresholds were recorded in quiet, task two was initiated in which participants completed a series of informational masking trials measuring a 1000 Hz threshold in the presence of a variable-frequency, multi-component masker. Before testing, listeners were informed that they would hear a series of three bursts: a pure tone presented alone followed by two identical bursts of noise. The presentation of the 1000 Hz tone at the beginning of every trial was intended to remind participants of what they were listening for prior to the presentation of the masker. Participants were instructed to listen to the presentation press “tone” or “no tone” to indicate whether or not the pure tone was present in the bursts of noise. As demonstrated in the first task, immediate feedback was displayed on the screen as “correct” or “wrong.” Task two also utilized the maximum likelihood algorithm described in Gray et al. (2002). For the first trial, the 1000 Hz pure tone was presented at an audible level of 70 dB SPL. The level of each tone in the multicomponent masker was

Formatted: Right

held constant at 60 dB SPL for each trial. Each run in task two consisted of 20 real trials and 20 catch trials. Participants completed 12 informational masking runs or stopped after 90 minutes, whichever came first. In total, completed blocked ranged from 11 to 14 amongst participants. To avoid listening fatigue, participants were allowed to take breaks when needed.

Statistical Analysis

IBM SPSS Statistics, Version 28 for Windows was used to perform all statistical analyses. Repeated Measured Analysis of Variance (RM-ANOVA) was primarily used to investigate the effect of remote-frequency maskers on both speech reception threshold and speech recognition scores. To evaluate the assumption of sphericity, Mauchly's sphericity test was used. Pearson correlations were used to examine the relationship between informational masking effects exhibited with speech stimuli and pure tone stimuli.

Formatted: Right

III. Results

Effect of Low vs High Band-Passed Target Stimuli on Speech Reception Thresholds

For low-passed target speech SRT conditions, a repeated measures analysis of variance (RM-ANOVA) was performed with the dependent variable of SRT and the within-subject factor of masker condition. There were three levels under the masker condition including Quiet, Gaussian noise, and time reversed speech masker. A significant difference in SRT in quiet and masked conditions was revealed, ($F(1,12) = 9.99$, $p = 0.001$, $\eta^2_p = .455$. The average SRT for all participants was 45.69 dB SPL in quiet (SE = 1.45). Average SRT for masked conditions utilizing Gaussian noise band maskers and reverse speech makers was 39.58 dB SPL (SE = 2.11) and 41.03 dB SPL (SE = 1.68), respectively. *Figure 1* represents the average SRTs obtained for the three conditions using low band-passed stimuli. Interestingly, a pairwise comparison with Bonferroni correction revealed thresholds in quiet were significantly higher than thresholds in Gaussian noise maskers ($p = 0.003$) and reverse speech maskers ($p = 0.004$). No significant difference was observed between thresholds obtained in both masked conditions ($p \geq .999$).

Formatted: Font: Italic

Formatted: Right

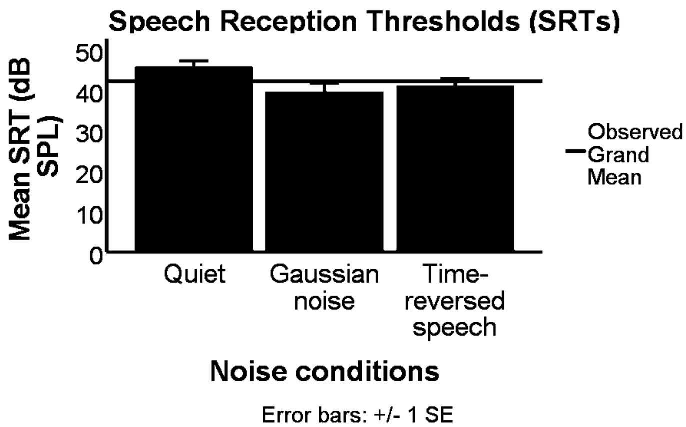

Figure* *1*: *SRT of low-frequency band speech in various masking conditions

A RM-ANOVA was conducted with the dependent variable of SRT and two within-subject independent variables including masking condition (Quiet versus Gaussian Noise) and frequency band of speech (low versus high band-passed speech). The analysis revealed a significant interaction between the SRTs obtained with low and high band-passed stimuli (F(1,12) = 8.99, p = 0.011, $\eta_p^2 = 0.428$). This relationship can be analyzed visually in *Figure 2*, which depicts the relationship between average SRT across participants for both noise conditions (quiet, masked) and both band-passed target stimuli (high, low). Despite the interaction observed in the RM-ANOVA, *Figure 2* visually indicates that SRTs were lower (i.e. better) in masked conditions than in quiet. Additionally, based on the larger slope of the high band-passed stimuli, a larger observed masking effect for high-band passed target speech can be inferred. A larger masking effect was observed for high band-passed target

Commented [NY-n1]:

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Sq
Noise	Sphericity Assumed	869.016	1	869
	Greenhouse-Geisser	869.016	1.000	869
	Huynh-Feldt	869.016	1.000	869
	Lower-bound	869.016	1.000	869
Error(Noise)	Sphericity Assumed	318.972	12	26
	Greenhouse-Geisser	318.972	12.000	26
	Huynh-Feldt	318.972	12.000	26
	Lower-bound	318.972	12.000	26
SpchFreq	Sphericity Assumed	1391.592	1	1391
	Greenhouse-Geisser	1391.592	1.000	1391
	Huynh-Feldt	1391.592	1.000	1391
	Lower-bound	1391.592	1.000	1391
Error(SpchFreq)	Sphericity Assumed	1314.825	12	109
	Greenhouse-Geisser	1314.825	12.000	109
	Huynh-Feldt	1314.825	12.000	109
	Lower-bound	1314.825	12.000	109
Noise * SpchFreq	Sphericity Assumed	55.325	1	55
	Greenhouse-Geisser	55.325	1.000	55
	Huynh-Feldt	55.325	1.000	55
	Lower-bound	55.325	1.000	55
Error(Noise*SpchFreq)	Sphericity Assumed	73.826	12	6
	Greenhouse-Geisser	73.826	12.000	6
	Huynh-Feldt	73.826	12.000	6
	Lower-bound	73.826	12.000	6

Formatted: Font: Italic

Formatted: Right

speech than low band-passed target speech, as indicated by differences in the slopes of the lines in *Figure 2*.

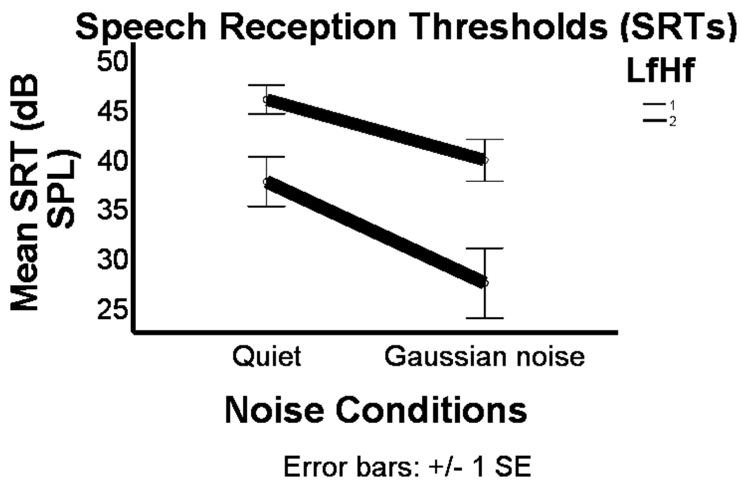

***Figure 2**: SRT of Low Band Speech in various masking conditions*

Effect of Masking Condition on Speech Recognition Scores of Low Band-Passed Target Stimuli

A RM-ANOVA with the dependent variable of SRS measured using low band-passed target speech and the within-subject independent variable of masking condition revealed a significant difference in SRSs obtained in quiet and masked conditions (F(3, 3) = 6.48, p = 0.001, η_p^2 = 0.351. A posthoc pairwise comparison with Bonferroni correction revealed that only the forward speech masker produced significantly worse performance on SRS tasks (*p* = 0.029), as seen visually in *Figure 3*. No masking

Commented [NY-n21]:

Tests of Within-

Measure: MEASURE_1

Source		Type III Sum of Squares	d
Noise	Sphericity Assumed	.137	
	Greenhouse-Geisser	.137	1
	Huynh-Feldt	.137	2
	Lower-bound	.137	1
Error(Noise)	Sphericity Assumed	.254	
	Greenhouse-Geisser	.254	22
	Huynh-Feldt	.254	26
	Lower-bound	.254	12

Formatted: Right

effect was observed with Gaussian noise band or reverse speech maskers (*p* values > .999). *Figure 3* reveals overall performance for speech recognition tasks in each of the three conditions. Average performance for adults is 64.56% (SE = 0.03) in quiet, 61.04% (SE = 0.04) in the presence of a reverse-speech masker, 49.99% (SE = 0.03) in the presence of a forward-speech masker, and 61.81% (SE = 0.03) in the presence of a Gaussian noise band masker. The grand mean for all such conditions was 59.75%.

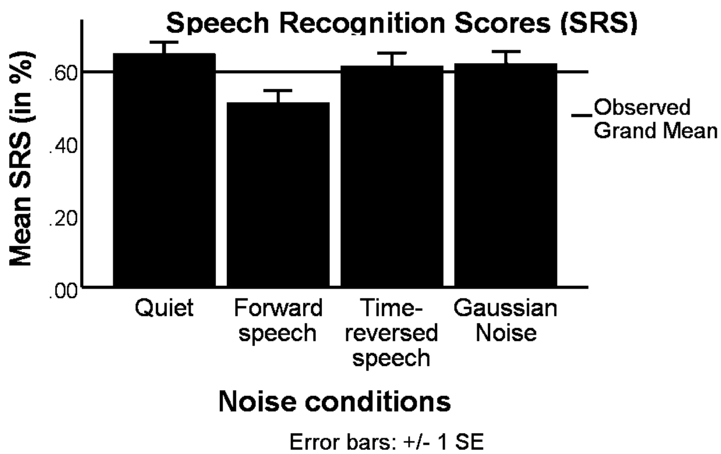

Figure 3*: Average performance on Speech Recognition Scores for various masking conditions*

Commented [AT3]: Address why performance in quiet was lower than expected, as per prospectus feedback

Commented [NY-n4]: I'm just attaching what I calculated based on what I have. Perhaps my data are not up-to-date? Below are the descriptive statistics from 13 participants.

Descriptive Statistics

	N	Mean		Std. Deviation
	Statistic	Statistic	Std. Error	Statistic
LFSRSQ	13	.6453	.02985	.10762
LFSRSF	13	.5099	.03273	.11801
LFSRSR	13	.6106	.03544	.12778
LFSRSG	13	.6184	.03200	.11537
Valid N (listwise)	13			

Formatted: Right

Correlation between Masking Effect Observed with Nonspeech vs Speech Stimuli

Across participants, average threshold in quiet was -0.83 dB SPL (SE = 1.81). The average threshold across participants in the presence of a multi-tonal masker was 17.49 dB SPL (SE = 1.81). Masked pure tone thresholds for each participant were calculated by finding the average amongst all repetitions with the multi-tonal masker present. The masking effect observed with non-speech stimuli was defined as the difference in 1 kHz pure-tone threshold in the presence of a multi-tonal, frequency-remote masker, subtracted by the 1 kHz pure tone threshold in quiet Average pure tone threshold difference (i.e. masking effect) was 18.32 dB SPL (SE = 1.52). The masking effect observed with speech stimuli was defined as the difference in the SRS in quiet and reverse speech masked conditions and the difference in the SRS in quiet and Gaussian noise band conditions, or as the difference in the SRT in Gaussian noise and in quiet for the low band-passed target speech condition. A Pearson correlation revealed no significant correlation between the masking effect observed with speech and non-speech stimuli when using the difference in SRS between quiet and reverse speech maskers (r = -0.452) or when using the difference in SRS between quiet and Gaussian noise band maskers (r = 0.081). It also did not reveal a significant correlation of the masking effect between the non-speech stimuli and SRT (r = -0.121). Of note, one outlier was identified, wherein the obtained masked pure tone threshold was greater than three standard deviations above the mean. As indicated in *Figure 4*, correlations were analyzed without (*Figure 4a)* and with (*Figure 4b)* this participant.

Formatted: Right

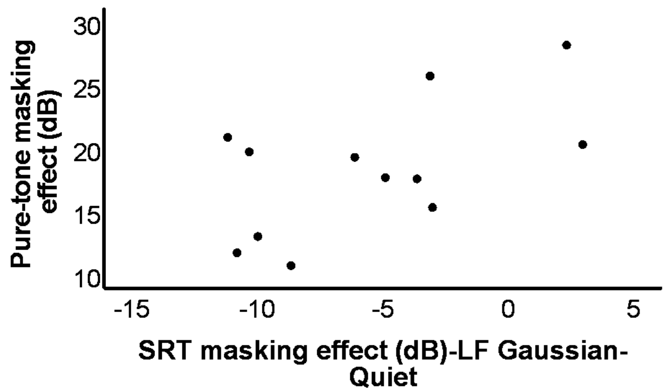

Figure 4a: *SRT masking effect (LF-GNB – quiet) and pure tone masking effect (masked – quiet) without the outlier.*

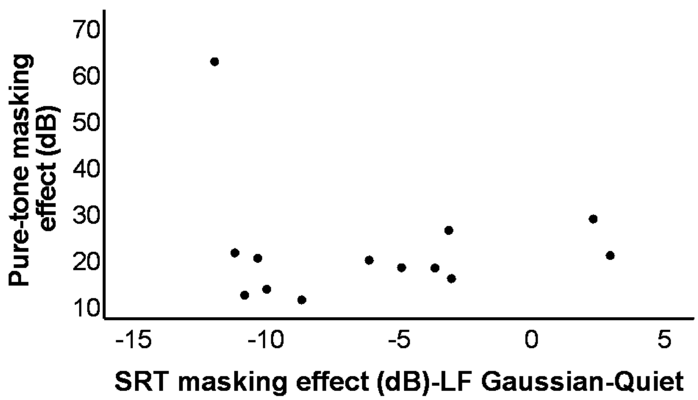

Figure 4b: *SRT masking effect (LF-GNB – quiet) and pure tone masking effect (masked – quiet) with an outlier.*

Formatted: Right

IV. Discussion

Overall, study findings were inconsistent with the hypothesis that SRT would be worse or unchanged in masked conditions when compared to quiet conditions. One possible explanation for this finding lies in the concept of spectral restoration, as explored in Warren (1997). Spectral restoration refers to the ability of a listener to utilize narrow spectral gaps in masking stimuli to catch “glimpses” of target speech, ultimately aiding in performance. Warren (1997) suggests that filling temporal gaps in speech sentences with noise produces appropriate phonemic restorations that can supplement part of the missing speech spectrum, ultimately aiding in intelligibility. It is important to note that differences exist in the methodologies of this study and that of Warren (1997), so direct comparison of findings may be limited.

Greater improvement in SRT was observed for high-band stimuli than low-band stimuli in noise versus in quiet. This possibly indicates that the 2500 Hz frequency region is more important for speech recognition than the 500 Hz frequency region. As thoroughly documented in the literature, the auditory system has a higher sensitivity to mid-frequency stimuli (i.e. 2-4 kHz) than low-frequency stimuli (Pickles, 2008), which likely contributes to the interaction observed in this study.

Overall, the SRTs obtained in quiet conditions were higher than clinically expected in reference to pure tone average (PTA). This can most clearly be explained by the exclusion of important frequency regions in the presence of a bandpass filter. To rule out equipment errors, SRT was calculated from one participant using full-spectrum speech stimuli in quiet, which revealed a much more appropriate SRT/PTA agreement. Participant PTA was [insert] and SRT using full-spectrum speech was 20 dB SPL.

Commented [AT5]: Get these numbers from Husna

Formatted: Right

As hypothesized, findings for Speech Recognition Scores agreed with the Youngdahl et al. (2018) findings. That is, no masking effect was observed in adults when using a narrow-band masker. The concept of informational masking explains the significant reduction in performance observed in speech recognition tasks in the presence of a forward speech masker. Due to the minimization of energetic masking ensured by the separation of target stimuli and masking stimuli in the frequency domain, there appears to be a more central masking component resulting in the reduction in performance. Additionally, the only significant masking effect for SRS tasks was found in the forward speech masker condition. This can perhaps be attributed to the greater similarity in target and masker. As demonstrated by Kidd et al (2002), it is shown that increasing the similarity between signals and maskers can create a greater informational masking effect. Ample literature exists demonstrating this topic. It has been shown that informational masking tends to increase as the masker goes from narrow-band noise to speech to same-sex talker to the same talker (Brungart, 2001).

Results to the pure tone stimuli task were consistent with the findings of a large number of studies (e.g., Neff, 1995). That is, the multi-tonal masker produced a large information masking effect on the detection of the target pure tone. Our study found no correlation between the informational masking effect with pure tone masking and speech-on-speech masking computed in the SRS difference between quiet and forward-speech masking conditions. This was not unexpected, considering the masking effects were in different domains involving different levels of processing. Firstly, the stimuli were different in the two experiments assessing the masking effect. Secondly, the two experiments assessed different levels of auditory processing. Detection methods were

Formatted: Right

utilized for the non-speech stimuli and recognition methods were employed for the speech stimuli. When these disparities were reduced, a trend of correlation was observed between pure tone informational masking and SRT masking effect, although not significant. Additional data will be collected to further analyze this trend. Of note, one explanation for the lack of correlation between the masking effect observed with pure tone and speech stimuli lies in the inherent differences in the task. The pure tone stimuli task is one of detection; the participant is instructed to respond when stimuli presence is simply detected. In contrast, the speech stimuli task is one of recognition; the participant is asked to detect and discriminate the word presented.

An important factor to consider pertains to the influence of the characteristics of the masker on the detection or recognition of target stimuli. Oster & Werner (2017) demonstrate that the effects of the masker used depends on the spectrotemporal properties of the target speech. That is, stimuli with more constant intensity, such as a vowel, yield more useful information when catching glimpses of the target speech, as compared to an entire word with a larger spectral variation. The influence of the characteristics of the masker are different for adult and infant populations. Therefore, caution should be exercised when comparing the results of various studies. Greater consistency amongst the methodologies used across studies would allow for a greater comparison of performance both within and between adult and pediatric populations.

Formatted: Right

V. Conclusion

This study has several future directions aiming at further exploration of informational masking using speech reception thresholds. First, we hope to extend the study to include the pediatric population. Results can be compared to those of Warner and Bargones (1991), Leibold and Neff (2011), and Youngdahl et al. (2018), to examine if a similar developmental release of masking for SRT occurs around age 7. Additionally, we aim to retrospectively analyze the data collected in this study using psychometric functions on trial-by-trial data to assess SRT to rule out procedural explanations for unexpected SRT findings.

Formatted: Right

Appendix

Extended Literature Review

Informational Masking Using Pure Tone Stimuli

Unlike frequency sensitivity, which reaches developmental maturity as early as five months of age, there are several aspects of auditory perception that develop at a slower pace (Olsho, 1985). Informational masking refers to an elevation in signal threshold that occurs in the presence of a novel masker, particularly when temporal or spectral characteristics of the masker are uncertain or unpredictable (Watson, Kelly, & Wroton, 1976). Informational masking differs from energetic masking in that it cannot be explained by spectral overlap of target and masker stimuli or traditional filtering models of the peripheral auditory system. Existing literature heavily documents that the phenomenon of informational masking is more prevalent in the pediatric population, as these individuals are more susceptible to distractions by novel or unpredictable stimuli in their environment.

While between-group differences exist, there is also a large degree of variation in the amount of informational masking observed between individuals in both the adult (Neff and Dethlefs, 1995) and pediatric populations (Oh et al., 2001). In an informational task used by Kidd et al. (1994), masking effects ranged from 0-40 dB amongst adult participants, with an average effect of 15 dB. Various explanations for this have been proposed, including differences in processing strategies and the number of auditory filters used in the detection process (Lutfi et al., 2003). For instance, Oh et al. (2001) demonstrated that children are more likely to place more emphasis on non-signal frequency channels than adults, suggesting that they adopt a broader listening strategy and are more susceptible to

Formatted: Right

informational masking effects. Other studies, such as that from Allen et al. (1989), suggest that non-auditory factors, including attention, may contribute to performance differences.

Various characteristics of masking stimuli have been shown to influence the overall masking effect in both age groups. Among these are masker uncertainty and similarity. Masker uncertainty refers to the predictability of the masker regarding its temporal and spectral qualities. Masker similarity refers to how similar the masker is to the target stimuli. Existing research demonstrates that higher levels of uncertainty lead to a larger informational masking effect (Oh and Lutfi, 2000). Likewise, the more similar the masking stimuli are to the target stimuli, the larger the observed masking effect (Kidd et al., 1994).

Neff (1995) explored this further by investigating the effect of signal type, duration, and presentation mode on the overall masking effect. This study revealed that amplitude modulated and narrowband noise signals resulted in improved performance compared to a quasi-modulated masker. Masking effect was reduced when utilizing a dichotic presentation with narrowband maskers. Additionally, the greatest and most consistent reductions in masking were observed when the 1000 Hz pure tone target duration was shortened. This effect is most likely due to masker-frequency uncertainty. This is an important finding, as it underscores the importance of a standardized methodology when comparing results between or within groups across studies. Leibold and Buss (2016) found that within-subjects, a greater masking effect was observed when the masker was gated such that its onset was synchronize with the onset of the target signal. This suggests that the difficulty of the detection task is increased when listeners must segregate synchronous, as opposed to asynchronous, sounds.

Formatted: Right

Informational Masking Using Speech Stimuli

While early work exploring this topic, including the aforementioned studies, utilized a pure tone detection task to measure performance and quantify the degree of informational masking present, other studies have used speech stimuli to evaluate performance. One considerable limitation with pure tone detection paradigms is that the task is not particularly realistic or generalizable to everyday listening tasks. That is, pure tone detection tasks are relatively static and maskers often possess a spectral quality that is considerably different than that of the target signal. Research has shown that when listening tasks utilize a paradigm in which the distractor or masking stimuli are modified to be less predictable and more uncertain, source segregation and attention become more imperative and the task becomes more difficult (Whitman and Kistler, 2005). For these reasons, speech tasks utilizing maskers with variable and more uncertain characteristics may produce results that are more generalizable to everyday listening situations.

Newman, Morini, and Chatterjee (2013) evaluated performance in infants to respond to their own name and a different name in the presence of spectrally remote and overlapping maskers. Results indicate that the infants listened longer to their own name (versus a different name) in spectrally remote noise, but not in overlapping noise. This may be suggestive that infants can take identify and take advantage of a spectrally remote masker in certain listening situations. When considered in conjunction with other research, however, it is apparent that this skill is not fully developed.

The Coordinate Response Measure (CRM) paradigm, chosen for use in this study, was developed by Bolia et al. (2000) at the Air Force Research Laboratory and has been used in several studies to evaluate performance on informational masking tasks using

Formatted: Right

speech stimuli. CRM sentences consist of a spoken target using the sentence structure "Ready *call sign*, go to *color number* now." The *call sign* used remains consistent throughout trials and is chosen from a list of eight possible names. The "color" and "number" are picked randomly from a pre-selected set of four colors and eight numbers. Masking, or "distractor," stimuli may vary but are temporally aligned so that they begin and end at the same time.

Brungart (2001) suggests that the masking produced by the CRM paradigm is predominately informational in nature. This study found that performance improved when the target and masker voices were of opposite sex, suggesting that the masking effect observed was likely due to informational masking, given that the two voices were perceptually different in frequency. Studies using CRM to measure speech performance yield findings in agreement to studies utilizing pure tone stimuli. That is, children still demonstrate a greater masking effect than adults. More specifically, one study using CRM sentences found that this effect was greatest for children aged 4-5 years of age. This age group demonstrated a masking effect greater than 15 dB than that of adults (Wightman and Kistler, 2005).

Some studies have utilized a dichotic listening paradigm to evaluate the effect of informational masking using speech stimuli in children. These studies confirmed that children do not perform as well as adults in dichotic listening tasks. While this is valuable information regarding understanding the central auditory development, this situation is unlikely in real life, as it is an unnatural listening environment for a signal to be presented to one ear while a distractor presented to the other ear (Doyle, 1973). Other studies have evaluated performance on speech tasks in children when the target and distractor are presented to the same ear. Results consistently indicate a similar pattern to dichotic

Formatted: Right

conditions in that children experience more difficulty with speech understanding than adults under the same conditions (Fallon, Trehub,and Schneider, 2000; Hall et al., 2002).

Youngdahl et al (2008) were among the first to evaluate the time course of development regarding informational masking observed with speech stimuli. Developmental effects were consistent with those of earlier studies, including Werner and Bargones (1991) and Leibold and Neff (2011), which both utilized pure tone stimuli to discern a developmental timeline for these effects. That is, speech recognition was reduced in the presence of frequency remote noise for the 5-year old group, but not for the 7-year old or adults groups. Results are suggestive that the presumptive release from informational masking, and subsequent reduction in the need for a higher SNR to understand speech in noise, occurs at around seven years of age.

As we continue to extend research to further evaluate this phenomenon, it is of interest to examine if these trends extend from pure tone detection to word recognition to speech reception thresholds. Speech reception thresholds (SRT) are a very common clinical tool used to measure threshold levels of speech detection. By collecting data in the adult population, this study documents an important step in ultimately comparing adult and pediatric performance in informational masking paradigms using threshold levels of speech detection.

Formatted: Right

References

Allen, P., Wightman, F., Kistler, D., & Dolan, T. (1989). Frequency resolution in children. Journal of Speech, Language, and Hearing Research, 32(2), 317–322. https://doi.org/10.1044/jshr.3202.317

Bargones, J. Y., & Werner, L. A. (1994). Adults listen selectively; infants do not. Psychological Science, 5(3), 170–174. https://doi.org/10.1111/j.1467-9280.1994.tb00655.

Bolia, Robert & Nelson, W. & Ericson, Mark & Simpson, Brian. (2000). A speech corpus for multitalker communications research. The Journal of the Acoustical Society of America. 107. 1065-6. 10.1121/1.428288.

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. https://doi.org/10.1121/1.1345696

Dai, H., Scharf, B., & Buus, S. (1991). Effective attenuation of signals in noise under focused attention. The Journal of the Acoustical Society of America, 89(6), 2837–2842. https://doi.org/10.1121/1.400721

Doyle, A.-B. (1973). Listening to distraction: A developmental study of selective attention. Journal of Experimental Child Psychology, 15(1), 100–115. https://doi.org/10.1016/0022-0965(73)90134-3

Fallon, M., Trehub, S. E., & Schneider, B. A. (2000). Children's perception of speech in Multitalker Babble. The Journal of the Acoustical Society of America, 108(6), 3023–3029. https://doi.org/10.1121/1.1323233

Formatted: Right

Gray, L. C., Breier, J. I., Foorman, B. R., & Fletcher, J. M. (2002). Continuum of impulsiveness caused by auditory masking. International Journal of Pediatric Otorhinolaryngology, 66(3), 265–272. https://doi.org/10.1016/s0165-5876(02)00251-3

Hall, J. W., & Grose, J. H. (1991). Notched-noise measures of frequency selectivity in adults and children using fixed-masker-level and fixed-signal-level presentation. Journal of Speech & Hearing Research, 34(3), 651–660. https://doi.org/10.1044/jshr.3403.651

Hall, J. W., 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. https://doi.org/10.1097/00003446-200204000-00008

Leibold, L. J., & Buss, E. (2016). Factors responsible for remote-frequency masking in children and adults. The Journal of the Acoustical Society of America, 140(6), 4367–4377. https://doi.org/10.1121/1.4971780

Leibold, L. J., & Neff, D. L. (2011). Masking by a remote-frequency noise band in children and adults. Ear & Hearing, 32(5), 663–666. https://doi.org/10.1097/aud.0b013e31820e5074

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. https://doi.org/10.1121/1.1873913

Lutfi, R.A., Kistler, D.J., Oh, E.L. et al. (2003). One factor underlies individual differences in auditory informational masking within and across age groups. Perception & Psychophysics 65, 396–406 https://doi.org/10.3758/BF03194571

Formatted: Right

Kidd, Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994). Reducing informational masking by sound segregation. The Journal of the Acoustical Society of America, 95(6), 3475–3480. https://doi.org/10.1121/1.410023

Kidd, G., Mason, C. R., & Arbogast, T. L. (2002). Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns. The Journal of the Acoustical Society of America, 111(3), 1367–1376. https://doi.org/10.1121/1.1448342

Neff, D. L. (1995). Signal properties that reduce masking by simultaneous, random-frequency Maskers. The Journal of the Acoustical Society of America, 98(4), 1909–1920. https://doi.org/10.1121/1.414458

Neff, D. L., &; Dethlefs, T. M. (1995). Individual differences in simultaneous masking with random-frequency, Multicomponent Maskers. The Journal of the Acoustical Society of America, 98(1), 125–134. https://doi.org/10.1121/1.413748 Neff, D. L. (1994). ‘‘Signal properties that reduce masking by simultaneous, random-frequency maskers,’’ J. Acoust. Soc. Am. 98, 1909–1920

Newman, R. S., Morini, G., & Chatterjee, M. (2013). Infants' name recognition in on- and off-channel noise. The Journal of the Acoustical Society of America, 133(5). https://doi.org/10.1121/1.4798269

Oh, E. L., & Lutfi, R. A. (2000). Effect of Masker harmonicity on informational masking. The Journal of the Acoustical Society of America, 108(2), 706–709. https://doi.org/10.1121/1.429603

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–

Formatted: Right

2895. https://doi.org/10.1121/1.1371764 Scharf, B., Quigley, S., Aoki, C., Peachey, N. & Reeves, A. (1987). Focused auditory attention and frequency selectivity. Perception & Psychophysics, 42(3), 215–223. https://doi.org/10.3758/bf03203073

Olsho, L. W. (1985). Infant auditory perception: Tonal masking. Infant Behavior and Development, 8(4), 371–384. https://doi.org/10.1016/0163-6383(85)90002-5

Oster, M.-M., & Werner, L. A. (2017). The influence of target and Masker characteristics on infants' and adults' detection of speech. Journal of Speech, Language, and Hearing Research, 60(12), 3625–3631. https://doi.org/10.1044/2017_jslhr-h-16-0464

Pickles, J. O. (2008). An introduction to the physiology of hearing. Brill.

Scharf, B., Quigley, S., Aoki, C., Peachey, N., &; Reeves, A. (1987). Focused auditory attention and frequency selectivity. Perception & Psychophysics, 42(3), 215–223. https://doi.org/10.3758/bf03203073

Warren, R. M., Hainsworth, K. R., Brubaker, B. S., Bashford, J. A., & Healy, E. W. (1997). Spectral restoration of speech: Intelligibility is increased by inserting noise in spectral gaps. Perception & Psychophysics, 59(2), 275–283. https://doi.org/10.3758/bf03211895

Watson, C. S., Kelly, W. J., & Wroton, H. W. (1976). Factors in the discrimination of tonal patterns. II. selective attention and learning under various levels of stimulus uncertainty. The Journal of the Acoustical Society of America, 60(5), 1176–1186. https://doi.org/10.1121/1.381220

Werner, L.A., & Bargones, J.Y. Sources of auditory masking in infants: Distraction effects. Perception & Psychophysics 50, 405–412 (1991). https://doi.org/10.3758/BF03205057

Formatted: Right

Wightman, F., & Allen, P. (n.d.). Individual differences in auditory capability among preschool children. Developmental Psychoacoustics., 113–133. https://doi.org/10.1037/10119-004

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. https://doi.org/10.1121/1.2082567

Youngdahl, Carla & Healy, Eric & Yoho, Sarah & Apoux, Frederic & Holt, Rachael. (2018). The Effect of Remote Masking on the Reception of Speech by Young School-Age Children. Journal of Speech Language and Hearing Research. 61. 10.1044/2017_JSLHR-H-17-0118.

Formatted: Right