THE EFFECT OF SEMANTIC INTERFERENCE

OF BABBLE MASKERS

ON THE PERCEPTION OF AZBIO SENTENCES

by

Anna Willis

A Senior Honors Project Presented to the

Honors College

East Carolina University

In Partial Fulfillment of the

Requirements for

Graduation with Honors

By

Anna Willis

Greenville, NC

April 30th, 2025

Approved by:

Andrew J. Vermiglio, AuD (Thesis director)

Department of Communication Sciences and Disorders, College of Allied Health Sciences

Abstract

Speech in noise testing is conducted to evaluate speech perception in everyday listening conditions. Speech in noise testing involves speech recognition in the presence of noise or talker babble. A babble masker contains semantic (meaningful) information that may affect the results of the speech perception test. The goal of this study was to determine the effect of semantic masker content on speech perception. Semantic interference is a phenomenon in which semantic content of the background babble interferes with intelligibility of the target speech.

In the present study, AzBio sentences were used to measure speech perception ability. The maskers for this study included 4-talker babble and 10-talker babble presented forward and reversed. Semantic interference was calculated as the AzBio score for the reversed babble condition minus the score for the forward babble condition. It was hypothesized that semantic interference would be greater for the 4-talker babble than for the 10-talker babble listening condition. Poorer speech perception ability was found in the presence of forward babble than in reversed babble for each talker condition. Semantic interference for the 4-talker babble condition was 19.69 percentage points, and it was greater than the semantic interference for the 10-talker condition, which was 9.88 percentage points. The difference in semantic interference between these conditions was 9.81 percentage points, which was statistically significant ($p < 0.01$).

The secondary purpose of this study was to investigate the relationship between pure-tone threshold averages vs. AzBio performances for the four listening conditions. It was hypothesized that a weak and statistically non-significant relationship would be found between these two variables. There was no statistically significant relationship found between bilateral pure-tone averages and AzBio scores.

It was concluded that it is more difficult to perceive speech when there is semantic content in the background babble than when there is not. It was also concluded that speech recognition in noise (SRN) ability cannot be predicted by pure-tone thresholds. This study can be applied clinically, as speech in noise testing can provide audiologists with information about a patient's hearing ability that is unavailable from the standard pure-tone threshold testing. Detecting the amount of semantic interference and understanding how it may impact a patient's hearing allows for more useful interpretations of test results and provides for proper counseling and recommendations for hearing disorders.

Introduction

Speech perception is the ability to recognize spoken language. Speech is perceived in both quiet and noisy environments. Speech perception in noise difficulties may exist in the presence of normal pure-tone thresholds (Middelweerd et al., 1990; King, 1954). A person's ability to recognize speech in the presence of background noise has been referred to as their speech recognition in noise (SRN) ability (Vermiglio et al., 2012).

A Brief History of Hearing Screening in the Royal Air Force

According to Dickson and Bateman (1946), the whisper test was administered to candidates for the Royal Air Force (RAF) to determine if they were suitable for the job. To conduct this test, an examiner would place their hand on the non-test side of the patient to steady their head and shield them from lip reading. The examiner would then place their middle finger into the external auditory meatus of the non-test ear and shake it to produce background (masking) noise. At the same time, a different examiner would stand 20 feet away from the candidate, facing the test ear. This examiner would whisper a combination of letters and numbers. The candidate passed this examination by correctly restating three out of six of the target words/numbers. The whisper test was eventually deemed to be an unreliable measure of one's hearing ability, so clinicians began to use pure-tone threshold testing instead. Pure-tone threshold testing indicates the lowest level a person can hear a tone across different frequencies to provide an outline of their overall hearing ability. This type of hearing screening is good for diagnosing decreased hearing sensitivity at specific frequencies; however, it does not provide a complete assessment of an individual's ability to hear speech in noise (Middelweerd et al., 1990).

Dennis Butler Fry was a professor of experimental phonetics and a member of the RAF acoustics lab during World War II. Fry (1942) suggested that pure-tone threshold testing was not satisfactory for the selection of RAF flying personnel. Instead, Fry recommended using a functional test that would indicate the candidate's ability to recognize speech in background noise. A noisy listening environment would be used to simulate conditions similar to those in which the candidate would have to work, making for a more realistic testing environment for their hearing ability. Dickson et al. (1946) developed the "efficiency test" based on the recommendations made by Fry.

Following the development of the efficiency test, it was recommended that aviation candidates be evaluated in three stages (Dickson et al., 1946). First, candidates would take the efficiency test, which consisted of speech in the presence of simulated aircraft noise. If the candidate failed at this stage, there was no further testing. If the candidate passed, they would proceed to the second stage of testing, which consisted of pure-tone threshold testing. All candidates from the second stage who reached the required standard in the efficiency test would be sent on to the Ear, Nose, and Throat (ENT) specialist. For the third stage of testing, the ENT specialist would analyze the candidate's pure-tone threshold scores and determine if there was a “clinical reason” to reject the candidate despite having passed the efficiency test (Dickson, 1946).

Hinchcliffe (1992) coined the term “King-Kopetzky syndrome” (KKS) for the manifestation of hearing difficulties in noise with no evidence of hearing loss, as indicated by pure-tone threshold testing. He highlighted the multiple possible underlying causes for this symptom set, particularly the combination of external factors (e.g., psychological or social) that can lead to the experience of hearing difficulties. Pryce et al. (2010) stated that at least 10% of people who present with “hearing difficulties” will be found to have normal hearing thresholds. This symptom set has been clinically categorized as KKS (Hinchcliffe, 1992), obscure auditory dysfunction (Saunders and Haggard, 1989), or auditory processing disorder (Pryce et al., 2010).

Factors That May Impact Speech Recognition in Noise Ability

Many factors may impact speech recognition in noise (SRN) ability. SRN ability may be impacted by physical characteristics, including the integrity of the inner hair cells and the central auditory nervous system pathway (Kujawa and Liberman, 2009; Liberman et al., 2016). Kujawa and Liberman (2009) investigated the effect of noise exposure on threshold and suprathreshold measures in animals. They found that following noise exposure, there was a temporary reduction in outer hair cell function (threshold measure) that would restore itself over time. They also found a reduction in one of the suprathreshold measures (ABR at 32,000 Hz) with no recovery over time. Additionally, they found that there was a reduction in synaptic ribbon counts in the inner hair cells immediately following noise exposure with no recovery. Based on their findings on animals, they hypothesized that following noise exposure, human patients could experience a reduced SRN ability even with normal pure-tone thresholds.

Liberman et al. (2016) investigated the effects of noise exposure on a group of individuals with normal pure-tone thresholds with and without a history of noise exposure. Participants were divided into a high-risk group (people with high levels of noise exposure) and a low-risk group (people with little to no noise exposure). For this study, the Northwestern University Auditory Test No. 6 (NU-6) word lists were used as the target speech. The word lists were presented monaurally at 35 dB HL in quiet and in the presence of white noise. When the target speech was presented in quiet, speech recognition in noise ability was excellent (> 96% correct) for both groups. However, when the target speech was presented in the presence of ipsilateral white noise, the high-risk group performed more poorly than the low-risk group at both a signal-to-noise ratio of 5 dB ($p < 0.05$) and 0 dB ($p < 0.01$). They hypothesized that this could be due to potential synaptic ribbon loss (cochlear synaptopathy), as reported by Kujawa and Liberman (2009), which cannot be detected with pure-tone threshold testing.

According to Anderson et al. (2010), SRN ability may also be impacted by neural timing delays, which cannot be detected by pure-tone threshold testing. These authors investigated the relationship between neural timing and SRN ability in children. For this study, participants were administered the Hearing in Noise Test (HINT; Nilsson et al., 1994) and were then divided into the "top speech-in-noise (SIN) group" and the "bottom SIN group" based on their SRN ability. Anderson et al. also measured auditory brainstem responses to a synthetic speech syllable ("da"). They compared the neural brainstem responses to "da" presented in quiet and in noise. For the quiet condition, they found no significant difference in neural timing between the top SIN group and the bottom SIN group. Using a two-way mixed-model ANCOVA, they found that background noise significantly delayed brainstem responses in all participants ($p < 0.01$). The authors reported greater neural timing delays for the bottom SIN group as compared to the top

SIN group. For example, post hoc analyses indicated significant group by noise effects for peak 43 ($p < 0.01$). For the peak occurring at about 43 milliseconds (ms), there was an average latency of 0.18 ms for the top SIN group and an average latency of 0.97 ms for the bottom SIN group. This shows that both groups experienced a delay in the presence of noise. It also shows that there was a greater delay for the bottom SIN group as compared to the top SIN group. This study demonstrated that poor SRN ability, as measured by the HINT, is associated with greater neural timing delays.

Semantic Interference

Semantic (meaningful) content in background babble is another factor that can impact speech perception ability. Semantic interference, also referred to as perceptual masking, is the phenomenon in which semantic content in background babble interferes with the understanding of target speech in the foreground (Carhart, 1969). A review of the literature on semantic interference is presented in Table 1. For each study in Table 1, the details of the test are included. Test information including: the target speech and noise conditions used are indicated, along with information on the participants and their mean pure-tone thresholds. The scores from each study for forward and reversed maskers are included in the table and semantic interference was calculated for each test group. Semantic interference is calculated as the score of correct responses in the presence of reversed babble minus the score of correct responses in the presence of forward babble.

Sperry (1997) experimented with a group of individuals who presented with no hearing loss. For this experiment, 6-talker babble forward and reversed was used as the masking noise, and the target speech was the female-talker version of NU-6 50-item word list. Sperry found that participants performed better in the presence of reversed babble (without semantic content) than in the presence of forward babble (with semantic content). The average performance for the forward babble condition was 40.3% words correct. For the reversed babble condition, the average performance was 56.1% words correct. In Sperry's study, the semantic interference was 15.8 percentage points, which was statistically significant ($p < 0.01$). This supports the idea that semantic content in the masking noise (babble) can make it more difficult to detect the target speech.

Hoen et al. (2007) compared semantic interference for three different noise conditions: 4-talker babble, 6-talker babble, and 8-talker babble. The participants for this study were French listeners, and the target speech was a list of 288 French mono-syllabic, tri-phonemic words. Each noise condition was presented forward and reversed to investigate the impact of semantic content in background babble on SRN ability. For the 4-talker babble condition, there was a six-percentage point difference in lexical retrieval for forward and reversed maskers, with better performance in the presence of reversed babble ($p < 0.01$). For the 6-talker babble condition, there was a two-percentage point difference in lexical retrieval between the forward and reversed babble, with better performance with forward babble (no statistically significant difference). For the 8-talker babble, there was no difference in lexical retrieval for forward and reversed babble (no statistically significant difference).

Wilson et al. (2012) compared semantic interference for young participants without hearing loss to that of older participants with hearing loss. Each participant was presented with a 6-talker babble masker at 80 dB SPL and the Words-in-Noise (WIN) word list starting at 104 dB SPL and decreasing to 80 dB SPL by 4 dB steps. In young participants with no hearing loss, they found that performance was 0.3 dB better in the presence of reversed babble than in forward babble (not statistically significant). In older participants with hearing loss, they found a 0.3 dB difference in performance for forward vs. reversed masker results. Older participants exhibited better performance in the presence of forward babble than in reversed babble (not statistically significant).

Ormond et al. (2022) compared the amount of semantic interference for 4-talker babble and conversational babble. AzBio sentences (Spahr et al., 2012) were used as the target speech. The 4-talker babble was from the same source file as the babble used for the Quick Speech-in-Noise (QuickSIN: Killion et al., 2004) test. It should be noted that the standard masker for the AzBio test is a 10-talker babble. Each babble masker was presented both forward and reversed to measure the effects of semantic interference. For the 4-talker babble condition, participants performed an average of 22.12 percentage points better when the masking noise was reversed (no semantic content) than when it was forward (with semantic content). As for conversational babble, there was no statistically significant difference between results for the forward and reversed babble conditions. In a separate study, Vermiglio et al. (2022) investigated the relationship between bilateral pure-tone average and AzBio scores using a 10-talker babble forward masker. The authors reported a very weak and statistically non-significant relationship between these two values ($r = 0.13$, $p = 0.39$).

The primary purpose of the current study was to investigate the semantic interference of 4-talker babble and 10-talker babble maskers. Both babble maskers were presented forward and reversed. Based on the studies conducted by Hoen et al. (2007) and Ormond et al. (2022), it was hypothesized that semantic interference would be greater for the 4-talker babble than for the 10-talker babble listening condition. The secondary purpose of this study was to investigate the relationship between pure-tone average vs. AzBio performances for the four listening conditions. Based on the study by Vermiglio et al. (2022), it was hypothesized that a weak and statistically non-significant relationship would be found between bilateral pure-tone average and the AzBio scores for each of the listening conditions.

Study	SRN Test	Noise Conditions	Level of Background Noise	Target Speech	Level of Target Speech	Participants	Age	Mean Pure tone average (PTA)	Performance of Forward Masker Result	Performance of Reversed Masker Result	Semantic Interference*	p-value
Sperry (1997)	Northwestern University Auditory Test No. 6 (NU-6)	Multitalker competing message (forward and reversed)	Presentation level 40 dB SL re:	Female-talker version of NU-6-50-item word lists	Presentation level 40 dB SL re: PTA for 1.0, 2.0, 4.0 kHz	11 women and 7 men w/ no history of ear disease or excessive noise exposure	18-30 yr	No poorer than 10 dB HL	40.3% correct	56.1% correct	15.8 percentage points	$p < 0.01$
Hoen et al (2007) (4-talker)	French words in noise	4-talker babble (forward and reversed)	Not clearly stated	288 (French momo-syllabic, tri-phonemic) target words mixed together with 4 seconds of silence	Not clearly stated	36 volunteers, native French speakers w/ no known HL or language disorders	18-32 yr	Not reported	57% lexical retrieval rate	63% lexical retrieval rate	6 percentage points	$p < 0.01$
Hoen et al (2007) (6-talker)	French words in noise	6-talker babble (forward and reversed)	Not clearly stated	288 (French momo-syllabic, tri-phonemic) target words mixed together with 4 seconds of silence	Not clearly stated	36 volunteers, native French speakers w/ no known HL or language disorders	18-32 yr	Not reported	63% lexicial retrieval rate	61% Lexical retrieval rate	-2 percentage points (better performance w/ forward babble)	n.s
Hoen et al (2007) (8-talker)	French words in noise	8-talker babble (forward and reversed)	Not clearly stated	288 (French momo-syllabic, tri-phonemic) target words mixed together with 4 seconds of silence	Not clearly stated	36 volunteers, native French speakers w/ no known HL or language disorders	18-32 yr	Not reported	59% lexical retrieval rate	59% Lexical retrieval rate	0 percentage points	n.s
Wilson et al (2012) (young w/o HL)	Words-in-Noise test (WIN)	Multi-talker (6-talker) babble (forward and reversed)	80 dB SPL	List 1 & List 2 of WIN	104-80 dB SPL (decreased in 4 dB incriments)	16 listeners w/o HL	Mean: 23.9 yr	8.8 dB HL	4.2 dB S/N (MEAN)	3.9 dB S/N (MEAN)	0.3 dB difference (better w/ backwards babble)	$p = 0.45$
Wilson et al (2012) (old w/ HL)	Words-in-Noise test (WIN)	Multi-talker (6-talker) babble (forward and reversed)	80 dB SPL	List 1 & List 2 of WIN	104-80 dB SPL (decreased in 4 dB incriments)	64 listeners w/ Hearing loss	Mean: 65.6 yr	27.8 dB HL	11.7 dB S/N (MEAN)	12.0 dB S/N (MEAN)	0.3 dB difference (better w/ forwards babble)	$p = 0.5$
Ormond et al. (2022) (4-talker)	AzBio with standard and alternative maskers	4-talker babble speech masker (forward and reversed)	65 dBA	AzBio sentences	65 dBA	52 Native English speakers	Mean: 21.33 yr	$\leq$ 25 dB HL, 0.25- 8.0 kHz (with the exception of 2 participants who had thresholds of 30 dB at 8.0 kHz)	mean score: 44.79%	mean score: 66.92%	22.12 percentage points	$p < 0.01$
Ormond et al. (2022) (conversational babble)	AzBio with standard and alternative maskers	Conversational babble speech masker (forward and reversed)	65 dBA	AzBio sentences	65 dBA	52 Native English speakers	Mean: 21.33 yr	$\leq$ 25 dB HL, 0.25- 8.0 kHz (with the exception of 2 participants who had thresholds of 30 dB at 8.0 kHz)	mean score: 64.35%	mean score: 65.62%	1.27 percentage points	$p = 0.41$

Table 1. A review of the literature on semantic interference. *= reversed masker result minus forward masker result

Methods

Thirty-five young, female, native English speakers were tested for this study. The mean age for participants was 21 years, with a range from 19 to 30 years. All participants had clear ear canals and normal pure-tone thresholds (≤ 25 dB HL, 0.25-4.0 kHz), with the exception of one participant who had a threshold of 30 dB HL at 4 kHz for the left ear. The AzBio sentences were presented in the presence of 4-talker and 10-talker babble conditions.

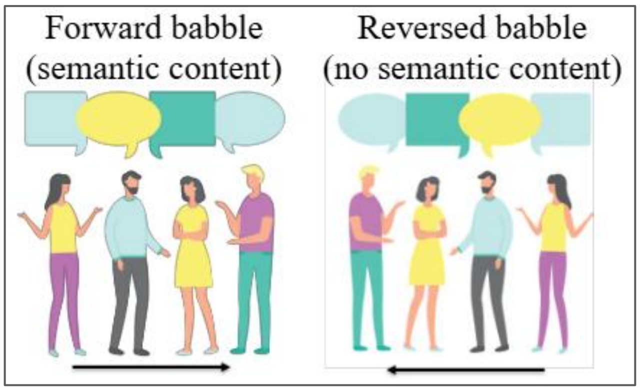

Figure 1. AzBio Sentences Presented in Forward and Reversed Babble Conditions.

As shown in Figure 1, the AzBio test was used to evaluate binaural speech perception in the presence of forward and reversed babble maskers. The forward babble maskers contained semantic content and the reversed babble maskers had no semantic content. The test conditions used for the current experiment were: 4-talker babble forward, 4-talker babble reversed, 10-talker babble forward, and 10-talker babble reversed. All test stimuli (AzBio sentences and babble maskers) were presented at 65 dBA (0 dB SNR) through supra-aural headphones in a sound-treated booth. Participants were asked to report back the sentences they perceived, and the AzBio score was calculated as the percentage of correctly repeated words. Semantic interference was then calculated as the AzBio score for the reversed condition minus the score for the forward condition. All test conditions were randomized to prevent order effects.

Results

Table 2 shows the descriptive statistics for all variables. The highest mean AzBio score was 71.27%, which was found for the 4-talker babble reversed condition. The lowest mean score was found for the 4-talker babble forward condition, and it was 51.58%. The largest range in AzBio scores was 71.05 percentage points, which was found for the 4-talker babble forward condition. The smallest range was 41.64 percentage points, which was found for the 4-talker babble reversed condition. Figure 2 shows that the mean semantic interference for 4-talker babble was 19.69 percentage points, and the mean semantic interference for 10-talker babble was 9.88 percentage points. The difference between these values was 9.81 percentage points, which was statistically significant ($p < 0.01$).

Masker Type	Mean (%)	SD	Max (%)	Min (%)	Range
4-Talker Babble Forward	51.58%	16.86	99.32%	28.26%	71.05*
4-Talker Babble Reversed	71.27%	9.04	92.75%	51.11%	41.64*
Semantic Interference 4-Talker Conditions	19.69*	14.82	43.66*	-19.13*	62.79*
10-Talker Babble Forward	59.05%	11.14	82.22%	39.86%	42.37*
10-Talker Babble Reversed	68.93%	10.50	97.83%	47.41%	50.42*
Semantic Interference 10-Talker Conditions	9.88*	10.39	32.61*	-17.90*	50.51*

Table 2. Descriptive Statistics for all AzBio Scores (n = 35). * = percentage points

A repeated-measures ANOVA was conducted to investigate the effects of masker condition (4-talker babble forward, 4-talker babble reversed, 10-talker babble forward, and 10-talker babble reversed) on AzBio scores. This analysis revealed that the main effect was statistically significant (F value is 27.35 for masker condition, $p < 0.01$). A post-hoc analysis using t-tests revealed statistically significant differences for all comparisons ($p < 0.01$) with the exception of 10-talker babble reversed vs. 4-talker babble reversed (not statistically significant).

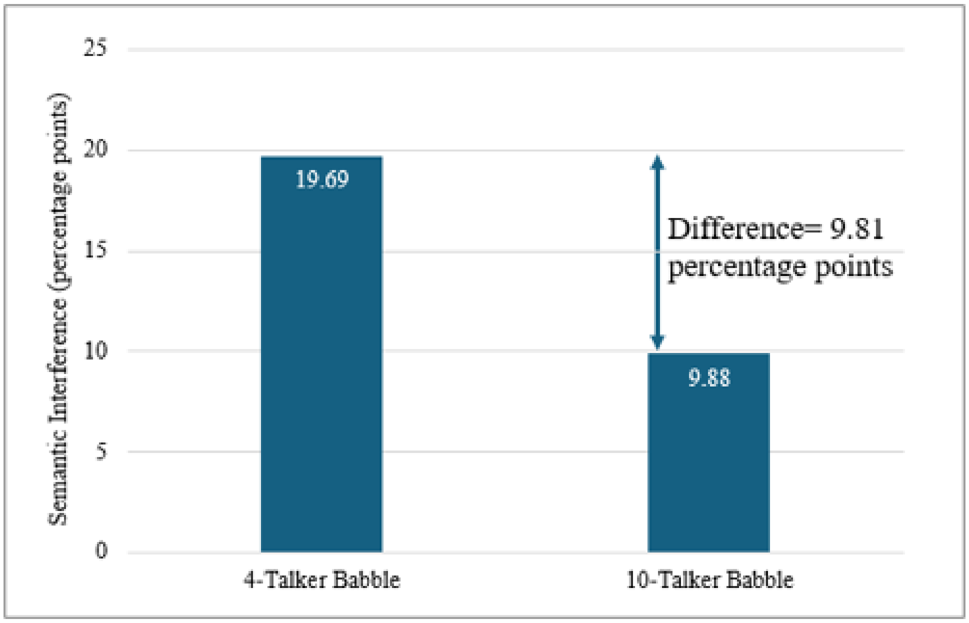

Figure 2. Difference in Semantic Interference Between Test Conditions. This difference was statistically significant ($p < 0.01$).

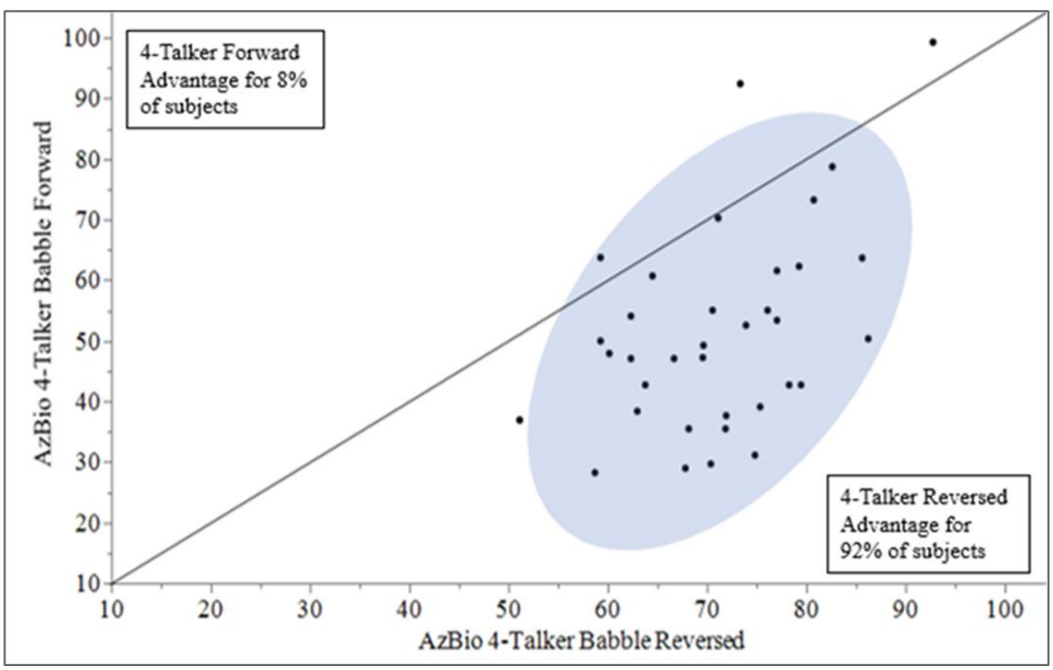

Figure 3. AzBio 4-talker Babble Forward vs. 4-talker Babble Reversed.

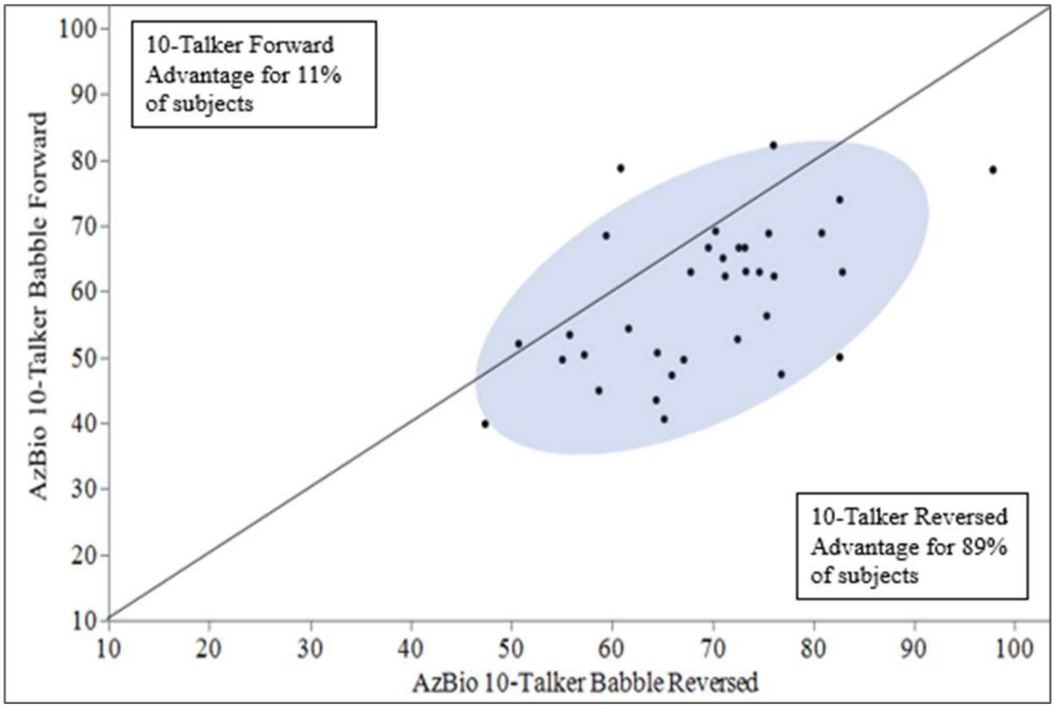

Figure 4. AzBio 10-talker Babble Forward vs. 10-talker Babble Reversed.

Figures 3 and 4 show speech intelligibility performance for forward vs. reversed babble results for 4-talker and 10-talker conditions, respectively. Figure 3 indicates that 92% of subjects performed better with reversed babble and 8% performed better with forward babble. Figure 4 indicates that 89% of subjects performed better with reversed babble and 11% performed better with forward babble. There was a moderate and statistically significant relationship between forward and reversed scores for 4-talker babble ($r = 0.42$, $p = 0.01$) and 10-talker babble ($r = 0.48$, $p < 0.01$).

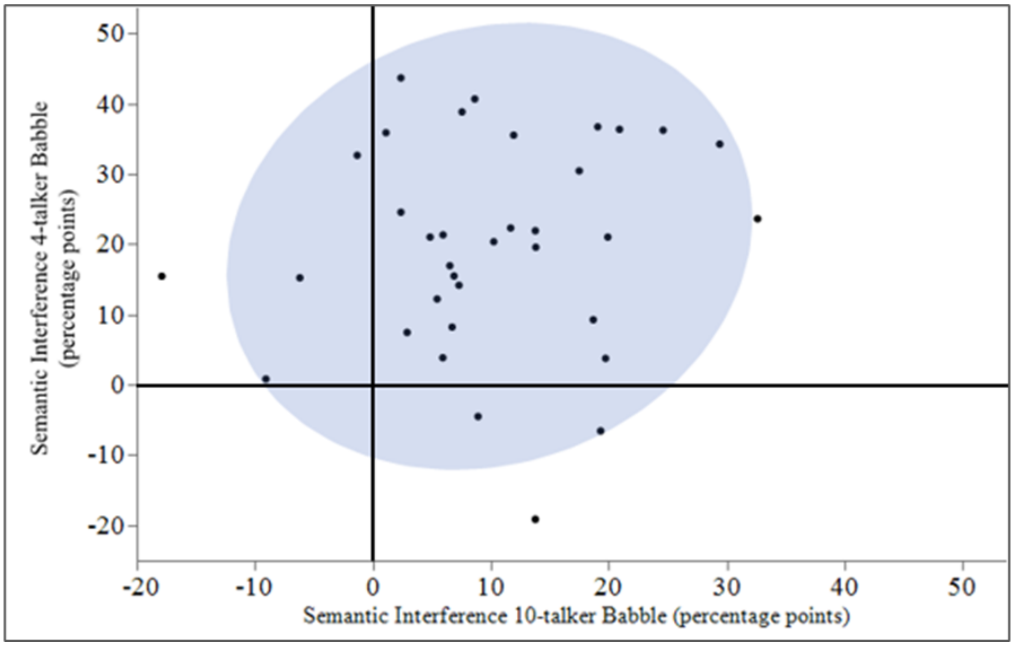

Figure 5. Semantic Interference 4-talker Babble vs. 10-talker Babble.

Figure 5 shows semantic interference for 4-talker babble compared to 10-talker babble. Semantic interference is indicated by any value greater than zero. Semantic advantage is a phenomenon in which the listener has better performance in forward babble rather than reversed babble, this is indicated by any value less than zero. There was a very weak and statistically non-significant relationship between semantic interference for 4-talker babble vs. 10-talker babble (r = 0.14, p = 0.41).

Bilateral Pure-Tone Average vs. AzBio Results

The relationships between bilateral pure-tone average (BPTA) and AzBio scores for all four listening conditions were determined. The following BPTAs were used: BPTA 500, 1000, 2000 Hz; BPTA 3,000, 4,000, 6,000 Hz; BPTA 250 - 8,000 Hz. All the relationships between BPTA vs. AzBio scores were weak and statistically non-significant ($p > 0.05$).

Discussion

Semantic Interference

Recall that the first hypothesis of this study was that there would be greater semantic interference for the 4-talker babble condition than the 10-talker babble condition. For the 4-talker babble condition there was a semantic interference of 19.69 percentage points. For the 10-talker babble condition, there was a semantic interference of 9.88 percentage points. The semantic interference for the 4-talker babble condition was 9.81 percentage points greater than that of the 10-talker babble condition, which was statistically significant ($p < 0.01$). This supports the first hypothesis of this study. Likewise, Ormond et al. (2022) reported greater semantic interference found for the 4-talker babble condition than the semantic interference found for the conversational masker condition. They reported a semantic interference of 22.12 percentage points for the 4-talker babble condition, and a semantic interference of 1.27 percentage points for the conversational masker condition. The semantic interference reported by Ormond et al. (2022) for the 4-talker babble condition was 20.85 percentage points greater than that of the conversational masker condition.

Contrary to the present study, Hoen et al. (2007) reported a semantic interference of only 6 percentage points for 4-talker babble ($p < 0.01$). This value was lower than the semantic interference calculated for the current study and the study conducted by Ormond et al. (2022). The difference in these results indicates that there was greater sematic content in the 4-talker babble used for the current study and the study by Ormond et al. (2022) than in the study conducted by Hoen et al. (with French listeners). For each of these studies, a comparison across talker babble conditions revealed the greatest semantic interference for 4-talker babble.

Comparison of AzBio Results Across Studies

Study	4-Talker Babble Forward AzBio Mean	4-Talker Babble Forward AzBio Range	10-Talker Babble Forward AzBio Mean	10-Talker Babble Forward AzBio Range	20-Talker Babble Forward AzBio Mean	20-Talker Babble Forward AzBio Range
Current Study	51.58%	71.05 percentage points	59.05%	42.37 percentage points		
Ormond et al. (2022)	44.79%	60.29 percentage points				
Vermiglio et al. (2022)			63.27%	55.31 percentage points		
Holder et al. (2018)					90.80%	Approximately 13 percentage points

Table 3. A comparison of AzBio results across studies. All test stimuli were delivered at 0 dB SNR.

Table 3 compares the AzBio results across studies. The participants for each study had normal pure-tone thresholds, and all the test stimuli were delivered at 0 dB SNR. For the 4-talker babble forward condition, the current study reported an average AzBio performance of 51.58% and Ormond et al. (2022) reported an average AzBio performance of 44.79%. The mean AzBio performance reported by Ormond et al. (2022) for the 4-talker babble forward condition was 6.79 percentage points lower than that reported in the current study. The current study reported a range of 71.05 percentage points for the 4-talker babble forward condition and Ormond et al. (2022) reported a range of reported a range of 60.29 percentage points. The range reported by Ormond et al. (2022) for the 4-talker babble forward condition was 10.76 percentage points less than that reported by the current study.

For the 10-talker babble forward condition, the current study reported an average AzBio performance of 59.05%. Vermiglio et al. (2022) reported an average AzBio performance of 63.27% for the 10-talker babble forward condition. The mean performance reported by Vermiglio et al. (2022) was 4.22 percentage points more than that reported by the current study.

The current study reported a range of 42.37 percentage points for the 10-talker babble forward condition. Vermiglio et al. (2022) reported a range of 55.31 percentage points for the 10-talker babble forward condition. The range reported by Vermiglio et al. (2022) was 12.94 percentage points more than that reported by the current study.

Holder et al. (2018) conducted the AzBio test on a group of participants between 20 and 79 with normal pure-tone thresholds. Participants were divided into groups based on their ages. They administered the AzBio test in the presence of 20-talker babble at different signal-to-noise ratios (SNRs). All test materials were presented through a single loudspeaker. When the AzBio was administered in quiet, the group of participants between ages 21-49 years scored an average of 99.85%. As the SNR decreased, the AzBio performances decreased among all age groups. For the 0 dB SNR condition, Holder et al. (2018) reported a maximum score of approximately 98% and a minimum score of approximately 85% for the group of participants between the ages 21-49 years. The range of performances for the group of participants between the ages 21-49 years was estimated to be 13 percentage points. The differences in the mean and range of AzBio performances indicate that there was most likely far less detectable semantic content in the 20-talker babble masker used by Holder et al. (2018) than in the 4-talker and 10-talker babble maskers used in the current study. The wide range in AzBio results found in the three studies in Table 3 indicate that pure-tone threshold testing alone does not provide the full picture of a patient's hearing ability.

Pure-tone Average vs. AzBio Performance

The secondary purpose of this study was to investigate the relationship between BPTA and AzBio performances for all four listening conditions. All the relationships between BPTA and AzBio Scores were weak and statistically non-significant ($p > 0.05$). The results from this study support the hypothesis that there is no significant relationship between BPTAs and AzBio scores. These results were consistent with the results of Vermiglio et al. (2022), who reported a very weak and non-statistically significant relationship between BPTA (0.5-6.0 kHz) and AzBio scores ($r = 0.13$, $p = 0.39$). These results indicate that SRN ability cannot be predicted by pure-tone thresholds; therefore, it must be measured directly rather than being inferred from pure-tone threshold testing.

Importance of SRN testing

Fry (1942) claimed that one's ability to interpret signals depended on both hearing and intelligence. He concluded that measuring hearing acuity alone is not enough to fully understand one's SRN ability. Similarly to Fry (1942), the American Academy of Otolaryngology (1979) argued that auditory function should be assessed not only with pure-tone thresholds, but also with some measure of SRN ability. According to the World Health Organization (WHO, 2021) pure-tone thresholds are an accurate measure of hearing ability in daily life. This establishes pure-tone threshold testing as the gold standard; however, the current study indicates that individuals with normal pure-tone thresholds may exhibit poor SRN abilities as shown by the wide range of SRN scores (from very poor to very good) in the present study and previous investigations (Middelweerd et al., 1990; Vermiglio et al., 2012; Vermiglio et al., 2022). For this reason, SRN ability must be measured directly rather than being inferred from pure-tone threshold testing results.

Conclusion

Generally, participants displayed better speech perception in noise ability in the presence of reversed babble compared to forward babble. This is due to the effect of semantic interference. When there is semantic content in the background masker, participants have a more difficult time perceiving target speech. Greater semantic interference was observed in the presence of 4-talker babble compared to 10-talker babble. This is due to the amount of semantic content present in the masker. It is evident that 4-talker babble has more detectable semantic content than the 10-talker babble. It was also concluded from this study that there is no significant relationship between bilateral pure-tone threshold averages and AzBio scores.

Clinical Implications

Speech-in-noise testing can provide important information about a patient's hearing that pure-tone threshold testing cannot. Speech-in-noise test results may be used for employment considerations, recommendations for listening technologies (such as hearing aids and FM systems), and for accommodations in academic settings. Individuals are impacted by semantic interference differently, and it can be attributed to factors such as attention and listening conditions. Detecting the amount of semantic interference and understanding how it may impact a patient’s hearing allows for more useful interpretation of test results and provides for proper counseling and recommendations for hearing disorders.

References

American Academy of Ophthalmology and Otolaryngology. (1979). *Guide for the evaluation of hearing handicap*. American Academy of Ophthalmology & Otolaryngology.

Anderson, S., Skoe, E., Chandrasekaran, B., & Kraus, N. (2010). Neural timing is linked to speech perception in noise. *Journal of Neuroscience*, *30*(14), 4922-4926.

Carhart, R., Tillman, T. W., & Greetis, E. S. (1969). Perceptual masking in multiple sound backgrounds. *The Journal of the Acoustical Society of America*, *45*(3), 694-703.

Dickson, E. D. D., & Bateman, G. H. (1944). A Survey of the Aural Health of Aircrew Candidates for Service with the Royal Air Force. *The Journal of Laryngology & Otology*, *59*(12), 437-442.

Dickson, E. D. D., Simpson, J. F., Fry, D. B., Swindell, G. E., & Brown, R. E. C. (1946). A new method of testing the hearing efficiency of aviation candidates. *The Journal of Laryngology & Otology*, *61*(3), 139-203.

Fry, D. B. (1942). A suggestion for a new method of testing hearing in aviation candidates. *The Journal of Laryngology & Otology*, *57*(1), 11-13.

Hinchcliffe, R. (1992). King-Kopetzky syndrome: An auditory stress disorder. *Journal of Audiological Medicine*, *1*(2), 89-98.

Hoen, M., Meunier, F., Grataloup, C. L., Pellegrino, F., Grimault, N., Perrin, F., ... & Collet, L. (2007). Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. *Speech communication*, *49*(12), 905-916.

Holder, J. T., Levin, L. M., & Gifford, R. H. (2018). Speech recognition in noise for adults with normal hearing: Age-normative performance for AzBio, BKB-SIN, and QuickSIN. *Otology & Neurotology*, *39*(10), e972-e978.

King, P. F. (1954). Psychogenic deafness. *The Journal of Laryngology & Otology*, *68*(9), 623-635.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *116*(4), 2395-2405.

Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. *Journal of Neuroscience*, *29*(45), 14077-14085.

Liberman, M. C., Epstein, M. J., Cleveland, S. S., Wang, H., & Maison, S. F. (2016). Toward a differential diagnosis of hidden hearing loss in humans. *PloS one*, *11*(9), e0162726.

Middelweerd, M. J., Festen, J. M., & Plomp, R. (1990). Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram: Original papers. *Audiology*, *29*(1), 1-7.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The journal of the acoustical society of America*, *95*(2), 1085-1099.

Ormond, A. (2022). The Effect of Semantic Interference on Speech Perception in Noise Ability.

Pryce, H., Metcalfe, C., Claire, L. S., & Hall, A. (2010). Causal attributions in King-Kopetzky syndrome. *International Journal of Audiology*, *49*(7), 482-487.

Saunders, G. H., & Haggard, M. P. (1992). The clinical assessment of C "Obscure Auditory Dysfunction" (OAD) 2. Case control analysis of determining factors. *Ear and hearing*, *13*(4), 241-254.

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., ... & Cook, S. (2012). Development and validation of the AzBio sentence lists. *Ear and hearing*, *33*(1), 112-117.

Sperry, J. L., Wiley, T. L., & Chial, M. R. (1997). Word recognition performance in various background competitors. *Journal-American Academy of Audiology*, *8*, 71-80.

Vermiglio, A. J., Osborne, H. R., Bonilla, E., Leclerc, L., Thornton, M., & Fang, X. (2022). The Relationship between Speech Perception in Quiet and in Noise for Young Adults with Pure-Tone Thresholds≤ 25 dB HL. *Journal of the American Academy of Audiology*, *33*(07/08), 390-395.

Vermiglio, A. J., Soli, S. D., Freed, D. J., & Fisher, L. M. (2012). The relationship between high-frequency pure-tone hearing loss, hearing in noise test (HINT) thresholds, and the articulation index. *Journal of the American Academy of Audiology*, *23*(10), 779-788.

Wilson, R. H., Trivette, C. P., Williams, D. A., & Watts, K. L. (2012). The effects of energetic and informational masking on the Words-in-Noise Test (WIN). *Journal of the American Academy of Audiology*, *23*(07), 522-533.

World Health Organization. (2021). *World report on hearing*. World Health Organization.