Bilingual Speech Perception in Noise:

Effects of Target Predictability, Noise Type, and Language Background

By

Sarah Varley

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Bachelor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Sarah R. Varley

It was defended on

November 18, 2020

and approved by

Dr. Barbara Kucinski, Lecturer II, Department of Psychology

Dr. Eleonora Rossi, Assistant Professor, Department of Psychology and Linguistics

Dr. Natasha Tokowicz, Associate Professor, Department of Psychology

Dr. Melinda Fricke, Assistant Professor, Department of Linguistics

Bilingual Speech Perception in Noise:

Effects of Target Predictability, Noise Type, and Language Background

Sarah R. Varley, BPhil

University of Pittsburgh, 2020

To examine how different types of noise and potential context benefits impact various groups of listeners (English monolinguals, English-Spanish L2 bilinguals, and English-Spanish heritage bilinguals), we utilized a word transcription task where participants heard English sentences in four different noise conditions (clear, speech-shaped noise, English two-talker babble and Spanish two-talker babble) and in high or low predictability sentences. Participants were instructed to type the final, target word of the sentence they heard into a text box. Results suggested that the ability to use sentence context as a SPIN resource is related not only to different processing demands, but also varying cognitive control abilities, strength of L2 activation and whether the L2 was learned natively (i.e. heritage bilinguals) or non-natively (L2 bilinguals). Findings also displayed that the overall accuracy in each noise condition varied across speaker groups. More specifically, even when speaker groups performed with the same global accuracy, differences in accuracy between the high and low predictability conditions reflect differences in how each group approached and coped with the task. Findings from the present study display intricate differences in how monolingual, L2, and heritage speakers can cope with noise and use context to their advantage. Ultimately, findings suggest that L2 bilinguals, who developed an English representation completely independent of Spanish, appear to be less adversely impacted by English noise than heritage bilinguals, whose English representations have never existed entirely independent of their Spanish representations. Importantly, this underscores a difference in how two groups that both identify as native English speakers bilingual in Spanish are impacted differently during speech perception in environments of native L1 noise.

Keywords: speech perception in noise, informational masking, context, bilingual

Table of Contents

List of Tables

List of Figures

1.0 Introduction

1.1 Bilingual Speech Perception in Noise

Previous research on bilingualism has produced consistent results displaying a deficit in bilinguals' speech perception in noise (SPIN) such that word recognition occurs relatively easily in quiet conditions but becomes increasingly challenging in environments where background noise deteriorates verbal cues (Bronkhorst, 2000; Lecumberri & Cooke, 2006; Shi, 2010; Tabri, Chacra, & Pring, 2011). In a widely cited study by Cooke, Lecumberri and Barker (2008), native English monolinguals and native Spanish bilinguals identified target keywords within English sentences in quiet environments and in environments of increasing noise up to 6 decibels louder than the target speech. Results displayed the non-native English-speaking bilinguals as being disproportionately affected by the increasing noise when compared to the native counterparts, indicating that noise can disproportionately hinder non-native speakers' speech perception. The possible reasons for this phenomenon are numerous and are summarized in Figure 1 from Cooke et al. (2008).

1.2 Energetic and Informational Masking

As displayed in Figure 1, the two overarching factors that have been suggested to explain bilinguals' weaker performance in SPIN are *energetic masking* and *informational masking*. When a speech signal is physically interfered with by the noise 'masking' or covering the signal, this is referred to as *energetic masking* (Pollack, 1975). Put simply, this is the idea that when you have two separate streams of sound, the acoustic energy of one stream degrades the other target acoustic stream such that it becomes difficult to accurately decode. The other factor contributing to bilinguals' SPIN deficit is *informational masking*, a catch-all term used to encompass the additional difficulty, above and beyond energetic masking, of a second speech signal containing meaningful information that interferes with the target speech identification.

Multiple factors have been suggested to account for the difficulty that comes with informational masking as displayed in the aforementioned figure. Of note is the roles of interference from a known-language masker, higher cognitive load, and target predictability. Recently, Krizman et al. (2017) have suggested that the disproportionate effects of noise bilinguals face may be due in part to cross-language activation. The proposal is that bilingual participants perform at lower levels than monolinguals because they have to cope with activation of the non-target language, which may incur additional cognitive taxation. This explanation of higher cognitive load affecting SPIN is based on the idea that the presence of a noise signal, which deteriorates the target signal, increases the cognitive resources it takes the listener to identify the target signal. Because cognitive resources are limited, this additional cognitive taxation that comes with being bilingual (suggested to stem at least in part from cross language activation) decreases the resources bilinguals can allocate to the target speech.

Regarding the role of context, results from Bradlow and Alexander (2007) support the idea that bilinguals gain less information from context when listening in their second language (L2) due to more difficultly decoding information at multiple levels, including phonetic, lexical, and semantic. When these effects are compounded, bilinguals are unable to reap full contextual benefits as compared to native monolingual speakers who can, in performance, successfully use context to make accurate predictions about the masked target speech in the L2. In a study examining SPIN performance across different speaker types, Tabri, Chacra, and Pring (2011) found that highly proficient bilinguals who regarded their L2 as their dominant language displayed the bilingual SPIN deficit. This finding suggests that an L2 may remain impaired no matter how proficient it becomes because of the nature of its later acquisition than the L1. An alternative explanation would be that regardless of order of acquisition or dominance, all of a bilingual's languages become impaired. Additional findings by Weiss and Dempsey (2008) displayed that less experience in the L2 correlated with better speech-in-noise perception in the L1. Taken together, these findings suggest that bilinguals being disproportionately affected by noise in L2 SPIN may extend to performance in L1 SPIN.

1.3 Impact of Different Types of Noise on SPIN

Of additional interest is what the non-target noise signal consists of. Simpson and Cooke (2005) have shown that noise with linguistic information in it affects listeners differently than non-linguistic noise. Building on this idea, Van Engen and Bradlow (2007) explored the impact of different linguistic noise conditions. They found that in two-talker babble, where two different speakers with competing speech streams make up the noise (i.e. two people sound like they are trying to talk over each other), native English-speaking participants' performance at SPIN was hindered significantly more by English babble than Mandarin babble. This implies that not only does it matter whether or not the noise consists of linguistic information, but it also matters what those specific signals are and how they interact with the listener's own language history.

Importantly, however, participants in this study were either monolingual English speakers or native English speakers bilingual in a language other than Mandarin.

Following up on these results, Van Engen (2010) conducted another study examining Native English speakers and L2 English speakers whose L1 was Mandarin. Participants were tested on English target recognition in both English and Mandarin two-talker babble. Results displayed that both groups of participants experienced greater difficulty in English babble versus Mandarin babble, but importantly that the native Mandarin speakers experienced a smaller improvement in performance in Mandarin two-talker babble noise compared to English two-talker babble noise. This suggests that similarity between the target and noise *and* the unique language experience of the listener contribute to how much interference is experienced while listening to target speech in the presence of two-talker noise. Considering findings of both studies, the question still remains how variability in L2 proficiency might impact performance in different types of noise when the listener is a speaker of both the target language and noise language.

1.4 Heritage vs L2 Bilinguals

In addition to the various types of noise that can impact speakers' SPIN abilities, there are various types of bilinguals who must cope with SPIN. In particular, there exist those bilinguals who have always been fluent in two languages and those who have learned the second language later in life as a conscious effort. Individuals who grew up speaking a language at home that is different from the dominant language in their community are called *heritage speakers*. For example, someone who grew up in the United States speaking Spanish at home and English at school and everywhere else would be considered a heritage speaker of Spanish. In contrast, if someone grew up as a monolingual English speaker, started taking Spanish classes in high school, and developed proficiency in Spanish such that they were bilingual, they would be considered an L2 speaker of Spanish.

Work from Bialystok, Craik, Green and Gollan (2009) and Green and Abutalebi (2013) displays that use of multiple languages plays a role in shaping an individual's performance in non-verbal tasks measuring cognitive control. It has been suggested that this occurs because the additional cognitive demands that come with controlling not one but two (or more) languages lead to changes in cognitive skills. Because heritage bilinguals have used more than one language for significantly longer than L2 speakers, they have had to deal with controlling two competing languages (for example dealing with cross language activation) significantly more than L2 speakers. As such, it's possible that heritage bilinguals have more highly developed cognitive skills and resources than L2 bilinguals. Because cognitive load is one of the cited factors that make up informational masking, how much cognitive load different types of bilingual speakers can handle (stemming from differences in language profile) may produce differences in SPIN performance.

1.5 The Present Study

The current literature on bilinguals' speech perception in noise suggests that bilinguals are disproportionately affected by noise because of (1) inability to reap full contextual benefits, (2) the type of noise deteriorating the target signal, and (3) the interaction between their L1 and L2 knowledge. However, whether the inability to benefit from context in SPIN in the L2 extends to deficits in the L1 has not been explicitly tested, nor has it been examined in the context of varying noise conditions, which could reasonably be expected to tax the language processing system in different ways. Further, it remains of interest whether different types of bilinguals, such as heritage speakers versus L2 speakers (i.e. bilingual English-Spanish speakers both native in the same L1 are affected in the same ways) by SPIN.

The goal of the present study is to unpack and understand these questions and gain a deeper understanding of how the bilingual experience impacts speech processing in adverse listening environments where noise deteriorates the target signal. Specifically, we seek to understand whether bilinguals' weaker performance in SPIN can be attributed to how different processing demands impact the ability to make accurate predictions about target words. In the present study we look specifically at the L1 where there exists little controversy that listeners utilize context to make predictions about target speech. We are also interested in the interaction of noise type and listener's language profile (monolingual vs L2 vs heritage). These factors were tested by measuring accuracy in transcribing the final target word of sentences varying in high and low contextual predictability and embedded in four different types of noise. The findings of the present research questions will help to bring about a deeper understanding of the effects of different types of noise, different types of linguistic experience, and more broadly, the way a bilingual speaker's L1 and L2 uniquely interact.

2.0 Methods

2.1 Design

The present study examines the effects of context and various types of noise in both monolinguals' and bilinguals' SPIN by utilizing a word transcription task administered online using Gorilla Experiment Builder (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020). Using a mixed repeated measures design, three groups of participants heard high versus low contextually constraining sentences in four noise conditions: in the clear, and embedded in speech-shaped noise, English two-talker babble, and Spanish two-talker babble.

The sentence context manipulation, including the sentences, keywords, and design, were adapted from Bradlow and Alexander (2007). Half of the trials were contextually restrictive, leading the listener to predict the target word, while the other half of the trials were non-restrictive (but notably did *not* lead participants to make an inaccurate prediction regarding the target). 120 critical target words were rotated across the different noise conditions for different participants, for a total of 30 trials per noise condition (consisting of 15 high predictability and 15 low predictability sentences, the "context" manipulation). Context was counterbalanced across participants: each of the 120 critical target words were embedded within both a high predictability and low predictability sentence for a total of 240 developed sentences, but each participant heard each of the 120 critical target words only once. Items were rotated in this manner to make sure that any effects did not depend on the specific words or sentences in a given version of the experiment. Because the item rotations help guard against the possibility that any effects are being driven by particular items or sentence contexts, in this paper we present only the by-participant analyses. For a visual guide to see how items were rotated across noise conditions and participants see Table 1.

Table 1 Rotation of the 120 critical target words across different noise conditions for different participants.

Version	Clear Set	Speech-shaped Noise Set	English Babble Set	Spanish Babble Set
1	A1	B1	C1	D1
2	A2	B2	C2	D2
3	B1	A1	D1	C1
4	B2	A2	D2	C2

2.2 Materials

The present experiment examines word recognition in the clear as well as three different types of noise: speech-shaped noise, English two-talker babble and Spanish two-talker babble. The target words and sentences were adapted from Bradlow and Alexander (2007), Fallon, Trehub and Schneider (2002), and Munro (1998), and target word frequency was controlled across the different sets of words (see Table 2 below for example target words and their corresponding high and low predictability sentences). Sentences were recorded by a female native speaker of central American Spanish who learned English as a second language and (impressionistically) spoke it with a native English-sounding accent.

Table 2 Example target words and their corresponding high and low predictability sentences.

Target Word	Predictability	Sentence
Month	High	July is a summer month.
Month	Low	Mom talked about the month.
Sun	High	She talked about the sun.
Sun	Low	When it is cloudy, I can't see the sun.
Toys	High	The little girl played with her toys.
Toys	Low	She heard about the toys.
Belt	High	To hold his pants up he wears a belt.
Belt	Low	She heard about the toys.
Mirror	High	Mom talked about the month.
Mirror	Low	We looked at the mirror

The two-talker babble conditions were created using English and Spanish podcasts, respectively. Podcasts were selected based on their sound quality, presence of only a single (female) talker's voice at a time, and similar speaking style and semantic content for the two languages. Two English-language and two Spanish-language news podcasts were chosen and all non-speech sounds and pauses longer than 500 ms were removed using the Praat software for acoustic analysis (Boersma & Weenink, 2018). Each podcast's average intensity was scaled to 72 dB, while the target sentences' average intensity was scaled to 70 dB, for a signal-to-noise ratio (SNR) of -2 dB. For each target sentence and each language condition (English versus Spanish two-talker babble), a random excerpt was selected from each of the two podcasts and combined with the target sound file such that the two-talker babble started 500 ms prior to the onset of the target sentence and lasted until 500 ms after the target word's offset.

To create the speech-shaped noise, all four podcasts that were used to make the two-talker babble were used to generate a long-term averaged spectrum. This spectrum was then used as a filter for white noise: when noise with sound distributed randomly across all frequencies (i.e. white noise) is filtered through a spectral envelope that has the same shape as speech (i.e. sound with a very particular, non-random distribution of frequencies), it then becomes "speech-shaped" noise. Speech-shaped noise was then added to the target sentences such that it also started 500 ms prior to the onset of the target sound file and continued until 500 ms following the offset of the target word.

2.3 Procedure

The study was divided between two testing sessions. The first session acted as a screening session (see the section on Participants for more details on screening). All participants first completed the English LexTale proficiency measure (Lemhofer & Broersma, 2012), a lexical decision task that measures vocabulary knowledge, followed by the Spanish LexTale measure (Izura, Cuetos, & Brysbaert, 2014). The final task in the first session asked participants to complete an extensive language history questionnaire (LHQ) (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007) which included self-assessment of their fluency, speaking, understanding, and reading abilities in English and Spanish. The participants' data were then used to categorize them as monolinguals, heritage bilinguals, L2 bilinguals, or those who based on LHQ/LexTale data did not proceed to the experiment. Only participants whose detailed language histories matched the criteria for the study were then invited to participate in the main experiment session.

In the second session, all participants completed the experiment. Participants were instructed to listen to the sentences and transcribe the last word in each sentence. Participants transcribed the critical word into a text box and entered a "0" if there was a sound file error causing the recording to not play properly. This task consisted of four experimental blocks differing only by the unique noise conditions in each block. In block one, participants did the task in the clear without any noise that would deteriorate the target signal. Block two consisted of the main task in speech-shaped noise, followed by English two-talker babble and Spanish two-talker babble in blocks three and four. Participants were given three practice trials before the first block to orient themselves to the task, and these trials were not included in analysis. The order of the noise blocks was kept constant across participants. While ordering effects are expected, the is not the focus of the present study. Additionally, if noise block condition were to be rotated across participants, the number of participants required to look at the data in a systematic way would likely be more than is realistically possible to recruit. Following the word transcription task, participants went on to complete two other experimental SPIN tasks and the AX-CPT task (a measure of non-linguistic cognitive control), though data collected from these tasks were not analyzed for our purposes.

2.4 Participants

Participants belonged to one of three different speaker types: English monolinguals, English-Spanish heritage bilinguals, and English-Spanish L2 bilinguals. Participants were recruited through Prolific, a third-party online research participant recruitment platform (https://www.prolific.co). All participants independently created a Prolific profile based on their demographics and were recruited for the current study based on the following criteria: must live within the United States currently, must have grown up in the United States, must be (self-reported) native speakers of English, within 18-35 years of age, and must not have any known cognitive, language, learning, hearing or visual impairments. Additional criteria were also used to help identify participants whose language background might place them in one of the three participant groups for the study. To identify monolinguals, we searched for native English speakers who reported learning only English from birth and who specified that they did not speak any additional languages. For L2 bilinguals, we searched for native English speakers who reported learning only English from birth but who reported being fluent in Spanish. For heritage bilinguals, we searched for native English speakers who reported learning more than one language from birth and who also reported being fluent in Spanish.

The monolingual speaker group consisted of 18 speakers ranging in age from 19 to 34 years, with an average age of 29.2 years. The group consisted of 7 males, 10 females and 1 participant that didn't specify gender identity. Monolinguals' average age of English acquisition was 4 years and 100% of monolinguals self-reported as fluent in English. Using a scale of 0-10, with 0 meaning "no ability at all" and 10 "like a native speaker", monolinguals' average self-reported English reading was 9.94, English understanding was 9.4 and English speaking was 10. Because it is so common in the United States to have some type of second language learning during middle and/or high school, most monolingual recruited participants did report some level of second language exposure. For our purposes, recruited participants were included in the monolingual speaker group only if they reported that their fluency, understanding, speaking, and reading in a second language were all a 3 or less on the scale of 0-10. Monolinguals reported being exposed to English 99% of the time, Spanish 6.72% of the time, and any other languages 3.44% of the time.

The L2 late learners of Spanish group consisted of 17 speakers ranging in age from 19 to 35 years, with an average age of 28.88 years. The group consisted of 7 males and 10 females. L2 speakers' average age of English acquisition was 0.53 years and average age of Spanish acquisition was 13.43. L2 participants on average reported reaching fluency in Spanish at age 21, and fluency in English at age 3.63 (participants who specified "partial fluency" in Spanish were not included in the Spanish averages). They reported beginning to read in Spanish on average by age 14.63 and achieved reading fluency in Spanish by age 21.46. In English, they reported beginning to read on average by age 4.34 and achieved English reading fluency by age 7.69. 100% of L2 bilinguals' self-reported as fluent English speakers, while 11.76% self-reported full Spanish fluency, 82.35% self-reported partial fluency in Spanish, and 5.89% reported not being fluent in Spanish. Using a scale of 0-10, with 0 meaning "no ability at all" and 10 "like a native speaker", English reading was rated 10, English understanding was 10, and English speaking was 9.94. The Spanish counterpart averages were 6.82 for Spanish reading, 6.47 for Spanish understanding, and 6.06 for Spanish speaking. L2 bilinguals reported being exposed to English 90.76% of the time, Spanish 13.17% of the time, and any other languages 1.94% of the time.

The Heritage bilingual speaker group consisted of 16 speakers ranging in age from 18 to 35 years, with an average age of 25.33 years. The group consisted of 11 males and 5 females. Heritage speakers' average age of English acquisition was 1.93 years and average age of Spanish acquisition was 3.4 years. Heritage participants on average reported reaching fluency in Spanish at age 7.2, and fluency in English at age 4.8. They reported beginning to read in Spanish on average by age 8.76 and achieved reading fluency in Spanish by age 9.2. In English, they reported beginning to read on average by age 4.66 and achieved English reading fluency by age 6.60. 100% of heritage-bilinguals self-reported that they were fluent in English, while 37.5% self-reported partial fluency in Spanish and 43.75% self-reported full Spanish fluency. Using a scale of 0-10, with 0 meaning "no ability at all" and 10 "like a native speaker", average English reading ability was 6.60, English understanding was 9.38, and English speaking was 9.19. Their Spanish counterpart averages were 5.69 for Spanish reading, 7.13 for Spanish understanding, and 6.25 for Spanish speaking. 25% of heritage Speakers self-identified as a heritage speaker when given the option, while 81.25% reported Spanish as a language used growing up. The average reported number of years spent in a Spanish speaking household was 19.46. Heritage bilinguals reported being exposed to English 85.5% of the time, Spanish 29.44% of the time, and any other languages 2.75% of the time. See Table 3 below for a summarized breakdown of relevant participant self-report data.

Table 3 Breakdown of participants' average self-reported demographic information.

	Monolingual	**L2**	**Heritage**
N	18	17	16
Average Age	29.2 years	28.88 years	25.33 years
Age of English Acquisition	0.24 years	0.53 years	1.93 years
Age of English Fluency	3.72 years	3.63 years	4.80 years
English Understanding (Scale of 0 – 10)	9.4	10	9.38
Age of Spanish Acquisition	n/a	13.43 years	3.40 years
Age of Spanish Fluency	n/a	21.00 years	7.20 years
Spanish Understanding (Scale of 0 to 10)	n/a	6.47	7.13

3.0 Overview of Analyses

To maximize interpretability, we did not run the full 3 (language backgrounds) x 4 (noise conditions) x 2 (predictability conditions) ANOVA. Instead, we first compared accuracy across groups for each noise condition. We then analyzed accuracy within each group for differences across noise conditions, and for effects of context.

3.1 Testing for Interactions of Language Background and Noise Type

All speaker types' performance in each of the four noise conditions was analyzed using a 3 (monolingual vs L2 vs heritage) x 4 (clear vs speech-shaped noise vs English babble vs Spanish babble) mixed-design ANOVA with Bonferroni-corrected *t* tests for post hoc comparisons. A total of four post hoc comparisons were carried out in order to compare differences between speaker groups, resulting in a Bonferroni-corrected α level of .0125.

3.2 Within Group Analyses

Each speaker group's performance in each of the four noise conditions and in high versus low context was then analyzed using a repeated measure 4 (noise conditions) x 2 (predictability conditions) ANOVA with follow-up Bonferroni-corrected *t* tests for planned comparisons. A total of seven planned comparisons (all paired-samples *t* tests) were carried out: four to compare accuracy in the high and low predictability conditions within each noise condition, and three to compare overall accuracy (i.e. ignoring any effects of context) for clear versus speech-shaped noise, English versus Spanish two-talker babble, and speech-shaped noise versus Spanish two-talker babble. This resulted in a Bonferroni-corrected α level of .007. Due to the highly unlikely (and unobserved in our study) event where high predictability would correspond with a decrease in accuracy, *t* tests examining context effects were all one-tailed, corresponding to the prediction that the high predictability context would be associated with greater accuracy as compared to low predictability.

4.0 Results

4.1 Testing for Interactions of Language Background and Noise Type

The mixed-design ANOVA revealed a significant main effect of noise condition, [$F(3, 144)=168.445$, $p<.001$], a significant main effect of speaker type, [$F(2,48)=4.043$, $p=.024$], and a significant interaction between the two, [$F(6, 144)=3.492$, $p=.003$]. See Figure 1 below for a visual summary of this information.

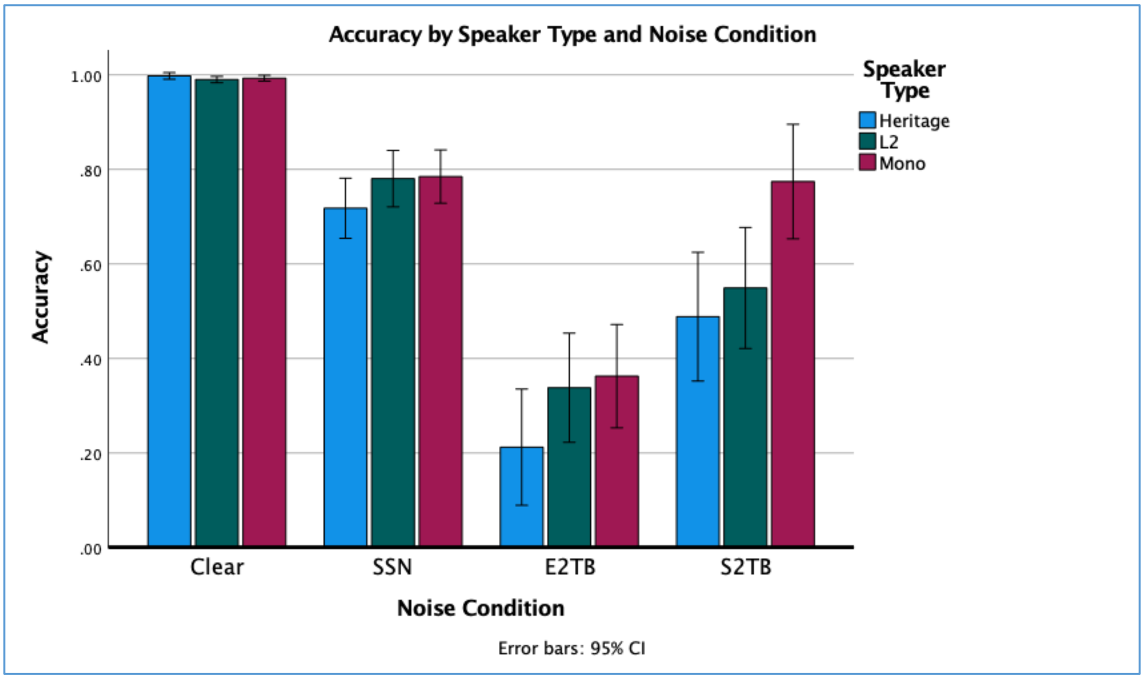

Figure 1: Participant accuracy by speaker type and noise conditions.

4.2 Effects of Speech-Shaped Noise Across Different Listener Groups

In comparing speaker accuracy in speech-shaped noise, there was no significant difference between L2 speakers' performance (M=.78, SD=.12) and heritage speakers' performance (M=.70, SD=.13); $t(31) = 1.913$, p =.0325.

4.3 Analysis of English Two-Talker Babble Across Different Listener Groups

In comparing L2 speakers' accuracy (M=.34, SD=.23) and heritage speakers' accuracy (M=.21, SD = .22) in the English two-talker babble condition, there was no significant difference between two groups; $t(31) = 1.655$, $p = .054$.

4.4 Analysis of Spanish Two-Talker Babble Across Different Listener Groups

In comparing monolinguals' accuracy (M=.77, SD = .08) to L2 speakers' accuracy (M=.55, SD=.35) in Spanish two-talker babble noise there was a significant difference between speaker groups; $t(34) = -2.759$, $p = .0045$. Additionally, in comparing heritage speakers' (M=.49, SD=.30) to L2 speakers' (M=.55, SD=.35) performance, there was no significant difference between the groups' accuracy; $t(30) = -.527$, $p = .301$. In sum, both bilingual groups performed with the same accuracy in Spanish two-talker babble, while the monolingual English group outperformed the bilinguals.

4.5 Monolinguals Within Group Results

Monolingual participants' performance in each of the four noise conditions and in high versus low predictability was analyzed using a repeated measures ANOVA. The ANOVA revealed a significant main effect of noise condition, [$F(3, 54)=85.570$, $p<.001$], a significant main effect of context, [$F(1,18)=22.469$, $p<.001$], and a significant interaction between the two, [$F(3, 54)=2.886$, $p=.033$]. See Figure 2 below for visual summary of the main effects and interaction.

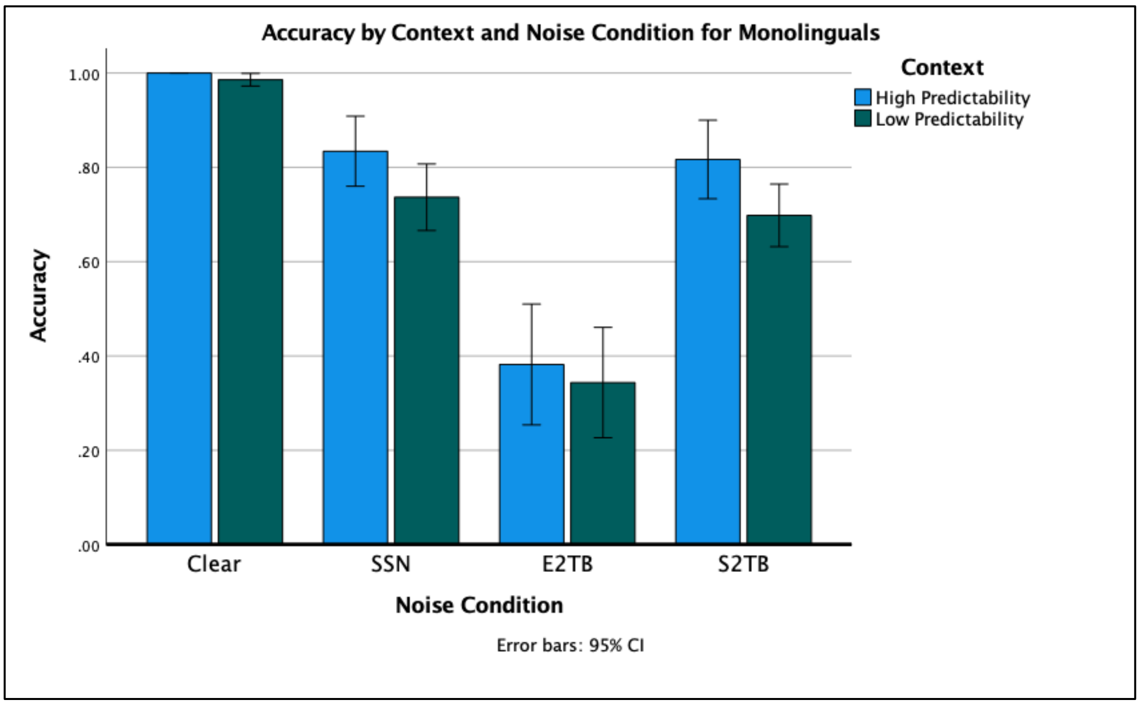

Figure 2. Breakdown of monolinguals' accuracy performance in both high and low context predictability and across all four noise conditions.

There was no significant difference in accuracy for high predictability (M = 1.00, SD = .00) and low predictability (M = .99, SD = .03) in the clear noise condition; $t(18) = -2.191$, $p = .021$. For speech-shaped noise, high predictability (M=.03, SD=.15) and low predictability (M=.74, SD=.15) displayed a significant difference; $t(18)= -3.552$, $p = .001$. In the English two-talker babble condition there was *not* a significant difference between high (M=.38, SD=.27) and low (M=.34, SD=.24) predictability; $t(18)= -1.529$, $p=.072$. High predictability (M=.81, SD=.14) and low predictability (M=.70, SD=.17) in Spanish two-talker babble displayed a significant difference in accuracy; $t(18)= -2.736$, $p = .007$. To summarize, the effect of context predictability for monolinguals was significant in speech-shaped noise, marginal in Spanish two-talker babble, and absent in the clear and English two-talker babble.

There was a significant difference in monolinguals' accuracy in the clear noise condition (M=.99, SD=.01) as compared to the speech-shaped noise condition (M = .78, SD = .14) such that accuracy was higher in the clear condition; $t(18) = 6.794$, $p < .001$. There was also a significant difference between their accuracy in the English two-talker babble condition (M = .36, SD = .25) and the Spanish two-talker babble condition (M = .77, SD = .08) such that accuracy was higher in Spanish two-talker babble noise; $t(18) = -8.830$, $p < .001$. However, analysis revealed there was no significant difference in monolinguals' accuracy in the speech-shaped noise condition (M = .78, SD = .14) and the Spanish two talker babble condition (M = .77, SD = .08); $t(18) = .392$, p = .35.

4.6 L2 Within Group Results

L2 participants' performance in each of the four noise conditions and in high versus low predictability was analyzed using a repeated measures ANOVA. Results revealed a significant main effect of noise condition, [$F(3, 48) = 40.770$, $p<.001$], a significant main effect of predictability, [$F(1,16)=9.909$, $p<.003$], and a significant interaction of predictability and noise condition, [$F(3, 48)=2.879$, $p<.034$]. See Figure 3 below for a visual summary of the main effects and interaction.

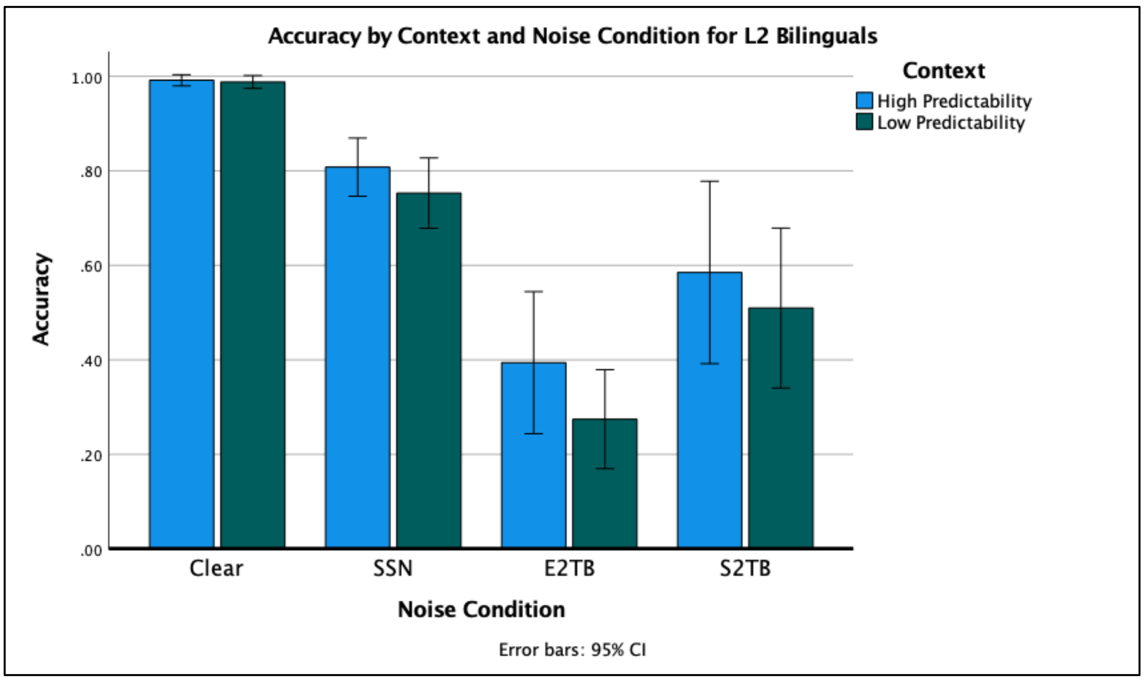

Figure 3. Breakdown of L2 bilinguals' accuracy performance in both high and low context predictability and across all 4 noise conditions.

In the clear noise condition there was *not* a significant difference between L2 speakers' high (M=.99, SD=.02) and low (M=.99, SD=.03) predictability accuracy; $t(16)=.339$, $p = .3475$. There was no significant difference in accuracy for high predictability (M = .81, SD = .12) and low predictability (M = .75, SD = .14) in the speech-shaped noise condition; $t(16)= 1.869$, $p =.04$. High predictability (M=.39, SD=.29) and low predictability (M=.27, SD=.20) in the English two-talker babble condition yielded marginally different accuracy; $t(16)=2.616$, $p = .0095$. High (M=.59, SD=.38) and low (M=.51, SD=.33) predictability in the Spanish two-talker babble condition did not result in a significant difference in accuracy; $t(16)= 2.391$, $p = .0145$. In summary, L2 bilinguals displayed a marginal effect of context in the English two-talker babble condition. There was no significant observed effect of context predictability in the clear, speech-shaped noise or Spanish two-talker babble conditions.

There was a significant difference in L2 bilinguals' accuracy in the clear noise condition (M=.99, SD=.02) and the speech-shaped noise condition (M = .78, SD = .12) indicating that accuracy was higher in the clear; $t(17) = 7.384$, $p < .001$. There was also a significant difference between their accuracy in the English two-talker babble condition (M = .34, SD = .23) and the Spanish two-talker babble condition (M = .55, SD = .35) such that higher accuracy was observed in the Spanish two-talker babble noise; $t(17) = -3.687$, $p < .001$. Finally, there was a significant difference between the speech shaped noise condition (M = .78, SD = .12) and Spanish two-talker babble condition (M = .55, SD = .35); $t(17) = 2.797$, $p = .0065$ such that the speech-shaped noise condition yielded higher accuracy.

4.7 Heritage Within Group Results

Heritage participants' performance in each of the four noise conditions and in high versus low predictability was analyzed using a repeated measures ANOVA. Results revealed a significant main effect of noise condition, [$F(3, 42) = 58.606$, $p<.001$], a significant main effect of predictability, [$F(1,14)=7.043$, $p<.01$], and a significant interaction of predictability and noise condition, [$F(3, 42)=2.608$, $p<.047$]. See Figure 4 below for visual summary of the main effects and interaction.

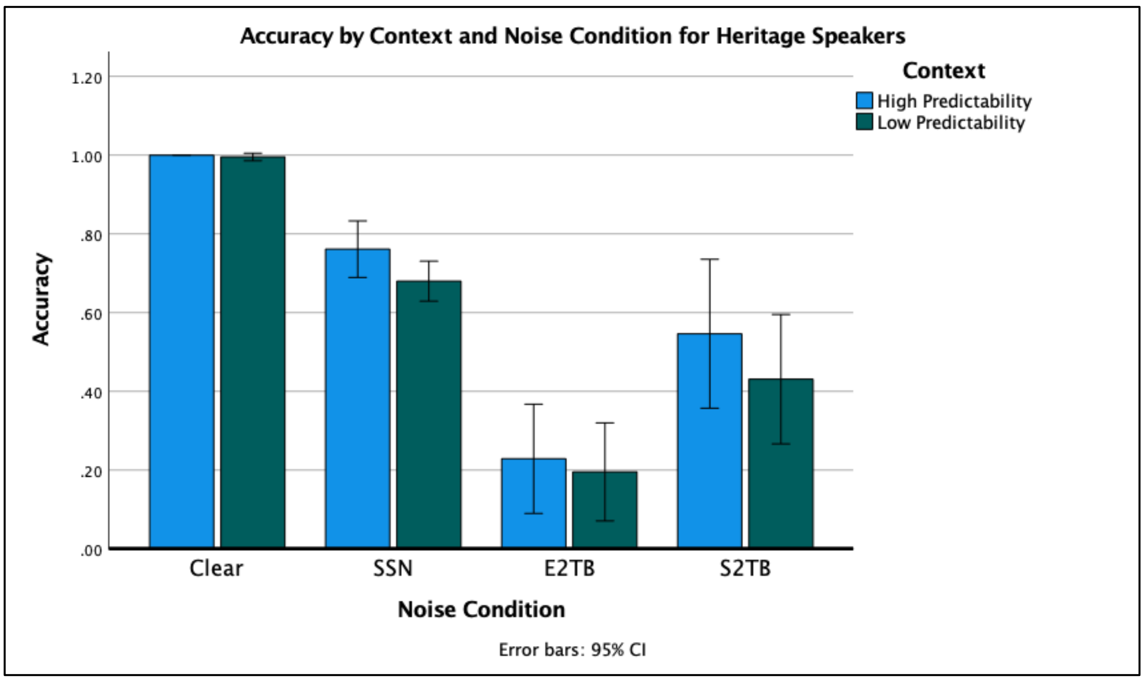

Figure 4. Breakdown of L2 bilinguals' accuracy performance in both high and low context predictability and across all 4 noise conditions.

In the clear noise condition there was not a significant difference between high (M=1.0, SD=.00) and low (M= 1.0 SD=.02) predictability accuracy; $t(15) = -1.00$, $p = .1665$. There was a significant difference in accuracy for high predictability (M = .74, SD = .15) and low predictability (M = .66, SD = .11) in the speech-shaped noise condition; $t(15) = -3.446$, $p = .002$. There was no observed significant difference between high (M=.22, SD=.24) and low (M=.19, SD = .22) predictability in the English two-talker babble noise condition; $t(15) = -.974$, $p = .1725$, nor was there a significant difference between high predictability (M=.55, SD=.34) and low predictability (M=.43, SD=.30) in the Spanish two-talker babble condition; $t(14) = -2.128$, $p = .026$. In summary, heritage participants only showed a significant effect of context predictability in the speech-shaped noise condition.

All comparisons between noise conditions revealed differences in accuracy in each condition. In the clear noise condition (M=1.0, SD=.01) heritage participants performed with higher accuracy than in the speech-shaped noise condition (M=.70, SD=.13); $t(15) = 9.25$, $p < .001$. In the English two-talker babble condition (M=.21, SD=.23), participants performed with lower accuracy than in the Spanish two-talker babble condition (M=.49, SD=30); $t(14) = -5.650$, $p < .001$. Finally, in the speech-shaped noise condition (M=.70, SD=.13) participants performed with marginally higher accuracy than in the Spanish two-talker babble condition (M=.49, SD=30); $t(14) = 2.759$, $p = .0075$.

5.0 Discussion

5.1 Summary of Findings

The results indicate that noise type and predictability interacted in complex ways in the study. Examining across speaker-type, the clear, speech-shaped noise, and English two-talker babble conditions all displayed no significant difference between overall performance among speaker groups. The Spanish two-talker babble condition displayed monolinguals outperforming L2 and heritage speakers who performed at the same accuracy in this condition.

For listeners of all language backgrounds, accuracy was higher in the clear than in speech-shaped noise. Monolingual results indicate that accuracy was higher in Spanish babble than English babble, and equal between speech shaped noise and Spanish babble. L2 results illustrate that accuracy was higher in Spanish babble than English babble and, distinct from monolinguals, accuracy was higher in speech-shaped noise than Spanish babble. Finally, heritage results show that accuracy was higher in the Spanish babble than English babble and accuracy was marginally higher in speech-shaped noise than in Spanish babble.

As for the effects of context, none of the groups displayed a context effect in the clear and only monolinguals displayed a context effect in speech-shaped noise. In the English two-talker babble condition, L2 participants showed a marginal context effect while heritage speakers showed a significant context effect. In the Spanish two-talker babble condition, monolinguals displayed a marginal difference between high and low predictability trials, while L2 and heritage speakers showed no context effect at all.

It is worth noting that the sample size for each of the speaker groups is smaller than we had hoped for, which may have caused the analyses to be underpowered. We plan to continue collecting data in order to increase our statistical power and clarify marginal significance values. In the meantime, given the significant main effects and interactions throughout the experiment, marginal differences will be discussed to a limited extent.

5.2 Informational Masking Hinders Monolingual Ability to Reap Context Benefit

Monolingual participants were significantly worse on average in English babble than in speech-shaped noise (which was equal in accuracy to Spanish two-talker babble performance). This reflects the effects of informational masking on monolingual SPIN in English babble. The English babble noise condition is the only non-clear condition where monolinguals were unable to reap contextual benefits such that they could make an accurate prediction about the target speech. This suggests that monolinguals know to use context to aid in SPIN difficulty since they used it in other conditions, but they are not cognitively capable of successfully utilizing context in the English babble condition where they have to deal with a large amount of competing informational masking. Coping with the informational masking comes from the fact that monolingual participants can take away information from the babble which distracts them from the target signal. This likely lowers their cognitive load such that they are no longer able to maintain all information at once in working memory, cope with the competing babble, *and* successfully make an accurate prediction about the critical target word.

5.3 Context Effect in the Presence of Energetic Masking

In the clear noise condition, all groups are performing the same and do not show a context benefit because accuracy is so high. With the presence of energetic masking in speech-shaped noise, all groups perform the same except L2 speakers who show no effect of context. Because the L2 group displays a drop in accuracy between clear and speech-shaped noise, it appears that the lack of a context effect for the L2 group implies that they are unable to reap contextual benefits. However, L2 speakers show that they are able to take advantage of context benefits in more challenging environments with informational masking (English two-talker babble). One interpretation of these findings is that it's not that L2 speakers *can't* reap context benefits in speech-shaped noise, but that they simply don't *need* to because they don't need the additional predictability benefit to cope with the demands of the task. Other groups do show a need to use context to cope with task demands as evidenced by the fact that monolinguals and heritage speakers see a deficit in the low context that is not present for L2 speakers (see Figures 2 and 4 for a visual display of the deficit in low context for monolingual and heritage participants). This is to say that all three groups are performing roughly the same in the high context condition, but only monolingual and heritage speakers perform with less accuracy in the low context condition (i.e. L2 are not negatively affected by the absence of high predictability context).

5.4 The Interaction of Noise Type and Linguistic Profile

Additionally, L2 speakers' accuracy in speech-shaped noise is not significantly different from heritage speakers' performance in this condition. This could be surprising if we expect that heritage speakers' longer history coping with cross-language activation throughout their life (as compared to later L2 bilinguals) could lead to more highly-developed cognitive resources. We expected this language profile difference to strengthen their cognitive control abilities such that they outperform the L2 group because they can allocate more cognitive resources to reaping contextual benefits. However, it is possible that the presence of speech-shaped noise is enough to trigger cross-language activation in heritage speakers but not L2 speakers. This would mean that heritage speakers suffer in noise in a way that L2 speakers do not, offsetting any potential cognitive advantage. Additionally, those who become fluent in a second language (the L2 group) may be those who are more cognitively capable to begin with (making it easier for them to reach L2 fluency). This could boost the accuracy of the L2 group relative to that of heritage speakers. It is important to note that heritage speakers did show a significant context benefit in speech-shaped noise, but L2 speakers did not. Considering that their overall accuracies are the same, this suggests that heritage speakers *must* rely on context to perform at the same accuracy as L2 speakers, who do not need to utilize that resource. Additionally, there's simply more variation in the L2 group's accuracy performance, so it's possible that this comparison doesn't rise to the level of significance due to low statistical power. This further supports the idea that the presence of speech-shaped noise is enough to trigger cross-language activation in heritage speakers but not L2 bilinguals, which overloads their cognitive abilities such that they *must* resort to using context to their advantage.

Because heritage speakers show a significant effect of context in speech-shaped noise, we know that they are capable of using context to their advantage. However, in English babble (the first noise condition with linguistically meaningful noise and informational masking), heritage speakers lose the significant effect of context they maintained in speech-shaped noise and L2 speakers gain the context effect that they did not maintain in speech-shaped noise. This suggests that the increase in informational masking that comes with English babble (and not speech-shaped noise) is something that an L2 speaker can cope with by using context to make predictions about the target, but not a heritage speaker.

Potential reasons for this suggestion are twofold. First, heritage speakers have had to code switch and cope with cross-language activation between English and Spanish for almost their entire lives, making their English and Spanish representations frequently interrelated. In contrast, L2 speakers' representation of English was acquired independent of Spanish and remained that way until much later in life. Once again, this supports the idea that heritage bilinguals are facing deficits in the English babble condition related to cross-language activation that L2 bilinguals are not. Taken together, this provides evidence for our hypothesis that SPIN does not affect all Spanish-English bilinguals who are native speakers of English equally, and disproportionately hinders heritage speakers' performance for multiple reasons related to the relative intertwinedness of their English and Spanish representations.

Additionally, the L2 group has the same overall accuracy as monolinguals in the English two-talker babble condition. However, we see that L2 speakers are suffering more when in low context than monolinguals who are not even able to reap contextual benefits in English babble. This implies that despite overall accuracy showing monolinguals and L2 bilinguals performing the same, L2 bilinguals are relying more on context and the two groups are using different approaches to the task.

In the Spanish two-talker babble condition, while both groups of bilinguals had the same accuracy rate, L2 speakers showed a stronger effect of context than Heritage speakers. This means that heritage speakers were unable to reap full contextual benefits such that they could successfully make an accurate prediction of the target word when the babble is in English or Spanish. In contrast, L2 speakers could better utilize context in both English and Spanish babble. We present two possible explanations for this finding. First, findings suggest that it is not simply knowing two languages that contributes to bilinguals' weaker ability to utilize context in SPIN, but something about the process of acquiring a language non-natively that contributes to L2 speakers' better performance (perhaps they acquire more developed and fine-tuned cognitive skills through the process of learning to control the more dominant L1). Findings also suggest that our L2 speaker group might have had higher cognitive control abilities to begin with such that they were able to cognitively cope better with all aspects of the task at hand. This would allow them to maintain enough resources to successfully take on the cognitive load of analyzing context and making predictions regarding the target speech, whereas the heritage group could not cope as well. This would once again make sense according to the idea that those who become fluent in a second language, the L2 group, may be those who are more cognitively capable.

Further, in the Spanish two-talker babble condition monolinguals performed with much higher accuracy than the two bilingual groups (who didn't differ from one another). This is as we had expected, since bilinguals experience significantly more informational masking in a Spanish babble environment than monolinguals who cannot derive any meaningful information from Spanish babble. In fact, the accuracy with which monolinguals performed in Spanish two-talker babble is the accuracy rate they performed at in the speech-shaped noise condition. In both noise conditions, accuracy is the same and there is a significant effect of context. Because monolinguals lack Spanish knowledge, the Spanish babble has the same observed SPIN effects as speech-shaped noise, both of which seem to function as energetic masking for monolinguals, since Spanish babble effectively has no informational content for an English monolingual.

5.5 Implications and Conclusion

Overall the present study sought to examine three overarching research questions. Question 1 examines the effect of language background on the ability to take advantage of contextual predictability during SPIN. Results from the current study show that context deficits in L2 SPIN can extend to the L1. However, results suggest more specifically that deficit extension to the L1 depends on factors such as cognitive control abilities, whether the L2 is strongly activated, and whether the L2 was learned natively such that the L1 and L2 mappings are not entirely independent (potentially as in the case of self-identified native English speaking heritage bilinguals).

Question 2 asks how different noise conditions affect monolinguals, heritage speakers and L2 speakers differently and whether heritage and L2 bilinguals are affected by noise differently. As we've seen, one key difference between speaker types is that Spanish babble and speech-shaped noise impact monolingual SPIN the same but bilingual SPIN differently due to monolinguals' not gaining any information out of Spanish babble and minimizing the informational masking taxation that bilinguals must deal with. Consequently, it appears that for monolinguals, speech-shaped noise and Spanish babble both consist of energetic masking noise. In speech-shaped noise, heritage and L2 bilinguals showed the same accuracy, but only heritage speakers needed to use context as a resource to cope with the SPIN deficit. In English babble, both bilingual groups struggled with informational masking and utilized context predictability to make accurate target predictions. However, in Spanish babble L2 bilinguals were better able to use sentence context to their advantage, importantly underscoring a difference in how Spanish noise affects different types of English-Spanish bilinguals that both identify as native English speakers. Additionally, Spanish babble affected monolinguals to the same extent that speech-shaped noise did, displaying that monolinguals are much less affected by Spanish babble than bilinguals are due to the lack of informational masking monolinguals experience from the babble.

Altogether, the insight into the first two research questions is highly related to the final research question that more broadly asks how L1 and L2 knowledge interact during SPIN. We can see that for monolinguals, a lack of L2 knowledge allows listeners to perform just as accurately identifying an L1 target during Spanish babble as in speech-shaped noise babble. In contrast, bilinguals in general who have knowledge of both the L1 and L2 face effects of informational masking in Spanish babble, which significantly hinder their ability to achieve the target as compared to monolinguals. More specifically, L2 bilinguals who developed an English representation completely independent of Spanish appear to be less adversely impacted by English babble than heritage bilinguals whose English representation has never existed entirely independent of their Spanish representation. L2 bilinguals and heritage bilinguals also show that differences in L2 acquisition (the difference between the two types of bilinguals) are related to different strategies in the speech-shaped noise condition, implying that speech-shaped noise results in the same global accuracy for both bilingual groups but affects them in different ways.

Taken together, findings related to the present research questions reveal that bilinguals' SPIN deficit is highly individualized based on the bilingual's specific language and cognitive profile. Results suggest that among bilinguals there is a large amount of variation in the extent of informational masking that listeners must cope with, and in their ability to cope with such masking. Results also suggest not only that L2 and heritage bilinguals do not perform identically in noise, but that they use different methods and resources to cope with the demands of SPIN. The origins of these differences appear to be based in differences in cognition between L2 and heritage speakers, suggesting that something about learning an L2 natively or as a second language is related to cognitive differences that impact SPIN processing.

Bibliography

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research, Methods, 52*(1), 388-407.

Bialystok, E., Craik, F. I., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological science in the public interest*, *10*(3), 89-129.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0. *URL http://www.praat.org/.*

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. The Journal of the Acoustical Society of America, 121(4), 2339-2349.

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. Acta Acustica United with Acustica, 86(1), 117-128.

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. The Journal of the Acoustical Society of America, 123(1), 414-427.

Fallon, M., Trehub, S. E., & Schneider, B. A. (2002). Children's use of semantic cues in degraded listening environments. *The Journal of the Acoustical Society of America*, *111*(5), 2242-2249.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515-530.

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, *35*(1), 49-66.

Krizman, J., Bradlow, A. R., Lam, S. S. Y., & Kraus, N. (2017). How bilinguals listen in noise: linguistic and non-linguistic factors. Bilingualism: Language and Cognition, 20(4), 834-843.

Lecumberri, M. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. The Journal of the Acoustical Society of America, 119(4), 2445-2454.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325-343.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940-967.

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition*, *20*(2), 139-154.

Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America*, *57*(S1), S5-S5.

Shi, L. F. (2010). Perception of acoustically degraded sentences in bilingual listeners who differ in age of English acquisition. Journal of Speech, Language, and Hearing Research, 53(4), 821-835.

Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *The Journal of the Acoustical Society of America*, *118*(5), 2775-2778.

Tabri, D., Chacra, K. M. S. A., & Pring, T. (2011). Speech perception in noise by monolingual, bilingual and trilingual listeners. *International Journal of Language & Communication Disorders*, *46*(4), 411-422.

Weiss, D., & Dempsey, J. J. (2008). Performance of bilingual speakers on the English and Spanish versions of the Hearing in Noise Test (HINT). *Journal of the American Academy of Audiology*, *19*(1), 5-17.

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native-and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, *121*(1), 519-526.

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication*, *52*(11-12), 943-953.