

## Article

# Predictive Processing and Inhibitory Control Drive Semantic Enhancements for Non-Dominant Language Word Recognition in Noise

Melinda Fricke <sup>1,\*</sup>  and Megan Zirnstein <sup>2</sup><sup>1</sup> Department of Linguistics, University of Pittsburgh, Pittsburgh, PA 15260, USA<sup>2</sup> Department of Linguistics and Cognitive Science, Pomona College, Claremont, CA 91711, USA

\* Correspondence: melinda.fricke@pitt.edu

**Abstract:** Auditory word recognition in the non-dominant language has been suggested to break down under noisy conditions due, in part, to the difficulty of deriving a benefit from contextually constraining information. However, previous studies examining the effects of sentence constraints on word recognition in noise have conflated multiple psycholinguistic processes under the umbrella term of “predictability”. The present study improves on these by narrowing its focus specifically on prediction processes, and on whether the possibility of using semantic constraint to predict an upcoming target word improves word recognition in noise for different listener populations and noise conditions. We find that heritage, but not second language, Spanish listeners derive a word recognition-in-noise benefit from predictive processing, and that non-dominant language word recognition benefits more from predictive processing under conditions of energetic, rather than informational, masking. The latter suggests that managing interference from competing speech and generating predictions about an upcoming target word draw on the same cognitive resources. An analysis of individual differences shows that better inhibitory control ability is associated with reduced disruption from competing speech in the more dominant language in particular, revealing a critical role for executive function in simultaneously managing interference and generating expectations for upcoming words.

**Keywords:** bilingualism; speech perception; speech-in-noise perception; semantic constraint; predictive processing; inhibitory control; individual differences



**Citation:** Fricke, Melinda, and Megan Zirnstein. 2022. Predictive Processing and Inhibitory Control Drive Semantic Enhancements for Non-Dominant Language Word Recognition in Noise. *Languages* 7: 239. <https://doi.org/10.3390/languages7030239>

Academic Editors: Georgia Zellou and Anne Pycha

Received: 21 April 2022

Accepted: 30 August 2022

Published: 15 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Bilingual Speech Perception in Noise

There is broad agreement in the literature that coping with the effects of background noise during speech perception is more difficult in the non-native than in the native language (see [Garcia Lecumberri et al. \(2010\)](#) and [Scharenborg and van Os \(2019\)](#) for reviews). There is also general agreement that, at least for sufficiently proficient listeners, recognition deficits may be less apparent at earlier stages of processing as compared with later stages. While recognition of individual speech sounds, such as consonants and vowels or CV sequences, shows similar deficits in noise for native versus non-native listeners ([Cutler et al. 2004](#); [Hazan and Simpson 2000](#)), studies examining later processing stages have been more likely to report disproportionate deficits in the non-native language. [Krizman et al. \(2017\)](#), for example, found that while early Spanish–English bilinguals performed better than English monolinguals in a non-linguistic tone perception task, they performed on par with monolinguals in a word perception task and worse than monolinguals in a sentence perception task. In the same vein, [Kousaie et al. \(2019\)](#) tested three groups of bilinguals and found that, contrary to simultaneous and early bilinguals, late bilinguals listening in the less dominant language were unable to take advantage of constraining sentence information to improve word recognition in noise.

These findings support the idea that speech-in-noise recognition deficits are more likely to emerge at later stages of processing. However, the picture remains complex given the variability across studies and the variety of factors that are likely to influence the ability to take advantage of constraining sentence contexts. Several studies have suggested age of acquisition (AoA) as an important determinant, with an earlier AoA in some studies associated with a greater accuracy in high constraint contexts (Coulter et al. 2021; Kousaie et al. 2019; Mayo et al. 1997; see also MacKay et al. 2001; Meador et al. 2000). Other studies have indicated that proficiency is likely to play a role. Kilman et al. (2014) showed that a higher proficiency in the non-native language (as measured by a written test of grammatical and lexical proficiency) was associated with a higher noise tolerance, for multiple noise types, in an oral sentence repetition task. Work by Bradlow and Alexander (2007) comparing L2 to native listeners has argued that sufficient ability to decode the phonetic signal in particular is a necessary prerequisite for semantic constraint effects. However, given the variety of proficiency measures used in the literature, and given that L2 proficiency and AoA are often highly correlated (though see Birdsong 2018; Fricke et al. 2019 for more nuanced discussion), it remains unclear to what extent AoA versus various aspects of proficiency may have distinct effects. Indeed, Gor (2014) found that high-proficiency heritage listeners (as measured by a standardized oral proficiency interview) outperformed both lower-proficiency heritage listeners and L2 listeners of high and low proficiency in terms of their ability to benefit from constraining information. This is consistent with the view that proficiency and AoA may have *additive* effects on L2 processing in general, and L2 speech perception in noise in particular.

An additional consideration that remains poorly understood in this context concerns the informational content of the masking noise itself. While some studies have examined purely *energetic masking* (i.e., the ability of co-occurring noise to render portions of the target speech signal inaudible due to overlapping time-frequency characteristics), others have investigated the role of *informational masking* (i.e., any additional decrement in intelligibility that remains once energetic masking has been accounted for). While it is known that a competing talker incurs less masking than stationary noise presented at the same signal-to-noise ratio (Festen and Plomp 1990), the masking effect of competing speech is also quite susceptible to variation. Previous work has suggested both that native listeners find competing speech to be more disruptive when it is in a language that they understand (Van Engen and Bradlow 2007) and that proficiency in the language of both the target (Kilman et al. 2014) and masker (Fricke 2022) influences the degree of perceptual disruption.

Moreover, Van Engen (2010) argued that the effects of linguistic similarity versus familiarity can help explain the extent of informational masking. In an experiment examining English keyword recognition in the context of English versus Mandarin two-talker babble, English two-talker babble was more disruptive than Mandarin two-talker babble for both native English and L1 Mandarin listeners, suggesting an effect of linguistic similarity for both listener groups (i.e., English competing speech may be especially disruptive when the target speech is also English). However, the difference in recognition performance for the two masker types was reduced for the L1 Mandarin group, suggesting a separate effect of linguistic familiarity (i.e., English competing speech was less disruptive for L1 Mandarin listeners as compared with English natives).

More broadly speaking, Cooke et al. (2008) identified at least four distinct sources of potential informational masking, including errors introduced by mistaking the masker for the target; the ability of masking noise to capture the listener's attention; linguistic interference caused by the simultaneous processing of the target and masking noise; and the increased cognitive load associated with all of these effects, either in isolation or when combined. Together, these factors are all likely to impact speech-in-noise recognition and, importantly, the degree of disruption incurred by each can reasonably be expected to interact with language proficiency (for either the target or masker languages, in the case of cross-linguistic masking), and with the ability to efficiently engage cognitive resources to mitigate the effects of disrupted linguistic processing.

### 1.2. Predictability versus Prediction

The picture regarding sentence context effects in speech-in-noise recognition can also be clarified in an additional respect; namely, the context manipulation itself. While a handful of studies have specifically examined semantic priming effects by comparing the processing of semantically related versus unrelated pairs of words (Golestani et al. 2009; Hervais-Adelman et al. 2014), many others have relied on sentence stimuli designed to maximize the difference in cloze probability between so-called high-predictability and low-predictability sentence contexts (Bilger et al. 1984; Bradlow and Alexander 2007; Coulter et al. 2021; Kalikow et al. 1977; Kousaie et al. 2019). Importantly, however, the term “predictability” here should be understood as referring to the end result of a number of distinct psycholinguistic processes. A sentence context can be “high predictability” because certain words within it elicit semantic or associative priming, and/or because the sentence-final word constitutes part of a phrase with high collocational strength. Compare, for example, sentences (1–3) below, which have been used in previous work to examine predictability effects in speech-in-noise perception:

1. The team was trained by their coach.
2. Eve was made from Adam’s rib.
3. The good boy is helping his mother and father.

While (1–3) may all elicit high cloze probability, they do so for different reasons. In (1), the target word *coach* is semantically related to both *team* and *trained* in the preceding context. In (2), the target *rib* is associated with both *Eve* and *Adam* by way of the biblical story, though the semantic relationship between *Eve/Adam* and *rib* is not particularly strong. Furthermore, in (3), the context is constraining in part because *mother and father* is a multiword unit with high collocational strength.

On the other hand, sentences eliciting low cloze probability can also have qualitatively distinct properties. Sentences (4–5) below would both result in low cloze probability because the sentence context is almost completely uninformative, but notably, the context of (5) actually results in the target word *throat* violating whatever weak predictions may have been made, since throats are not typically something one considers (à propos of nothing).

4. Mom talked about the pie.
5. He is considering the throat.

These examples therefore demonstrate the potential conflation of effects of *predictability* (i.e., whether a specific target word fits reasonably well within a given sentence’s context) and *prediction* (i.e., whether the semantic constraint of the sentence’s context leads one to actively predict the meaning and form of an upcoming target word). Prediction relies specifically on the process of incrementally constructing a sentence or discourse representation during comprehension. To examine how such processes influence word recognition in noise, one would need sentence stimuli that are moderate to high in semantic constraint, but also avoid associative primes and multiunit word collocations, as in (6) below (the Spanish translation of which was used in the present study).

6. The party next door lasted into the night and I couldn’t sleep because of all the noise.

Importantly, the target word *noise* in (6) is only predictable to the extent that the listener is able to build up a meaningful interpretation of the preceding context. Using sentences such as (6) therefore helps to constrain the set of processes that may or may not be helpful during speech perception in noise. Indeed, such sentence stimuli are frequently used in the reading comprehension literature (paired with matched, low-constraint sentence contexts) to understand the role of prediction processes during online processing (for a review, see Kutas et al. 2014). A large body of work in the psycholinguistics literature has now suggested that prediction is a normal and active part of sentence comprehension, although it may not occur in all contexts; for example, when stimuli are presented too quickly for prediction processes to unfold (Wlotko and Federmeier 2015), when the cognitive load is too high (Ito et al. 2018), when L2 proficiency is too low (Martin et al. 2013), when the ability to regulate interference from the dominant language is too low (Zirnstein et al.

2018), and during the course of healthy aging (Federmeier et al. 2002, 2010). Critically, when non-dominant language listeners are presented with the challenge of perceiving speech-in-noise, they must also overcome similar obstacles. Listeners must simultaneously deal with potential interference (whether via energetic or informational masking); they cannot control the presentation speed of continuous, naturalistic speech; and even if they are able to both manage interference and rapidly attend to a specific talker, they may not have sufficient proficiency in the L2 to benefit from the sentence context that they are perceiving.

This suggests that, to the extent that lexical and/or semantic prediction is likely to occur for bilingual readers of an L2, it may similarly be present when perceiving speech-in-noise. While it is not possible to say for certain whether previously observed predictability effects in speech-in-noise perception involved prediction processes per se, the relative ubiquity of prediction in multiple comprehension contexts suggests that prediction effects may also be occurring under these circumstances. More specifically, for listeners who are able to overcome the interference imposed by energetic and/or informational masking when perceiving speech-in-noise, actively generating semantic predictions for the meanings of upcoming words may be a viable strategy that supports effective perception and comprehension. The question that remains is which listener and noise characteristics will enable such a strategy.

### 1.3. The Current Study

Given the literature reviewed above, a primary motivation of the current study is to specifically isolate the role of prediction processes in improving speech-in-noise recognition in the non-dominant language. We also seek to better understand how individual differences in prior linguistic experience and cognitive control ability impact the ability to take advantage of sentence constraint when listening in noise and, moreover, to better understand the extent to which these factors interact with the informational content of the masker. With regard to the latter point, we are specifically interested in the question of how competing speech in the more- versus less-dominant language might affect listeners with different cognitive-linguistic profiles.

#### Exploring the Role of Inhibitory Processes in Managing Acoustic and Linguistic Interference

As one of the primary goals of the current study is to investigate prediction processes in the context of speech-in-noise perception, it is important to outline how inhibitory control may assist non-dominant language listeners in a speech-in-noise paradigm. Speech-in-noise paradigms necessarily involve the use of masking, whether it be energetic or informational, to create interference with the target signal. This interference necessarily makes the target speech stream more difficult to perceive, owing to the activation of non-target representations (see Scharenborg et al. (2018) for recent evidence that even purely energetic masking increases the number of non-target representations that are activated). The resulting competition between target and non-target representations may be mediated by the deployment of inhibitory control skills. That is, the domain general skill of suppressing or inhibiting irrelevant information should apply to any speech-in-noise paradigm, but is especially relevant for bilinguals or L2 learners perceiving speech in the non-dominant language when the masker consists of competing speech in the dominant language.

It has long been observed in the bilingual cognition literature that L2, or non-dominant, language processes rely heavily on the ability to inhibit or regulate the L1, or dominant, language, which is understood to always be active in parallel (for further discussion, see Kroll and Dussias 2013 and Kroll et al. 2022). The ability to regulate the dominant L1 has also been demonstrated to support the generation of predictions in the L2 during reading (Zirnstein et al. 2018). Together, this suggests that inhibitory control ability may play a special role in assisting bilingual listeners with the dual-task of simultaneously managing

interference and engaging in resource-demanding comprehension strategies, such as predicting the meanings of upcoming words. In the current study, we therefore explore the hypothesis that bilingual and L2 listeners perceiving non-dominant language speech in noise may employ inhibitory skills to manage interference from competing representations, and that the ability to do this efficiently may be associated with greater so-called semantic enhancements from constraining sentence contexts. To index inhibitory skills, we used the inhibition-demanding portion of the AX continuous performance task (i.e., the interference version described in [Braver et al. 2001](#)), which we describe in Section 2.4.

## 2. Materials and Methods

This study received ethical approval from the Institutional Review Board of the University of Pittsburgh (protocol code STUDY19010203, approval date 17 February 2020).

### 2.1. Overview of Experimental Session

All experimental procedures were administered using Gorilla Experiment Builder ([www.gorilla.sc](http://www.gorilla.sc); [Anwyl-Irvine et al. 2020](#)), and participants completed all tasks via internet in a location of their choosing. Participation was restricted to users of desktop computers to maximize the probability that listeners would be seated in a static location conducive to concentration.

The experiment reported here formed part of a larger study, and constituted the first task within the third and final testing session. All participants first completed a short screening session comprised of a language history questionnaire (a modified version of the LEAP-Q; [Marian et al. 2007](#)) and the English ([Lemhöfer and Broersma 2012](#)) and Spanish ([Izura et al. 2014](#)) versions of the LexTALE vocabulary test. Each version of the LexTALE takes approximately three to five minutes to administer, and consists of an untimed visual lexical-decision task which was independently developed for each language. The English version of the task has been found to correlate well with other measures of proficiency ([Lemhöfer and Broersma 2012](#)).

The second testing session generally took place within a few days of screening, and consisted of an English version of the present task (English keyword transcription in noise), followed by an English lexical-decision task (reported in [Fricke 2022](#)) and then a consonant-identification task in noise. The third, Spanish testing session took place within a few days of the English session, and included the same three experimental tasks (keyword transcription, lexical decision, and consonant identification), followed by the AX variant of the continuous performance task ([Braver et al. 2001](#)) to assess inhibitory control. Both the English and Spanish sessions were preceded by a six-trial headphone check ([Woods et al. 2017](#)) to verify that participants were wearing headphones as instructed. The Spanish session took approximately 90 min to complete, and participants were compensated \$15 for their time, with a \$1 bonus for successfully completing all tasks in the session.

### 2.2. Participants

Participants were recruited using Prolific ([www.prolific.co](http://www.prolific.co)). A total of 33 bilinguals took part in the experiment (16 heritage and 17 L2 Spanish listeners). One heritage and one L2 participant's data were excluded due to low accuracy in the clear condition of the main experimental task (63% and 60%, respectively). All remaining participants accurately transcribed a minimum of 25 out of 30 trials (~83% accuracy) in the clear condition.

A summary of participant characteristics is provided in Table 1. Participants were categorized as either heritage Spanish or L2 Spanish bilinguals according to their questionnaire responses. All participants reported growing up in the U.S. and listed English as their more dominant language. All reported an English AoA of 5 or less; 14 out of 16 L2 Spanish bilinguals and 7 out of 15 heritage bilinguals gave an English AoA of 0. All heritage bilinguals reported that Spanish was spoken in their home while growing up, and gave a Spanish AoA of 3 or less; 12 out of 15 heritage participants reported a Spanish

AoA of 0. L2 Spanish bilinguals gave Spanish AoAs ranging from 6 to 28, and did not list Spanish as a language spoken in their home while growing up.

**Table 1.** Summary of participant characteristics.

	Heritage Spa		L2 Spa	
N (n female)	15 (6 F)		16 (13 F)	
Age (SD)	<b>25.2 (4.3)</b>		<b>28.6 (4.7)</b>	
	Eng	Spa	Eng	Spa
Oral comprehension, self-rating (SD)	10.0 (0.0)	<b>9.1 (1.0)</b>	10.0 (0.0)	<b>7.6 (2.1)</b>
Speaking ability, self-rating (SD)	9.9 (0.5)	7.4 (1.5)	10.0 (0.0)	6.6 (2.0)
Reading ability, self-rating (SD)	9.9 (0.3)	6.7 (2.7)	10.0 (0.0)	7.8 (2.1)
% daily exposure (SD)	<b>71% (15%)</b>	23% (13%)	<b>84% (15%)</b>	15% (15%)
Age began acquiring (SD)	<b>2.2 (2.2)</b>	<b>0.5 (1.0)</b>	<b>0.2 (0.5)</b>	<b>14.2 (4.9)</b>
Age became fluent (SD)	5.5 (3.3)	<b>10.8 (7.1)</b>	4.8 (3.8)	<b>22.2 (4.3)</b>
LexTALE (SD)	0.90 (7%)	<b>0.71 (15%)</b>	0.95 (7%)	<b>0.61 (7%)</b>
AY Trials Accuracy, % (SD)	86% (15%)		86% (15%)	
AY Trials RT, ms (SD)	574 (178)		571 (173)	

Standard deviations are reported in parentheses. Participant characteristics that were significantly different between groups are highlighted in gray.

Two-tailed Welch-corrected *t* tests were used to compare participant characteristics for the heritage versus L2 Spanish groups. The groups differed in terms of age ( $t(29.0) = -2.1, p = 0.047$ ), self-reported daily exposure to English ( $t(28.9) = -2.3, p = 0.027$ ), English AoA ( $t(15.5) = 3.3, p = 0.004$ ), self-rated oral comprehension ability in Spanish ( $t(22.2) = 2.5, p = 0.02$ ), Spanish AoA ( $t(16.3) = -11.0, p < 0.001$ ), Spanish age of fluency ( $t(20.1) = -4.8, p < 0.001$ ; 2 heritage and 5 L2 participants reported they were not fluent in Spanish), and performance on the Spanish LexTALE test ( $t(20.0) = 2.1, p = 0.047$ ). No other differences between groups were significant.

### 2.3. Main Experimental Task

#### 2.3.1. Trial Procedure

The sentence perception-in-noise task consisted of a keyword transcription task loosely modeled on that of Bradlow and Alexander (2007). Participants heard either a high- or low-constraint Spanish sentence and were asked to transcribe the last word of the sentence using their computer keyboard. We recorded both the transcribed responses and reaction times.

#### 2.3.2. Design and Materials

There were a total of 240 trials, evenly split across four noise conditions, which were blocked and presented in a fixed order: all participants transcribed keywords first in the clear, followed by speech-shaped noise (SSN), then Spanish two-talker babble (S2TB), and finally in English two-talker babble (E2TB). We presented the E2TB block after the S2TB block in order to minimize any carryover effects of non-target language activation (e.g., Misra et al. 2012). Each block was preceded by the same three practice sentences, which were unrelated to the critical sentences.

Each noise block therefore comprised 60 critical trials, 30 in each of the high- versus low-constraint conditions. Presentation order within each block was fully randomized by the experimental program, and no participant ever responded to the same target twice throughout the experiment. In total there were eight sets of target words, which were rotated across conditions to create four versions of the experiment; word sets were matched according to the log-transformed word frequency (mean = 3.3, SD = 0.6), length in phonemes (mean = 5.7, SD = 1.4), total number of phonological neighbors (mean = 5.6, SD = 6.7), number of phonological neighbors with a higher frequency than the target word (mean = 0.9, SD = 2.2), and word duration (mean = 500 ms, SD = 88 ms). These lexical statistics were obtained from the CLEARPOND database (Marian et al. 2012). We also avoided using cognate words as targets; half of the eight sets had no cognate targets, while half had just one or two cognate targets. Target word durations in the high- versus low constraint-conditions did not differ significantly (498 vs. 502 ms, respectively; paired one-sided  $t(119) = 1.5, p = 0.07$ ).

As mentioned previously, an important difference in this study as compared with prior studies is the way in which sentence constraint was manipulated. High-constraint sentence stimuli were created and edited by native, heritage speakers of Spanish who were born and lived in the United States. Sentences were created to support the expectation and prediction of an upcoming sentence's final word, while avoiding high semantic associations, multi-word units, and high collocational strength between content words in the sentence context and the target word itself. Sentence construction also took into account dialect-specific word usage that would be more common for Spanish speakers learning and using Spanish in the United States. To verify the effectiveness of the constraint manipulation, a separate norming study was conducted using Prolific. Fifteen heritage bilingual participants were recruited using the same recruitment procedures as the main study. Participants were presented with written versions of the high-constraint sentences and asked to supply the most likely final word. On average, 11 out of 15 participants supplied the intended target word, and this did not vary across the eight subsets of stimuli (mean cloze probability = 0.76, SD = 0.24;  $F(7,112) = 1.1, p = 0.35$ ).

We also assessed the semantic association strength between target words and all content words in the high-constraint sentence stems using the University of Salamanca Spanish Free Association Norms database (Normas de Asociación libre en castellano de la Universidad de Salamanca (NALC); Diez et al. 2018; an updated version of the NIPE: Normas e Índices para Psicología Experimental database; Fernandez et al. 2004). Specifically, we calculated forward association strengths between prior content words in the high-constraint context and the later, expected target word in each sentence. Eighty-five percent of the context words and 100% of the target words were found in the database (or a close associate, such as the same verb in a different tense, or the singular version of a plural word). Associations between content words in the context and target words were infrequent, with only 10.8% of the expected targets having an associate of 0.3 or above. The mean association strength between all content words and the target word was calculated for each high-constraint sentence item. On average, the mean association strength was 2.5% (SD = 5.7%).

Low-constraint sentences were modeled on those in previous studies, and consisted of brief, uninformative frames such as *Ella leyó sobre el/la . . .* ("She read about the . . ."; Bradlow and Alexander 2007; Kalikow et al. 1977). This constraint-manipulation method created a difference in sentence length across conditions; high-constraint sentence frames were 13 words long on average, versus just 4 words for low-constraint frames. It is possible that this difference in frame length across conditions could have affected the estimate of the sentence-constraint effect obtained. For example, to the extent that longer, more complex sentences require more processing resources in order to derive a context benefit, the stimuli could underestimate the typical effects of the sentence constraint. On the other hand, given that the stimuli provide numerous, reinforcing cues to the upcoming target, they could overestimate typical sentence-constraint effects (although the analysis of the

association strength above suggests this may not be the case). For our purposes, we simply underscore that our constraint-manipulation provided a clear contrast between the low- and high-constraint conditions, while reducing the likelihood of any unintended effects of semantic association or collocational strength.

### 2.3.3. Auditory Stimuli

More details concerning preparation of the auditory stimuli are available in [Fricke \(2022\)](#), but we also summarize the stimulus creation procedure here. The sentences were recorded by a female native speaker of Central American Spanish who was chosen for her ability to produce native-sounding stimuli for both the Spanish and English versions of the experiments. Recordings were made in a sound-attenuated booth at a sampling rate of 44.1 kHz and 16-bit depth, and were later downsampled to 22,050 Hz and converted to .mp3 format to minimize loading delays over the internet.

Speech-shaped noise was created by filtering white noise through the long-term average spectrum of the speech samples that were used to create the two-talker babble conditions. The two-talker babble conditions were created using freely available news podcasts, all with female presenters (two English-speaking and two Spanish-speaking). First, all non-speech noise and pauses longer than 500 ms were manually edited out of the news podcasts. Next, for each babble condition, one random clip from each of the two podcasts was combined with each target sentence. For all three noise conditions, the masking noise extended 500 ms beyond the onset and offset of the target speech. Based on previous literature (e.g., [Garcia Lecumberri et al. 2010](#)) and informal pilot testing, a signal-to-noise ratio of  $-2$  dB was chosen, with the root mean square (RMS) amplitude for the target speech scaled to 70 dB and the amplitude for the noise scaled to 72 dB.

### 2.4. AX-Continuous Performance Task

The AX-continuous performance task (AX-CPT; [Nuechterlein 1991](#)) is a version of the continuous performance task (CPT; [Rosvold et al. 1956](#)) that combines continuous participant responses to visual events with stimulus manipulations that impact the engagement of cognitive control. In most versions of the task, participants are asked to attend to a particular task goal that may change the button used for their continuous response. The task is designed to be biased towards the maintenance of this goal, leading to specific opportunities for researchers to examine (a) when participants engage in proactive goal maintenance, and (b) when they must recruit reactive or inhibitory control. The latter skill is frequently difficult for participants to engage in during the task, due to the intentionally designed proactive bias inherent in the task itself. As such, performance on this task has been used as a metric for the successful, if effortful, engagement of proactive goal maintenance and reactive or inhibitory control across multiple populations, including schizophrenic patients ([Cohen et al. 1999](#); [Edwards et al. 2010](#)), the elderly ([Braver et al. 2001](#); [Paxton et al. 2006](#)), and, importantly for the current study, bilingual speakers ([Morales et al. 2013, 2015](#)). Performance specifically on the inhibition-demanding portion of the AX-CPT has been used to predict the engagement of cognitive control during visual sentence-processing in young and older adult monolinguals and bilinguals ([Zirnstein et al. 2018, 2019](#)), but has not yet been used to establish the extent to which cognitive control is engaged during speech perception in noise.

In the AX-CPT, participants are instructed to respond as quickly and accurately as possible to letters that are presented on the center of the screen, one at a time. In the interference version of this task (see [Braver et al. 2001](#)), which was used in the current study, letters appear in sequences of five, starting with a red letter cue, followed by three white letter distractors, and ending with a red letter probe. Each letter is presented for 300 ms, with a 1000 ms inter-stimulus interval (ISI), with no pause between letter sequences (with the exception of during accuracy feedback in the practice phase). Thus, participants are continuously responding to the presentation of letter stimuli using a button response. Participants are also given the goal of monitoring for a certain cue-probe letter combination

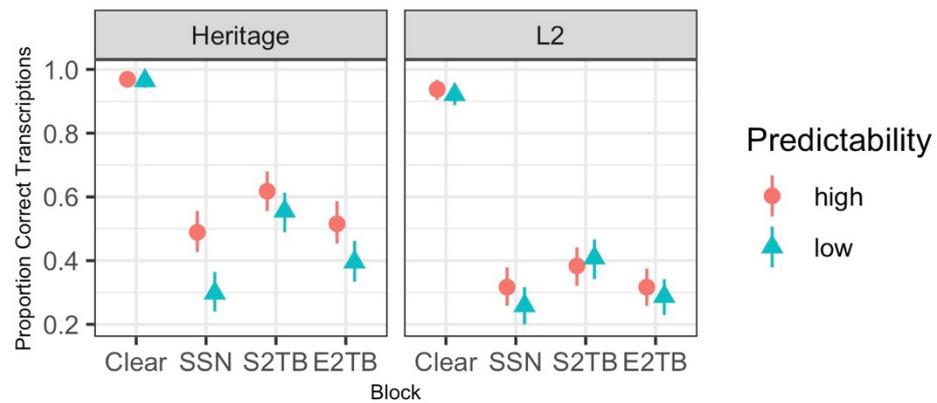
of “A” followed by “X”. When that pattern occurs, they are asked to change their response button to a “Yes” (the “0” key, in our version of the experiment) for the red letter “X” at the end of the sequence. For all other letters, including cues, distractors, and any other probe letters that do not meet this task goal requirement, participants are asked to continue to respond using a “No” button (the “1” key, in our case). Distractor letters in this version of the task are any letter other than “A” or “X”. Letters such as “K” and “Y” are also not presented, due to their high visual similarity to “X”. In addition, “X” is never presented as a cue, and “A” is never presented as a probe, in order to avoid confusion when monitoring for the “A followed by X” cue-probe combination. Response times and accuracy for each 5-letter trial sequence were recorded.

Trials on this task were in one of four conditions: AX (when “A” cues were followed by “X” probes, and when participants should change their response to “Yes” after seeing the probe letter), AY, BX, and BY. Prior to the start of the task, participants were given instructions and an opportunity to practice the task and receive feedback on their trial-level performance. Following the completion of eight practice trials, participants completed the main task. The main task included 100 cue-probe trial sequences in total. A total of seventy percent of those trials were in the AX condition, in order to bias participants towards maintaining the task goal over time. In contrast, the remaining 30% of the trials were divided equally across the three remaining conditions: AY, BX, and BY. AY and BX trials were designed to be more difficult for participants, due to their similarity to the AX task goal. For the AY trials, participants tended to respond more slowly and less accurately to the Y probe, as the previous A cue led to the expectation that X may be the upcoming probe letter. For the BX trials, participants tended to respond more slowly and less accurately to the X probe if they had failed to maintain the task goal (i.e., failing to correctly determine that the presence of the B cue indicated that any subsequent probe letter would not match the task goal).<sup>1</sup> The BY trials, in contrast, served as a contrast condition to the AY and BX trials, respectively, as neither the cue nor the probe were related to the AX task goal. Performance on the AY trials has previously been utilized as an index of the effortful engagement of reactive or inhibitory control, especially for young adult monolinguals and bilinguals (Zirnstein et al. 2018). Based on this prior research, we selected reaction times to correct trials in the AY condition as an index of inhibitory control for our individual difference analyses.

### 3. Results

#### 3.1. Accuracy Analysis

Prior to analysis, all transcriptions were checked for obvious typos and missing or extraneous accent marks; a total of 54 trials (1.9% of the data) were hand-corrected at this stage. Figure 1 plots the proportion of correct transcriptions for the four blocks of the experiment, as a function of both predictability (high vs. low) and participant group (heritage vs. L2 bilinguals). Because both participant groups were at or near ceiling in the clear (97% overall accuracy for heritage bilinguals and 93% for L2 bilinguals, with no discernible effect of predictability), the accuracy analysis focused on the three noise blocks of the experiment. Block was Helmert-coded such that the first contrast compared Spanish two-talker babble (“S2TB”) with English two-talker babble (“E2TB”), and the second contrast compared the speech-shaped noise condition (“SSN”) with the average of the two two-talker babble conditions (“2TB”). The participant group was contrast-coded, such that the model intercept corresponded roughly to the grand mean of the two bilingual groups. Predictability was treatment-coded, with low-predictability trials as the reference level. The intercept and each of the simple effects in the coefficient table therefore represent the low-predictability trials only, while interactions involving the “Predictability [high]” term tested whether any effects differed in the high-predictability condition, i.e., whether the presence of a constraining sentence context modulated the other effects.



**Figure 1.** Raw accuracy data from the speech-in-noise transcription task. (SSN = speech-shaped noise, S2TB = Spanish two-talker babble, E2TB = English two-talker babble).

We fit a mixed-effects logistic regression using the lme4 R package (Bates et al. 2015; version 1.1.26) that included the three experimentally manipulated variables (Block, Group, Predictability) and all of their two- and three-way interactions, plus the maximal random effects structure that would converge; this consisted of only by-participant random intercepts. The fitted model is given in Table 2.

**Table 2.** Fitted model examining keyword transcription accuracy for the speech-in-noise recognition task.

Coefficient	Fixed Effect Estimates			
	Odds Ratio	SE	CI (95%)	p
Intercept (=low Predictability)	1.85	0.24	1.44–2.39	<0.001
Group	0.64	0.16	0.38–1.05	0.079
<b>Predictability [high]</b>	<b>0.71</b>	<b>0.06</b>	<b>0.60–0.83</b>	<b>&lt;0.001</b>
<b>Block c1 (S2TB vs. E2TB)</b>	<b>0.72</b>	<b>0.05</b>	<b>0.63–0.83</b>	<b>&lt;0.001</b>
<b>Block c2 (SSN vs. 2TB avg)</b>	<b>1.24</b>	<b>0.05</b>	<b>1.14–1.35</b>	<b>&lt;0.001</b>
<b>Group * Predictability [high]</b>	<b>0.63</b>	<b>0.10</b>	<b>0.46–0.87</b>	<b>0.005</b>
Group * Block c1	0.93	0.13	0.70–1.23	0.601
Group * Block c2	1.14	0.10	0.96–1.35	0.123
Predictability [high] * Block c1 (S2TB vs. E2TB)	1.14	0.11	0.93–1.38	0.200
<b>Predictability [high] * Block c2 (SSN vs. 2TB avg)</b>	<b>0.88</b>	<b>0.05</b>	<b>0.78–0.99</b>	<b>0.030</b>
Group * Predictability [high] * Block c1	1.00	0.20	0.67–1.48	0.994
Group * Predictability [high] * Block c2	0.93	0.11	0.74–1.18	0.567
<b>Random Effects</b>				
$\sigma^2$			3.29	
$\tau_{00}$ Participant			0.41	
N Participant			31	
Observations			2790	
Marginal R <sup>2</sup> /Conditional R <sup>2</sup>			0.066/0.169	

The simple effect of Predictability indicates that, averaged across participant groups and noise conditions, high-predictability contexts yielded higher accuracy as compared with low-predictability contexts. The simple effect of Group was not significant, but there was an interaction between Group and Predictability, reflecting greater predictability effects for heritage bilinguals as compared with the L2 group. For Block, both contrasts were significant, reflecting the fact that in low-predictability contexts, accuracy was higher in 2TB as compared with SSN, and higher in S2TB as compared with E2TB. However, there was additionally an interaction between Predictability and Block indicating that the effect of Predictability differed in SSN as compared with the 2TB conditions.

Post-hoc comparisons of the estimated marginal means using the emmeans package (Lenth 2021; version 1.6.3) were used to explore the interactions among the variables. When averaging across all noise conditions, the effect of Predictability was significant for heritage (estimate = 0.6,  $z = 4.9$ ,  $p < 0.0001$ ) but not L2 bilinguals (estimate = 0.1,  $z = 1.0$ ,  $p = 0.34$ ). Pairwise comparisons were then conducted within each group, with a familywise adjustment of  $p$  values using a multivariate  $t$  distribution. For heritage bilinguals, the effect of Predictability was reliable in the SSN (estimate = 0.9,  $z = 4.3$ ,  $p < 0.001$ ) and E2TB (estimate = 0.5,  $z = 2.7$ ,  $p = 0.02$ ) conditions, but not in S2TB (estimate = 0.3,  $z = 1.4$ ,  $p = 0.40$ ). For L2 bilinguals, the Predictability effect was not reliable in any condition (all  $z$ s  $< 1.5$ , all  $p$ s  $> 0.37$ ).

Finally, given that accuracy for the low-predictability trials was higher in S2TB than E2TB, and higher in 2TB than in SSN, we also sought to determine whether accuracy in E2TB was higher than in SSN; this comparison was significant when averaging across both groups (estimate = 0.3,  $z = 2.1$ ,  $p = 0.03$ ), indicating a cline in accuracy for the low-predictability trials (SSN  $<$  E2TB  $<$  S2TB).

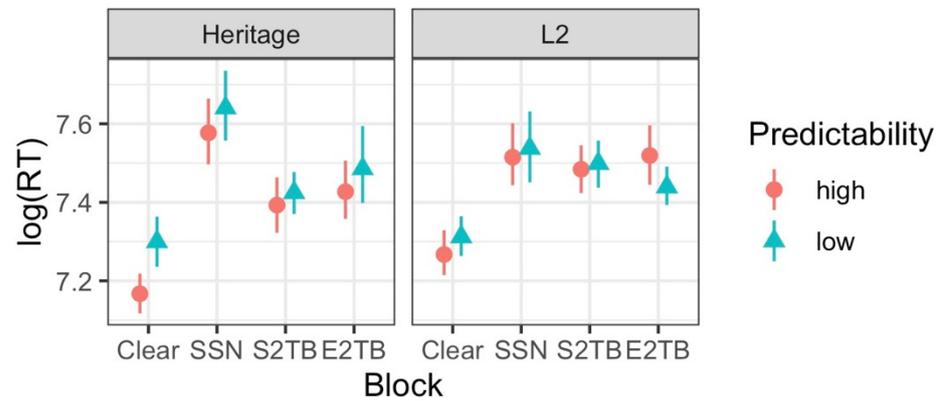
### 3.2. Reaction Time (RT) Analysis

Reaction times (RTs) were calculated from the onset of the target word to the first keystroke pressed during transcription. We calculated RTs from the target word onset (rather than the offset) in line with previous research suggesting that the engagement of cognitive control processes occurs quite early during bilingual auditory word recognition (Beatty-Martínez et al. 2021; Blumenfeld and Marian 2013). We considered only the timing of the first keystroke on analogy with the literature on (bilingual) picture and word naming, in which the onset of word articulation is taken as an index of the speed of lexical access (e.g., Meuter and Allport 1999). Our RT measure thus maximized comparability to previous studies of bilingual lexical access, and can be interpreted as a measure of how quickly the target lexical representation can be selected in response to incoming acoustic information.

The analysis was restricted to accurate trials with RTs slower than 300 ms, under the assumption that RTs faster than 300 ms were likely to reflect accidental key presses; 45 trials were excluded at this stage (~2% of the data). We also removed (log-transformed) RTs more than three SDs away from each participant's mean, comprising 28 trials (~1% of the data). The remaining 1927 RTs were log-transformed and scaled prior to analysis. Figure 2 plots the RT data as a function of the three experimental variables. Participant Group was again contrast-coded, and Predictability was again treatment-coded, with low-predictability trials as the reference level. In contrast to the accuracy analysis, the Clear condition was not excluded from the RT analysis, which necessitated a different variable coding scheme; Block was contrast-coded such that the first contrast compares the Clear condition with the average of the three noise conditions, indexing the overall noise deficit. The second contrast compares SSN with the average of the two two-talker babble conditions ("2TB"), indexing any overall differences in energetic vs. informational masking, while the third contrast compares S2TB with E2TB, assessing the impact of the babble language.

We fit a mixed-effects linear regression to the RT data that again included the three experimentally manipulated variables (Block, Group, Predictability) and all of their two- and three-way interactions, plus the best-fitting random effects structure that would converge. This included random intercepts for both Participant and Target Word, plus random by-participant slopes for the effect of Block; a model containing by-participant slopes for both Block and Predictability did not converge, while a model that included just by-participant slopes for Predictability had a significantly higher model AIC as compared with the model in Table 3 (4686 vs. 4665, respectively). To control for any effects of target word duration while maximizing statistical power (Matuschek et al. 2017), we started with a model that also included as covariates the duration of the target word (log-transformed and scaled) and its two-way interactions with Block, Group, and Predictability, then used leave-one-out comparisons of model log likelihood with  $\alpha_{LRT} = 0.10$  to determine which of

these improved model fit. The two-way interactions of target word duration with Block and Group were removed, while the interaction with Predictability was retained.



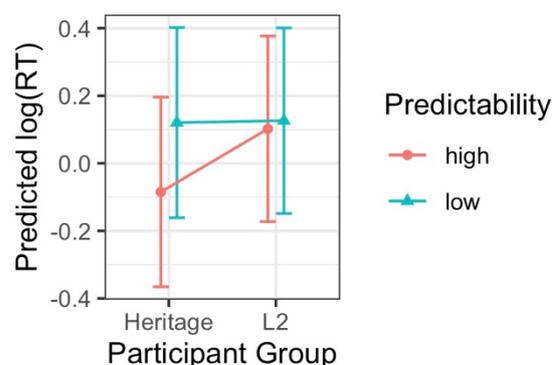
**Figure 2.** Raw response time data from the speech-in-noise transcription task. (SSN = speech-shaped noise, S2TB = Spanish two-talker babble, E2TB = English two-talker babble).

**Table 3.** Fitted model examining reaction times for accurate keyword responses in the speech-in-noise recognition task.

Coefficient	Fixed Effect Estimates			
	Estimate	SE	CI (95%)	p
Intercept (=low Predictability)	0.24	0.10	0.05–0.44	<b>0.015</b>
Target Duration	0.02	0.03	−0.04–0.08	0.555
Group	−0.03	0.19	−0.42–0.35	0.858
<b>Predictability [high]</b>	<b>−0.08</b>	<b>0.04</b>	<b>−0.16–−0.01</b>	<b>0.031</b>
<b>Target Duration * Predictability [high]</b>	<b>0.11</b>	<b>0.04</b>	<b>0.04–0.18</b>	<b>0.002</b>
<b>Block c1 (Clear vs. All Noise)</b>	<b>−0.57</b>	<b>0.07</b>	<b>−0.69–−0.44</b>	<b>&lt;0.001</b>
<b>Block c2 (SSN vs. 2TB avg)</b>	<b>0.39</b>	<b>0.11</b>	<b>0.17–0.61</b>	<b>0.001</b>
Block c3 (S2TB vs. E2TB)	0.01	0.08	−0.15–0.17	0.898
<b>Predictability [high] * Group</b>	<b>0.19</b>	<b>0.08</b>	<b>0.03–0.34</b>	<b>0.016</b>
Group * Block c1	0.12	0.12	−0.11–0.36	0.302
Group * Block c2	−0.34	0.21	−0.76–0.08	0.111
Group * Block c3	0.24	0.16	−0.08–0.56	0.134
<b>Predictability [high] * Block c1</b>	<b>−0.17</b>	<b>0.07</b>	<b>−0.31–−0.03</b>	<b>0.021</b>
Predictability [high] * Block c2	−0.08	0.11	−0.29–0.13	0.450
Predictability [high] * Block c3	−0.06	0.11	−0.29–0.16	0.581
Predictability [high] * Group * Block c1	0.03	0.14	−0.24–0.31	0.812
Predictability [high] * Group * Block c2	0.05	0.21	−0.37–0.47	0.826
Predictability [high] * Group * Block c3	−0.31	0.22	−0.75–0.13	0.167
<b>Random Effects</b>				
$\sigma^2$			0.57	
$\tau_{00}$ TargetWord			0.04	
$\tau_{00}$ Participant			0.27	
$\tau_{11}$ Participant.Blockc1			0.03	
$\tau_{11}$ Participant.Blockc2			0.14	
$\tau_{11}$ Participant.Blockc3			0.01	
$\rho_{01}$ Participant.Blockc1			0.59	
$\rho_{01}$ Participant.Blockc2			0.30	
$\rho_{01}$ Participant.Blockc3			0.11	
N Participant			31	
N TargetWord			120	
Observations			1927	
Marginal R <sup>2</sup> /Conditional R <sup>2</sup>			0.128/0.461	

The final fitted model is given in Table 3. Target word duration did not predict RTs on low-predictability trials, but shorter word durations were associated with faster RTs on high-predictability trials. Similar to the accuracy model, the RTs model included simple effects of Predictability and Block, with interactions of Group with Predictability and of Predictability with Block. However, the two models differed somewhat in terms of which Block comparisons were significant: the RT model returned a significant difference between the SSN and 2TB conditions (but no difference for S2TB vs. E2TB), while in the Predictability by Block interaction, only the coefficient for the Clear vs. All Noise comparison (and not SSN vs. 2TB) was different from zero.

Figure 3 depicts the two-way interaction of Predictability with Group using the estimated marginal means. Post-hoc comparisons indicated that when averaging across all noise conditions, the effect of Predictability was significant for heritage (estimate = 0.2,  $t = 3.4$ ,  $p = 0.001$ ) but not L2 bilinguals (estimate =  $-0.0$ ,  $t = -0.1$ ,  $p = 0.99$ ). Separate pairwise comparisons for the heritage bilinguals with a familywise-adjusted  $p$  value indicated that the Predictability effect for RTs was only reliable in the Clear condition (estimate = 0.3,  $t = 4.2$ ,  $p = 0.0001$ ; all other  $t$ s  $< 1.7$ ,  $p$ s  $> 0.34$ ).



**Figure 3.** Estimated marginal means from the fitted RT model depicting the interaction between Predictability and Group, with other effects held constant.

### 3.3. Individual Differences Analysis

The literature reviewed above indicates that proficiency in the non-dominant language is an important determinant of speech-in-noise recognition. It also indicates that inhibitory control ability is likely to explain some variance in individual bilinguals' ability to manage interference. To test these predictions in the context of the present experiment, we sought to augment the basic version of the RT model presented in Section 3.2. with both Spanish LexTALE and AX-CPT performance as individual-difference predictors. In Section 3.3.1. below, we first describe the AX-CPT results in order to establish the validity of our inhibitory control measure. We then describe the model fitting procedure and the results of the individual differences analysis in Section 3.3.2.

#### 3.3.1. AX-CPT Results

Before proceeding to individual difference analyses, we first wanted to establish whether the AX-CPT results from the current study replicated findings from the same task in prior research. Our dependent variables of interest were reaction times (RTs) to correct trials and proportions of errors to probe letters, both of which should increase (in some conditions) as difficulty with engaging cognitive control also increases. We focused our analyses on the AY, BX, and BY conditions, as each condition was equally proportioned in the task.

Prior to analysis, trials with probe RTs below 100 ms and over 2.5 standard deviations above the mean for each participant were removed, comprising 4% of the data. Participants' mean condition response times and error proportions were also removed as outliers if they were above or below 2.5 standard deviations of the group mean across conditions. This

resulted in the removal of BX and BY RTs for one participant, and the removal of some error proportions for five participants (AX = 1 removed; AY = 3; BX = 1; BY = 1). Errors were then log linear corrected using the following equation:  $(\text{error} + 0.5) / (\text{frequency of trials} + 1)$ , to account for trials where no errors were made (for reference, see Braver et al. 2009, Supplementary Materials).

Table 4 reports the means and standard deviations for all participants across each condition on the AX-CPT, as well as performance of each participant group separately. Heritage bilinguals and L2 listeners performed very similarly across conditions, and demonstrated the expected increase in RTs to correct AY trials in comparison with BY trials. We conducted repeated measures ANOVA to examine the interaction between group (heritage vs. L2) and condition (AY, BX, and BY) on RTs to correct trials and the proportion of errors produced by participants. For error proportions, there was a main effect of condition ( $F(2,44) = 11.096, p < 0.001$ ), but no significant group-by-condition interaction ( $F(2,44) = 1.550, p = 0.224$ ). Similarly, an analysis of the RTs revealed a main effect of condition ( $F(2,44) = 54.260, p < 0.001$ ) and no group-by-condition interaction ( $F(2,44) = 0.787, p = 0.461$ ). Following the significant main effects of condition for both RTs and errors, we conducted planned comparisons between each condition of interest (AY and BX, respectively) and the control condition (BY). AY trials were significantly more difficult than BY trials for both RTs ( $t(28) = 9.107, p < 0.001$ ) and errors ( $t(25) = 4.811, p < 0.001$ ). No significant effects were observed for the contrast between BX and BY trials, for either RTs ( $t(28) = 0.779, p = 0.443$ ) or errors ( $t(27) = 1.982, p = 0.058$ ). These results, in summary, indicate that the AY trials were consistently more effortful for participants to correctly respond to than the BY trials, and that this effect was similar across heritage bilinguals and L2 listeners. This replicates prior studies on the AX-CPT (Braver et al. 2001; Gonthier et al. 2016; Paxton et al. 2006), which have demonstrated similar effects for young adults across multiple versions of this task. For the following individual difference analyses, we utilized RT performance on correct AY trials (range: 331 to 1011 ms) as an index of inhibitory control, with faster RTs reflecting more successful and efficient engagement of control as compared with slower RTs.

**Table 4.** Mean proportions of errors and reaction time performances on the AX-CPT for heritage bilingual and L2 speakers of Spanish.

	Errors			Reaction Times (ms)		
	Total	Heritage	L2	Total	Heritage	L2
<i>AX-CPT Condition</i>						
AX	0.049 (0.004)	0.049 (0.003)	0.049 (0.005)	439 (155)	477 (197)	405 (101)
AY	0.054 (0.008)	0.053 (0.008)	0.055 (0.008)	572 (172)	573 (178)	571 (173)
BX	0.050 (0.007)	0.051 (0.008)	0.049 (0.002)	380 (170)	419 (207)	343 (121)
BY	0.047 (0.001)	0.046 (0.000)	0.048 (0.001)	396 (187)	430 (241)	364 (119)

Standard deviations are reported in parentheses.

### 3.3.2. Individual Differences Results

To optimize model fit and statistical power, we first fit a saturated model and used backwards selection with nested model log likelihood comparisons to remove predictors in a stepwise fashion, until all predictors were significant at a predetermined alpha-level of  $\alpha_{LRT} = 0.10$  (Matuschek et al. 2017). The initial, saturated model included all of the experimentally manipulated variables (Group, Block, and Predictability) and their two- and three-way interactions, plus LexTALE performance (centered and scaled) and its two- and three-way interactions with Block and Predictability, as well as the AYRTs predictor (log-transformed, centered, and scaled) and its two- and three-way interactions with Block and Predictability, and finally the four-way interaction of LexTALE, AYRTs, Block,

and Predictability. The maximal converging random effects structure included both by-participant and by-word intercepts.

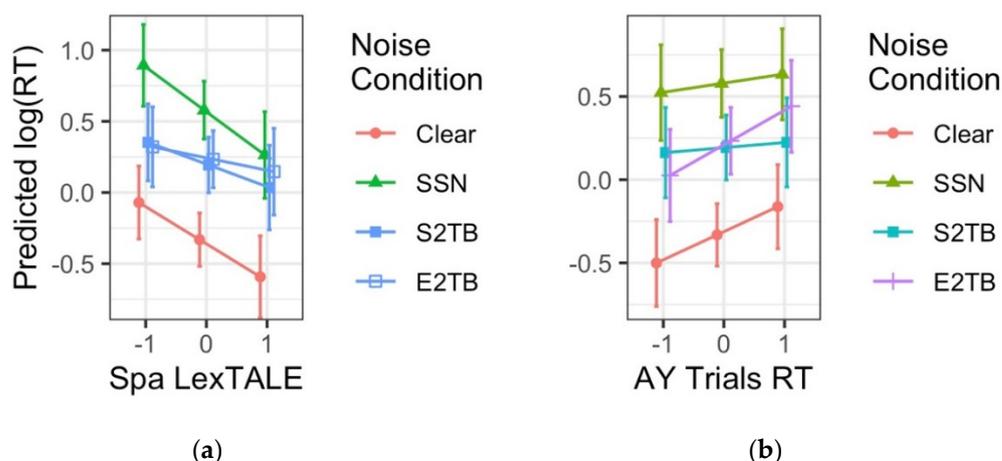
The final, fitted model examining individual differences (“IDs”) is given in Table 5. Similar to the basic RTs model, the IDs model included simple effects of Predictability and Block, as well as interactions of Predictability with Target Duration, Group, and Block. In general, the same Block comparisons were significant across the two models: the Clear vs. All Noise and SSN vs. 2TB comparisons were significant with Predictability at the reference level (i.e., low), and the coefficient for the interaction of Predictability with Clear vs. All Noise was also significant. The addition of the IDs predictors did result in one change in the simple effects when comparing across models: while the coefficient for the interaction of Group with SSN vs. 2TB was not significant in the basic RTs model, it was significant in the IDs model; for low-predictability trials, RTs were slower in SSN as compared with the 2TB babble conditions for the heritage listeners, but not for the L2 listeners. This effect can be seen in Figure 2.

**Table 5.** Fitted model examining individual differences in reaction times for accurate keyword responses in the speech-in-noise recognition task.

Coefficient	Fixed Effect Estimates			
	Estimate	SE	CI (95%)	p
Intercept (=low Predictability)	0.22	0.09	0.03–0.40	0.023
Target Duration	0.02	0.03	−0.04–0.08	0.522
Group	−0.18	0.20	−0.56–0.21	0.369
<b>Predictability [high]</b>	<b>−0.09</b>	<b>0.04</b>	<b>−0.17–−0.02</b>	<b>0.018</b>
AY Trials RT	0.12	0.09	−0.06–0.29	0.189
<b>Spa LexTALE</b>	<b>−0.21</b>	<b>0.10</b>	<b>−0.40–−0.01</b>	<b>0.038</b>
<b>Target Duration * Predictability [high]</b>	<b>0.11</b>	<b>0.04</b>	<b>0.04–0.19</b>	<b>0.002</b>
<b>Block c1 (Clear vs. All Noise)</b>	<b>−0.59</b>	<b>0.06</b>	<b>−0.70–−0.47</b>	<b>&lt;0.001</b>
<b>Block c2 (SSN vs. 2TB avg)</b>	<b>0.42</b>	<b>0.09</b>	<b>0.24–0.60</b>	<b>&lt;0.001</b>
Block c3 (S2TB vs. E2TB)	−0.03	0.08	−0.20–0.13	0.695
<b>Predictability [high] * Group</b>	<b>0.20</b>	<b>0.07</b>	<b>0.06–0.34</b>	<b>0.005</b>
Group * Block c1	0.08	0.08	−0.08–0.24	0.333
<b>Group * Block c2 (SSN vs. 2TB avg)</b>	<b>−0.50</b>	<b>0.12</b>	<b>−0.74–−0.25</b>	<b>&lt;0.001</b>
Group * Block c3	−0.01	0.13	−0.27–0.25	0.958
<b>Predictability [high] * Block c1 (Clear vs. All Noise)</b>	<b>−0.15</b>	<b>0.07</b>	<b>−0.30–−0.01</b>	<b>0.040</b>
Predictability [high] * Block c2	−0.13	0.11	−0.35–0.09	0.244
Predictability [high] * Block c3	−0.01	0.12	−0.24–0.21	0.910
Block c1 * AY Trials RT	0.07	0.04	−0.00–0.15	0.060
Block c2 * AY Trials RT	−0.06	0.06	−0.18–0.05	0.256
<b>Block c3 * AY Trials RT (S2TB vs. E2TB)</b>	<b>−0.18</b>	<b>0.06</b>	<b>−0.29–−0.06</b>	<b>0.003</b>
Block c1 * Spa LexTALE	−0.07	0.04	−0.15–0.01	0.071
<b>Block c2 * Spa LexTALE (SSN vs. 2TB avg)</b>	<b>−0.19</b>	<b>0.06</b>	<b>−0.31–−0.07</b>	<b>0.002</b>
Block c3 * Spa LexTALE	−0.07	0.06	−0.20–0.05	0.268
<b>Random Effects</b>				
$\sigma^2$			0.58	
$\tau_{00}$ TargetWord			0.04	
$\tau_{00}$ Participant			0.23	
N Participant			30	
N TargetWord			120	
Observations			1841	
Marginal R <sup>2</sup> /Conditional R <sup>2</sup>			0.185/0.441	

The IDs model also returned significant effects of several of the IDs predictors. There was a simple effect of Spanish LexTALE such that listeners with higher Spanish LexTALE scores tended to have faster RTs; this effect is depicted in the left panel of Figure 4, which plots the estimated marginal means for listeners at the mean +/- 1 SD for Spanish LexTALE

performance. The interaction of Spanish LexTALE with the coefficient comparing SSN vs. 2TB is also visible here. Post-hoc comparisons with familywise-adjusted  $p$  values indicated that listeners with the lowest LexTALE scores were slower in SSN as compared with 2TB (estimate = 0.5,  $t = 5.7$ ,  $p < 0.0001$ ), while listeners with the highest LexTALE scores showed no difference between SSN and 2TB (estimate = 0.2,  $t = 1.9$ ,  $p = 0.11$ ). The right panel of Figure 4 depicts the interaction of AYRTs with noise condition; the simple effect of AYRTs was not significant, but the interaction between AYRTs and the coefficient for S2TB vs. E2TB was. As seen in Figure 4b, AYRTs did not predict the speed of word recognition in S2TB, but they did predict word recognition RTs in E2TB: as AYRTs get slower, so do word recognition times in E2TB. A post-hoc comparison went in opposite directions for listeners with the lowest vs. highest AYRTs (estimate =  $-0.4$ ,  $t = -2.9$ ,  $p = 0.003$ ).



**Figure 4.** Estimated marginal means from the fitted IDs model depicting (a) the interaction between Noise Condition and Spanish LexTALE performance (an index of Spanish proficiency), and (b) the interaction between Noise Condition and AX-CPT performance (AY Trials RT, an index of inhibitory control). In (a), listeners with higher Spanish proficiency had faster RTs overall, and showed no difference between the SSN and 2TB conditions. In (b), listeners with less-efficient inhibitory control (i.e., higher values for AY Trials RT) were disproportionately affected by E2TB. (SSN = speech-shaped noise, S2TB = Spanish two-talker babble, E2TB = English two-talker babble).

## 4. Discussion

### 4.1. Summary of Findings

The primary goal of the present study was to better understand the role of predictive processes in mitigating speech-perception-in-noise deficits when listening in the non-dominant language (Spanish, for our listeners). We also wanted to understand how the ability to take advantage of sentence constraints during speech-in-noise perception might be modulated by listeners' linguistic and cognitive profiles, and also by the content of the masking noise itself.

The analyses of both keyword transcription accuracy and response times showed that listeners were able to take advantage of sentence constraints to improve word recognition in noise. However, these effects were carried by the heritage Spanish listeners; as a group, L2 Spanish listeners did not show predictability effects. The magnitude of the predictability effect was also modulated by noise type. In the group-based analyses, constraining contexts were associated with both larger improvements in transcription accuracy and faster RTs in speech-shaped noise as compared with two-talker babble, though the largest effect of predictability on RTs was found when *no* background noise was present. Predictability did not enter into any three-way interactions, but we did find interactions of the individual difference predictors (i.e., Spanish proficiency and inhibitory control ability) with the type of masking noise. Higher Spanish proficiency was associated with faster RTs overall,

and also with a reduction in the difference between the SSN and 2TB conditions. Finally, listeners with the least-efficient inhibitory control ability showed the greatest interference from E2TB in the form of particularly slow RTs in that condition. In the sections that follow, we consider these findings in turn.

#### 4.2. Listeners Can Benefit from Constraining Sentence Contexts under Some Conditions

As anticipated in Section 1.2, our results suggest that—similar to previous findings regarding comprehension in the written modality—the ability to make predictions about an upcoming target word can facilitate word recognition. Moreover, given our manipulation of semantic constraint in particular (i.e., avoiding influence from factors such as associative priming or collocational strength), we can be reasonably confident that the listeners who benefited from higher-predictability contexts in this study were benefiting specifically from predictive processing.

It is notable, then, that the predictability effects in this study were carried by the heritage listeners, who differed from the L2 listeners along multiple dimensions (AoA for Spanish and English, Spanish vocabulary size, and others; see Table 1). Given the differences between the groups, we cannot state with certainty which aspects of language experience promote an ability to take advantage of constraining sentence contexts. However, we do note that our objective proficiency measure, the Spanish LexTALE vocabulary test, was directly pitted against the binary group classification in the analysis of individual differences, and this measure of Spanish proficiency did not add any additional predictive power beyond the group classification (at least in terms of its association with predictability effects). Given the moderate overlap in Spanish proficiency across groups, combined with the fact that the main difference between participant groups was Spanish AoA (or some latent variable associated with AoA), the present results may align with previous suggestions that AoA may be particularly important in the context of speech-in-noise perception (Fricke 2022; Kousaie et al. 2019; Mayo et al. 1997). These results also suggest that earlier exposure to and use of a language, as was the case for Spanish in the heritage listener group, may confer advantages in the ability to generate predictions online when under cognitive load.

We also found that the magnitude of the predictability effect differed across noise types. The finding that predictability effects were largest in the clear aligns with Bradlow and Alexander (2007) in demonstrating that context effects can be helpful when listening in a non-dominant language, but a sufficiently clear signal appears to be necessary in order for listeners to take full advantage of them. However, the present results diverge from previous findings in demonstrating that predictability effects were greater in speech-shaped noise as compared with two-talker babble. Importantly, the particularly low accuracy in the low-predictability condition in SSN (see Figure 1) indicates that SSN was the condition in which it was hardest for listeners to extract acoustic information. The fact that performance in the low-predictability condition was considerably better in two-talker babble, but that there was *not* a concomitant improvement for high-predictability sentences, indicates that prediction processes in particular were negatively impacted in the two-talker babble conditions; the predictability boost in two-talker babble was not as great as we would otherwise expect. This suggests that generating predictions and coping with informational masking may draw on the same (limited) cognitive resources; when cognitive resources are being devoted to managing informational interference, fewer resources are available for making predictions. This interpretation is in line with previous research by Zirnstein et al. (2018), which demonstrated that bilingual second language prediction may rely on the ability to regulate the dominant language—a skill that may similarly be recruited when attempting to mitigate informational masking effects.

#### 4.3. Separating the Effects of Energetic versus Informational Masking

The analysis of individual differences indicated that Spanish proficiency and participant group each predicted the magnitude of the RT difference between speech-shaped

noise and two-talker babble. Given that the two participant groups showed equivalent performance on low-predictability trials in SSN (see Figure 1), this indicates that heritage listeners experienced more release from masking in two-talker babble, as compared with the L2 listeners; heritage listeners were thus better able to take advantage of glimpsing (Cooke 2006). To the extent that the residual group differences in this analysis were associated primarily with AoA, this could suggest that earlier AoA may specifically confer a benefit in terms of coping with masking that fluctuates in the temporal domain.

Interestingly, however, increasing Spanish proficiency was found to be associated with a decreasing release from masking in the two-talker babble conditions (Figure 4). Across both participant groups, listeners with the greatest Spanish proficiency experienced equivalent interference in the SSN and 2TB conditions. One possible interpretation of this finding is that proficiency in the non-dominant language is most predictive of word recognition performance in the clear and under conditions of energetic masking, and it is simply not as predictive of performance in informational masking. However, this finding should be interpreted with caution, given the correlation between Spanish proficiency and group membership in this dataset; it is possible that once the effects of group membership are partialled out, the residual effect of Spanish proficiency can no longer be interpreted in a straightforward manner.

Finally, our results comparing the two informational masking conditions run counter to the language similarity hypothesis (Van Engen 2010). Van Engen found that English competing speech was more disruptive than Mandarin speech for both native English and L1 Mandarin listeners, leading to the proposal that linguistic similarity between the target and masker speech was an important determinant of masking. For our participants, the analogous finding would have been if Spanish two-talker babble were more disruptive than English babble, since the target language in our study was Spanish. The fact that our participants experienced *less* disruption from Spanish than from English competing speech (combined with other recent findings; Calandruccio et al. 2018) calls for a reevaluation of the language similarity hypothesis. We propose instead a language immersion hypothesis: taken together, our findings and Van Engen's findings suggest that the greater immersion context within which the experiment takes place may be an important determinant of masking effects. In other words, English may have been uniquely disruptive for both Van Engen's and our participants because both experiments took place in the U.S., within the context of a predominantly English-speaking society. Importantly, the idea that the immersion context (or "predominant language"; Silva-Corvalán and Treffers-Daller 2016) of the experiment should figure into the interpretation of the results has been pointed out in other areas of bilingualism research (Beatty-Martínez et al. 2020; Jacobs et al. 2016; Linck et al. 2009; Zirnstein et al. 2018); our proposal is simply to extend this idea to account for the relative disruption caused by competing speech in different languages.

#### 4.4. Inhibitory Control Ability Modulates Interference from the Dominant Language

To our knowledge, this study is the first to report an association between inhibitory control ability and the ability to manage interference from competing speech. More specifically, we found that performance in the AY condition of the AX-CPT is associated with a modulation of the relative disruption incurred by competing English speech; on average, listeners with the least efficient inhibitory control (i.e., the slowest RTs on correct AY trials) were also the slowest to recognize Spanish words in the presence of English two-talker babble. Critically, this finding suggests a relationship between inhibitory control ability and the ability to manage interference from the more dominant language, an effect that has been reported widely in other circumstances of perception and processing of a non-dominant language. Our findings therefore support claims that managing dominant language interference draws upon domain general executive or inhibitory control abilities, and that more effective deployment of those abilities online allows listeners to engage in more cognitively demanding perception strategies, such as forming expectations for upcoming words in the non-dominant language.

## 5. Conclusions

Word recognition in the non-dominant language is notoriously difficult in noisy conditions. Recent research has begun to converge on the idea that such difficulties may be due in large part to an inability to take advantage of constraining contexts. The present study supports and augments this proposal by demonstrating that the ability to take advantage of constraining contexts is subject to modulation by both cognitive and linguistic factors. Future research should take care to separate out the effects of proficiency, dominance, and age of acquisition—all key factors in distinguishing the nuanced language histories of heritage bilingual speakers from late acquirers of an L2—as well as the greater societal context of the experimental setting. Heritage bilinguals, in particular, may benefit from early life experiences with the home language, in ways that are echoed later in life, even after their language dominance shifts to the majority or predominant language (e.g., akin to work by [Pierce et al. \(2014\)](#) demonstrating the enduring maintenance of early-formed language representations following attrition). Our results align with previous findings (e.g., [Kousaie et al. 2019](#); [Mayo et al. 1997](#)) in suggesting that age of acquisition may be particularly important for understanding speech-in-noise perception in the non-dominant language. While much work remains to be carried out in order to understand precisely where such AoA benefits could originate, the present results suggest that dominance and proficiency are unlikely to tell the full story.

**Author Contributions:** Conceptualization, M.F.; methodology, M.F. and M.Z.; statistical analysis, M.F. and M.Z.; data curation, M.F. and M.Z.; writing—original draft preparation, review, and editing, M.F. and M.Z.; visualization, M.F.; project administration, M.F.; funding acquisition, M.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by a *Language Learning* Early Career Grant to the first author.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of the University of Pittsburgh (approval date 17 February 2020).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** All data and supporting materials are publicly available via the Open Science Framework at <https://osf.io/7xpjm>.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Note

- <sup>1</sup> Poorer performance on the BX condition (in contrast with the BY condition, which serves as a general control comparison) is more likely to occur for individuals who experience difficulty with proactive goal maintenance, such as the elderly ([Braver et al. 2001](#)) and those diagnosed with schizophrenia ([Cohen et al. 1999](#)), and is less likely in neurotypical, young adult populations.

## References

- Anwyl-Irvine, Alexander L., Jessica Massonnié, Adam Flitton, Natasha Kirkham, and Jo K. Evershed. 2020. Gorilla in our midst: An online behavioral experiment builder. *Behavioral Research Methods* 52: 388–407. [[CrossRef](#)] [[PubMed](#)]
- Bates, Douglas, Martin Maechler, Ben Bolker, and Steven Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48. [[CrossRef](#)]
- Beatty-Martínez, Anne L., Christian A. Navarro-Torres, Paola E. Dussias, María T. Bajo, Rosa E. Guzzardo Tamargo, and Judith F. Kroll. 2020. Internacional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46: 1022–47. [[CrossRef](#)] [[PubMed](#)]
- Beatty-Martínez, Anne L., Rose E. Guzzardo Tamargo, and Paola E. Dussias. 2021. Phasic pupillary responses reveal differential engagement of attentional control in bilingual spoken language processing. *Scientific Reports* 11: 23474. [[CrossRef](#)] [[PubMed](#)]
- Bilger, Robert C., J. M. Nuetzel, William M. Rabinowitz, and C. Rzeczkowski. 1984. Standardization of a test of speech perception in noise. *Journal of Speech, Language, and Hearing Research* 27: 32–48. [[CrossRef](#)]
- Birdsong, David. 2018. Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology* 9: 81. [[CrossRef](#)]

- Blumenfeld, Henrike K., and Viorica Marian. 2013. Parallel language activation and cognitive control during spoken word recognition in bilinguals. *Journal of Cognitive Psychology* 25: 547–67. [CrossRef]
- Bradlow, Ann R., and Jennifer A. Alexander. 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustic Society of America* 121: 2339–49. [CrossRef]
- Braver, Todd S., Deanna M. Barch, Beth A. Keys, Cameron S. Carter, Jonathan D. Cohen, Jeffrey A. Kaye, Jeri S. Janowsky, Stephan F. Taylor, Jerome A. Yesavage, Martin S. Mumenthaler, and et al. 2001. Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General* 130: 746–63. [CrossRef]
- Braver, Todd S., Jessica L. Paxton, Hannah S. Locke, and Deanna M. Barch. 2009. Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences* 106: 7351–56. [CrossRef]
- Calandruccio, Lauren, Emily Buss, Penelope Bencheck, and Brandi Jett. 2018. Does the semantic content or syntactic regularity of masker speech affect speech-on-speech recognition? *The Journal of the Acoustical Society of America* 144: 3289–302. [CrossRef]
- Cohen, Jonathan D., Deanna M. Barch, Cameron Carter, and David Servan-Schreiber. 1999. Context-processing deficits in schizophrenia: Converging evidence from three theoretically motivated cognitive tasks. *Journal of Abnormal Psychology* 108: 120–33. [CrossRef] [PubMed]
- Cooke, Martin. 2006. A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America* 119: 1562–73. [CrossRef] [PubMed]
- Cooke, Martin, Maria Luisa Garcia Lecumberri, and Jon Barker. 2008. The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustic Society of America* 123: 414–27. [CrossRef]
- Coulter, Kristina, Annie C. Gilbert, Shanna Koussaie, Shari Baum, Vincent L. Gracco, Denise Klein, Debra Titone, and Natalie A. Phillips. 2021. Bilinguals benefit from semantic context while perceiving speech in noise in both of their languages: Electrophysiological evidence from the N400 ERP. *Bilingualism: Language and Cognition* 24: 344–57. [CrossRef]
- Cutler, Anne, Andrea Weber, Roel Smits, and Nicole Cooper. 2004. Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustic Society of America* 116: 3668–78. [CrossRef]
- Diez, Emiliano, María Angeles Alonso, Natividad Rodríguez, and Angel Fernandez. 2018. Free-Association Norms for a Large Set of Words in Spanish. [Database]. Available online: [https://iblues-inico.usal.es/iblues/nalc\\_about.php](https://iblues-inico.usal.es/iblues/nalc_about.php) (accessed on 20 April 2022).
- Edwards, Bethany G., Deanna M. Barch, and Todd S. Braver. 2010. Improving prefrontal cortex function in schizophrenia through focused training of cognitive control. *Frontiers in Human Neuroscience* 4: 32. [CrossRef]
- Federmeier, Kara D., Devon B. McLennan, Eesmerelda De Ochoa, and Marta Kutas. 2002. The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology* 39: 133–46. [CrossRef]
- Federmeier, Kara D., Marta Kutas, and Rina Schul. 2010. Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language* 115: 149–61. [CrossRef]
- Fernandez, Angel, Emiliano Diez, María Angeles Alonso, and María Soledad Beato. 2004. Free-association norms for the Spanish names of the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers* 36: 577–84. [CrossRef]
- Festen, Joost M., and Reinier Plomp. 1990. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *Journal of the Acoustic Society of America* 88: 1725–36. [CrossRef] [PubMed]
- Fricke, Melinda. 2022. Modulation of cross-language activation during bilingual auditory word recognition: Effects of language experience but not competing background noise. *Frontiers in Psychology* 13: 674157. [CrossRef] [PubMed]
- Fricke, Melinda, Megan Zirnstein, Christian Navarro-Torres, and Judith F. Kroll. 2019. Bilingualism reveals fundamental variation in language processing. *Bilingualism, Language and Cognition* 22: 200–7. [CrossRef] [PubMed]
- Garcia Lecumberri, Maria Luisa, Martin Cooke, and Anne Cutler. 2010. Non-native speech perception in adverse conditions: A review. *Speech Communication* 52: 864–86. [CrossRef]
- Golestani, Narly, Stuart Rosen, and Sophie K. Scott. 2009. Native-language benefit for understanding speech-in-noise: The contribution of semantics. *Bilingualism: Language and Cognition* 12: 385–92. [CrossRef]
- Gonthier, Corentin, Brooke N. Macnamara, Michael Chow, Andrew R. A. Conway, and Todd S. Braver. 2016. Inducing proactive control shifts in the AX-CPT. *Frontiers in Psychology* 7: 1822. [CrossRef]
- Gor, Kira. 2014. Raspberry, not a car: Context predictability and a phonological advantage in early and late learners' processing of speech in noise. *Frontiers in Psychology* 5: 1449. [CrossRef]
- Hazan, Valerie, and Andrew Simpson. 2000. The effect of cue-enhancement on consonant intelligibility in noise: Speaker and listener effects. *Language & Speech* 43: 273–94. [CrossRef]
- Hervais-Adelman, Alexis, Maria Pefkou, and Narly Golestani. 2014. Bilingual speech-in-noise: Neural bases of semantic context use in the native language. *Brain & Language* 132: 1–6. [CrossRef]
- Ito, Aine, Martin Corley, and Martin J. Pickering. 2018. A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition* 21: 251–64. [CrossRef]
- Izura, Cristina, Fernando Cuetos, and Marc Brysbaert. 2014. Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica* 35: 49–66. [CrossRef]
- Jacobs, April, Melinda Fricke, and Judith F. Kroll. 2016. Cross-language activation begins during speech planning and extends into second language speech. *Language Learning* 66: 324–53. [CrossRef] [PubMed]

- Kalikow, Daniel N., Kenneth N. Stevens, and Lois L. Elliott. 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America* 61: 1337–51. [CrossRef] [PubMed]
- Kilman, Lisa, Adriana Zekveld, Mathias Hällgren, and Jerker Rönnerberg. 2014. The influence of non-native language proficiency on speech perception performance. *Frontiers in Psychology* 5: 651. [CrossRef]
- Kousaie, Shanna, Shari Baum, Natalie A. Phillips, Vincent Grocco, Debra Titone, Jen-Kai Chen, Xiaoqian J. Chai, and Denise Klein. 2019. Language learning experience and mastering the challenges of perceiving speech in noise. *Brain and Language* 196: 104645. [CrossRef]
- Krizman, Jennifer, Ann R. Bradlow, Silvia Siu-Yin Lam, and Nina Kraus. 2017. How bilinguals listen in noise: Linguistic and non-linguistic factors. *Bilingualism, Language and Cognition* 20: 834–43. [CrossRef]
- Kroll, Judith F., and Paola E. Dussias. 2013. The comprehension of words and sentences in two languages. In *The Handbook of Bilingualism and Multilingualism*, 2nd ed. Edited by Tej K. Bhatia and William C. Ritchie. Malden: Wiley-Blackwell Publishers, pp. 216–43. [CrossRef]
- Kroll, Judith F., Kinsey Bice, Mona Roxana Botezatu, and Megan Zirnstein. 2022. On the dynamics of lexical access in two or more languages. In *The Oxford Handbook of the Mental Lexicon*. Edited by Anna Papafragou, John C. Trueswell and Lila R. Gleitman. New York: Oxford University Press, pp. 583–97. [CrossRef]
- Kutas, Marta, Kara D. Federmeier, and Thomas P. Urbach. 2014. The “negatives” and “positives” of prediction in language. In *The Cognitive Neurosciences*. Edited by Michael S. Gazzaniga and George R. Mangun. Cambridge: MIT Press, pp. 649–56.
- Lemhöfer, Kristin, and Mirjam Broersma. 2012. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavioral Research Methods* 44: 325–43. [CrossRef]
- Lenth, Russell V. 2021. Emmeans: Estimated Marginal Means, Aka Least-Squares Means. R Package Version 1.6.3. Available online: <https://CRAN.R-project.org/package=emmeans> (accessed on 20 April 2022).
- Linck, Jared A., Judith F. Kroll, and Gretchen Sunderman. 2009. Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science* 20: 1507–15. [CrossRef]
- MacKay, Ian R., Diane Meador, and James E. Flege. 2001. The identification of English consonants by native speakers of Italian. *Phonetica* 58: 103–25. [CrossRef] [PubMed]
- Marian, Viorica, Henrike K. Blumenfeld, and Margarita Kaushanskaya. 2007. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research* 50: 940–67. [CrossRef]
- Marian, Viorica, James Bartolotti, Sarah Chabal, and Anthony Shook. 2012. CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE* 7: e43230. [CrossRef] [PubMed]
- Martin, Clara D., Guillaume Thierry, Jan-Rouke Kuipers, Bastien Boutonnet, Alice Foucart, and Albert Costa. 2013. Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language* 55: 381–401. [CrossRef]
- Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, Harald Baayen, and Douglas Bates. 2017. Balancing type I error and power in linear mixed models. *Journal of Memory and Language* 94: 305–15. [CrossRef]
- Mayo, Lynn H., Mary Florentine, and Søren Buus. 1997. Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research* 40: 686–93. [CrossRef] [PubMed]
- Meador, Diane, James E. Flege, and Ian R. A. MacKay. 2000. Factors affecting the recognition of words in a second language. *Bilingualism, Language and Cognition* 3: 55–67. [CrossRef]
- Meuter, Renate F. I., and Alan Allport. 1999. Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language* 40: 25–40. [CrossRef]
- Misra, Maya, Taomei Guo, Susan Bobb, and Judith F. Kroll. 2012. When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory & Language* 67: 224–37. [CrossRef]
- Morales, Julia, Carlos J. Gómez-Ariza, and Maria Teresa Bajo. 2013. Dual mechanisms of cognitive control in bilinguals and monolinguals. *Journal of Cognitive Psychology* 25: 531–46. [CrossRef]
- Morales, Julia, Carlos J. Gómez-Ariza, and Maria Teresa Bajo. 2015. Bilingualism modulates dual mechanisms of cognitive control: Evidence from ERPs. *Neuropsychologia* 66: 157–69. [CrossRef]
- Nuechterlein, Keith H. 1991. Vigilance in schizophrenia and related disorders. In *Neuropsychology, Psychophysiology, and Information Processing*. Edited by Stuart R. Steinhauer, John H. Gruzelier and Joseph Zubin. Amsterdam: Elsevier, vol. 5, pp. 397–433.
- Paxton, Jessica L., Deanna M. Barch, Martha Storandt, and Todd S. Braver. 2006. Effects of environmental support and strategy training in older adults’ use of context. *Psychology and Aging* 21: 499–509. [CrossRef]
- Pierce, Lara J., Denise Klein, Jen-Kai Chen, Audrey Delcenserie, and Fred Genesee. 2014. Mapping the unconscious maintenance of a lost first language. *Proceedings of the National Academy of Sciences* 111: 17314–19. [CrossRef] [PubMed]
- Rosvold, H. Enger, Allan F. Mirsky, Irwin Sarason, Edwin D. Bransome, and Lloyd H. Beck. 1956. A continuous performance test of brain damage. *Journal of Consulting Psychology* 20: 343–50. [CrossRef] [PubMed]
- Scharenborg, Odette, and Marjolein van Os. 2019. Why listening in background noise is harder in a non-native language than in a native language: A review. *Speech Communication* 108: 53–64. [CrossRef]
- Scharenborg, Odette, Juul M. J. Coumans, and Roeland van Hout. 2018. The effect of background noise on the word activation process in nonnative spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 44: 233–49. [CrossRef]

- Silva-Corvalán, Carmen, and Jeanine Treffers-Daller. 2016. Digging into dominance: A closer look at language dominance in bilinguals. In *Language Dominance in Bilinguals. Issues of Measurement and Operationalization*. Edited by Carmen Silva-Corvalán and Jeanine Treffers-Daller. Cambridge: Cambridge University Press, pp. 1–14. [[CrossRef](#)]
- Van Engen, Kristin J. 2010. Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication* 52: 943–53. [[CrossRef](#)]
- Van Engen, Kristin J., and Ann R. Bradlow. 2007. Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of the Acoustic Society of America* 121: 519–26. [[CrossRef](#)]
- Wlotko, Edward W., and Kara D. Federmeier. 2015. Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex* 68: 20–32. [[CrossRef](#)]
- Woods, Kevin J. P., Max H. Siegel, James Traer, and Josh H. McDermott. 2017. Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics* 79: 2064–72. [[CrossRef](#)]
- Zirnsstein, Megan, Janet G. van Hell, and Judith F. Kroll. 2018. Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition* 176: 87–106. [[CrossRef](#)] [[PubMed](#)]
- Zirnsstein, Megan, Janet G. van Hell, and Judith F. Kroll. 2019. Cognitive control and language ability contribute to online reading comprehension: Implications for older adult bilinguals. *International Journal of Bilingualism* 23: 971–85. [[CrossRef](#)]