

CONSEQUENCES OF BILINGUALISM FOR SPEECH UNDERSTANDING IN NOISE

By

Jens Schmidtke

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Second Language Studies – Doctor of Philosophy

2015

ABSTRACT

CONSEQUENCES OF BILINGUALISM FOR SPEECH UNDERSTANDING IN NOISE

By

Jens Schmidtke

The present study sought to identify factors that would be associated with speech understanding in noise (SUN) ability in monolingual and bilingual listeners. The Ease of Language Understanding (ELU) model predicts that mismatches between the speech signal and phonological representations stored in long-term memory (LTM) will result in greater explicit processing effort and, as a consequence, decreased comprehension. Such mismatches can be the result of signal degradations or imprecise lexical representations in LTM. Based on the lexical quality hypothesis (Perfetti & Hart, 2002; Perfetti, 2007), it was hypothesized that the quality of lexical representations would differ within speakers as a function of word frequency and between speakers a function of overall language experience, operationalized here as vocabulary knowledge. From these assumptions it followed that bilingual speakers would have less precise lexical representations than monolinguals because of their reduced language experience as a result of speaking two languages. A second hypothesis was that the same relationship between vocabulary knowledge and SUN exists in monolingual and bilingual speakers.

The present study tested these predictions in a sample of 53 English monolingual and 48 early Spanish-English bilingual speakers with a mean age of 20.7 years ($SD = 2.6$, range = 18-31). All participants completed two subtests of verbal ability (picture vocabulary and verbal analogies) from the Woodcock-Muñoz Language Survey (WMLS), a standardized test of English. In addition, participants completed tests that were believed to be associated with SUN, a verbal WM test, a nonlinguistic test of auditory attention, and a consonant perception in noise

test. SUN was tested using sentences from a previously published test, the Speech Perception in Noise (SPIN) test (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984), at two signal-to-noise ratios (SNR; 3 dB and -2dB), using multi-talker background babble as the noise masker. The participants' task was to type the last word of the sentence, which was either predictable from context (e.g., *The ship sailed along the coast*) or nonpredictable (e.g., *Mrs. Brown did not consider the coast*).

When looking at group differences, the results replicated previous studies, showing that bilinguals recognized target words with lower accuracy relative to monolinguals. In addition, monolinguals benefitted more from a predictive context than bilinguals. The results from the WMLS showed that bilinguals scored significantly lower than monolinguals. When English proficiency was used as a covariate, higher proficiency was associated with higher SUN accuracy in both groups. In addition, an analysis of word frequency showed that group differences were largest for low frequency words. However, the frequency effect was modulated by English proficiency in the bilingual group. Assuming that both the frequency effect and language proficiency are closely related to exposure to English, the present results suggest that the bilingual disadvantage in SUN results from reduced exposure to English, which is a consequence of being exposed to two languages.

In conclusion, the results confirmed the predictions of the ELU, showing that both signal degradations and receiver limitations (less precise phonological representations of words in LTM) resulted in less accurate SUN ability.

ACKNOWLEDGMENTS

This past five years at Michigan State University have been an enormous learning experience for me, both inside and outside of the university and I would like to thank those who have invested their time and efforts in me and the friends I have made on the way.

First of all I would like to thank my advisor Aline Godfroid for always encouraging me to pursue my ideas while giving me direction on the way. I am also grateful for her excitement about my research and faith in me throughout these years that often exceeded my own.

My special thanks also go to the professors on my dissertation committee, Debra Hardison, Laura Dilley, and Paula Winke, who contributed their expertise to this dissertation but I also learned a lot by taking classes with them.

My gratitude also goes to Susan Gass for her leadership in the Second Language Studies, which always created a pleasant and productive work environment and I am thankful for the different teaching and research assistantships I received during my tenure in the program. Lastly, I would like to thank those with whom I have taken classes in these past five years, Diogo Almeida, Dr. Altman, Dr. Enbody, Debra Friedman, Shawn Loewen, Mr. McCullen, Charlene Polio, and Patti Spinner. This dissertation would not have been possible without the knowledge that they passed on to me. My thanks also go to professors Elizabeth Howard, Letti Naigles, and Manuela Wagner from the University of Connecticut without whom I would not have pursued this Ph.D.

Besides these amazing professors I also need to thank my colleagues in the Ph.D. program without whose encouragement I would not have endured without them, especially those in my cohort, Roman, Se Hoon, Ayman, Dominik, Le Anne, and Scott.

I am also especially grateful for my good friends Tim, Justin, Garrett, and Jessie. Although their areas of expertise are not in linguistics, their support and prayers were indispensable over the years and I don't know if I had been able to keep my sanity without them. The same is true for my friends from graduate intervarsity, especially Dan and Danielle, Camille, Laura and Chris and Priscilla.

Finally I would like to thank my family for all their support, I love you all. Danke, Oma und Opa, für all eure Gebete. And lastly I thank God without whom everything would be in vain.

This dissertation was financially supported by the College of Arts and Letters with a Dissertation Completion Fellowship and by the NSF with a Doctoral Dissertation Improvement Grant (NSF-DDIG 1349125).

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	ix
KEY TO ABBREVIATIONS	xiii
INTRODUCTION	1
CHAPTER 1: REVIEW OF THE LITERATURE	3
1.1 Speech perception	3
1.2 Speech perception under adverse listening conditions	6
1.3 Factors affecting speech perception in noise	7
1.3.1 Language background	7
1.3.2 Language proficiency	11
1.4 How does language proficiency influence SUN?	14
1.4.1 Less precise phonological representations	14
1.4.2 The Lexical Quality Hypothesis	18
1.4.3 Frequency effects	20
1.4.4 Activation, inhibition, and lexical knowledge	25
1.4.5 Word predictability	29
1.5 Speech perception in noise and cognition	30
1.5.1 Working memory	31
1.5.2 Working memory and Speech perception in noise	36
1.5.3 The Ease of Language Understanding model	39
1.6 Phonological quality hypothesis	41
CHAPTER 2: EXPERIMENT 1	45
2.1 Research questions and predictions	45
2.2 Methods	46
2.2.1 Participants	46
2.2.2 Materials	48
2.2.2.1 Background questionnaire	48
2.2.2.2 Speech perception in noise test	48
2.2.3 Procedure	49
2.3 Analysis	50
2.4 Results	53
2.4.1 The effects of noise and predictable context	53
2.4.2 The influence of lexical and sublexical variables on word recognition	55
2.5 Discussion	58

CHAPTER 3: EXPERIMENT 2	61
3.1 Methods	61
3.1.1 Participants	61
3.1.2 Materials	61
3.1.2.1 Woodcock Muñoz Language Survey - Revised	62
3.1.2.2 Test of attention in listening	62
3.1.2.3 Working memory	64
3.1.2.4 Consonant perception in noise	65
3.1.3 Relationship between predictor variables	66
3.2 Results	68
3.3 Discussion	74
CHAPTER 4: GENERAL DISCUSSION	78
CHAPTER 5: ANALYSIS OF INDIVIDUAL TESTS	85
5.1 Words in Noise	85
5.1.1 Methods	86
5.1.1.1 Participants	86
5.1.1.2 Materials	86
5.1.2 Results	87
5.1.2.1 English Words in Noise Test	87
5.1.2.2 Spanish Words on Noise Test	89
5.1.2.3 Individual differences analysis	90
5.1.3 Discussion	92
5.2 Verbal ability	96
5.2.1 Materials	98
5.2.2 Procedure	98
5.2.3 Results	99
5.2.4 Discussion	109
5.3 Working memory	115
5.3.1 Materials and procedure	115
5.3.2 Results	116
5.3.3 Discussion	117
5.4 Consonant perception in noise	119
5.4.1 Materials and Procedure	120
5.4.2 Results	121
5.4.3 Discussion	134
5.5 Test of Attention in Listening	138
5.5.1 The bilingual advantage	139
5.5.2 Methods	141
5.5.2.1 Materials	141
5.5.2.2 Procedure	142
5.5.3 Analysis	142
5.5.4 Results	143
5.5.5 Discussion	151

CHAPTER 6: CONCLUSION	155
APPENDIX	159
REFERENCES	164

LIST OF TABLES

Table 1. <i>Participant characteristics devided by language status.</i>	47
Table 2. <i>Summary of mixed-effects regression results for variables predicting accuracy on the Speech Perception in Noise test.</i>	54
Table 3. <i>Summary of the mixed-effects regression results of SPIN accuracy for monolinguals and bilinguals.</i>	58
Table 4. <i>Means and standard deviations of the predictor variables used in Experiment 2.</i>	66
Table 5. <i>Correlations and bivariate correlations between predictor variables and the four conditions of the Speech perception in Noise test.</i>	67
Table 6. <i>Results from the mixed-effects regression analysis of SPIN accuracy.</i>	69
Table 7. <i>Results of the mixed-effect regression analysis of the SPIN test for the monolingual and bilingual group.</i>	71
Table 8. <i>Word frequency of high, mid, and low frequency words on the SPIN test</i>	72
Table 9. <i>Mean language proficiency of the upper and lower half of the monolingual and bilingual group.</i>	72
Table 10. <i>Mean accuracy on the Words in Noise test.</i>	87
Table 11. <i>Mean number of Spanish speakers and percent exposure to Spanish</i>	101
Table 12. <i>Results of the regression analysis predicting picture vocabulary scores</i>	103
Table 13. <i>Differences in background variables between balanced and unbalanced bilingual participants.</i>	107
Table 14. <i>Confusion matrix - bilingual participants.</i>	124
Table 15. <i>Confusion matrix - monolingual participants.</i>	125
Table 16. <i>Typical consonant confusions by monolingual and bilingual participants.</i>	126
Table 17. <i>Results of the regression analysis of TAIL response times.</i>	147
Table 18. <i>Results of the regression analysis of response times on the TAIL.</i>	150
Table 19. <i>Mean item accuracy on the WIN.</i>	160
Table 20. <i>Items discrimination index for E-WIN words.</i>	161

LIST OF FIGURES

- Figure 1. Three different sources of adverse listening conditions. Based on Mattys et al. (2012); also see Mattys, Brooks, and Cooke (2009). 6
- Figure 2. Effect of word frequency on lexical decision times for Dutch (DLP), English (ELP), and French (FLP). From Keuleers, Diependaele, and Brysbaert (2010). Used with permission under the Creative Commons license. 21
- Figure 3. Results of the Speech Perception in Noise test divided by noise level and group. Error bars show the 95% confidence interval. 54
- Figure 4. Results of the Speech Perception in Noise test. Results are divided by condition and language group. Error bars show the 95% confidence interval. 55
- Figure 5. Effect of biphone probability on Speech Perception in Noise accuracy divided by group. Grey-shaded area shows the 95% confidence interval of the slope of the regression line. Each point represents the mean accuracy of a certain word. 56
- Figure 6. Effect of log10 word frequency on Speech Perception in Noise accuracy divided by group. Grey-shaded area shows the 95% confidence interval of the slope of the regression line. Each point represents the mean accuracy of a certain word. 57
- Figure 7. Relationship between oral language ability and accuracy on the SPIN test, depending on condition. HNHP=high noise-high predictability. HNLP=high noise-low predictability. LNHP=low noise-high predictability. LNLP=low noise-low predictability. 70
- Figure 8. The effect of frequency show for each of four groups. The monolingual and bilingual group were each divided into a high and low group based on a median split of their proficiency score. Whiskers show the 95% confidence interval. 74
- Figure 9. Mean accuracy on the SPIN test for each group (bilingual/monolingual) separated by noise level (high/low) and target word frequency (low/mid/high). The figure shows that in the bilingual group, the effect of noise was largest when frequency was low. 77
- Figure 10. Results of the English WIN test. Solid lines show the predicted values based on coefficients of the regression model described in the text. Dashed lines show the fitted values of this model. Whiskers show the 95% confidence interval. 88
- Figure 11. Results of the English and Spanish versions of the WIN test (bilingual participants only). Solid lines show the predicted values based on coefficients of the regression model described in the text. Dashed lines show the fitted values of this model. Whiskers show the 95% confidence interval. 90

- Figure 12. Effect of *Baseline RT* on WIN accuracy at each SNR. SNR = signal-to-noise ratio. Baseline RT is the mean response time on the Test of Attention in Listening (see text for further explanation). 91
- Figure 13. Effect of oral language ability on WIN accuracy at each SNR. SNR = signal-to-noise ratio. W-scores are arbitrary units with equal interval spacing. 92
- Figure 14. Relationship between language dominance and proficiency in English and Spanish. Language dominance was calculated by subtracting Spanish scores from English scores. Thus a positive score means English dominance and a negative score means Spanish dominance. 104
- Figure 15. Relationship between percent of exposure to Spanish and age in the bilingual sample. Participants were divided into a balanced and an unbalanced group based on the difference between their Spanish and English score on the WMLS (see text). 106
- Figure 16. Relationship between the picture vocabulary and the verbal analogies subtests of the WMLS. Compared to the monolingual participants, bilinguals performed lower on the picture vocabulary test as would be expected from the verbal analogies score. 108
- Figure 17. Relationship between working memory capacity and picture vocabulary scores. Grey-shaded area shows the 95% confidence interval of the regression line. 117
- Figure 18. Distribution of working memory scores when the effect of picture vocabulary was partialled out (residual variance). 119
- Figure 19. Mean accuracy on the consonant perception test divided by babble segment and speaker. Whiskers show the 95% confidence interval. Note the limited range of the y-axis to highlight the effects. 122
- Figure 20. Mean accuracy for each consonant on the consonant perception test. Whiskers show the 95% confidence interval. 123
- Figure 21. Relationship between accuracy on the consonant perception test and oral language ability. The regression line included one knot at 99.5. 130
- Figure 22. Accuracy on the consonant perception test as a function of group. The monolingual and bilingual groups were each divided into a high and low proficiency group based on a median split of their verbal ability score. Whiskers show the 95% confidence interval. 131
- Figure 23. Mean accuracy for each consonant on the consonant perception test. The monolingual and bilingual groups were each divided into a high and low proficiency group based on a median split of their verbal ability score. Whiskers show the 95% confidence interval. 132
- Figure 24. Relationship between mean accuracy on the consonant perception test and oral language ability. Consonant were divided into high and low phonotactic probability based on a median split. The interaction between phonotactic probability and language ability was significant. 134

- Figure 25. Mean accuracy on the TAIL in each of four conditions. Whiskers show the 95% confidence interval. Note the limited range of the y-axis to highlight the effect. 144
- Figure 26. Mean accuracy on the TAIL for monolinguals and bilinguals. The difference between same frequency and different frequency trials was larger for monolinguals than for bilinguals. Whiskers show the 95% confidence interval. Note the limited range of the y-axis to highlight the effect. 144
- Figure 27. Mean response time (RT) on the TAIL in each of four conditions. Whiskers show the 95% confidence interval. 145
- Figure 28. Mean response time (RT) in msec. on same and different frequency trials. Whiskers show the 95% confidence interval. Note the limited range of the y-axis. 146
- Figure 29. Mean response times (RT) in msec. in each of the four conditions of the TAIL. DF/SF = different/same frequency, DL/SL = different/same location. The difference between Version 1 and 2 was the location of the response keys (see Methods section in text). 147
- Figure 30. Effect of frequency difference between the first and second tone on response times (RT) in msec. The regression line shows the best fit with a polynomial function with three terms. 148
- Figure 31. Effect of language dominance on response times (RT, in msec.) and the location effect. Language dominance was calculated by subtracting Spanish proficiency scores from English proficiency scores so that scores above 0 indicate English dominance. 151
- Figure 32. Schematic representation of the results in this study. Arrows indicate significant relationships between variables. The two-way arrow indicates that more exposure to one language is associated with less exposure to the other language. SUN = speech understanding in noise. WM = working memory. CP = consonant perception. 156
- Figure 33. Mean accuracy on the English Words in Noise test for List 1 and 2. Whiskers show the 95% confidence interval. 163

KEY TO ABBREVIATIONS

AoA	Age of acquisition
CP	consonant perception
dB	decibel
DFDL	different frequency-different location
DFSL	different frequency-same location
ELU	Ease of Language Understanding
HNHP	high noise-high probability
HNLP	high noise-low probability
LNHP	low noise-high probability
LNLP	low noise-low probability
LQH	lexical quality hypothesis
LRM	lexical restructuring hypothesis
LTM	long-term memory
M	mean
ms	milliseconds
OL-SS/OL-W	oral language (standard/W-score)
PQH	phonological quality hypothesis
PV	picture vocabulary
RT	response time
SD	standard deviation
SFDL	same frequency-different location

SFSL	same frequency-same location
SNR	signal-to-noise ratio
SPIN	speech perception in noise
SRT	speech reception threshold
STM	short-term memory
SUN	speech understanding in noise
TAIL	test of attention in listening
VA	verbal ability
VC	vowel consonant
WIN	words in noise
WM	working memory
WMC	working memory capacity
WMLS-R	Woodcock-Muñoz Language Survey-Revised

INTRODUCTION

Many can attest to the difficulty of following a conversation in a noisy environment. Yet, while everyone is affected by noise, some people seem to be better able to cope with adverse listening situations than others. The aim of the research described in this dissertation is to find factors that would be able to explain some of these individual differences. Of special interest is the variable language experience given that many prior studies have found that listening in noise in a second language is more difficult than in one's first language (e.g., Mayo, Florentine, & Buus, 1997). Although this seems to be a robust finding, it is not yet clear what factors are responsible for these differences.

The foundation for the hypotheses generated and tested in the present investigations are the Ease of Language Understanding (ELU) model (Rönnberg et al., 2013) and the lexical quality hypothesis (LQH; Perfetti & Hart, 2002; Perfetti, 2007). The LQH was developed by Perfetti and colleagues to explain differences between skilled and less skilled readers but, according to Perfetti, it also applies to “spoken language with a focus on phonological representations and meaning” (2007, p. 361). The assumption of the hypothesis is that lexical representations will be more or less precise, whereby preciseness of phonological representations is defined as stronger connections between levels of representation (phonology, semantics, and orthography) and more distinguishing features of words that make similar sounding words less confusable. More experience with a word will strengthen its representations so that a high frequency word has more robust representations than a low frequency word. The LQH has direct consequences for bilingual speakers because they often have less experience with words in either of their languages compared to a speaker of only one language (cf. Gollan, Montoya, Cera, & Sandoval, 2008).

Whereas the LQH predicts differences in phonological processing because of differences in lexical representations, the ELU model provides a framework for investigating the influence of individual differences in executive functions on word recognition in noise. The model assumes that lexical access is effortless when there is a match between the speech signal and phonological representations in long-term memory. However, under sub-optimal listening conditions, when the speech signal is distorted, the resulting mismatch has to be resolved through explicit processing, which depends on working memory resources, to fill in information missing from the input. Thus the prediction is that individual differences in working memory are correlated with scores of word recognition in noise. At the same time, the model also predicts a greater mismatch between the signal and long-term memory representations when these representations are less precise, that is, fewer phonological attributes match the speech signal. Thus the ELU model complements the LQH and both make similar predictions regarding bilingual speakers.

The aim of this dissertation is to test the predictions generated by the two models to better understand how individual differences in language experience and executive functions affect language processing in noise. Results will help refine models and hopefully also inform interventions that aim at improving listening in noise ability in mono- and bilingual speakers.

CHAPTER 1: REVIEW OF THE LITERATURE

1.1 Speech perception

Speech perception is a complex process that, simply speaking, comprises the mapping of an acoustic signal (mechanic vibrations at different frequencies) to internal abstract representations in the brain (Giraud & Poeppel, 2012b, p. 225). What is remarkable about this process is that the recognition of words in the signal seems to be effortless despite the fact that, contrary to written language, no clear markers of word boundaries are present in the acoustic signal. What is more, the speech signal is surprisingly variable, that is, there is considerable variance in the production of single phonemes and words between and even within speakers (e.g., Ernestus & Warner, 2011; Pitt, Dilley, & Tat, 2011). Thus one important goal of research on speech perception, and word recognition, is to find the mechanism by which the brain decodes and recomposes the signal into the message that was intended by a speaker. The field of speech perception has thus been concerned with these two problems: How are words recognized (e.g., Dahan & Magnuson, 2006; McQueen, 2007) and how is invariant perception achieved despite a variable speech signal (e.g., Diehl, Lotto, & Holt, 2004; Liberman & Mattingly, 1985). In this review I will mostly focus on the word level.

Most models of spoken word recognition assume a process by which the acoustic-phonetic signal is mapped to phonological representations, or phonemes, that in turn activate matching words. Furthermore, because possible words are often embedded within words and can also cross word boundaries, most models agree that word recognition is a competitive process (Luce & Pisoni, 1998; Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris, 1994). For example, the sentence *The catalogue in a library* contains the embedded words *cat*, *cattle*, *login*, *lie*, and *eye* (Norris & McQueen, 2008, p. 361). These words are assumed to receive *activation*, a

metaphor often used in psycholinguistics. Evidence for activation of multiple words comes from cross-modal priming studies among others. For example, hearing /lai/ extracted from *library* facilitates recognition of both *lie* and *library* in a lexical decision task (i.e., deciding whether an orthographically presented stimulus is a word or nonword). Hearing /laib/, on the other hand, impedes recognition of *lie* (see Cutler, 2012). Thus it is assumed that words are only considered as possible candidates for word recognition as long as they match the speech signal. In the sentence above, *catalogue* receives more activation as the signal unfolds and in turn inhibit *cattle*.

Through this process of activation and inhibition of possible words, competition between lexical candidates is resolved and those lexical candidates that exceed a certain activation threshold are selected. The examples above show the interactive nature of word recognition. Words receive activation as the speech signal unfolds, that is, as soon as the signal partially matches a phonological representation, and activation also cascades down to the semantic level. This is well documented by studies using the visual-world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In this paradigm, participants are typically presented with four pictures on a computer screen and hear instructions to manipulate one of the pictures (e.g., by clicking on it or moving it on the screen), with the assumption being that fixation probabilities on pictures reflect lexical activation. A seminal study by Allopenna, Magnuson, and Tanenhaus (1998) showed that eye-movements are closely linked to the unfolding speech signal. In this study, participants saw, for example, a display with a beaker, a beetle, a speaker, and a baby carriage. When participants heard “Pick up the beaker” they were initially equally likely to look at any of the pictures. However, after the onset of the target word, in this case *beaker*, they were more likely to look at the beaker and the beetle until the two words disambiguated. A few

hundred milliseconds into the target word participants also looked more at the speaker than the unrelated object, suggesting that *speaker* had received activation despite the initial mismatch.

Of course, speech perception is a much more complex process than outlined here and models of spoken word recognition have to make many simplifying assumptions about the input. For example, most models take a pre-processed signal as input and omit the stage during which the signal is presumably decoded into phonemes. As a consequence, one can easily forget that speech outside the laboratory hardly ever consists of a stream of discrete phonemes in citation form. For example, competing noise, sloppy pronunciations and coarticulation, an unfamiliar accent, all make the signal that arrives at the ear less than optimal. Yet people usually succeed in decoding and understanding the message. In fact, research has shown that speech perception, and subsequently language comprehension, is still possible when the signal is deeply impoverished. For example, speech with reduced spectral information (e.g., voice-vocoded speech) that preserves the temporal structure of the speech signal can still be understood (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). At the same time, when temporal detail is removed from the signal through low-pass filtering, comprehension is still possible with detailed spectral information (Obleser, Eisner, & Kotz, 2008). This shows that the speech signal carries information that is seemingly redundant under optimal listening conditions.

What is evident from studies using acoustically degraded stimuli is that word recognition cannot simply be a process by which phonemes are mapped to lexical entries stored in long-term memory. Rather, it must be a probabilistic process with the parser settling on the most likely intended message given the evidence from the bottom-up signal but also top-down information such as the topic (Norris & McQueen, 2008; also see Obleser & Eisner, 2009).

1.2 Speech perception under adverse listening conditions

Mattys, Davis, Bradlow, and Scott (2012) identify three sources for adverse listening conditions. Source degradation refers to situations where the speech signal diverges from speech carefully produced by a member of the same speech community as the listener. Reasons for source degradation can be casual speech, a speech disorder, or an unfamiliar accent.

Transmission degradation, on the other hand, occurs during the transmission of the signal from the sender to the receiver. This can be as a result of energetic masking or non-energetic masking. Energetic masking refers to the masking of the speech signal by a competing signal. When the competing signal is another talker, the listener also has to selectively attend to one speaker and ignore the other, which will result in additional cognitive load. Non-energetic masking occurs through signal distortions such as reverberation but also telephone conversations. In the latter, frequencies below 400 Hz and above 3400 Hz are cut out which results in a smaller range than the one covered by typical speech (100 – 5000 Hz).

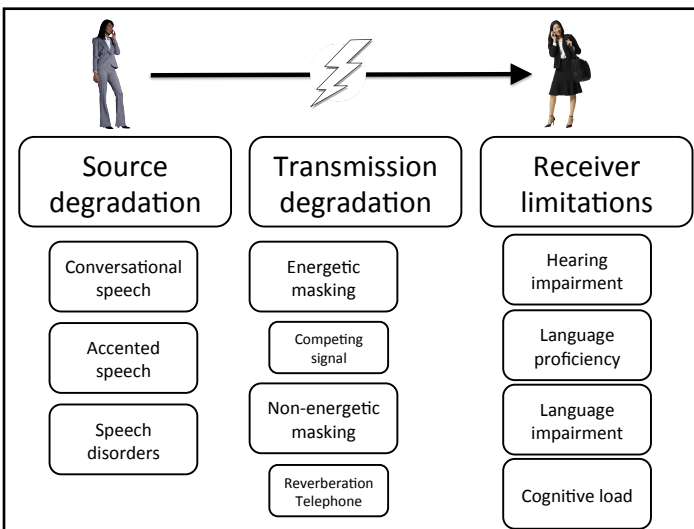


Figure 1. Three different sources of adverse listening conditions. Based on Mattys et al. (2012); also see Mattys, Brooks, and Cooke (2009).

Lastly, receiver limitations can also result in suboptimal listening situations. The cause can be a hearing impairment, insufficient proficiency in a language, a language impairment, for

example as a result of brain injury, and cognitive resource limitations (e.g., Mattys & Wiget, 2011; Mayo et al., 1997; Wilson, McArdle, & Smith, 2007).

The research described in this dissertation investigates the effects of one type of transmission degradation, energetic masking, and two potential receiver limitations, namely language experience/proficiency and individual differences in cognitive resources (executive functions).

1.3 Factors affecting speech perception in noise

Two broad factors influencing speech understanding in noise (SUN) will be reviewed here that are the focus of this dissertation. The first factor is verbal ability in relation to the language status of the tested language (first language vs. second language) and language experience (growing up with one language vs. two languages). The second factor is cognitive ability or executive functions, which are, broadly defined, “a set of general-purpose control mechanisms [...] that regulate the dynamics of human cognition and action” (Miyake & Friedman, 2012, p. 8). These two factors have been associated with SUN but have typically been studied in isolation. However, there may be interactions between verbal and cognitive abilities, which remain hidden if these factors are studied separately. Not included in this review are studies on SUN in clinical populations and the elderly. For example, deficits in SUN have been shown to be associated with dyslexia (e.g., Ziegler, Pech-Georgel, George, & Lorenzi, 2009) and language learning impairment (e.g., Ziegler, Pech-Georgel, George, Alario, & Lorenzi, 2005).

1.3.1 Language background

Many studies have investigated differences in speech perception in native and nonnative speakers. The usual finding is that speech perception in quiet is not different between first language (L1) and second language (L2) speakers but in noise L2 speakers typically perform

significantly worse (Bradlow & Alexander, 2007; Crandell & Smaldino, 1996; Mayo et al., 1997; Meador, Flege, & Mackay, 2000; Rogers, Lister, Febo, Besing, & Abrams, 2006; Schneider, Avivi-Reich, & Daneman, 2014; Shi & Sánchez, 2010, 2011; Shi, 2009, 2010; Van Engen, 2010). A few studies have also tested the same speakers in their L1 and L2 and found that L2 SUN is usually worse (Kilman, Zekveld, Hällgren, & Rönnberg, 2014; Rosenhouse, Haik, & Kishon-Rabin, 2006; Weiss & Dempsey, 2008). What is not always consistent across studies is whether noise has an additive or multiplicative effect on L2 listeners. Whereas Mayo et al. (1997) found an interaction between group and noise level (also see Shi, 2010; Tabri, Smith About Chacra, & Pring, 2011), other studies failed to find this interaction (Rogers et al., 2006). This may be due to differences in the tested participant population and noise conditions. For example, Rogers et al. (2006) used three fixed signal-to-noise ratios (SNRs) whereas Mayo et al. (1997) used an adaptive staircase procedure¹.

In studies on bilingual SUN samples are often divided into early and late bilinguals to test if age of acquisition has an effect on hearing in noise ability. An early onset of L2 acquisition is often defined as age 6 or younger whereas late commonly refers to 11 or older. These cutoff points are based on research on the critical period hypothesis that suggests a critical period for language acquisition roughly between 6 and puberty (e.g., Flege, Yeni-Komshian, & Liu, 1999; Johnson & Newport, 1989). For example, Meador et al. (2000) tested a group of early bilinguals (L2 onset ~ 7 years), a “mid” group (age of arrival ~14 years) and a “late” group with an age of arrival of ~19 years. They found a linear negative relationship between age of arrival and SUN

¹ In this procedure, the SNR is adjusted up or down depending on whether a participant correctly repeated a target word. This is done until the SNR is found at which a participant is able to repeat the target word 50% of the time (for a detailed explanation of this procedure see Mayo et al., 1997, p. 687).

performance, with all groups being worse than monolingual native speakers. In a subsequent regression analysis the authors found that age of arrival could explain 41.5% of the variance in SUN test scores. This same pattern was confirmed in other studies (e.g., Rogers et al., 2006; Shi & Sánchez, 2010), in which language background variables such as age of acquisition (AoA) and self-rated proficiency explained up to 80% of the variance in SUN (Shi, 2012). Thus the claim that AoA and other linguistic variables influence SUN is firmly established in the literature. What is still an open question is whether bilinguals who learned both languages from infancy, often called simultaneous bilinguals, will perform like monolinguals. Shi (2009) tested 12 simultaneous bilinguals who learned English between 1 and 3 and found no difference in performance to a group of 24 monolingual English speakers at an SNR of 0 dB in four different noise conditions (speech-weighted noise, multi-talker babble, and instrumental music played forward and reversed). Calandruccio and Zhou (2013) tested bilinguals growing up in a Greek-English bilingual environment in New York and found no difference compared to a group of monolingual English speakers when tested with three-talker background babble at an SNR of -5 dB. Interestingly, the bilingual group was also tested in Greek and no significant difference between the English and Greek test was found. However, the Greek version was not tested against a monolingual sample of Greek speakers and may therefore not be comparable to the English version. Nonetheless, the results show that the bilingual participants were proficient in both languages.

Shi (2010) also tested a group of eight simultaneous bilinguals and found no difference to a monolingual control group at SNRs of +6 and 0 dB and reverberation times of 1.2 and 3.6 seconds. The test Shi used included sentences with high and low predictability taken from the Speech Perception in Noise (SPIN) test (Bilger et al., 1984). In a high predictability sentence, the

final word, which participants have to recognize, can be inferred from the preceding context (e.g., “The ship sailed along the coast” vs. “Ms. Brown thought about the coast”). Shi (2010) found a significant difference between the monolinguals and the simultaneous bilinguals in the most unfavorable listening condition (high noise and high reverberation) in the predictable context condition with a large effect size (Cohen’s $d = 2.58$). This suggests that bilinguals did not benefit as much from predictive context (cf. Mayo et al., 1997) but such differences may only emerge in the most unfavorable listening conditions. A similar conclusion can be drawn from a study by Crandell and Smaldino (1996), who tested 20 monolingual and 20 early bilingual children matched in age (age range = 8 – 10 years). The bilingual participants had started to learn English before the age of 2, as reported by their parents, and were exposed to each language roughly 50% of the time. In quiet and at an SNR of +6 dB the authors found no significant differences between the groups but at more unfavorable SNRs (-6, -3, 0, and +3 dB) the bilingual group performed significantly worse than the monolingual group with the slope of the decline under increasing noise levels appearing to be steeper for bilinguals (though the authors did not state whether the group-by-noise-level interaction was significant).

However, AoA may not be the only linguistic variable influencing SUN. Shi and Sánchez (2011) tested Spanish-English bilingual speakers using SUN tests in English and Spanish. All participants had learned Spanish from birth but one group learned English early (~4 years) and became dominant in English whereas the other group learned English later in life (~13 years) and was dominant in Spanish. The authors found that both groups performed better on the test that measured their dominant language. This suggests that more exposure to a language has a positive effect on SUN as already mentioned above, but reduced language exposure over a lifetime may also have a negative effect on word recognition in noise.

An improvement of more recent studies over earlier studies including bilingual populations is that more background variables are usually reported, following a realization that bilinguals differ in many respects (Grosjean, 2001, 2008; von Hapsburg & Peña, 2002) as well as the publication of more standardized assessment instruments (Marian, Blumenfeld, & Kaushkanskaya, 2007). This makes comparisons across studies easier and may help explain why studies sometimes seem to find conflicting results. However, even more detailed information about the participants' background may be necessary. For example, even simultaneous bilinguals who were exposed to both languages from birth may differ in the relative exposure to each language. Parents may be monolingual or bilingual speakers and participants may spend more or less time in monolingual environments; for example, they may go to an English-only daycare from an early age on. These variables may determine whether simultaneous bilinguals differ from monolinguals on tests of SUN or not, as it has been shown that amount of early language exposure influences processing efficiency in monolingual and bilingual children (e.g., Gollan, Starr, & Ferreira, 2014; Hurtado, Grüter, Marchman, & Fernald, 2013; Weisleder & Fernald, 2013).

1.3.2 Language proficiency

Language proficiency is often included as a variable in research on SUN in bilingual speakers. It is often measured through self-assessment on a Likert-scale (e.g., Shi, 2012) or a proficiency test (e.g., Kilman et al., 2014). Language proficiency is often correlated with other language background variables such as AoA and length of residence in the country where the target language is spoken. Nevertheless, proficiency can sometimes explain additional variance above and beyond AoA (Shi, 2012). This may be because AoA and length of residence do not take into account how much a participant was exposed to each language. Two participants may

have come to the US at the same age but one may have been completely immersed in English whereas the other may have had more contact to other speakers of their native language (see Meador et al., 2000). Bilinguals with such different profiles will likely also differ in their proficiency in their two languages and so language proficiency may be a proxy variable for language exposure over a lifetime.

While self-rated proficiency can often explain substantial variance in a diverse sample of second language speakers (e.g., Shi & Farooq, 2012), it may be less sensitive to more nuanced differences in a sample of highly proficient or native speakers. For example, Shi and Sánchez (2011) tested English-Spanish bilingual speakers who were either dominant in English or Spanish. Participants were tested in both languages and the authors found that self-rated proficiency was only correlated with SUN performance in the non-dominant language. This may have been because of greater variance between subjects in the non-dominant language. In comparison, participants may tend to overestimate their proficiency in the dominant language, resulting in ceiling effects. It may also be that participants rate their ability to successfully communicate in everyday situations, which would be a more holistic measure and may be different from more fine-grained measures of verbal ability. In a more homogeneous sample, self-rated proficiency may therefore not be as good of a predictor as in very diverse samples.

A few studies have used standardized tests to measure proficiency. Rimikis, Smiljanic, and Calandruccio (2013) tested a diverse group of 102 nonnative speakers of English enrolled at two US universities. Participants took the Versant English test, which is a test designed for nonnative speakers. In addition, they also took a test of SUN that was specifically created for nonnative English speakers with limited proficiency. As in the studies cited above, the authors found a correlation between SUN and age of immigration and length of residence. In addition,

they found a high correlation ($r = .73$) between SUN performance and the Versant. Combined, Versant and age of immigration could explain 63% of the variance in SUN performance.

Similar results were obtained by Kilman et al. (2014). The authors tested native speakers of Swedish who had learned English as a foreign language in school. All participants completed a standardized test of English and an adaptive SUN test in English and Swedish in four different noise conditions to determine the SNR at which they perceived sentences with 50% accuracy (the Speech Reception Threshold, SRT). The noise conditions were stationary and fluctuating speech-shaped noise, English babble, and Swedish babble. The correlations between the SRT and English proficiency were $r = -.48, -.6, -.51$, and $-.65$, respectively, in the four conditions when the target language was English². When the target language was Swedish, no correlations were found.

An interesting question is whether vocabulary knowledge or overall verbal ability are also predictive of word recognition in noise performance in monolingual native speakers of a language. The answer to this question would show whether differences between first and second language listening are of a qualitative or quantitative nature. Monolingual speakers growing up in typical circumstances do not differ in age of first exposure to the language but there are differences in the amount and quality of input infants receive from their caregivers, which lead to great variability in vocabulary knowledge even at a very young age (Hart & Risley, 1995). This variability is likely to influence SUN given a diverse sample. Some evidence that vocabulary knowledge may be associated with SUN comes from a study by Tamati, Gilbert, and Pisoni (2013). The authors first tested a large sample of 121 healthy young-adult listeners on a SUN test and then asked those performing in the lower and upper quartile to come back for additional

² Note that these correlations are negative because a lower SRT means better SUN hearing)

tests. One of those tests was a word familiarity test on which participants rated how familiar they were with 150 words that were categorized as high, mid, and low familiarity based on a previous norming study. The good SUN listeners were significantly more familiar with low and mid familiarity words and marginally more familiar with high familiarity words. One limitation is that these results were based on self-ratings instead of a standardized test but they suggest nonetheless that the better SUN listeners had a larger vocabulary.

1.4 How does language proficiency influence SUN?

In the previous section it was shown that language proficiency may be an important variable that predicts SUN ability. In the following sections, I will review studies that may explain why language proficiency is correlated with SUN.

1.4.1 Less precise phonological representations

Language proficiency as measured by vocabulary size is correlated with many variables related to language processing. One model of word recognition and lexical development in children that has been very influential is the Lexical Restructuring Model (LRM) by Metsala, Walley and colleagues (Metsala & Walley, 1998; Walley, Metsala, & Garlock, 2003; Walley, 2008). This model proposes that infants' phonological representations of words in memory start out as crude, whole-word representations that lack phonemic detail. In contrast to theories that assume that infants have the same phonemic representations as adults (e.g., Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), the LRM assumes that phoneme categories only develop, or emerge, over time as a result of vocabulary growth. As children add more words to their mental lexicons, there is a need for those words to be represented with finer, segmental detail to ensure efficient processing (also see Charles-Luce & Luce, 1990). The model proposes that lexical restructuring, from crude representations to fine-grained segmental representations,

occurs on an item-by-item basis and is determined by lexical frequency, how often a word is encountered, and phonological neighborhood size. Thus lexical representations of high frequency words will be more precise or detailed than those of low frequency words. In addition, a word with many neighbors³ will be represented with more detail because there are more words that it sounds similar to. For example, *cap* is a neighbor of *cat*, as is *bat*, *cut*, and *mat*. If *cat* was only crudely represented in memory, it would be easily confusable with its neighbors. A word with no neighbors such as *idol*, on the other hand, would not have to be represented with as much detail because there are no words competing with it during recognition. Thus, according to the model, high frequency, high density words have the most precise phonological representations whereas low frequency, low density words have the least precise representations.

While there is evidence for the LRM (Metsala & Walley, 1998), subsequent studies have shown that infants may be more sensitive to phonetic detail than previously thought. Using eye-tracking, Swingley and colleagues have shown that mispronunciations affect word recognition in infants as young as 18 months (Swingley & Aslin, 2002, 2000, 2007; Swingley, 2003). The idea here is that a mispronunciation would be harder to detect if a heard word is matched to a stored representation in memory only on overall similarity compared to a word that is stored with segmental information. In Swingley and Aslin (2000) infants and toddlers saw two pictures and heard a sentence like “Where is the baby” (correct pronunciation condition) or “Where is the vaby” (mispronunciation condition). In both conditions children looked more to the target picture than the distractor but they also looked more to the target when hearing the correct pronunciation

³ A phonological neighbor is typically defined as a word that can be formed from another word by adding, deleting, or substituting a single phoneme. A word with many neighbors is said to come from a dense neighborhood whereas a word with no or few neighbors is said to come from a sparse neighborhood.

compared to the mispronunciation. These results show that children were thrown off by the mispronunciation and therefore must have been sensitive to the b/v distinction. The ability to distinguish between /b/ and /v/ was unrelated to vocabulary knowledge or age as would have been predicted by the LRM, which assumes that phonemic representations emerge as a result of vocabulary growth. However, Swingley and Aslin note (2000, p. 161) that the results do not necessarily provide evidence that infants have segmental representations or more holistic representations of words. It could be that non-phonemic representations are still quite detailed phonetically. Furthermore, the words used in studies like this one are usually words with which children are familiar. Even though children can detect mispronunciations in familiar words, the LRM predicts that less familiar words will be represented less precisely in the mental lexicon. Thus frequency of encounter with a word may be more important than neighborhood density for lexical restructuring, especially since studies have shown that while words in children's lexicons have fewer neighbors than words in adult lexicons, there are still many words in children's vocabularies that have neighbors (Coady & Aslin, 2003). To recap, the LRM posits that vocabulary acquisition drives a restructuring of phonological representations. As children add more words to their lexicons there is a need for more precise representations to be able to distinguish similar sounding words. To come back to the questions of how vocabulary size is related to spoken word recognition, one could hypothesize that speakers with larger vocabularies have more precise representations of these words, which, in turn, results in more efficient retrieval from amidst a more densely populated neighborhood.

Several objections have been raised against the LRM. Instead of positing that representations of words are qualitatively different in young children and adults, observed differences in experiments could result from the fact that children are just less familiar with

words because they have not heard them as many times as an older person. The more experience someone has with a word, the more phonological detail may be stored for this word. Frequency effects are well documented in the literature (e.g., Grosjean, 1980; Monsell, 1991; Murray & Forster, 2004; Oldfield & Wingfield, 1965; Rubenstein & Pollack, 1963) and are a powerful predictor of speed and accuracy of word recognition. In addition, frequency effects appear early, before the offset of a word, suggesting that less phonetic information is needed for successful recognition of high frequency words compared to low frequency words (Dahan, Magnuson, & Tanenhaus, 2001; Grosjean, 1980).

A recent study tested the hypothesis that frequency of encounter with a word determines the precision of a phonological representation of that word. White, Yee, Blumstein, and Morgan (2013) used an artificial lexicon paradigm (see Magnuson, Tanenhaus, Aslin, & Dahan, 2003) in which participants learned mappings between artificial words and geometric figures. The authors manipulated frequency by presenting word-object pairings once, five, or eight times during the learning phase. In the testing phase, participants saw a familiar and a novel shape and heard a familiar word, a mispronounced familiar word (e.g., *gav* instead of *bav*), or a novel word while their eye-movements were tracked. The eye-movement results showed that participants were less sensitive to mispronunciations after one exposure than after five or eight exposures, as evidenced by looks to the familiar object. The authors assumed that the strength of a lexical representation could explain these results. Because low frequency words require more acoustic input to be recognized, competitor words will receive more activation and may be less efficiently inhibited. Whatever the underlying mechanism may be, the main point of the White et al. (2013) study is that the results from adults are very similar to those obtained from children, that is, both look more at the familiar than the unfamiliar object when presented with a mispronunciation of the

label for the familiar object. In other words, they do not take the mispronounced word as a label for the unfamiliar object, which suggests that they did not notice the mispronunciation. Therefore we may assume that child and adult word recognition is not qualitatively different. The fact that young children behave differently when tested with familiar words (see studies by Swingley and colleagues cited above) may just reflect the fact that they have less experience with these words compared to adults or older children and thus weaker phonological representations.

To come back to the relationship between vocabulary size and word recognition, if the quality of lexical representations is dependent on frequency of encounter, then we may assume that people with a larger vocabulary also have more language experience in general. In this case, the relationship between vocabulary size and word recognition would not be causal but mediated by language experience. For example, we may assume that people with a larger vocabulary hear and read words in a greater variety of contexts. This may be better illustrated for reading than for listening but is certainly true for both modalities. Someone who regularly reads newspapers, novels, and scientific journals will learn many words by reading but they will also encounter all words, and especially low frequency words, much more often than someone who seldom reads (Kuperman & Van Dyke, 2013). This view is expressed in the Lexical Quality Hypothesis (LQH; Perfetti & Hart, 2002) that I will talk more about in the next section.

1.4.2 The Lexical Quality Hypothesis

Perfetti and colleagues (Perfetti & Hart, 2002; Perfetti, 2007) developed the LQH to explain individual differences between low-skill and high-skill readers. The assumption is that entries in the mental lexicon of a given reader differ in the quality of their representations, from words that are well-known to the individual to others that are only rarely encountered and of which the individual only has rudimentary knowledge. Quality then refers to the precision in the

representation of a word's form and meaning. Perfetti (2007) identifies five features that may distinguish high from low quality representations: orthography, phonology, grammar, meaning, and constituent binding. For example, high-quality phonological representations differ from low quality representations in the amount of phonological redundancy that is stored and the stability of the phonological representation; a less stable representation may not always be retrieved successfully. In the meaning dimension, high-quality lexical representations are less dependent on context and can be readily distinguished from related words. For example, one individual may know that barley, wheat, oat, and rye are grains but they may not know any attributes that distinguish among them. That person would have low-quality meaning representations of these words. Important for the present study is also the feature Perfetti calls *constituent binding*, which is “the degree to which the first four features [orthography, phonology, morpho-syntax, and meaning] are bound together” (2007, p. 360). High-quality constituent bindings are characterized by stronger connections between the different features, especially meaning and orthographic and phonological form. A stronger connection between phonology and meaning will make the meaning accessible faster upon hearing the phonological form of a word. Less tightly bound constituents, on the other hand, may lead to slow retrieval or retrieval failures. For example, someone might recognize the phonological form of a word but not remember its meaning.

While the LQH was developed to explain individual differences in reading in monolingual speakers, the model can easily be extended to second language speakers. According to the LQH, word knowledge is essential to reading skill. More skilled readers have better knowledge of all constituents of a word. Thus reading skill develops with experience. For example, the LQH states that high frequency words, words that are encountered often, have more precise representations compared to low frequency words. Someone who reads a lot will

encounter all words more frequently, thus all words will be of higher absolute frequency for this individual compared to someone who reads seldom.

1.4.3 Frequency effects

Word frequency is usually determined by tallying up the number of occurrences of words in large corpora of language. For example, Brysbaert and New (2009) based their word frequency database on a corpus of subtitles from American movies; the British National Corpus is based on 100 million words extracted from different written and spoken (transcribed) texts. A word that occurs once in the corpus may be encountered more or less frequently by someone who reads a lot but never by someone who does not read. Thus the objective word frequency would not be accurate for these two individuals, the subjective frequency for them would be higher and lower, respectively.

To understand how subjective, or actual, word frequency influences word recognition, it is important to understand frequency effects in general. As stated earlier, the frequency effect is the most robust variable known to predict lexical access (Murray & Forster, 2004). High frequency words are named faster, read faster, and recognized faster in spoken word recognition. However, the relationship between frequency and lexical access is not linear. Differences in frequency in the low frequency range have a much bigger impact on response times (RTs), for example in lexical decision, than changes in the high frequency range. However, when the $\log_{10}$ frequency is used as a predictor instead of frequency per million, the relationship becomes linear up to the very high frequency range where RTs reach asymptote. This is shown in Figure 2, which is adopted from Keuleers, Diependaele, and Brysbaert (2010) and shows RTs in lexical decision across a wide range of word frequencies in three languages, Dutch, English, and French. Because of the logarithmic relationship between frequency and RTs, a change in magnitude at

the low end of the scale, for example from 1 to 10 occurrences per million, will have the same effect as a change in magnitude at the high end, say, from 100 to 1000 occurrences per million. This suggests that in terms of individual differences, differences in reading experience, or language experience in general, will only have small effects on words from the high frequency range. However, we can expect large differences at the low end, especially for the least frequent words that may almost never be encountered by some people.

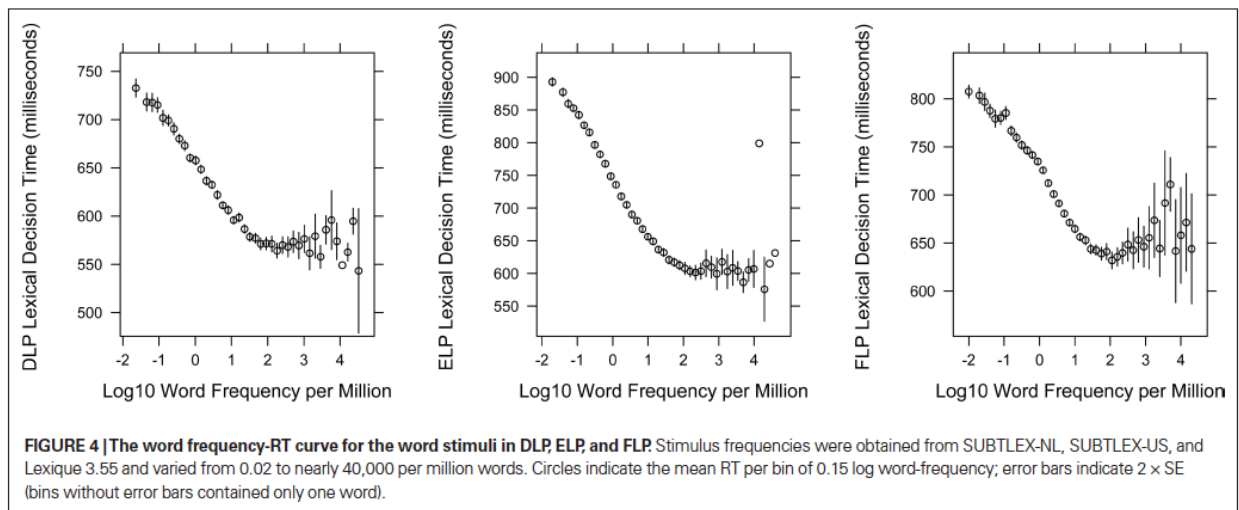


Figure 2. Effect of word frequency on lexical decision times for Dutch (DLP), English (ELP), and French (FLP). From Keuleers, Diependaele, and Brysbaert (2010). Used with permission under the Creative Commons license.

There is evidence for the hypothesis that individual differences in print exposure and vocabulary knowledge are associated with the size of the frequency effect. Chateau and Jared (2000) estimated reading exposure with a test called the famous author recognition test. In this test, participants are presented with a list of famous authors and foils and they check all the authors they recognize. Chateau and Jared found that on a lexical decision test, the frequency effect was larger for participants who recognized fewer authors compared to those who recognized more. This finding stands in contrast to a study by Lewellen, Goldinger, Pisoni, and Greene (1993). They divided participants into two groups (high verbal-low verbal) based on their

familiarity ratings of words, a vocabulary test, and a language experience questionnaire. The authors found that high verbal participants were consistently faster on three different tests, visual naming, lexical decision, and semantic classification. However, they did not find the critical interaction between lexical variables (frequency and neighborhood density) and group (high verbal/low verbal) on any of the tests. To reconcile these conflicting findings, Sears, Siakaluk, Chow, and Buchanan (2008) replicated both studies with a sample of university students that they divided into two groups based on their performance on the author recognition test. In two experiments, they used the same target words for a lexical decision test but manipulated the types of nonwords, regular nonwords (Exp. 1) and pseudohomophones (Exp. 2). Pseudohomophones are words that sound like real words, for example, *brane*, and are therefore harder to reject as nonwords. When pseudohomophones were used, the frequency by group interaction found by Chateau and Jared (2000) was replicated and when regular nonwords were used, no interaction was found as in Lewellen et al. (1993). Sears et al. (2008) suggest that the low print-exposure group relied more on phonological processes to compensate for less efficient orthographic processing. And so when the nonwords used in Exp. 2 sounded like real words, the task became more difficult for them.

Yap, Balota, Sibley, and Ratcliff (2012) reanalyzed data from a large scale project, the English Lexicon Project (Balota et al., 2007), in which the authors collected RTs for a wide range of words from a large sample of participants on two tasks, speeded naming (reading of single words) and lexical decision. Participants also completed a vocabulary test and the authors analyzed correlations between an individual's vocabulary score and the frequency effect in his or her RTs, as estimated by the regression coefficient. For speeded naming, they found a correlation such that higher vocabulary knowledge was associated with a smaller frequency effect but this

relationship was not found for lexical decision. The authors explained this discrepancy in findings in terms of task demands. They speculated that lexical decision involves two stages, a lexical access stage and a decision stage. Vocabulary knowledge more likely affects lexical access but if frequency effects mostly occur at the decision stage, then individual differences in vocabulary knowledge would be unrelated to the frequency effect.

Especially interesting with regards to individual differences in the frequency effect is a recent study by Diependaele, Lemhöfer, and Brysbaert (2013). These authors used a wide range of words of different frequencies and investigated the shape of the frequency curve. The task they used was a gated word identification test, in which participants saw words alternating with a visual mask on a computer screen. The visual form of the word appeared incrementally on the screen and participants hit a key and typed the word as soon as they recognized it. Participants were drawn from four groups, monolingual English speakers, and native speakers of Dutch, French, and German, who had learned English as a second language. All participants also completed a vocabulary test in English, in which participants had to decide whether a presented word was a real word in English. Because the authors were interested in the shape of the frequency curve, they used frequency-per-million counts from the Brysbaert and New (2009) subtitle corpus and fitted those to the RTs using a natural spline with 2 knots⁴ to account for the nonlinear relationship between RTs and frequency. Diependaele and colleagues found that for frequencies below ~100 per million the regression line was steep whereas it reached asymptote for frequencies above 100. Differences between groups only emerged in the lower frequency range, with the slope being steeper for the nonnative speakers compared to the native speakers.

⁴ A natural spline function allows the regression line to break at certain points to allow for nonlinear relationships between the predictor variable and the outcome variable.

Importantly, a proficiency-by-frequency interaction fitted the data better than a group by frequency interaction, suggesting that the differences between participants can be better explained in terms of English language proficiency rather than language status (L1 vs. L2). Moreover, the proficiency-by-frequency interaction was significant for the native English speakers, which confirms the results of the above cited studies and shows that even within a small restricted sample of college students sharing the same language experience (monolingual speakers) there is enough variance in proficiency scores to explain individual differences in lexical access.

All the studies cited in this section so far dealt with visual word recognition and found that print exposure or vocabulary knowledge, which is also related to print exposure (Lewellen et al., 1993), is related to the size of the frequency effect. A ready explanation for this finding is that for someone who reads a lot, all words will be of higher subjective frequency compared to someone who reads little. The weaker-links hypothesis developed by Gollan and colleagues (Gollan & Acenas, 2004; Gollan et al., 2008; Gollan, Montoya, & Werner, 2002; Gollan & Silverberg, 2001) is based on a similar logic in an attempt to explain differences in language production between monolingual and bilingual speakers. The hypothesis was originally put forth to explain why bilingual speakers experienced more tip-of-the-tongue states, which are situations in which the speaker knows a word but is unable to produce it (Gollan & Silverberg, 2001). The assumption underlying the weaker-links hypothesis is that while monolingual and bilingual speakers have the same overall language experience, the experience of bilinguals with either of their languages will be reduced. For example, a Spanish-English bilingual student may speak only Spanish at home and only English at school. Because of this reduced language experience, the hypothesis assumes that the links between semantic and phonological representations are

weaker compared to an age-matched monolingual speaker. This is similar to constituent binding feature of Perfetti's (2007) LQH. Evidence for the weaker-links hypothesis comes from a picture naming study that showed that the frequency effect was larger for bilingual speakers compared to monolingual speakers, even when bilinguals were tested in their dominant language (Gollan et al., 2008). In addition, the frequency effect in their nondominant (but first acquired) language was even larger. The same pattern was found by Ivanova and Costa (2008) who tested early bilingual speakers of Catalan and Spanish in their first acquired and currently dominant language. They found that compared to Spanish monolingual speakers, bilingual speakers were slower to name pictures with low frequency names. This is in line with a frequency account of the bilingual disadvantage because, due to the nonlinearity of the frequency effect, reduced exposure will affect low frequency words more than high frequency words.

1.4.4 Activation, inhibition, and lexical knowledge

What is common to the LRH and the frequency account is that words differ in their phonological representations as a result of language experience. Lexical representations of high frequency words may consist of more redundant (phonetic) information (Perfetti & Hart, 2002, p. 190). Thus, during spoken word recognition, more precise phonological representations may receive more activation from the acoustic signal because of a better match and thus its memory location is found faster, resulting in faster retrieval. When the speech signal is distorted, more redundancy in phonological representations will make lexical retrieval more robust. Consequently words with high quality lexical representations should be recognized more accurately and more efficiently under suboptimal listening conditions than low quality words. In line with this prediction is the finding that high frequency words are recognized more accurately under adverse listening conditions (Howes, 1957). Further evidence for the link between quality

of phonological representations and word recognition in noise comes from a recent study by Sommers and Barcroft (2011). In this study, native English speakers learned 24 novel words in Spanish, a language they had had no prior exposure to. All participants heard six repetitions of each word but half of the participants heard the words spoken by the same speaker whereas the other half heard each word spoken by six different speakers. In a subsequent testing phase, participants heard the Spanish words presented in white noise at four different SNRs, spoken by two speakers unfamiliar to either group, and had to provide the English translation. The results showed that the multiple-talker group performed this test more accurately and also faster. The authors concluded that by hearing a word from multiple speakers, listeners may form more robust lexical representations. This study also shows that it may not only be the frequency of encounter with a word that matters but also the contextual diversity of encounters that influences lexical representations.

To recap, more precise lexical representations may receive more activation from a distorted speech signal. Higher activation of the target word may then result in more efficient inhibition of competitor words. In many models of word recognition, segmentation of the speech signal is achieved by boosting activation of words matching the speech signal. For example, in TRACE (McClelland & Elman, 1986) the speech signal activates sub-lexical units that then send activation to those lexical units they are connected to. Because many words will temporarily match the speech signal, many words receive activation in parallel with the ones matching the speech signal best receiving the most activation. However, speech segmentation is achieved not only by activation but also inhibition. The more activation a word receives, the more it inhibits competitor words. Thus having less precise phonological representations may also result in less efficient inhibition, which in turn would make it more difficult for the parser to settle on the

correct segmentation of the signal. For example, going back to the sentence given in 1.1, a listener who hears *The catalogue in a library* may initially parse the signal as *the cattle*. However, as the speech signal unfolds further, *catalogue* would receive more activation than *cattle* and thus send inhibition to *cattle*. If inhibition is less efficient, a listener will be led down a garden path longer and will take longer to recover from it.

Experimental studies have provided some evidence for the link between less precise phonological representations and less efficient inhibition of competitors. In a study using an artificial lexicon similar in design to the one described in section 1.4.1, Magnuson et al. (2003) had participants learn mappings between novel words and arbitrary shapes. Because the researchers used an artificial language, they could tightly control phonological similarity between words. The study took place over two days and consisted of a learning phase and a testing phase using the visual world paradigm. Each word in the artificial lexicon had an onset competitor and a rhyme competitor. For example, the word *pibo* had the onset competitor *pibu* and the rhyme competitor *dibo*. In the testing phase on the first day, both rhyme and onset competitor effects were present. However, rhyme effects were larger on the first day compared to the second day. This suggests that the inhibition of competitor items had become more efficient with increased training.

An alternative way of thinking about the development of competition between words comes from Norris and McQueen's (2008) model of spoken word recognition called Shortlist B, which is built on Bayesian principles. Simply put, as the speech signal unfolds, the model evaluates the probability for a specific word in the lexicon given the evidence from the perceptual input and the prior probability of that word occurring in the language (based on subjective frequency of words but also more local factors such as context). In this model, words

do not inhibit each other directly but if the probability of a certain word increases, the probability of competitor words necessarily decreases. With newly learned or very infrequent words, there may be less certainty that the perceptual input refers to this word. Going back to the Magnuson et al. study, after only a few exposures to *pibo* and *dibo*, the prior probability of either word will be low given the perceptual evidence. Thus there will be more competition from similar sounding words.

Regardless of whether we think of word recognition in terms of interactive activation models or Bayesian models, more precise lexical representations will result in a more efficient parsing of the speech signal. In addition, speech perception studies have shown that listeners rely to a great extent on lexical cues (top-down word knowledge) and only to a lesser degree on sublexical cues (bottom-up cues from the signal) when segmenting the speech signal. Examples of sublexical cues to word boundaries are stress, biphone probability, and coarticulation. In a series of experiments, Mattys, White, & Melhorn (2005) tested those cues against each other and found that lexical cues (i.e., lexicality of the preceding segment) were the cues most strongly used by listeners to segment the speech. Sublexical cues only received greater weight when lexical cues were not informative or when the speech signal was severely distorted. One possible hypothesis that emerges from these results is that better lexical knowledge, that is, more precise lexical representations, will result in better segmentation of the signal. This hypothesis received some evidence from another study conducted by Mattys and colleagues. Mattys, Carroll, Li, and Chan (2010) compared native English speakers to native Cantonese speakers who had learned English as a second language and had attained advanced proficiency. The authors found that the L2 speakers relied more on sublexical cues for speech segmentation compared to the native

English speakers. This suggests that when language proficiency is relatively low, top-down information, that is, lexical knowledge, will play a smaller role in speech segmentation.

This section considered three mechanisms through which the quality of lexical representations may influence word recognition. High quality memory representations, defined as representations that contain more redundant phonetic information about a word, may receive more activation, exert more inhibition on competitor words, and provide better cues to speech segmentation compared to low quality representations. It is important to note that the quality of lexical representations will differ within speakers, as a function of frequency of encounter of individual words, and between speakers, as a function of overall language experience.

1.4.5 Word predictability

When listening to speech in noise, listeners make use of sentence context to compensate for a degraded speech signal. As a result, target words embedded in sentences with predictable contexts are recognized with greater accuracy compared to the same words in unpredictable contexts. The Speech in Noise test (SPIN; Bilger et al., 1984), for example, uses sentences in which the target word, the last word in the sentence, is either predictable or not from the preceding context (compare low predictability: Ms. Brown might consider the coast; with high predictability: The boat sailed along the coast). When the target is predictable, listeners can use top-down information (their knowledge of the world) and are therefore less reliant on the perceptual information from the speech signal. With regard to bilingual SUN, research suggests that bilingual speakers may not make as much use of a predictive context as monolinguals (Bradlow & Alexander, 2007; Mayo et al., 1997) with the effect being modulated by age of acquisition of the second language (Shi, 2010). It may be that second language speakers in general form weaker expectations during language processing in the second language (Martin et

al., 2013; but see Gollan et al., 2011) or that listening in noise consumes more attentional resources so that fewer resources can be devoted to exploiting a predictive context. This question will be revisited in section 1.5.1 on working memory.

1.5 Speech perception in noise and cognition

Section 1.4 dealt with the influence of linguistic factors on SUN. This section will consider how other cognitive variables contribute to individual differences in SUN. Early perception studies were often conducted and interpreted under the assumption that speech is special (e.g., Liberman & Mattingly, 1989), that is, the speech perception system is separate from other cognitive functions in the brain. More recently some researchers have adopted the view that speech perception and word recognition may also depend on domain-general cognitive, nonlinguistic resources (see Arlinger, Lunner, Lyxell, & Pichora-Fuller, 2009; Holt & Lotto, 2008; Mattys et al., 2012). For example, Mattys and Wiget (2011) tested the effect of cognitive load, operationalized as a visual search task, on phoneme identification. Participants heard an ambiguous phoneme on a /g-k/ continuum in contexts such as /?ift/ or /?iss/ that favored either a /g/ or /k/ response. The typical finding in this paradigm is that listeners give more /g/ responses in a /?ift/ context and more /k/ responses in a /?iss/ context. When participants concurrently performed the visual search task, this effect increased, suggesting that participants relied more on lexical knowledge than fine phonetic detail as a result of the greater task demands. This finding shows that perception is not impervious to cognitive load.

Other researchers have adopted a correlational approach and compared performance on cognitive tests, for example working memory, with performance on SUN tests (see, e.g., Akeroyd, 2008). Because I also adopted a correlational approach for the research described in

this dissertation, in the following sections I will concentrate on working memory and attentional control, which are thought to be components of executive functions (e.g., Miyake et al., 2000).

1.5.1 Working memory

The most influential model of working memory (WM) is that of Baddeley (Baddeley, 1992, 2012). Baddeley and colleagues proposed that WM is a multi-component construct, originally consisting of a central executive, a phonological loop, and a visuo-spatial sketchpad. It was thought that phonological processing and visual processing was done in different systems and that attention resources were of limited capacity that could lead to attentional overload when the capacity was exhausted. Later Baddeley added another component, the episodic buffer, and gave a greater importance to the interaction of WM with long-term memory. The research in the field of WM has been very fruitful with thousands of articles appearing since the first model was proposed by Baddeley and Hitch (1974). The focus in the review is not on a specific model of WM but on individual differences in WM and how these relate to outcomes on a range of different tests. Research in individual differences in WM started with a seminal paper by Daneman and Carpenter (1980). They developed the concept of reading span, which they determined by having participants read sentences with set sizes of 2 to 6 sentences. After each set of sentences, participants were asked to recall the last word of each sentence. The number of words that a participant could successfully recall was their reading span. The test was based on the logic that WM consists of storage and processing capacity. Reading sentences required participants to process them for meaning while remembering the last word tapped into storage. Surprisingly, an individual reading span was highly correlated with their verbal SAT score ($r=.59$), and passage comprehension (fact retrieval, $r=.72$, and pronoun reference, $r=.9$).

WM is not independent of long-term memory (LTM); rather, the two memory systems interact during processing (Cowan, 1993). Evidence for this hypothesis comes from studies showing that short-term memory (STM) for words is influenced by lexical and semantic variables such as frequency, familiarity, phonotactic probability, and imageability (e.g., Hulme, Maughan, & Brown, 1991; Roodenrys, Hulme, Alban, Ellis, & Brown, 1994). For example, high frequency words are remembered better than low frequency words. The fact that these variables influence recall of words suggests that LTM representations must become active. To account for these data, Hulme and colleagues hypothesized that “word frequency influences the redintegration of partially decayed traces retrieved from a short-term store” (Hulme et al., 1997, p. 1227). The same authors assume that the effect of word frequency manifests itself via more accessible or better-specified phonological representations of those words compared to low frequency words. Baddeley (2012) states that “the phonological loop, the simplest component of WM, is likely to depend on phonological and lexical representations within LTM as well as procedurally based language habits for rehearsal” (Baddeley, 2012, p. 18). Cowan’s (1999) model of WM differs from Baddeley’s in that Cowan does not assume a separate STM storage system. Rather, information in STM differs from LTM in the state of activation. Capacity limits in Cowan’s (1999) model arise from attention limits rather than storage limits since LTM is believed to be of unlimited capacity.

Researchers differ in their view of WM as being either domain general or depending on domain specific storage capacity. Conway and his collaborators, for example, view WM as a general capacity store that can hold information of any kind (see Conway et al., 2005). Conway and Engle showed that not only is reading span correlated with measures of verbal aptitude but also operation span, a measure derived in a similar way as reading span but requiring

mathematical operations as the processing component (Conway & Engle, 1996). Kane et al. (2004) also ascribe to a general capacity theory, hypothesizing that WM primarily consists of a domain-general executive attention component and a only secondarily of a domain-specific storage component. By taking a psychometric approach, the authors tested a large number of participants on a wide range of tests thought to tap into different WM components and other constructs such as fluid intelligence. Based on model comparisons, they concluded that a one-component model of WM fits their data best, that is, different WM tests such as verbal and spatial WM all shared a common variance. Kane et al. (2004) took this as evidence for the hypothesis that WMC is domain general.

Others have posited that individual differences in verbal working memory, as measured for example, by the reading-span test, depend on language experience plus differences in biological factors (MacDonald & Christiansen, 2002). In this latter view, a reading-span test and a text comprehension test correlate because both rely on language skill, which develops with language experience. The difference between MacDonald and Christiansen (2002) on one side and Conway, Kane, and colleagues (1996; 2004) on the other side may just be one of focus. Whereas the first focuses on the storage component, the latter focuses on the executive attention component. However, it seems that the differences are more fundamental, with MacDonald and Christiansen (2002) stating that “capacity is an intrinsic part of the language comprehension system, not a separately modulated resource” (p.50). Assuming that WM tests measure domain-general attention limits and domain-specific storage limits, one question arises regarding individual differences research in SUN: what component we are looking at when we observe correlations between WMC and SUN tests? It is beyond the scope of this dissertation to give a

definite answer to this question; however, we need to keep this issue in mind and I will come back to it in the discussion.

One test that is often used to assess phonological STM, which corresponds to the phonological loop in Baddeley's model (Baddeley, 1992), is the nonword repetition test. In this test, participants are asked to repeat nonwords of varying lengths. Nonword repetition has been shown to be a good predictor of vocabulary growth in children (S. E. Gathercole & Baddeley, 1989) and adults (Baddeley, Gathercole, & Papagno, 1998). However, nonword repetition ability is not independent of language experience. For instance, nonword repetition accuracy is related to the phonotactic probability of the nonword, that is, how probable its phoneme sequence is in the participant's L1 (Majerus, Linden, Mulder, Meulemans, & Peters, 2004). In one study (Majerus et al., 2004) participants were exposed to an artificial language with certain phonotactic rules. In a subsequent nonword repetition test following the brief exposure, participants were better able to remember nonwords that were in agreement with the phonotactic pattern of the artificial language compared to those that violated the phonotactic pattern. This suggests that phonotactic sensitivity emerges through language experience. To investigate this hypothesis further, Edwards, Beckman, and Munson (2004) tested children between 3 and 8 years and adults with a nonword repetition test in which they manipulated the phonotactic probability. They found that high-frequency sequences were repeated with greater accuracy than low-frequency sequences. In addition, there was an effect of age such that accuracy increased with age and an interaction between age and frequency, showing that the frequency effect, the difference between high and low frequency sequences, decreased as a function of age. Importantly, expressive vocabulary size explained 29% of the variance in accuracy scores after accounting for age effects. Although these results do not allow causal inferences, they imply that nonword repetition

ability is strongly related to language experience and specifically vocabulary knowledge. In the view of Edwards et al., listeners induce more generalized, abstract representations of sequences from the phonological patterns of words that they have encountered and learned. This will help the listener to quickly access similar patterns in other words and the fine-grained phonological knowledge becomes more precise as more instances of a patterns are accumulated (Edwards et al., 2004, p. 434). More evidence for the view that nonword repetition ability improves as a result of an individual speaker's experience with a language comes from a recent study by Parra, Hoff, and Core (2011). The authors tested a sample of English-Spanish bilingual 22-month-old children who had been exposed to both languages from birth. Phonological STM in each language, as measured by a nonword repetition test, was related to the relative exposure of children to English and Spanish. Children's exposure to English was positively correlated with their English nonword repetition score and negatively with their score for Spanish-like nonwords. Together these results suggest that both vocabulary and phonological STM develop as a function of language experience. Thus, verbal WM, of which phonological STM is a component, is not independent of language experience but is dependent on the quality, or precision, of phonological representations in LTM (Acheson, Hamidi, Binder, & Postle, 2011). The relationship between phonological STM and vocabulary acquisition therefore seems to be interactive rather than unidirectional (Thorn & Gathercole, 1999). A larger vocabulary is associated with better phonological STM and a better STM is associated with more efficient vocabulary acquisition (Gupta & Tisdale, 2009).

In section 1.3.4, I argued that phonological representations are, on average, less precise in bilinguals as a result of their reduced language experience. Given the relationship between phonological representations and verbal WM, it comes as no surprise that bilinguals often score

below monolinguals on tests of verbal WM. Studies have shown that verbal WM is usually better in an L1 than an L2 (e.g., Service, Simola, Metsänheimo, & Maury, 2002) and even highly proficient bilinguals may have poorer verbal WM compared to monolingual speakers, while visual WM is not affected (Luo, Craik, Moreno, & Bialystok, 2013). Thus, bilinguals do not seem to be impaired in general WMC but memory processes that rely on LTM representations will be less efficient. As described above, decaying phonological representations in STM are thought to be restored (redintegrated) from their LTM representations (Hulme et al., 1997). The bilingual disadvantage on verbal WM tests may therefore have the same underlying cause as the word frequency effect on tests of serial recall. Low frequency words are recalled less accurately than high frequency words, so if all words are of lower experienced frequency in bilinguals, performance on verbal memory tests should be commensurate with a bilingual's language experience in each language.

1.5.2 Working memory and Speech perception in noise

Recent studies have highlighted the role of WMC in SUN (for a review see Akeroyd, 2008). Several studies have investigated the role of WM in the encoding and recall of acoustically degraded speech. Brännström, Zunic, Borovac, and Ibertsson (2012) found a positive correlation between listeners' working memory span and the acceptable background noise level in their study. Pichora-Fuller, Schneider, and Daneman (1995) administered a verbal working memory test in which participants listened to sentences of variable length of which they had to remember the last word. The authors found that a SNR of +8 dB did not affect recall compared to the quiet condition; however, for SNRs of +5 and 0 dB, they found an interaction between set size and noise level showing that participants were able to recall fewer words as the SNR decreased. Piquado, Cousins, Wingfield, and Miller (2010) presented participants aurally

with word lists and found that when a word was masked so that it was just above the perceptual threshold, recall of that word and the preceding words was impeded. Ljung, Israelsson, and Hygge (2012) administered a WM test and in addition presented participants with word lists under different SNRs that they repeated immediately and also recalled later. The authors found that whereas individual differences in WM did not predict speech intelligibility in noise (i.e., word recognition), there was an interaction between WMC and SNR for the delayed recall, showing that recall was affected by SNR in low but not in high-span individuals. Tamati, Gilbert, and Pisoni (2013) tested a large sample with a SUN test and then asked those participants who fell into the upper or lower quartile to come back for additional testing. The authors found that backward-digit and forward digit spans of participants in the upper quartile were significantly longer than the digit spans of participants in the lower quartile group.

Lastly, Obleser, Wöstmann, Hellbernd, Wilsch, and Maess (2012) investigated the effects of memory load and acoustic degradation on WM by looking at behavioral and neuroimaging data. They presented participants with 2, 4, or 6 digits in one of three levels of degradation (voice-vocoded speech with 4, 8 or 16 frequency bands). After a brief pause, participants were then presented with one digit and had to decide whether it was in the list or not. The authors found that while accuracy was high (above 90%), there was an effect on RTs. Both larger set sizes and higher levels of degradation resulted in longer RTs. In addition, the authors investigated alpha oscillations, a frequency band associated with WM load, during the retention interval (the interval between encoding and recall) and found that alpha power did not only increase as a function of set size but also as a function of acoustic degradation. This finding suggests that the rehearsal of degraded verbal stimuli in WM is associated with extra effort and fits well with the ELU model that I will describe in the next section.

Underlying these accounts is a limited capacity view of attentional resources. When attentional resources are taken up by decoding and processing degraded speech, fewer resources are available for other mental processes such as rehearsal. However, as outlined in section 1.5.1, verbal WM tests do not only measure attention capacity available to a subject but also correlate with lexical knowledge. It is therefore not clear whether the storage or the attention component of verbal WM tests, or both components share variance with speech processing in noise when lexical knowledge is not controlled for. Two studies are suggestive of this hypothesis. Kilman et al. (2014) administered an English and a Swedish SUN test, an English and a Swedish WM test (reading span), and an English proficiency test to Swedish native speakers. The English SRT correlated more strongly with English proficiency than English reading span, and Swedish reading span was even more weakly associated with the English SRT. Sörqvist, Hurtig, Ljung, and Rönnerberg (2014) tested Swedish native speakers who had learned English as an L2. Participants performed an English reading proficiency test, a WM test in L1 and L2, and an English listening proficiency test with three different reverberation times. L1 and L2 WMC and L2 proficiency correlated highly with the listening test results. For further analysis, the authors ran a regression model in which they used the listening test results with the longest reverberation time as the dependent variable and the results from the shortest reverberation time as control variable. When L2-WMC was entered in a next step, it was significant but when L2 reading proficiency was entered, WM was no longer significant.

Results from these two studies show that WM and language proficiency may not independently contribute to SUN. The fact that WM was no longer predictive in Sörqvist et al. (2014) does not necessarily mean that individual differences in general WMC did not contribute to listening under reverberation. However, if language proficiency and verbal WM predict SUN

because both are indicative of the quality of phonological representations, then it may be difficult to disentangle the contributions of verbal WM and language proficiency. Using a nonverbal WM test or a composite score based on more than one WM test may be necessary to gauge the unique contribution of WMC during SUN. It may also be that individual differences in WMC become more important in older listeners.

The degree to which individual differences in WMC are predictive of SUN may also depend on the type of SUN test used. For example, the Words-in-Noise (WIN) test (Wilson, Carnell, & Cleghorn, 2007) has few attentional demands as listeners only have to repeat single words that are preceded by the carrier phrase “Say the word”. In the SPIN, on the other hand, target words are embedded in sentences with predictive and unpredictable contexts and so the onset of the target word is less predictable. In addition, if participants want to use the semantic context to predict the last word, they need to maintain representations of the preceding words in STM. Thus the test places greater attentional demands on the listener. I mentioned above that second language speakers and early bilinguals may not benefit as much from a predictive context as monolingual speakers (Bradlow & Alexander, 2007; Mayo et al., 1997; Shi, 2010). This may be because processing a sentence takes up more attentional resources when phonological representations are less precise. Thus listening in noise may take up more attentional resources in bilinguals and L2 speakers so that they have fewer resources left to predict the target word. In the next section I will describe a model that brings together WMC, quality of lexical representations, and SUN.

1.5.3 The Ease of Language Understanding model

The ELU model (Rönnberg et al., 2013; see Introduction) was developed to describe the interplay of bottom-up (the perceptual input) and top-down (lexical knowledge, WM) processes

during language processing. The broader context of the model is that of Cognitive Hearing Science (e.g., Arlinger et al., 2009), which developed out of a realization that domain-general higher-order cognitive processes interact with perceptual processes and therefore speech perception cannot be studied separately from the rest of the cognitive sciences.

The model assumes that sublexical information at the level of the syllable is buffered in a temporary storage system called RAMBPHO (rapid, automatic, multi-modally bound phonological representations). These syllabic units are then compared to phonological representations in LTM. The model assumes that phonological representations consist of multiple attributes and for successful lexical access the speech signal has to activate a minimum number of attributes. If the threshold for lexical retrieval is not reached, similar sounding words may be retrieved instead. However, contextual information may often be sufficient for a lexical item to be retrieved even when the speech signal is too distorted. In such cases when information in RAMBPHO cannot be matched with a LTM representation, explicit processing that involves WM is needed to resolve the mismatch, causing a delay in lexical access. Otherwise lexical access occurs automatically.

Mismatches between the speech signal and LTM representations can occur for speaker external reasons (e.g., distorted speech or an unfamiliar accent) or internal (imprecise phonological representations; Rönnberg et al., 2013, p. 3). According to this model, listening in noise will take up more attentional resources than listening in quiet because the perceptual information will often be too distorted to be effortlessly matched to LTM representations. Individual differences in WMC relate to SUN because individuals with greater WMC are thought to resolve mismatches with greater ease (c.f. Pichora-Fuller et al., 1995). At the same

time, the model predicts greater processing effort for individuals with less precise phonological representations in LTM, for example, second language learners.

The role of attention in language processing has recently been highlighted in a brain imaging study. Wild et al. (2012) tested participants with a complex task that required them to attend to one of three simultaneously presented stimuli, namely aurally presented sentences, auditory distracters, or visual stimuli. The intelligibility of the sentences was manipulated by reducing the spectral information present in the signal. The results showed that when participants heard undistorted sentences while attending to the auditory or visual information, frontal regions associated with speech comprehension showed activation and participants were later able to recall information from these unattended sentences. However, when unattended sentences were distorted, activation of frontal regions was not greater than in the control condition (unintelligible sentences) when participants were instructed to attend to the distracters. In addition, the level of activation of frontal regions correlated with the degree of acoustic distortion when participants attended to the sentences. Activation of auditory cortex, on the other hand, was not modulated by attention. The results from this study fit well with the ELU model in that processing of clear speech seemed to be effortless and not dependent on top-down attention; distorted speech, on the other hand, required attention for it to be processed and remembered. In addition, the finding that speech intelligibility was correlated with activation of frontal regions fits well with the ELU models emphasis on processing effort: the greater the distortion, the greater the need for explicit processing.

1.6 Phonological quality hypothesis

The phonological quality hypothesis (PQH) will be the overarching hypothesis of this dissertation. It is derived from the literature review presented here and related to the LQH put

forward by Perfetti (2007), the weaker links hypothesis (Gollan et al., 2008; Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Gollan et al., 2002), the phonological mismatch hypothesis (Imai, Walley, & Flege, 2005), the representation quality hypothesis (Sommers & Barcroft, 2011), and the ELU (Rönnberg et al., 2013). The PQH makes further assumptions regarding the nature of phonological representations. The LQH was developed to explain individual differences in reading and is therefore not directly translatable to spoken word recognition. I will make the same general assumption as the LQH, namely that words differ in the quality of their representation within a single speaker and between speakers. as a function of frequency of encounter. With each encounter, connections between phonological and semantic representations will be strengthened. High frequency words are encountered more often and in more diverse contexts than low frequency words. Therefore, the lexical representations of high frequency words are assumed to be more precise and semantic information can more easily be integrated to extract the meaning/gist of an utterance. The ELU makes the assumption that phonological representations differ in the number of attributes with which they are stored in LTM. Thus more precise representations will consist of a higher number of attributes compared to less precise representations. More attributes will result in a better match between the acoustic signal and phonological representations and thus more efficient and robust lexical retrieval. In the PQH, I further assume that representations in LTM do not only consist of abstract phonetic information but that each encounter with a word leaves a memory trace (cf. Goldinger, 1996). The hypothesis thus builds on exemplar theories of word recognition (Goldinger, 1996, 1998; Hintzman, 1986; Pierrehumbert, 2001). Exemplar-based models of the mental lexicon differ from models that assume that only abstract representations of words are stored and that the speech signal is normalized and stripped of all indexical information (e.g., speaker voice, gender, etc.) prior to

lexical access (e.g., K. Green, Kuhl, Meltzoff, & Stevens, 1991). The exact nature of the mental lexicon is still an active area of research but exemplar theories are especially useful in the present context because they make specific predictions about the frequency of encounter with words. Lexical items that are encountered more frequently will be associated with more episodic memory traces and a match between the signal and a memory representation will be more likely.

The same logic that can explain frequency effects within speakers can also be extended to explain individual differences in word recognition between speakers. Individuals who overall have more language experience will encounter all words, and especially low frequency words (Kuperman & Van Dyke, 2013), more often compared to other individuals. For example, some people may interact with a greater variety of people and in a greater variety of contexts. This is especially true for speakers of two or more languages who, on average, will spend less time listening to and speaking each language (for an application of exemplar theory to nonnative speakers see Hardison, 2003).

Whereas bilinguals may be able to estimate quite accurately what percentage of the time they speak and hear each language on average, it is certainly much harder for monolingual speakers to reliably estimate the number of hours they listen to language, how many speakers they interact with regularly, and the type of contexts in which they encounter language. Therefore the assumption is made in the present study that language experience is closely related to verbal ability or language proficiency (these terms will be used interchangeably throughout this dissertation). Language proficiency can easily be measured by a standardized test. Such tests were developed by testing a large sample representative of the general population and have high reliability. Standardized tests are thus preferable to data based on self-report. Proficiency is related to experience because individuals who are generally exposed to language more and

interact with more people are more likely to hear words more often, especially less frequent ones, than individuals who have fewer interactions.

CHAPTER 2: EXPERIMENT 1

2.1 Research questions and predictions

This experiment was designed to test the effect of noise, predictability and language status (bilingual/monolingual) on word recognition in noise. Secondly, I tested the influence of lexical and sublexical variables on word recognition in noise and whether the effect of these variables is different for monolinguals and bilinguals. Results could provide evidence for or against the hypothesis that the previously reported bilingual disadvantage in SUN is related to a bilingual's generally reduced exposure to each of their languages compared to someone who speaks only one language. As discussed in Section 1.4.3, the word frequency effect is a result of language exposure. Words that are more frequent in the language are encountered more often and are therefore recognized faster and with greater accuracy. As I described in Section 1.4.1, phonological representations of high frequency words are assumed to be more precise, including more redundant information, compared to low frequency words and thus recognition of high frequency words is more robust to the effect of background noise. If reduced exposure to each language due to bilingualism is one factor underlying the bilingual disadvantage in SUN, then differences between groups are expected to be disproportionately larger in the low frequency range because of the logarithmic nature of frequency effects (see Section 1.4.3)⁵. Thus, we would expect a frequency by group interaction. Originally I also intended to investigate the effect of neighborhood frequency on word recognition as previous studies have found this variable to be a good predictor of word recognition in noise (Luce & Pisoni, 1998). However, many words in the present study behaved differently than would have been expected based on their neighborhood

⁵ As described in Section 1.4.3, an individual with less exposure to the language that he or she speaks will encounter low frequency words disproportionately less often compared to someone with more language exposure.

frequency. This made the results difficult to interpret and so only the results for lexical frequency are reported here.

2.2 Methods

2.2.1 Participants

The study includes 53 monolingual and 48 bilingual participants. The inclusion criteria for monolinguals were that they did not learn a second language before 10. Some monolinguals had learned a second language in foreign language classes in school but they were not fluent in their second language and had not spent more than a short vacation in a non-English speaking country. Bilinguals had to have learned Spanish from birth and English before the age of 8. Four bilinguals reported to have learned English later than 8 but they were included in the study because they were born in the US and attended school in the US from kindergarten. They reported that they attended a Spanish-English bilingual program but that little English was taught. However, they likely had some exposure to English. Thirty-seven (77%) bilinguals were born in the US. Of the remaining bilinguals, all but five arrived in the US before the age of 6. Four of those immigrated at the age of 7 and one at the age of 13. The latter participant was included because her mother was a native speaker of English and she had learned both English and Spanish from birth and attended a bilingual school. In addition, participants had to be between 18 and 35 years old. I tested six additional monolinguals and five additional bilinguals but they were not included in the final sample because they did not meet the definition of monolingual (5), early bilingual (4), or were too old (1) or too young (1) to be included in the study. Detailed participant information can be seen in Table 1.

The bilingual participants were mostly second-generation immigrants from Mexico. As might be expected in this population (Capps et al., 2005), the parental education level was lower

than that of the monolingual group. Participant groups also differed on other variables, most notably English proficiency. I will come back to this point in the results section. Importantly, participants were matched on age, years of formal education, and self-rated hearing ability.

Table 1. *Participant characteristics divided by language status.*

	Monolingual	Bilingual
Age in years	20.6 (2.4)	20.8 (2.8)
Number of males	18 (34%)	16 (33%)
Years of formal education	14.9 (1.6)	14.4 (1.4)
Primary caregivers education level:		
- Less than high school	0%	40%
- High school	11%	46%
- Some college	30%	8%
- College	32%	4%
- Some Graduate school	4%	0%
- Graduate school	23%	2%
Self-rated hearing ability (out of 10)	8.6 (1.0)	8.6 (1.1)
Years of musical experience	4.7	1.0
Oral language W-score	533.2 (8.9)	515.6 (11.4)
Oral language Standard Score	105 (7.7)	90 (8.8)
Picture Vocabulary W-score	537.1 (11.0)	516.1 (13.5)
Picture Vocabulary Standard score	101 (7.6)	86 (8.4)
Verbal Analogies W-score	529.5 (9.2)	515.3 (11.8)
Verbal Analogies Standard Score	109 (7.3)	98 (9.0)
Oral language W-score - Spanish	-	503.0 (11.9)
Oral language Standard-score - Spanish	-	81 (9.3)
Picture Vocabulary W-score – Spanish	-	500.8 (11.8)
Picture Vocabulary Standard score – Spanish	-	77 (7.9)
Verbal Analogies W-score - Spanish	-	505.3 (14.2)
Verbal Analogies Standard Score –Spanish	-	90 (10.8)
Age of Acquisition: English	-	4.4 (2.5)
Age of Acquisition: Spanish	-	0
Age of Arrival in USA	-	1.3 (2.8)
Listening to English	-	64.6% (18.4)
Speaking English	-	65.5% (17.4)
Reading English	-	81.3% (16.7)

Participants were recruited through flyers. Monolinguals and bilinguals were tested at a large rural university in Michigan and additional bilinguals were tested at a large urban university in Illinois. Most bilinguals tested in Michigan were originally from Texas whereas those tested in Illinois were mostly from Chicago.

2.2.2 Materials

2.2.2.1 Background questionnaire

Participants' background information was collected with a questionnaire created for this study, administered by the experimenter. The instrument was loosely based on Marian et al. (2007) but included additional information about parental education and use of English and bilingual participants' use of English and Spanish during their childhood and adolescence. It took about 6 to 10 minutes to administer. The questionnaire can be seen in the Appendix.

2.2.2.2 Speech perception in noise test

The revised Speech Perception in Noise test (SPIN; Bilger et al., 1984) was used in a modified form. The test consists of 200 target words and each word is recorded in a predictive and unpredictable context. For example, the word *coast* could be preceded by *Ms. Brown might consider the coast* (low predictability) or by *The boat sailed along the coast* (high predictability). The original SPIN recordings were obtained on CD and were cut in Audacity so that each sentence could be saved in a separate sound file. For the background babble, a short sequence from the original track was chosen and mixed with each sentence in Praat (Boersma & Weenink, 2014) at two different SNRs (-2 dB and 3 dB). The sound intensity of the sentence was held constant and so the intensity of the babble differed for the two SNRs.

In the present study, 128 sentences from the test were chosen and divided into four lists of 32 words. Words in each list were matched on lexical frequency, based on subtitle frequencies

from Brysbaert and New (2009), and on neighborhood density. Each participant heard the first half at 3 dB SNR and the second half at -2 dB. Within each SNR, half of all words were played in a predictable context and the other half in an unpredictable context in a randomized order. Across all participants, each word was administered in all four conditions in a Latin-square design. After each sentence, the participant was prompted to type the last word of the sentence. The next trial started when a participant pressed Enter. Before the actual experiment, 10 sentences were administered at a SNR of 8 dB to ensure that participants had understood the task. Participants were also told to check the word they typed on the screen for any spelling errors before going to the next trial. This test was administered in Eprime 2.0 (Psychology Software Tools, Sharpsburg, PA).

Information about lexical variables was taken from different sources. Information about lexical frequency was taken from Brysbaert and New (2009). These norms are based on a large corpus created from subtitles of American movies and TV shows. The mean $\log_{10}$ word frequency of the stimuli used in the present study was 2.71 ($SD = 0.45$) and the mean FpM was 17.63 ($SD = 25.40$). Information about phonotactic probability came from Vitevitch and Luce (2004). This database provides the summed probabilities of each phoneme in a word and the summed probability of each biphone. The correlation between biphone probability and log-frequency was $r = .14$.

2.2.3 Procedure

In this experiment, participants heard recordings of spoken sentences. After each sentence, the participant was prompted to type the last word of the sentence. The next trial started when a participant pressed Enter. Before the actual experiment, 10 sentences were administered at a SNR of 8 dB to ensure that participants had understood the task. Participants

were also told to check the word they typed on the screen for any spelling errors before going to the next trial. This test was administered in Eprime 2.0 (Psychology Software Tools, Sharpsburg, PA).

2.3 Analysis

For the analysis, mixed-effects regression modeling was used (Baayen, Davidson, & Bates, 2008; Gelman & Hill, 2007). Models were run in *R* (R Core Team, 2014) using the package *lme4* (Bates, Maechler, Bolker, & Walker, 2014). Mixed-effects models have the advantage over ANOVA that they allow for crossed-random effects of subjects and items. That eliminates the necessity to run separate analyses for subjects and items. At the same time, models can be run with continuous predictors as in linear regression while controlling for multiple observations from the same subject. This is an advantage over ANOVA because naturally continuous variables such as word frequency can be used as a continuous predictor and do not have to be factorized into low and high frequency items. In addition, mixed-effects models allow to test for interactions between subject-level and item-level variables. Lastly, a great advantage over ANOVA is that mixed-effects models can handle continuous and dichotomous outcome variables. In ANOVA, when variables are dichotomous, such as accuracy data, researchers usually average across subjects and use percentage correct as the outcome variable, using a transformation to correct for the typically non-normal distribution. This traditional method has certain shortcomings as described in Jaeger (2008), which can be circumvented by using a generalized mixed-effects model with a binomial error distribution. This also obviates the need of averaging data.

Interpreting the output from a mixed-effects model is somewhat different from the output of an ANOVA. To report the significance of main effects and interactions, likelihood-ratio tests

will be reported. These tests compare the log-likelihood of a model excluding a variable with one including the variable. The change in log-likelihood has a chi-square distribution with the degrees of freedom corresponding to the difference in number of variables between the models (i.e., one). In addition, I will also report the model estimates for the effect sizes of each variable. For logistic regression, these are on the logit scale and thus not easily interpretable but the logit values can be transformed into odds-ratios by taking the exponent. An odds-ratio describes the likelihood of one event occurring over another event occurring. In the present analysis there are three categorical variables, Noise (high/low), Predictability (high/low), and Group (monolingual/bilingual). For each variable, one will be the reference category, for example, Low Noise. In the model output, the regression coefficient of the variable Noise then shows the change on the logit scale when noise is high. Because the logit scale is nonlinear, the actual values are meaningless. However, the sign of the coefficient will tell us whether the probability of recognizing a word increases or decreases relative to the reference category. So if the sign of the Noise coefficient is negative and Low Noise is the reference category, we know that the probability of recognizing a word is lower when noise is high compared to when it is low. It is important to know what the intercept in a regression model represents because a coefficient of a continuous variable will show the effect size for the baseline condition, which is the intercept. For example, if the reference categories for our three categorical variables are High Noise, Low Predictability, Monolingual, then the coefficient of the intercept will give us the probability of recognizing the word in this condition. If we include an interaction between Group and Frequency, the main effect of frequency will give us the effect size for the monolingual group and the coefficient of the interaction between Group and Frequency will indicate the change in the effect size for the bilingual group. For example, if the main effect of Frequency was 5% (that

is, a one unit change in Frequency is associated with a 5% increase in recognition), and the coefficient of the interaction between Group and Frequency is 4%, then the effect size for the bilinguals is 9%. If the interaction is significant, it means that the slopes of the frequency effect for monolinguals and bilinguals are statistically different. Because the actual values will be on the logit scale, the sign of the coefficients can help us determine again whether the effect is smaller or larger for bilinguals compared to monolinguals. Instead of adding up the coefficients, we could also change the reference category for group to bilingual. The main effect of Frequency would then show us the effect size for the bilinguals. Note that this is different from doing multiple comparisons because we run the same model and just change the reference category.

The test was scored automatically by Eprime and an answer was counted as correct, when it matched the target word. All answers that were coded as incorrect were manually checked for any spelling mistakes. A misspelled word was counted as correct in the following cases: letter transposing (e.g., *theif* for *thief*), wrong letter when the correct letter was adjacent to it on the keyboard and the resulting word was not a word in English (e.g., *ahore* for *shore*), when a letter was missing and the resulting word was not a word in English, or when the answer was a homophone of the target word, regardless of whether the typed word was a real English word (e.g., *gyn* or *jin* for *gin*). In total, 286 (2.2%) instances were corrected this way, which is comparable to 2.5% in Luce and Pisoni (1998) who used a similar procedure. In three instances, participants started typing before the prompt. In this case, the letters that were typed before the prompt were not recorded. This typically resulted in very short RTs (measured from the start of the prompt to the point where a participant hit enter). For example, one participant seemingly only typed *t* for *pet* with an RT of only 435 ms (a typical RT would be between 1500 and 3000 ms). These trials were coded as missing data.

2.4 Results

For the analysis, I ran one regression model but I will report the results separately for each research question, that is, does the effect of noise and predictable context differ for each group and what is the effect of frequency and phonotactic probability on monolingual and bilingual speakers.

2.4.1 The effects of noise and predictable context

Words in low noise ($M = 85.5\%$, $SD = 35.2$) were recognized with higher accuracy than words in high noise ($M = 67.6\%$, $SD = 46.8$; $\chi^2(1) = 712.4$, $p < .001$), and words in a predictable context ($M = 88.7\%$, $SD = 31.6$) better than words in an unpredictable context ($M = 64.4\%$, $SD = 47.9$; $\chi^2(1) = 1059.3$, $p < .001$). The effect of predictability was 28.2% when noise was high and 20.5% when noise was low and this interaction was significant ($\chi^2(1) = 30.7$, $p < .001$). Monolinguals recognized words more accurately ($M = 80.8\%$, $SD = 39.4$) than bilinguals ($M = 71.8\%$, $SD = 45.0$; $\chi^2(1) = 76.7$, $p < .001$). The effect of noise was smaller for monolinguals (16.1%) than bilinguals (19.9%), but this interaction was not significant ($\chi^2(1) = 3.3$, $p = .068$, see Figure 3). The effect of predictability was larger for monolinguals (24.8%) than bilinguals (23.8%; $\chi^2(1) = 46.7$, $p < .001$). The effect of predictability can best be seen in Figure 4. When noise was low, the effect of predictability was larger for bilinguals (22.7%) than monolinguals (18.6%), likely because monolinguals were at ceiling in the low noise-high predictability condition ($M = 98.2\%$, $SD = 13.4\%$). When noise was high, on the other hand, the effect was larger for monolinguals (31.1%) than bilinguals (24.9%), but the three-way interaction was not significant ($\chi^2(1) = 0.1$, $p = .809$). Expressed as Cohen's d , the effect sizes of group differences were as follows in the four conditions: HNLP: 0.16, HNHP: 0.37, LNHP: 0.25, LNLP: 0.21. The model estimates for the SPIN test are summarized in Table 2.

Table 2. *Summary of mixed-effects regression results for variables predicting accuracy on the Speech Perception in Noise test.*

	Estimate (β)		SE
	Odds ratio	Logit scale	
Intercept (High noise, low predictability, bilingual)	1.03	0.03	0.15
Noise (high vs. low)	3.50	1.25	0.09
Predictability (low vs. high)	4.25	1.45	0.09
Group (bilingual vs. monolingual)	1.54	0.43	0.11
Noise*Predictability	2.00	0.69	0.16
Predictability*Group	2.20	0.79	0.14
Noise*Group	1.20	0.18	0.13
Noise*Predictability*Group	1.07	0.07	0.27
Frequency (scaled)	1.41	0.35	0.13
Frequency*Group	0.86	-0.15	0.05
Biphone probability (scaled)	1.22	0.20	0.13
Biphone Prob.*Group	1.09	0.09	0.06

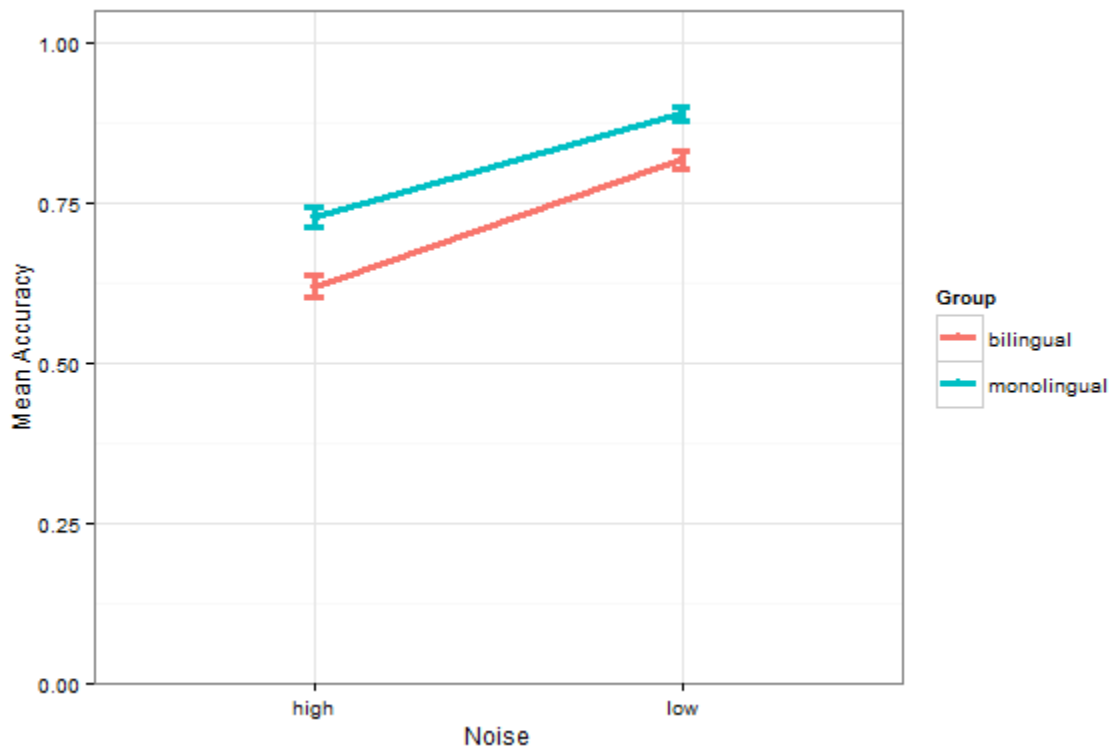


Figure 3. Results of the Speech Perception in Noise test divided by noise level and group. Error bars show the 95% confidence interval.

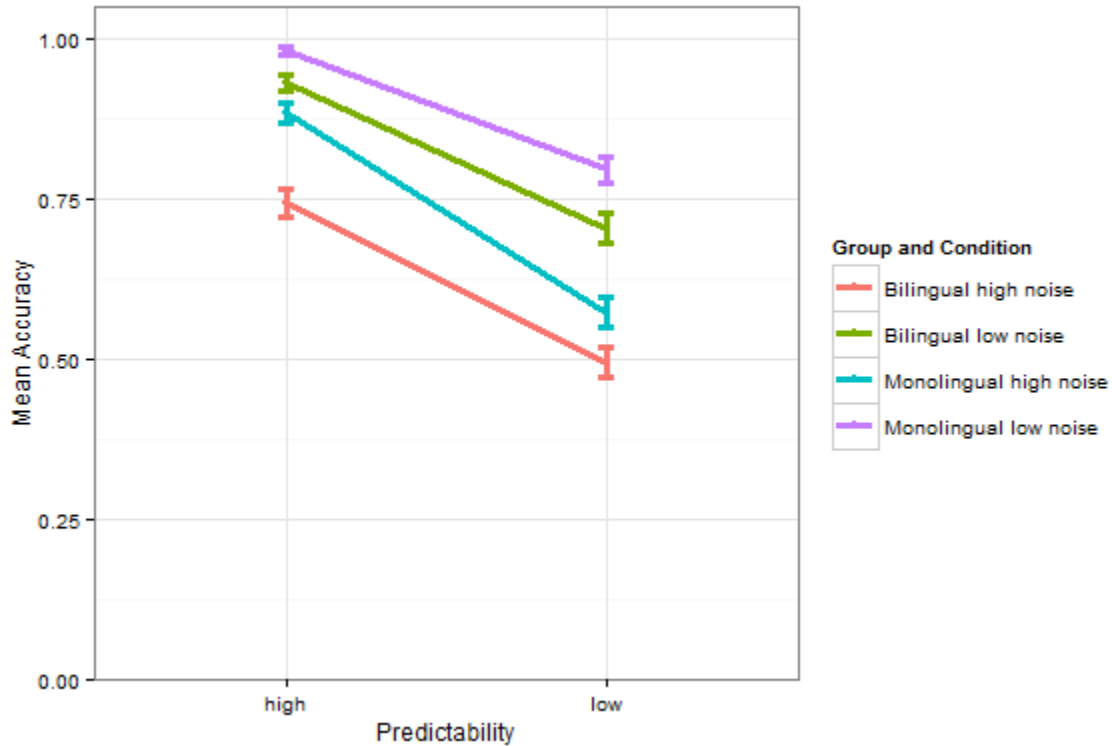


Figure 4. Results of the Speech Perception in Noise test. Results are divided by condition and language group. Error bars show the 95% confidence interval.

2.4.2 The influence of lexical and sublexical variables on word recognition

Phonotactic probability: The effect of phonotactic probability can be seen in Figure 5.

There appears to be an effect of phonotactic probability such that words with higher probability were recognized with higher accuracy than low probability words but when the variable was entered into the model along with frequency, the effect was only marginally significant ($\chi^2(1) = 3.3, p = .068$). From Figure 5 it appears that the effect was the same for both groups and model estimates confirmed this, showing that the interaction with group was not significant ($\chi^2(1) = 2.3, p = .127$).

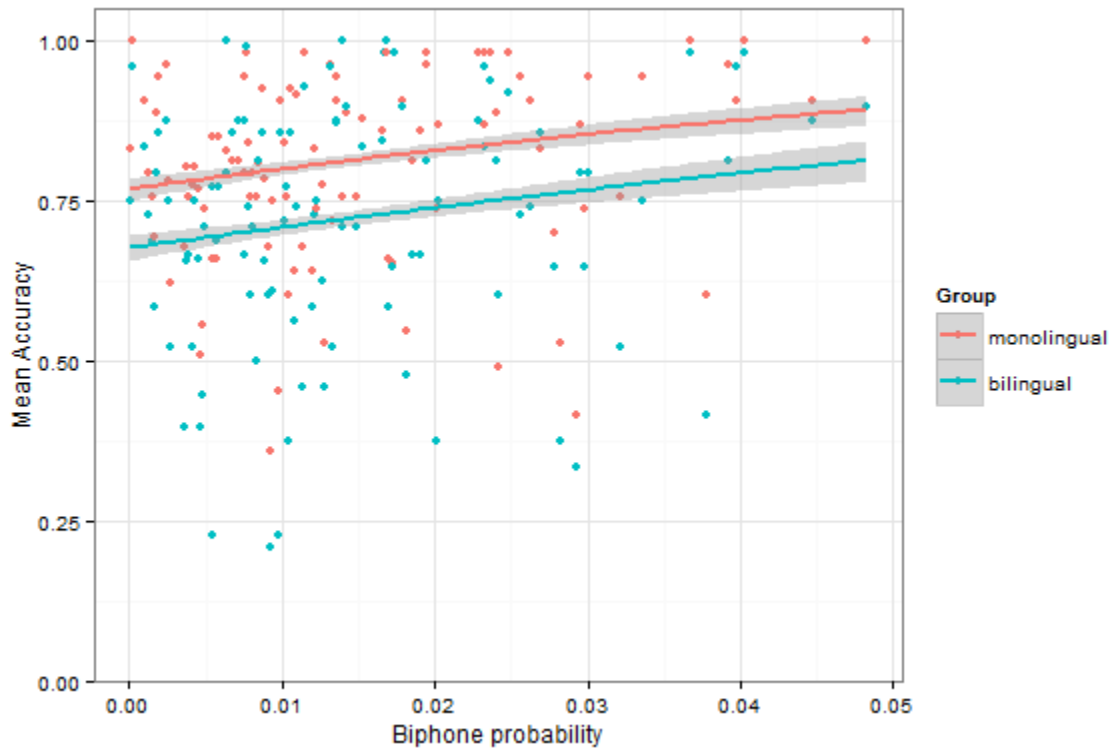


Figure 5. Effect of biphone probability on Speech Perception in Noise accuracy divided by group. Grey-shaded area shows the 95% confidence interval of the slope of the regression line. Each point represents the mean accuracy of a certain word.

Lexical frequency: The word frequency effect is shown in Figure 6. Recognition of high frequency words was better than recognition of low frequency words ($\chi^2(1) = 4.6, p = .033$) and the interaction between frequency and group was significant ($\chi^2(1) = 7.9, p = .005$). Table 2 shows a negative sign for the interaction between group and frequency, which indicates that the effect of frequency was smaller for monolinguals than bilinguals.

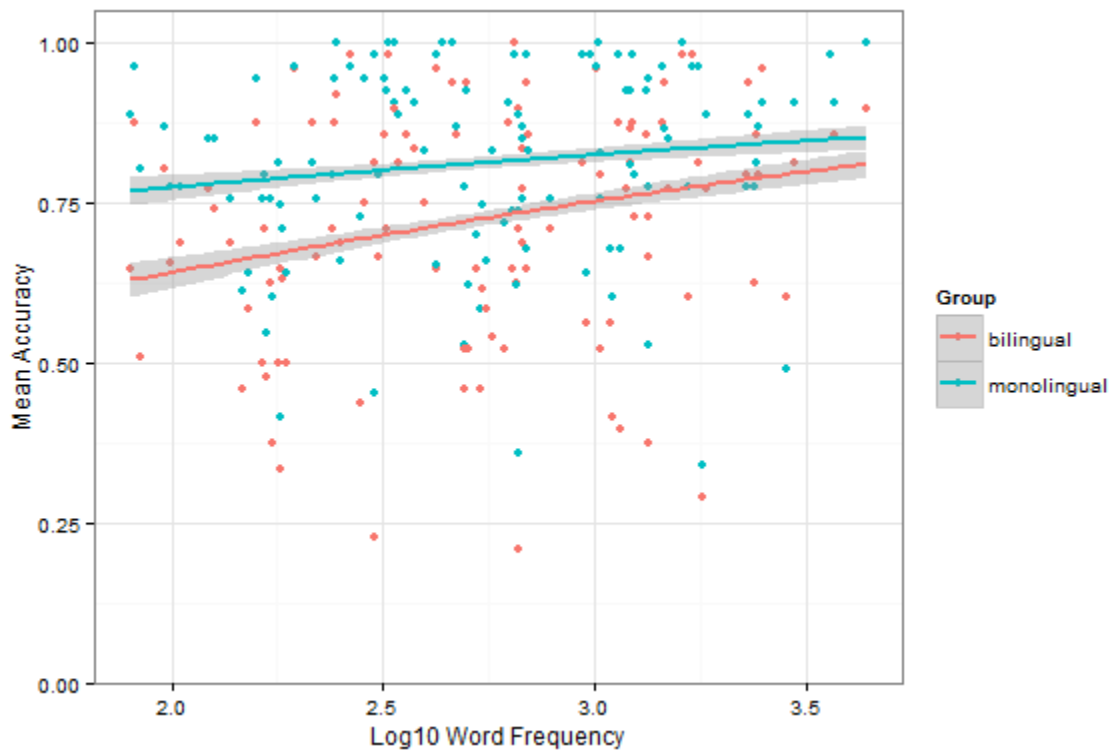


Figure 6. Effect of log10 word frequency on Speech Perception in Noise accuracy divided by group. Grey-shaded area shows the 95% confidence interval of the slope of the regression line. Each point represents the mean accuracy of a certain word.

In order to better understand the interactions reported here, I also ran separate models for each group. As can be seen in Table 3, the effect of predictability was higher for monolinguals than bilinguals. The effect of frequency was slightly larger in the bilingual group than in the monolingual group and the effect only reached significance in the bilingual group. The opposite pattern was found for biphone probability, which was only significant in the monolingual group. Although frequency and biphone probability were not highly correlated ($r = .14$), the effects are likely not independent and so the fact that each variable was only significant in one group but not the other may be the result of this correlation.

Table 3. *Summary of the mixed-effects regression results of SPIN accuracy for monolinguals and bilinguals.*

	Monolinguals			Bilinguals		
	OR	LS	SE	OR	LS	SE
Intercept	1.6	0.48**	0.15	1.0	0.01	0.14
(High noise, low predictability)						
Noise (high vs. low)	4.3	1.46***	0.09	3.4	1.24***	0.09
Predictability (low vs. high)	9.7	2.23***	0.11	4.2	1.43***	0.09
Noise*Predictability	2.1	0.74***	0.22	2.1	0.73***	0.16
Frequency (scaled)	1.2	0.20	0.14	1.4	0.34**	0.13
Biphone probability (scaled)	2.1	0.74*	0.22	1.2	0.19	0.13

*** $p < .001$; ** $p < .01$; * $p < .05$. OR = odds ratio. LS = logit scale. SE = standard error.

2.5 Discussion

The first part of the results replicated previous studies that showed that the effect of a predictable context is smaller for bilinguals compared to monolinguals (Bradlow & Alexander, 2007; Mayo et al., 1997). However, contrary to some previous studies (Mayo et al., 1997), the interaction between Noise level and Group was only marginally significant. This suggests that the effect of noise was the same for both groups.

The second research question asked whether differences between groups observed on the SPIN could be explained by reduced language exposure of the bilinguals. The assumption was that because bilinguals speak two languages, they will hear and speak each language less frequently. As a result, phonological representations of words will be weaker or less precise, and a bilingual person, on average, will know fewer words in each language compared to an age-matched monolingual person. The prediction was that if reduced language experience is the cause for less accurate word recognition, it would affect low frequency words more than high frequency words because low frequency words will be heard disproportionately less as a result of reduced exposure (see section 1.4.3). This prediction was borne out by the results. Frequency

effects were larger in bilinguals compared to monolinguals, with bilinguals performing close to monolinguals for high frequency words but different for low frequency words.

Phonotactic probability has a facilitative effect on word recognition such that words with high phonotactic probability are recognized with higher accuracy than those with low phonotactic probability and this was confirmed in the present study, although the effect was only marginally significant. An additional prediction for phonotactic probability was that bilinguals would be more negatively affected by low-phonotactic-probability words because they would be less sensitive to phonotactic patterns that are less common in the language. The prediction for the present study was based on the fact that sensitivity to phonotactic probability is a result of language experience and vocabulary knowledge (Edwards et al., 2004). Speakers extract the probabilities for phoneme sequences by generalizing across all lexical items in their mental lexicons. Because bilinguals may know fewer words, they have a smaller basis to abstract the sound patterns from and so they may be less sensitive to those sublexical units. This prediction was not confirmed in the present results. When separate models were run for each group, phonotactic probability was only significant in the monolingual group and frequency was only significant in the bilingual group. This may suggest that the monolinguals relied more on sublexical information whereas the bilinguals relied more on lexical information. However, when comparing the coefficients of these effects between groups, the effects appear to be very similar and the different significant levels may be a result of the fact that the two effects are not completely independent, that is, high frequency words also tend to have higher phonotactic probability. In addition, the interaction between phonotactic probability and group was not significant, whereas the interaction between group and frequency was highly significant. This

suggests that differences between groups were mostly present in the frequency effect, suggesting that frequency of exposure to English may be one factor affecting group differences on the SPIN.

CHAPTER 3: EXPERIMENT 2

Results of Experiment 1 replicated previous studies and provided some evidence that bilinguals are more affected by noise than monolinguals and that they benefit less from a predictive context under adverse listening conditions (i.e., high noise). The experiment also provided some evidence for the hypothesis that the cause of differences between monolinguals and bilinguals is the quality of phonological representations. According to the phonological quality hypothesis, reduced language experience is the reason for less precise phonological representations and a generally smaller vocabulary. If this hypothesis is true, we would not only expect group differences between monolinguals and bilinguals but also individual differences within each group as a result of language experience and vocabulary knowledge. Thus, the purpose of Experiment 2 was to investigate factors that could explain individual variation in the sample. Besides language experience, the influence of individual differences in aspects of cognition were investigated, namely WM and auditory attentional control.

3.1 Methods

3.1.1 Participants

The results analyzed in Experiments 2 come from the same participants as those reported in Experiment 1 (see section 2.2.1).

3.1.2 Materials

In experiment 2, the results from experiment 1 will be reanalyzed using an individual differences design instead of a group design. To assess individual differences on different dimensions, participants completed several tests that will be described in the following sections. In chapter 5, these tests will be analyzed in more detail; here, they just serve as predictor variables for the SPIN test.

3.1.2.1 Woodcock Muñoz Language Survey - Revised

The Woodcock-Muñoz Language Survey - Revised (WMLS-R; Woodcock, Muñoz Sandoval, Ruef, & Alvarado, 2005) is a norm-referenced, standardized test of English and Spanish. Both versions were normed on a large sample of speakers in the US and Latin America in the case of the Spanish version. The raw-score on the test can be transformed into a standard score with a population mean of 100 and a standard deviation of 15 through software that is provided with the test (Schrank & Woodcock, 2005). In addition, scores can be expressed as W-scores, which are based on an equal interval scale and are therefore suitable for statistical analyses and group comparisons. Unlike standard scores, W-scores are not corrected for participant age at testing.

The WMLS-R consists of seven tests, two of which were administered in the present study. The first one is called Picture-Vocabulary (PV) test. Participants are shown pictures in sets of six and are asked to name them one by one as the experimenter asks them “What is this” and points at the picture. The second administered test is called Verbal Analogies (VA). Participants are asked to solve “riddles” such as *In is to out as down is to ...?* Scores from both tests can be combined into a single score with the provided software, which the test developers call Oral Language Ability (OL). This score correlates highly with the cluster score that is based on all tests of the WMLS-R ($r = .9$). The standard error of the mean for all tests is between 5.55 and 5.93 and the internal consistency reliability coefficients were around $r_{11} = .9$ (Alvarado & Woodcock, 2005).

3.1.2.2 Test of attention in listening

The Test of Attention in Listening (TAIL) was adapted from Zhang, Barry, Moore, and Amitay (2012). In this test, participants have to decide whether two tones were played to the

same ear or different ears. What makes this test challenging is that the frequency of the two tones is sometimes the same and sometimes different. Because participants are only supposed to respond based on the location of the tones, response conflict arises on trials on which the location is different but the frequency the same or the location the same and the frequency different. The manipulation of frequency and location results in four conditions, same-frequency same-location (SFSL), same-frequency different-location (SFDL), different-frequency same-location (DFSL), different-frequency different-location (DFDL). The original test also has a second condition where frequency is the task-relevant dimension and location is the irrelevant dimension that has to be ignored. However, only the first condition was used in the present study to reduce the time needed to administer the test.

Three different measures can be derived from the TAIL, baseline RT, involuntary orientation, and conflict resolution. Baseline RT is the mean RT in the SFSL condition. In Zhang et al. (2012), baseline RT correlated with the RTs in a separate test that did not involve response conflict and therefore the authors suggested that this measure reflects information processing speed. Involuntary attention can be calculated by subtracting RTs on trials with the same frequency from those of different frequency ($[DFDL+DFSL] - [SFSL+SFDL]$). Conflict resolution can be calculated by subtracting the mean RTs on trials where location and frequency were both different or both the same (no response conflict) from those where they were different ($[SFSL+DFDL] - [SFDL+DFSL]$).

The tones were created in Praat (Boersma & Weenink, 2014) as pure tones with a length of 100 ms. The frequency ranged between 500 and 1400 Hz in 100 Hz intervals, which resulted in ten different sound files. There were a total of 96 experimental trials, 24 trials in each condition. The experiment was programmed in E-Prime.

3.1.2.3 Working memory

The WM test used for this study comes from the NIH Toolbox. The NIH toolbox is a collection of different tests in the areas of cognition, emotion, motor function, and sensation. All tests are available freely and are administered online. In the WM test, participants see pictures and their labels and hear their names. The set-size differs from 2 to seven pictures. Pictures are either animals or food items. After each set of pictures, participants are asked to repeat what they just saw in size order from smallest to biggest. For example, if they saw a bear, a duck, and an elephant, they would say duck, bear, elephant. To establish the size order, participants have to pay attention to the size of the object on the screen but in most cases, the relative proportions on the screen corresponded to real life. The test has two parts. In the first part, sets consist only of animals or only of food items. In the second part, sets consist of animals and food and participants are asked to repeat the food first from smallest to biggest and then the animals from smallest to biggest. Both parts start with two practice sets to ensure that participants understood the directions. If they made a mistake in either practice set, the instructions were repeated and the set was administered again. After the practice items, the test starts with a set size of two. If a participant correctly repeats all pictures, the set size of the next trial increases by one. If the participant makes an error, another set of the same size but different items is administered. Testing stopped when a participant could not correctly repeat two sets in a row or when the last set was administered. Responses were recorded on a paper sheet and a score for each participant was calculated by counting the total number of items of all correctly repeated sets. Thus the total score for each part is 27 ($2+3+4+5+6+7$) and the total possible score is 54. This test was only administered in English.

Recently, the reliability of the test was established (Tulsky et al., 2014). The test-retest interclass correlation coefficient was .77. The test also correlated with other established WM tests ($r = .57$) and tests of executive function ($r = .43 - .58$). The correlation with a test of receptive vocabulary, on the other hand, was low ($r = .24$). Also interesting with respect to the present study was the finding that Hispanic participants scored, on average, .41 SDs below Caucasian participants.

3.1.2.4 Consonant perception in noise

In the consonant perception test (CP), participants heard 16 different consonants in a /VCV/ cluster and were asked to identify them by clicking on one of 16 options on the computer screen. The consonant recordings came from Shannon, Jensvold, Padilla, Robert, and Wang (1999). The original recordings done by Shannon and colleagues included 25 consonants in three different vowel contexts /u/, /a/, and /i/ in medial /VCV/ and initial position /CV/. Following Garcia Lecumberri and Cooke (2006), stimuli were reduced to 16 consonants (/p b t d k g tʃ b f v s z ʃ b m n l r/) in only one vowel context (aCa) and one consonant position. Two male speakers (M2 and M3) were chosen from the original set of 5 male and 5 female speakers and each token was repeated four times for a total of 128 items. The experimental items were mixed with background noise (multi-talker babble) taken from the original SPIN recording. Three different sections from the babble noise track were cut and mixed at a SNR of -4 dB in Praat (Boersma & Weenink, 2014). One of those babble segments was repeated once and the other two were played once. The SNR was chosen based on a pilot study. Participants in a pilot study performed at about 85% accuracy at an SNR of -2 dB. To avoid ceiling effects, the SNR was lowered to -4 dB in the present study. Participants also heard each token in silence at the beginning of the experiment so they could adapt to the pronunciation of each speaker. These trials were only used

as practice trials and were not scored. When a participant made a mistake on those practice trials, the same token was repeated until the participant made a correct response.

3.1.3 Relationship between predictor variables

The predictor variables used were oral language ability, WMC, consonant perception (henceforth CP) in noise (mean accuracy), and attention. The attention test provided different variables such as baseline RT, involuntary orientation, and conflict resolution (see 3.1.2.2).

There was no clear hypothesis for which of those measures, if any, would predict accuracy on the SPIN so the analysis was exploratory and results need to be interpreted with some caution.

Accuracy on the TAIL was not considered as a variable because there was very little variation in accuracy rates. The results of these tests can be seen in Table 4.

Table 4. *Means and standard deviations of the predictor variables used in Experiment 2.*

	Monolingual	Bilingual	Total sample
Oral language ability (W)	533.2 (8.9)	515.6 (11.4)	524.8 (13.4)
Working memory	37.6 (8.0)	32.4 (7.9)	35.2 (8.3)
Consonant perception (%)	76.9 (5.4)	66.9 (9.1)	72.2 (8.9)
TAIL measures			
Attention baseline RT (in ms)	680 (125)	702 (139)	690 (132)
Involuntary attention effect (in ms)	31 (40)	19 (49)	25 (44)
Conflict resolution effect (in ms)	46 (49)	29 (43)	38 (47)

Note. Oral language ability is reported in W scores, which are on an equal interval scale with an arbitrary unit. The maximum score on the working memory test was 54. TAIL = test of attention in listening. See text for an explanation of TAIL measures.

The influence of each variable was first investigated through simple and bi-partial correlations with accuracy on the SPIN test in each condition. Because some variables were intercorrelated (e.g., WM and verbal ability), the unique contribution of working memory was investigated by partialling out the covariance shared with oral language ability. The results are reported in Table 5. The correlation between WM and verbal ability was $r(53) = .43, p = .001$, in the monolingual group and $r(48) = .47, p < .001$, in the bilingual group. Consonant perception

and verbal ability were correlated in the bilingual group only $r(48) = .55, p < .001$ (monolinguals: $r(53) = .20, p = .154$).

Table 5. *Correlations and bivariate correlations between predictor variables and the four conditions of the Speech perception in Noise test.*

	LNHP		LNLP		HNHP		HNLP	
	Mono-lingual	Bilingual	Mono-lingual	Bilingual	Mono-lingual	Bilingual	Mono-lingual	Bilingual
Picture vocabulary	.37 <.01**	.60 <.01**	.10 .46	.34 .02*	.47 <.01**	.62 <.01**	.21 .13	.40 <.01**
Verbal analogies	.29 ^a .04*	.52 <.01**	.18 ^a .20	.37 <.01**	.36 ^a <.01**	.48 <.01**	.25 ^a .08 ⁺	.55 <.01**
Working memory	.18 .19	.19 .20	-.14 .33	.22 .12	.33 .02*	.36 .01*	.17 .23	.34 .02
Working memory (-verbal ability)	.03 .86	-.14 .33	-.23 .09 ⁺	.05 .73	.15 .27	.11 .47	.07 .64	.12 .41
Consonant Perception (CP)	.19 .18	.49 <.01**	-.01 .94	.37 .01*	.04 .75	.37 .01*	.12 .39	.41 <.01**
CP (-verbal ability)	.12 .38	.22 .13	-.04 .76	.20 .18	-.06 .68	.04 .81	.07 .61	.17 .25
Attention baseline	-.07 .62	-.20 .17	.02 .87	-.13 .36	-.08 .56	-.32 .03*	-.02 .88	-.28 .05 ⁺
Distraction effect	-.14 .30	.01 .93	.03 .85	.18 .22	-.37 <.01**	-.12 .40	-.24 .09 ⁺	.12 .43
Incongruency effect	-.24 .08 ⁺	-.11 .46	-.01 .94	-.03 .84	-.14 .33	-.03 .84	.10 .47	-.00 .98

^a n = 52.

Note. LNHP = low noise, high predictability; LNLP = low noise, low predictability; HNHP = high noise, high predictability; HNLP = high noise, low predictability. For each cell, the upper value shows the correlation coefficient (r -value) and the lower value shows the p -value. (-verbal ability) indicates that the effect of verbal ability was partialled out of the correlation.

From these correlation analyses it appears that WMC was only correlated with SUN in the HNHP condition but this effect disappeared when verbal ability was partialled out.

Consonant perception was correlated with SUN in the bilingual group only and the effect disappeared when verbal ability was controlled for. Because of these high correlations between

predictor variables, only verbal ability will be used as a predictor⁶. For the different attention measures, the results are somewhat difficult to interpret. For the bilinguals, a lower baseline RT was associated with higher accuracy in the HNHP condition. For the monolinguals, on the other hand, it was the distraction effect that was associated with higher accuracy in the HNHP condition. The analysis of the TAIL test showed some differences between the two groups on the test and this might be why different measures correlated with the SPIN test for the two groups (see section 5.5 for a detailed analysis of the TAIL). When the whole sample was considered, there was a small but significant correlation between baseline RT and HNHP mean accuracy ($r(101) = -.21, p = .032$) and so this measure will be used for further analyses. Following these preliminary analyses, a mixed-effects logistic regression model was run with verbal ability and baseline RT as additional predictor variables besides those entered in Experiment 1.

3.2 Results

The results of this analysis are reported in Table 6. As can be seen, those variables that interacted with Group in Experiment 1 also interacted with verbal ability measured as a continuous variable⁷. Specifically, there was a significant main effect of verbal ability and significant interactions between verbal ability and predictability on the one hand and verbal ability and frequency on the other hand. However, when these interactions were entered into the model, group was still a significant factor as well, suggesting that not all variance could be explained by verbal ability. Because many of the predictors are on continuous scales, the results

⁶ Working memory did not predict SPIN accuracy when entered together with verbal ability and consonant perception was only significant in the main analysis but not when groups were analyzed separately.

⁷ Because biphone probability did not interact with group in Experiment 1, the interaction between this variable and verbal ability was not entered into the model.

can be best interpreted by displaying them graphically. The main effect of verbal ability is shown in Figure 7.

Table 6. *Results from the mixed-effects regression analysis of SPIN accuracy.*

	Odds-ratio	Logit scale	SE	<i>p</i>
Baseline (HNLP, bilingual, ND=high, PhonemePr=high)	1.10	0.09	0.14	
Noise (high vs. low)	3.82	1.34	0.06	< .001
Predictability (low vs. high)	6.41	1.86	0.07	< .001
Noise*Predictability	2.23	0.80	0.14	< .001
Group (bilingual vs. monolingual)	1.36	0.31	0.09	< .001
Frequency (z-score)	1.31	0.27	0.13	.032
Phoneme Probability (z-score)	1.26	0.23	0.13	.072
Verbal ability (z-score)	1.20	0.17	0.05	< .001
Attention baseline RT (z-score)	0.92	-0.08	0.03	.012
Predictability*Verbal ability	1.46	0.38	0.07	< .001
Frequency*Verbal ability	0.94	-0.06	0.03	.024
Noise*Predictability*Language ability	1.21	0.19	0.13	.144

Note. Odds ratios are shown here because they are easier to interpret. For example, participants were 3.8 times as likely to recognize a word in the low noise condition compared to the high noise condition. For continuous variables, the coefficient shows that change in SPIN accuracy associated with a 1 SD increase in the predictor variable. Standard errors are on the logit scale. *P*-values were calculated using Type II likelihood ratio tests.

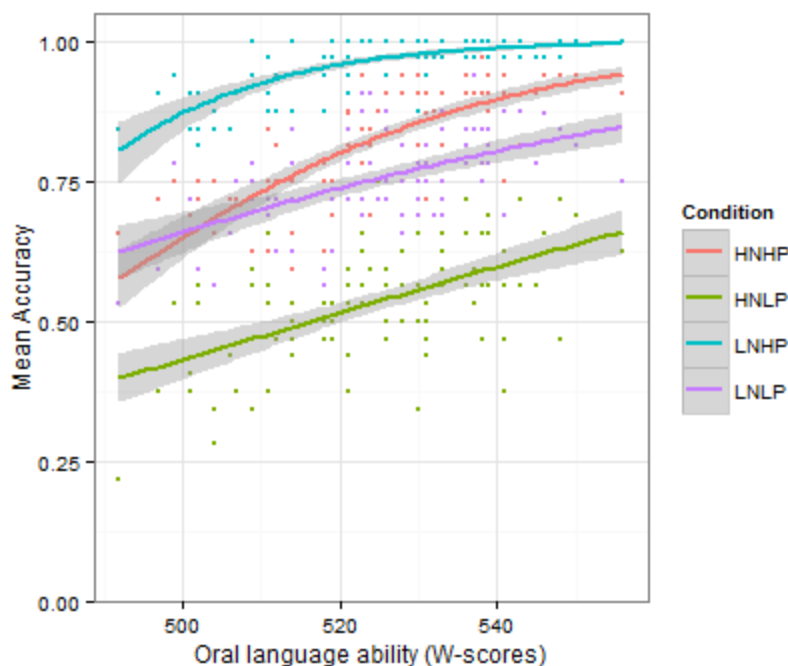


Figure 7. Relationship between oral language ability and accuracy on the SPIN test, depending on condition. HNHP=high noise-high predictability. HNLP=high noise-low predictability. LNHP=low noise-high predictability. LNLP=low noise-low predictability.

Because one of the crucial questions was whether the relationship between verbal ability and SPIN accuracy would be present in both groups, separate models were run for bilinguals and monolinguals. The results for each group are reported in Table 7. What is striking about these results is that many of the effect sizes are very similar when language proficiency is taken into account. For example, 1 SD increase in language ability was associated with a similar increase in accuracy for monolinguals and bilinguals. Likewise, the benefit of having a predictive context increased by roughly the same amount for 1 SD change in language ability.

The effect of attention was no longer significant for either group, although the effect size was the same for each group as in the previous analysis. This may indicate that the sample size was not large enough anymore to find the effect (i.e., there was too much uncertainty in the estimate).

Table 7. *Results of the mixed-effect regression analysis of the SPIN test for the monolingual and bilingual group.*

	Monolingual			Bilingual		
	OR	logit	SE	OR	logit	SE
Baseline (HNLP)	1.62	0.48	0.15	1.02	0.02	0.13
Noise (high vs. low)	4.28	1.45***	0.09	3.41	1.23***	0.09
Predictability (low vs. high)	9.92	2.29***	0.11	4.16	1.43***	0.09
Noise*Predictability	2.32	0.84***	0.24	2.25	0.81***	0.17
Frequency (z-score)	1.22	0.20	0.14	1.40	0.34**	0.13
Biphone Probability (z-score)	1.39	0.33*	0.14	1.21	0.19	0.13
Verbal ability (z-score)	1.15	0.14***	0.07	1.27	0.24***	0.07
Attention baseline RT (z-score)	0.93	-0.07	0.05	0.92	-0.09 ⁺	0.05
Predictability*Verbal ability	1.25	0.22**	0.10	1.14	0.13**	0.10
Frequency*Verbal ability	0.99	-0.01	0.03	0.97	-0.03	0.04
Noise*Predictability*Verbal ability	1.26	0.23	0.23	1.31	0.27	0.17

Note. See note in Table 5. *** $p < .001$; ** $p < .01$; * $p < .05$; ⁺ $p = .059$. p -values based on likelihood ratio tests.

One prediction that was not borne out by the data was that the frequency effect would be modulated by verbal ability. To further investigate this relationship, I divided the continuous frequency variable into three factors, low, mid, and high frequency, with each group containing an equal number of words. The reasoning behind this post-hoc analysis was that frequency effects may not be linear, that is, they may only be present at the low end of the scale. By dividing frequency into three factors, I can compare low frequency words to high frequency words, which may give more power to find effects. Using frequency as a factor is common in psycholinguistic studies, mainly because traditional ANOVAs do not allow continuous variables, and so the results will also be more easily comparable to other studies. The raw frequency of

each factor is shown in Table 8⁸. In addition to factorizing frequency, I also split both groups into a high and low proficiency group based on a median split of their oral language ability score. As a result, the bilingual high proficiency group was not significantly different from the monolingual low group ($p = .196$) and so the results will show if subgroups of monolinguals and bilinguals matched on proficiency will still perform significantly different⁹. The mean proficiency level of each group is shown in Table 9.

Table 8. *Word frequency of high, mid, and low frequency words on the SPIN test*

	High frequency	Mid frequency	Low frequency
Log10 frequency	3.22 (0.18)	2.72 (0.13)	2.22 (0.17)
Frequency per million	35.7 (16.6)	10.83 (3.33)	3.49 (1.31)

Note. Frequencies are based on Brysbaert and New's (2009) subtitle frequencies.

Table 9. *Mean language proficiency of the upper and lower half of the monolingual and bilingual group.*

	High group		Low group	
	OL-W	OL-SS	OL-W	OL-SS
Monolingual	541 (5.1)	112 (4.7)	527 (5.7)	100 (4.6)
Bilingual	525 (7.0)	98 (5.3)	507 (6.0)	83 (4.6)

Note. OL-W= oral language ability W-score. OL-SS = oral language ability standard score. See section 3.1.2.1 for further explanation. Standard deviations are shown in parentheses.

⁸ The results of ANOVA showed that the three frequency groups did not significantly differ in the number of neighbors ($F(2, 124) < 1$), biphone probability ($F(2, 124) = 2.1, p = .131$), or frequency-weighted neighborhood density ($F(1, 124) < 1$).

⁹ There are still some caveats, even when comparing participant groups matched on proficiency because these participants still differed on other dimensions such as parental education level. Nevertheless, results from this follow-up analysis can still be suggestive.

The results of this follow-up analysis can be seen in Figure 8. As can be seen, the decline from high to mid frequency is smaller than the decline from mid to low frequency, and this is the pattern for all groups. Of special interest was whether the monolingual low (ML) group would be significantly different from either the monolingual high (MH) or the bilingual high (BH) group. A mixed-effects logistic regression model with frequency (low/mid/high) and group (MH/ML/BH/BL) as predictor variables (all other variables were ignored for this analysis) showed a main effect of frequency ($\chi^2(2) = 7.7, p = .022$) and a main effect of group ($\chi^2(3) = 132.2, p < .001$) but the interaction was not significant ($\chi^2(6) = 8.0, p < .241$). To further investigate these group differences, follow-up analyses were run for each frequency level with the ML group as the reference category. When frequency was high, the BH group was marginally less accurate than the ML group ($b = -0.21, SE = 0.12, p = .084$) and the MH group was not significantly different from the ML group ($b = 0.19, SE = 0.13, p = .148$). At the mid frequency level, the BH group was not significantly different from the ML group ($b = -0.11, SE = 0.13, p = .343$) but the MH group was more accurate than the ML group ($b = 0.43, SE = 0.13, p = .001$). At the lowest frequency level, the BH group was less accurate than the ML group ($b = -0.42, SE = 0.10, p < .001$) and there was a trend for the MH group to be more accurate than the ML group ($b = 0.18, SE = 0.11, p = .095$).

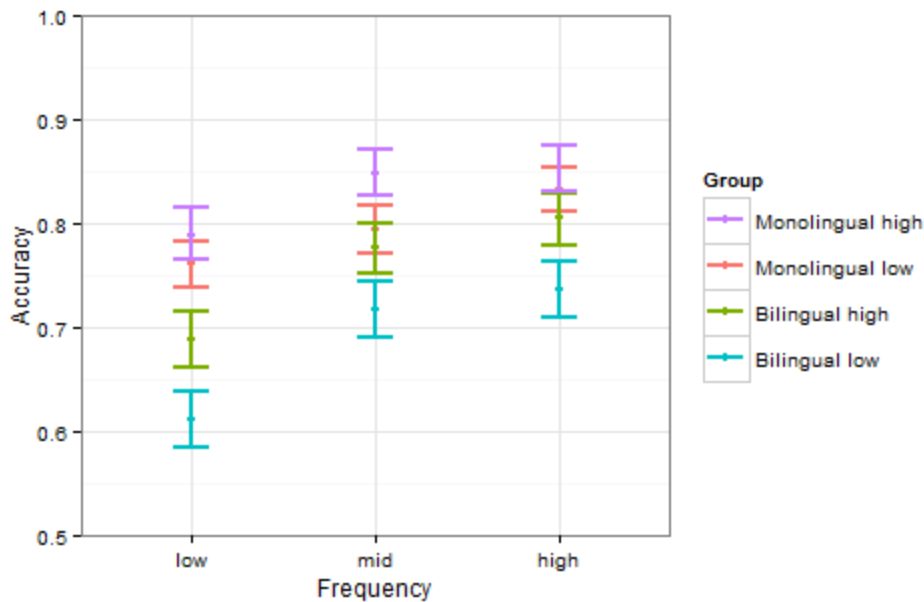


Figure 8. The effect of frequency show for each of four groups. The monolingual and bilingual group were each divided into a high and low group based on a median split of their proficiency score. Whiskers show the 95% confidence interval.

These results suggest that proficiency also had an effect in the monolingual group but effects were only statistically significant at the mid frequency level. More importantly, the subgroups of monolinguals and bilinguals that were matched on proficiency were not significantly different from each other at the mid and high frequency levels.

3.3 Discussion

Experiment 2 showed a large influence of individual differences on SUN. Both the main effect of group and the group by predictability interaction reported in Exp. 1 were modulated by language ability, as measured by the WMLS, in Experiment 2. This shows that differences between monolinguals and bilinguals previously reported in the literature may to a large part be attributable to differences in language experience. As described in the introduction, bilinguals often know fewer words in each of their languages compared to someone who only speaks one language (Bialystok & Luk, 2012; Gasquoine & Dayanira Gonzales, 2012; Portocarrero, Burright, & Donovanick, 2007). For example, Gasquoine and Dayanira Gonzales (2012) tested a

sample of 56 Mexican-American participants residing in the Rio Grande Valley region of South Texas using the same proficiency test that was used in the current study. These participants were more diverse than the current sample in terms of age (they ranged from 18 to 65 years) but were similar in terms of years of formal education (mean = 13.9 years). The age-adjusted standard score for the English picture vocabulary test was 86 in the Gasquoine and Dayanira Gonzales study, which is coincidentally the exact same figure as in the present study. The WMLS was also used in a study by Delgado, Guerrero, Goggin, and Ellis (1999). These authors tested 80 Spanish-English bilingual students in Texas. The sample differed somewhat from the present sample in that only half of the participants had received all of their formal education in the US. In this study, the mean W-score for picture vocabulary was 508.4 compared to 516.1 in the present study. The similarity in figures suggests that the bilingual participants in the present study were likely representative of the larger Spanish-English bilingual population in the US with a similar educational background. Given these differences in language proficiency compared to monolinguals, it is not surprising that bilinguals often perform less well on verbal tests. The monolinguals in the present study performed 1 SD higher¹⁰ ($d = 1.78$), which is a large effect. The present analysis showed that when group differences in language ability were controlled for, the difference in performance on the SPIN become much smaller. In addition, the relationship between accuracy on the SPIN and language ability was similar in both groups. As suggested by a previous study (Tamati et al., 2013), greater word knowledge is positively associated with better listening in noise ability in monolingual speakers. The present study confirms this result by using a standardized test of proficiency rather than self-ratings as in the Tamati et al. study.

¹⁰ That is, 1 standard deviation in the population sample, which is 15 for the WMLS.

In addition to the main effect of language ability, an interaction was also found between this variable and predictability. Participants with greater verbal ability were better able to make use of a predictive context (see Figure 7). Again, this was true for bilinguals and monolinguals, showing that the previously reported monolingual advantage (Bradlow & Alexander, 2007; Shi, 2010) may be better described as a general advantage associated with verbal ability.

So far, it seems that differences between groups can be best explained by differences in verbal ability. However, verbal ability by itself is not an explanatory factor but rather an observational factor. To test the hypothesis that individual differences in SUN are attributable to differences in language exposure, frequency effects were investigated. In the main analysis, frequency interacted with verbal ability as was predicted. However, in follow-up analyses of each group, the interaction was not found. This may be because of the more restricted range in language proficiency in each group but it may also suggest that the interaction was only significant because of group differences in verbal ability and so group status may be the actual cause of this interaction. In addition, the main effect of frequency was not significant in the monolingual group as in the previous group analysis reported in Experiment 1. However, a follow-up analysis with frequency as a factor with three levels, frequency effects also became apparent in the monolingual group. A wider range of word frequencies may be necessary, though, to find a more robust effect. The follow-up analysis may also provide some insight into the finding that group differences were still significant in the main analysis after controlling for proficiency and frequency effects. For the two subgroups of monolinguals and bilinguals that were matched on proficiency, differences in accuracy were small or nonexistent when frequency was in the medium to high range but became apparent when frequency was low. This finding suggests that even when monolinguals and bilinguals are matched on frequency, a bilingual

person may still have encountered low frequency words disproportionately less often compared to a monolingual person of the same overall language proficiency. Interesting in this respect is also the observation that the marginally significant noise-by-group interaction found in Experiment 1 seemed to have mostly been driven by the low frequency words as the following figure suggests (Figure 9), although the three-way interaction between noise, group, and frequency was not significant. This again shows the nonlinear nature of frequency effects and it suggests that the weakest lexical representations are those that are the most affected by noise.

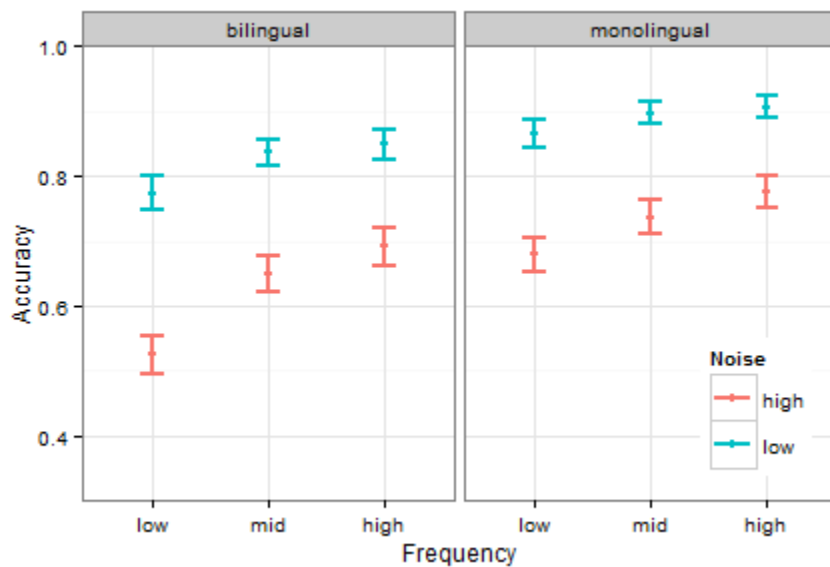


Figure 9. Mean accuracy on the SPIN test for each group (bilingual/monolingual) separated by noise level (high/low) and target word frequency (low/mid/high). The figure shows that in the bilingual group, the effect of noise was largest when frequency was low.

CHAPTER 4: GENERAL DISCUSSION

To summarize the main results again, Experiment 1 replicated earlier findings showing that the bilingual group performed below the monolingual group in all four conditions. The differences were especially large when noise and predictability were both high, replicating previous studies that found that bilinguals did not benefit as much from a predictive context as monolinguals (e.g., Bradlow & Alexander, 2007; Mayo et al., 1997; Shi, 2010).

The two-way interaction between frequency and group suggests that bilinguals' word recognition in noise may be especially affected when to-be-recognized words are of low frequency. These results are different from Imai et al. (2005) who also tested monolingual and bilingual participants on word recognition with low-level background noise (SNR = 12 dB) but did not find this interaction. Instead, they showed that the effect of neighborhood density was larger for the bilingual group compared to the monolingual group. The lack of an interaction in their study may have resulted from the corpus (Kucera & Francis, 1967) that their frequency counts were based on. As Brysbaert and New (2009) show, subtitle frequencies better reflect actual word frequencies, especially since the Kucera and Francis corpus is quite old and based on text word frequencies, something also noted by Imai et al. (2005). Bradlow and Pisoni (1999) also investigated the effects of lexical variables on native and nonnative word recognition in noise. These authors divided words into easy (high frequency, low neighborhood density) and hard (low frequency, high neighborhood density) words. They found that native and nonnative speakers of English recognized easy words better than hard words when there was single or multi-talker babble in the background. However, the effect was much larger for the nonnative speakers compared to the native speakers. Although the lexical variables investigated in the present study and the other two studies were not the same, the present results are nevertheless in

line with their results. Both studies suggest that nonnative speakers were more affected by these lexical variables than native speakers.

Previous SUN studies differ in whether the group-by-noise interaction is significant or not. Shi (2010) found a significant interaction whereas Rogers et al. (2006) did not. Shi (2010) compared a group of eight native bilingual speakers who learned English and another language before the age of two and another group of eight bilingual speakers who had learned English between five and seven to eight monolingual speakers (Shi also included two groups of late learners of English). At SNR +6 dB simultaneous bilingual group was not significantly different from the monolinguals but at SNR 0 the groups were different, suggesting an interaction between group and noise. The early bilingual group was different from the monolinguals at both SNRs (although this difference was not significant when correcting for multiple comparisons of which there were twenty). The results reported in Shi (2010) suggest that AoA is an important factor that predicts SUN. However, these results do not allow any conclusion about the origin of AoA effects. The present results shed some light on factors influencing the interaction between group and noise level as well as main effects of group. Because the bilingual participants in the present study were quite homogeneous in terms of AoA, the present results suggest that it is the amount of exposure to the tested language that determines SUN. Amount of exposure is closely related to AoA, of course, but what is striking about the present results is that the same relationship between proficiency and SUN was also found in the monolingual group who had acquired English from birth. As would be expected when amount of exposure is the determining factor, differences between groups were largest for low frequency words.

Reanalyzing the data from Experiment 1 with an individual differences design rather than a group design further confirmed the hypothesis that amount of exposure to English is the main

contributing factor to SUN. In Experiment 2, the effect of individual differences in different domains (linguistic vs. nonlinguistic) was explored. The largest difference between groups was in language proficiency and this variable also emerged as the strongest mediating variable between groups. In other words, differences between groups became smaller once language proficiency was taken into account. But language proficiency did not only mediate overall accuracy but also interacted with predictability. Both findings can be explained with the ELU. To reiterate, the basic assumption of the ELU (Rönnberg et al., 2013) is that speech information is bound into a phonological representation in an episodic buffer. This information, referred to as RAMBPHO (rapidly, automatically, and multimodally bound phonological representation) is assumed to operate at the syllable level and is matched to semantic representations in LTM (c.f. Giraud & Poeppel, 2012a). Listeners are assumed to constantly form predictions about upcoming acoustic information based on preceding suprasegmental, segmental, and semantic information (Pickering & Garrod, 2007). For example, when the context of an utterance is predictive of a certain word, this word may receive activation even before it is mentioned (Altmann & Kamide, 1999) and lexical access may happen even when the acoustic information is heavily degraded. Also, listeners have been shown to use distal prosodic information and context speech rate to make predictions about upcoming word boundaries to segment the speech stream (Brown, Salverda, Dilley, & Tanenhaus, 2011; Dilley & McAuley, 2008; Dilley & Pitt, 2010). When the speech signal is optimal, this process is effortless and proceeds rapidly. However, when mismatches between RAMBPHO and phonological LTM representations occur, lexical access is delayed and the feed-forward cycle is interrupted (Rönnberg et al., 2013, p. 3). Such mismatches can occur because of a poor speech signal or poorly specified phonological representations. In such cases, the assumption of the ELU model is that those mismatches have to be resolved

through explicit processes that operate on a larger time scale. According to the ELU, this is when individual differences in WMC will become visible. Those individuals with a larger WMC are assumed to have more processing resources available to make, for example, predictions based on the preceding context.

As is obvious from this discussion of the ELU is the great emphasis on WMC to explain individual differences in speech understanding in noise. In light of the present results, these assumptions may have to be modified to some extent. As predicted by the ELU, verbal WM was associated with better HIN. However, WM was correlated with verbal ability and when both variables were entered into a regression model to predict SPIN accuracy, only verbal ability was significant. The strong correlation between WM and verbal ability may be surprising because the WM used very common objects, namely animals and food items that participants can be expected to be very familiar with. As laid out in the introduction, verbal WM is not independent of LTM representations of words (Baddeley, 2012, p. 20; also see MacDonald & Christiansen, 2002). For example, studies have shown that high frequency words can be better remembered than low frequency words (e.g., Hulme et al., 1991), suggesting that stronger LTM representations may facilitate encoding and rehearsal of those words. The correlation between verbal ability and WM may have the same explanation as the frequency effect. For individuals with overall less precise lexical representations, all words may behave like low frequency words, that is, their representations are underspecified. Another aspect of the WM test used in the present study is that participants heard semantically related words (i.e., animals and foods) and thus had to inhibit previously activated words that were not relevant in the current set. For example, a participant may replace *elephant* with *bear* because a bear was the largest animal in

the previous set. Individuals with a larger vocabulary may be better able to inhibit previously activated words and thus prevent interference.

Given these interactions between verbal WM and verbal ability, these two variables may not be easily separated. The fact that WM was no longer a significant predictor when entered together with verbal ability does not necessarily mean that WM does not play a role in speech understanding in noise. However, individual differences in WMC may play a smaller role than individual differences in verbal ability. Such individual differences in verbal ability may influence speech understanding in noise in different ways. For an individual with more language experience, words in the mental lexicon may be better integrated semantically because they will have experienced words in more diverse contexts (Bolger, Balass, Landen, & Perfetti, 2008). For example, word collocations will be better entrenched because they are experienced more often and thus co-occurrences of words may be better predicted. In the sentence *The ship sailed along the coast* (taken from the SPIN test), both ship and sail may trigger associations with coast but for someone who has not experienced those words together much, the association with coast may only weakly exist and thus they would not predict coast and would have to rely more on the acoustic signal (i.e., bottom-up information). ERP studies of the N400 effect, an electrophysiological response that indicates semantic integration of words into the preceding context, have shown that the effect is modulated by vocabulary knowledge in monolingual and bilingual speakers (Moreno & Kutas, 2005; Newman, Tremblay, Nichols, Neville, & Ullman, 2012). These studies suggest that individuals with a larger vocabulary are better able to form predictions during listening.

Besides these semantic contributions to speech understanding, higher verbal ability may also help listeners to segment the speech stream into word units. Mattys et al. (2005) found that

listeners rely largely on lexical information for word segmentation and only revert to sublexical information such as word stress when the signal is heavily degraded (at SNRs of -5 dB). Thus stronger lexical knowledge may help listeners segment the speech stream more accurately and they may recover from false segmentations more rapidly.

Coming back to the ELU, the results of the present study suggest that individual differences in phonological representations in LTM may be more indicative of SUN difficulties than individual differences in WMC, at least in a sample of healthy young adults. It may be that in older people, individual differences in WMC become more important. Especially since vocabulary knowledge typically increases with age (and then decreases in old age; Kavé, Knafo, & Gilboa, 2010), it cannot be responsible for the common observation that SUN ability decreases as a function of age. The fact that the present study investigated a sample of healthy young adults may also explain why attentional control, measured by the TAIL, only had a small effect on recognition accuracy. A tentative interpretation of the TAIL effect is that individuals with better attentional control are better able to attend to the relevant speaker and ignore the background babble. Thus the temporal separation of the target and distractor signal may rely on attentional processes. Testing a wider range of age groups may reveal whether attentional control will correlate more highly with SUN in a younger or older sample. Future studies should also administer more than one test of attention and executive function to determine whether overall processing speed is more indicative of SUN or rather a specific component of executive function, that is, inhibition, updating, or shifting (Miyake & Friedman, 2012). Another way forward may be to manipulate attention load during SUN instead of employing a correlational design to be better able to establish causal relationships (cf. Mattys & Wiget, 2011). In any case, researchers need to make sure that the concurrent task or the task to be used as a predictor variable is not

dependent on verbal ability to avoid confounds. For example, Sommers and Danielson (1999) used a linguistic Stroop test to predict SUN performance. In this test, participants heard the words *mother*, *father*, and *person* spoken by a man or a woman. Inhibition was necessary when there was incongruence between the sex of the speaker and the gender of the spoken word, for example, when the word *mother* was spoken by a male speaker. An inhibition index was calculated by subtracting RTs in the incongruent condition from RTs in the neutral condition and this measure correlated with SUN for hard words (hard words were defined by the authors as low frequency and high neighborhood density words). Because a Stroop test using linguistic stimuli may not be independent of verbal ability, using a nonlinguistic auditory test of attention may be better.

CHAPTER 5: ANALYSIS OF INDIVIDUAL TESTS

In the previous section, the results from some of the administered tests were used as predictor variables in a regression analysis. The purpose of this section is to describe those tests plus one additional test, the Words in Noise (WIN) test, in more detail.

5.1 Words in Noise

Experiment 1 and 2 were designed to answer the question why monolinguals and early bilinguals are differentially affected by noise. However, the SUN test used was not administered in the way it is commonly administered to assess a hearing deficit (e.g., the noise levels were different than on the original test as described in Bilger et al., 1984). Therefore, a standardized hearing in noise test, the WIN, was also administered to all participants. This was done to investigate whether bilinguals may be wrongly diagnosed with a hearing deficit based on their bilingual status. The first research question I will answer is whether monolinguals and bilinguals are differentially affected by noise. We may expect results on the WIN to be different than those obtained with the SPIN because of the different make-ups of the two tests. On the SPIN, the onset of target words is unpredictable because the preceding context is different for each sentence. On the WIN, on the other hand, target words are always preceded by the same carrier phrase, which is *say the word*. In addition to making the target word onset predictable, the WIN places a lower processing load on participants compared to SPIN sentences for which the context is predictive of the target word. As was shown in Experiment 2, recognition of words in a predictable context is especially dependent on individual differences in verbal ability. When testing bilingual speakers for a hearing deficit, it may therefore be advisable to use a test that is not strongly correlated with verbal ability.

5.1.1 Methods

5.1.1.1 Participants

The same participants were tested as described above. One participant from the bilingual group was excluded from this analysis because the test could not be administered due to technical difficulties, which reduced the bilingual sample to 47.

5.1.1.2 Materials

The WIN was developed by Wilson and colleagues (Wilson, Abrams, & Pillion, 2003) and was also administered through the NIH Toolbox. The NIH toolbox is a collection of different tests in the areas of cognition, emotion, motor function, and sensation. All tests are available freely and are administered online. The test consists of two lists of 35 words each. Each list is divided into groups of five words that are played back with background babble (multiple speakers) at the same SNR. Participants hear a woman asking them to repeat words, for example, “Say the word dog”. The sound intensity of the background babble is fixed and the woman’s voice becomes increasingly softer starting at a SNR of 24 dB and decreasing to 0 dB in 4 dB decrements. Administration stops when none of five items at a particular SNR can be correctly repeated by a participant or when the end of the list is reached. Each list is administered monaurally to one ear only with ear of testing being counterbalanced. For example, one participant will hear List 1 presented to the right ear and List 2 to the left ear and another participant will hear List 2 presented to the right ear and List 1 to the left. The score for the test is derived from the inflection point of the psychometric function (which describes the relationship between accuracy and SNR) to determine at which SNR a participant recognized 50% of the words. This test was administered in English and Spanish. The Spanish version was administered at the end of the session after all English tests were done.

5.1.2 Results

5.1.2.1 English Words in Noise Test

A logistic mixed-effects regression model with a probit link-function and with Accuracy as the outcome variable and the main effects of Group and SNR and their interaction was run including random intercepts for words and subjects and random slopes for SNR within subjects. The descriptive statistics are shown in Table 10¹¹ and mean accuracy of all items can be found in APPENDIX. Because participants were at ceiling at SNR 24 dB to 12 dB, the regression model was only fit to the data between 12 dB and 0 dB. The results showed a main effect of SNR ($\chi^2 (1) = 90.1, p < .001$). The effect of group was not significant ($\chi^2 (1) = 2.8, p = .093$), nor was the interaction between SNR and group ($\chi^2 (1) = 2.4, p = .121$)¹².

Table 10. *Mean accuracy on the Words in Noise test.*

	SNR	24 dB	20 dB	16 dB	12 dB	8 dB	4 dB	0 dB
Group								
Monolingual	<i>M</i>	100%	100%	99.2%	100%	74.3%	53.0%	20.6%
	<i>SD</i>	(0.0)	(0.0)	(9.1)	(0.0)	(43.7)	(50.0)	(40.5)
Bilingual	<i>M</i>	99.4%	100%	96.9%	98.9%	71.6%	49.0%	21.7%
	<i>SD</i>	(8.0)	(0.0)	(17.3)	(10.3)	(45.2)	(50.0)	(41.3)

Note. SNR = signal-to-noise ratio.

The results of the regression model are shown in Figure 10 along with the observed values. The predicted values that are derived from the model estimates overestimate accuracy at SNR 8 dB and underestimate accuracy at SNR 4 dB. However, the actual fitted values that take

¹¹ An examination of individual items showed two outliers and these were excluded from the descriptive statistics. The word *time* at SNR 16 and *shawl* at SNR 20. See Table 19 for mean accuracy of all items.

¹² When the model was fit to the whole data set, the effect of group was significant ($\chi^2 (1) = 5.4, p = .020$) as was the interaction between group and SNR ($\chi^2 (1) = 13.8, p < .001$). However, it seems that this interaction was attributable to the lower performance of the bilingual group at SNR 16. At SNR 12, group differences were not significant and so the differences at SNR 16 are likely attributable to specific items that caused difficulty (also see Appendix 2).

into consideration subject and item variance are quite close to the observed values, suggesting that the model describes the data well.

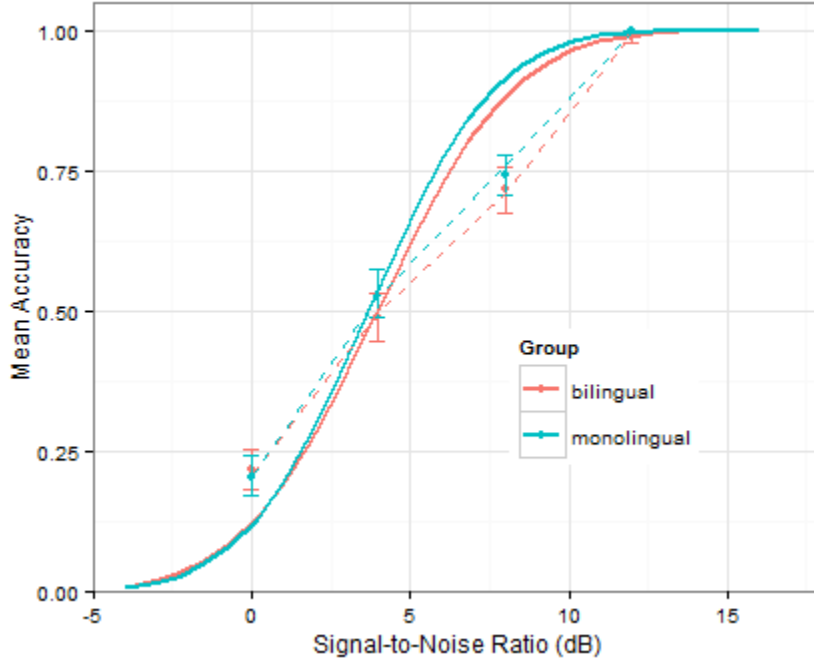


Figure 10. Results of the English WIN test. Solid lines show the predicted values based on coefficients of the regression model described in the text. Dashed lines show the fitted values of this model. Whiskers show the 95% confidence interval.

Another way to look at the data is to extract the SNR at which a participant achieved 50% accuracy. This can be done by running a logistic regression model for each participant. Using the predicted intercept and slope, we can calculate the SNR_{50} . The formula for this is

$$\log\left(\frac{y}{1-y}\right) = \alpha + \beta x$$

where x is SNR and y is percent accuracy at this particular SNR. Solving the equation for $y = .5$ gives

$$x = -\frac{\alpha}{\beta}$$

The regression coefficients can also be used to calculate the inclination of slope at the SNR needed to achieve 50% accuracy. The slope of a logistic regression model is nonlinear because the model tends to 0 and 1 at the extreme ends. The slope is the steepest at the central point where $\text{Pr}(x) = .5$, which is SNR_{50} . The formula for this function is

$$\text{slope}(\text{Pr}(x) = .5) = \frac{\beta e^{\alpha + \beta x}}{(1 + e^{\alpha + \beta x})^2}$$

We already established that $\alpha + \beta x$ equals 0 at the point of 50% accuracy and so the equation becomes

$$\text{slope}(\text{Pr}(x) = .5) = \frac{\beta e^0}{(1 + e^0)^2} = \frac{\beta}{4}$$

Thus we can simply divide the coefficient of the slope of the logistic regression by 4 to obtain the inclination of the slope at SNR_{50} , that is, the % change in accuracy for a change in 1 dB (for further explanation see Gelman & Hill, 2007, p. 82).

Using these formulae, the SNR_{50} for monolinguals is 3.66 dB and for bilinguals it is 3.93 dB. The slope at the inflection point is 8.00%/dB for monolinguals and 7.39%/dB for bilinguals.

5.1.2.2 Spanish Words on Noise Test

The bilingual participants also completed the Spanish version of the WIN, the S-WIN. Here I will compare performance on one versus the other test. As in the analysis of the E-WIN, a logistic mixed-effects regression model with a probit link-function was fit to the data. The model included random intercepts for subjects and items and the main effects of SNR and language (English/Spanish) and the interaction between the two variables were entered as fixed effects. As in the previous analysis, the model was only fit to data between SNR 12 and 0 dB.

The main effect of SNR was significant ($\chi^2(1) = 220.8, p < .001$). Neither the main effect of language ($\chi^2(1) = 1.4, p = .235$), nor the interaction between language and SNR ($\chi^2(1) = 0.1, p = .727$) were not significant. The SNR₅₀ on the S-WIN was 4.9 dB and the slope was 7.8%/dB. Given the nonsignificant results of language and the language-by-SNR interaction, the SNR₅₀ and slopes did not differ in either language.

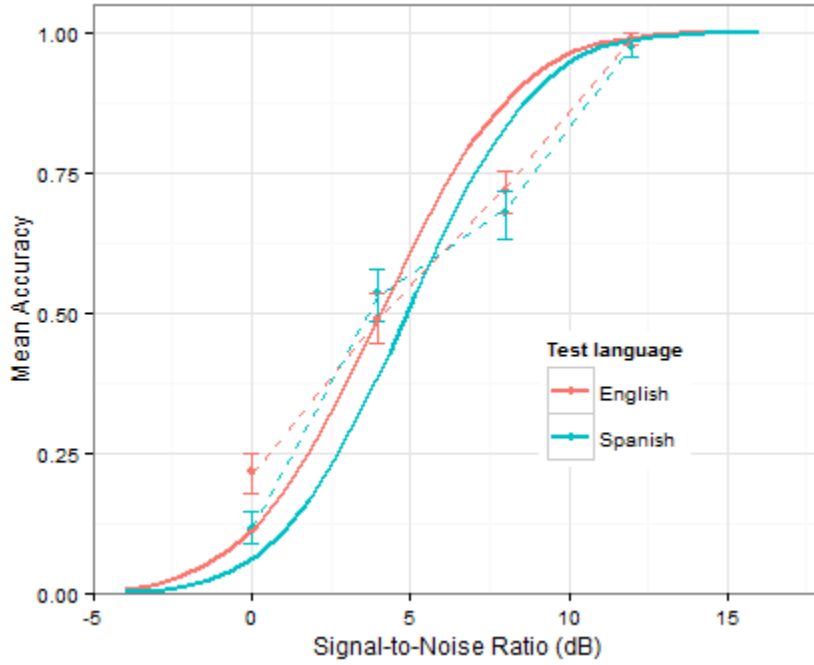


Figure 11. Results of the English and Spanish versions of the WIN test (bilingual participants only). Solid lines show the predicted values based on coefficients of the regression model described in the text. Dashed lines show the fitted values of this model. Whiskers show the 95% confidence interval.

5.1.2.3 Individual differences analysis

As in Experiment 2, the effect of individual differences was investigated. For this purpose, the variables oral language ability and Baseline RT (c.f. 3.1.2.2) were entered as continuous predictor variables into a regression model with the E-WIN as the outcome variable. Group was not entered as a predictor since it was not significant in the previous analysis. The results showed a main effect of language ability ($\chi^2(1) = 5.8, p = .016$) and a marginally

significant effect of Baseline RT ($\chi^2(1) = 3.3, p = .068$). The interaction between Baseline RT and SNR was also marginally significant ($\chi^2(1) = 2.8, p = .094$). To interpret these results, we can again calculate the SNR_{50} and the slope based on the model coefficients. For an individual 1 SD below the mean on Baseline RT, the predicted SNR_{50} is 3.58 dB and for an individual 1 SD above the mean, the predicted SNR_{50} is 3.99. Thus, faster processing speed (i.e., a lower baseline RT) was associated with a lower SNR_{50} . In addition, Figure 12 suggests that the effect of Baseline RT was largest at the lowest SNR. For language ability, the predicted SNR_{50} for individuals below and above 1 SD were 4.02 dB and 3.58 dB, respectively. Figure 13 suggests that this effect was most apparent at SNR 4 dB.

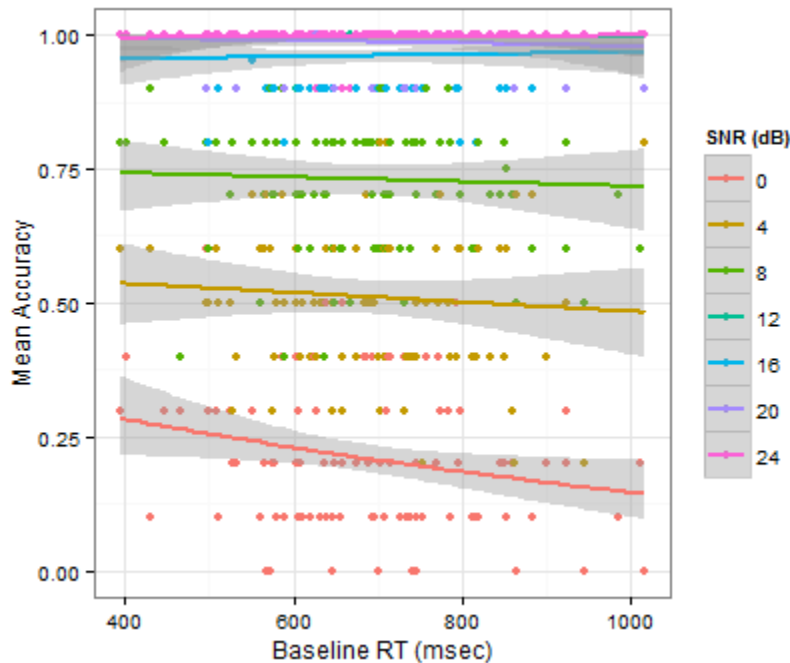


Figure 12. Effect of *Baseline RT* on WIN accuracy at each SNR. SNR = signal-to-noise ratio. Baseline RT is the mean response time on the Test of Attention in Listening (see text for further explanation).

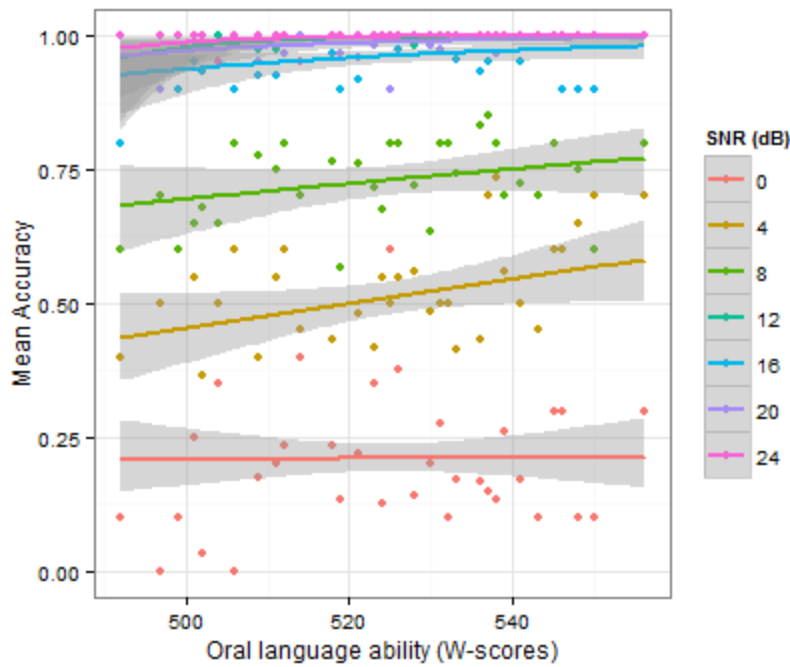


Figure 13. Effect of oral language ability on WIN accuracy at each SNR. SNR = signal-to-noise ratio. W-scores are arbitrary units with equal interval spacing.

Next, the effects of Spanish language ability and Baseline RT were investigated for the Spanish version. Results showed a significant main effect of language ability ($\chi^2(1) = 5.4, p = .021$) and SNR ($\chi^2(1) = 84.9, p < .001$). All other main effects and interactions were not significant ($ps > .150$). As in the English version, higher language ability was associated with a lower SNR_{50} . The predicted SNR_{50} s were 4.98 and 4.78 for individuals 1 SD above and below the mean on the language test, respectively.

5.1.3 Discussion

Are monolinguals and bilinguals differentially affected by noise? Returning to the research question of whether background babble at different SNRs differentially affected monolingual and bilinguals, the data suggest that both groups performed very similarly. The descriptive statistics showed that bilinguals were slightly less accurate but the psychometric

functions fit to the data showed that the SNR_{50} and the slope at the inflection point were very similar for both groups.

One concern with this test when interpreting the results is that test administration happened in a quiet but not sound insulated room. Also, the test was administered via the internet (following the NIH toolbox protocol) and no audiometer was used to adjust the sound pressure level (SPL). However, when comparing the present results to those of published results, they seem quite similar. Wilson, McArdle, and Smith (2007) compared normal hearing (NH) listeners and listeners with hearing loss (HL) on the WIN and other tests. The authors calculated the SNR_{50} and the slope of the psychometric function. In their study, the SNR_{50} for NH was 4.1 dB compared to 3.66 dB in the present study (monolinguals), which is quite similar. Performance of both groups using the 50% accuracy criterion was also within one standard deviation of the NIH toolbox norming study (NIH toolbox Technical Manual, p.25; available through NIHtoolbox.org), which were $M = 4.79$, $SD = 4.07$. The slopes for the monolinguals appear to be steeper in the present study (8%/dB) compared to 6.3%/dB in Wilson, McArdle, et al. (2007) but are similar to 8.4%/dB reported in Wilson, Carnell, and Cleghorn (2007). The NIH manual for the WIN does not report mean values of the slopes of the psychometric function for the norming population.

Using the criterion of the NIH manual, 91% of participants scored within the range for NH ($SNR_{50} \leq 6$ dB) and 9% within the range of mild hearing loss ($SNR_{50} < 8$ dB). In Wilson, McArdle, et al. (2007), the WIN was the best out of four SUN tests to distinguish listeners with HL from normal hearing listeners. Only 1% of the listeners with HL performed within the 95% CI of the normal hearing listeners and there were marked differences between groups at each SNR of the WIN. In the present study, pure-tone thresholds to measure hearing loss were not

obtained from participants but participants rated their hearing as good (8.6 out of 10 on average). Two participants rated their hearing as 6 but those two participants did not perform outside the range of the remaining participants. Furthermore, all participants were young adults and performance was similar to the study by Wilson, McArdle, et al. (2007; see above).

How does performance on the WIN compare to the results reported in Experiment 1? In the HNLP condition (SNR = -2 dB), bilinguals achieved around 50% accuracy, whereas the SNR on the WIN for 50% accuracy was around 4 dB. This could be because of differences in the speaker voice, differences in the babble noise used, and differences in the target words. At the same time, differences between groups seem to be much more pronounced on the SPIN. On the WIN, both groups performed very similar at each SNR but on the SPIN, group differences were significant in each condition, although effect sizes of group differences were small. The different performance of both groups relative to each other may be explained by different task demands. On the WIN, words are always presented with the same carrier phrase. This makes the onset of the target word predictable and thus puts low demands on word segmentation ability. On the SPIN, on the other hand, target word onset is not predictable and so listeners may be more affected by missegmentations. In addition, listeners also have to pay attention to sentence context if they want to exploit it to predict the target word and this places higher attentional demands on the listener. Because listening may be generally more demanding for bilingual speakers (Schmidtke, 2014), noise may disproportionately increase attentional demands. This may also explain why bilinguals did not benefit as much from a predictive context as monolinguals.

How does performance in one language relate to performance in the other language?

The results for the SNR₅₀ for the Spanish version ($M = 4.9$ dB) are similar to those obtained from the norming sample ($M = 5.53$ dB, $SD = 1.36$; NIH toolbox Technical Manual, p.25). The mean SNR₅₀ of the bilinguals reported in Carlo (2008) is somewhat higher with 6.2 dB ($SD = 1.3$). In Carlo (2008), the mean slopes of the psychometric functions were steeper in the Spanish version than the English version. In the present study, performance on the E-WIN was not significantly different from performance on the S-WIN, neither in the SNR₅₀ nor the slope of the psychometric function. This suggests that as a group, test language did not have an effect on hearing in noise ability. However, for the individual it may have an effect depending on the proficiency in English and Spanish as I will discuss in the next section.

Individual differences predicting WIN accuracy: As in the analysis of the SPIN in section 0, I also investigated whether individual differences in verbal ability and processing speed (Baseline RT on the TAIL test) would be associated with accuracy on the WIN test. The WIN test is supposed to place minimal attentional and memory demands on the listener in order to measure hearing ability and not some other skill. As was shown in the analysis of the SPIN, individuals with larger verbal ability can potentially compensate for their hearing loss by being less dependent on the bottom-up signal. Other tests such as the QuickSIN (Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004) have participants repeat a whole sentence and they receive one point for each of five keywords that they repeat per sentence. Thus participants with better STM may score higher because they are better able to remember the keywords. The present results suggest that even though the WIN test reduces the possibility to compensate for hearing loss by employing higher order cognitive skills, the test may still be sensitive to these individual differences. However, it should be noted that these effect sizes were small and they

may have greater theoretical than practical implications. For example, the fact that processing speed predicted WIN accuracy suggests that this may be one reason for greater hearing difficulty in older people. Further confirming the conclusion that individual differences in linguistic abilities did not play a big role on the WIN was the finding that monolinguals and bilinguals performed very similar. This suggests that the WIN may be a good test to use with nonnative speakers of English. When testing Spanish-English bilingual speakers, it may be best to test them in their stronger language because both English proficiency and Spanish proficiency was associated with higher accuracy on each respective test.

5.2 Verbal ability

The results from Experiment 2 in section 3.2 and the results reported in the previous section have shown that verbal ability is associated with higher accuracy on SUN tests. In this section, I am going to investigate which biographical variables predict verbal ability in bilingual speakers.

Many studies have found that vocabulary knowledge in bilinguals is lower than in age-matched monolinguals (Bialystok, Luk, Peets, & Yang, 2009; Bialystok & Luk, 2012; Portocarrero et al., 2007). The purpose of the present study was to find variables that would predict individual differences in verbal ability between monolingual and bilingual speakers. Previous studies found that exposure to each language is a good predictor of language development in children (Hammer et al., 2012; Hurtado et al., 2013; Place & Hoff, 2011). However, few studies have systematically investigated vocabulary knowledge in young adult bilinguals. Because the participants did all their schooling in the US, it may be that by the time they entered college, they had caught up with their monolingual peers. The present study shows that this was not the case and therefore it may be beneficial to identify variables that predict

proficiency in the dominant language. In addition, the present study contributes to the literature on heritage language maintenance (Peyton, Ranard, & McGinnis, 2001) by not only testing participants in their dominant language but also in their home language to investigate how different variables differentially affect proficiency in English and Spanish.

The predictor variables for verbal ability came from the background questionnaire that was administered to all participants. Participants were asked to estimate what percentage of the time they were exposed to English and Spanish growing up and the number of people who interacted with them during childhood and adolescence in each language on a regular basis (regular was defined as *at least once in two weeks*). Participants were given 5 different age periods (age 0 – 2; 3 – 5; elementary school; middle school; high school). The variable *number of speakers* was included based on a recent study that suggested that the number of speakers an individual interacted with predicted language proficiency above and beyond frequency of use (Gollan et al., 2014). Gollan et al. asked participants to estimate the number of speakers and percentage of use of the heritage language from birth through high school. In the present study, participants were asked to give more nuanced answers according to the five age-related categories mentioned above to see how the relative use of English and Spanish changed from birth to high school. In addition, participants were asked to estimate their current use of English and Spanish in three areas, speaking, listening, and reading.

A second purpose was to investigate the relationship between vocabulary knowledge and verbal reasoning. Most studies on bilingualism only include a test of vocabulary knowledge. However, bilinguals often know a word in one language but not the other because they do not use each language in the same contexts. For example, many of the participants in the present study reported speaking Spanish at home but English in most other situations. Therefore,

vocabulary knowledge in one language likely underestimates the total number of words that a bilingual speaker knows and thus vocabulary knowledge may not be a good indicator of general verbal ability. For example, a W-score of 500 is the average score that a 10-year-old is expected to achieve. In the present study, some bilingual participants performed below 500, yet they were studying at a major US university. This suggests that the true verbal ability of an individual scoring around 500 is most likely higher. Using the same tests of verbal ability as in the present study, I found in a previous study that monolinguals and early bilinguals did not perform significantly different on the verbal analogies test but bilinguals gave significantly fewer correct responses on the vocabulary test (Schmidtke, 2014). Thus the prediction follows that the bilinguals score on the PV test is significantly lower than would be expected based on their VA score compared to the monolingual group.

5.2.1 Materials

See section 3.1.2.1 for a description of the Woodcock-Muñoz Language Survey-Revised (WMLS-R).

5.2.2 Procedure

Following the standard procedures (Alvarado & Woodcock, 2005), both tests started from an age appropriate page. If a participant did not give six correct answers, the test was administered in backward order until the participants could correctly answer all six items from a set or until the first item was administered. Once the basal score was established, testing resumed from the first administered page. Testing stopped when a participant could not correctly name any item from a set of six.

5.2.3 Results

Woodcock-Muñoz Language Survey-Revised English: For group comparisons, it is most appropriate to use the age-corrected standard scores, which are normed on a large sample with a population mean of 100 and a standard deviation of 15. However, for subsequent statistical analyses it will be more appropriate to use the W-scores, which are not age-corrected, because for many research questions absolute vocabulary knowledge is of greater importance than relative vocabulary knowledge in comparison to peers of the same age.

The mean Picture Vocabulary (PV) standard score for monolinguals was 101 ($SD = 7.6$), which is right at the population mean and that for bilinguals was 86 ($SD = 8.4$), which is almost 1 standard deviation below the population mean. This difference was significant, $t(99) = 9.05$, $p < .001$, $d = 1.80$. The Verbal Analogies (VA) scores were also significantly different between groups, $t(98) = 6.90$, $p < .001$, $d = 1.38$. Monolinguals scored above the population mean ($M = 109$, $SD = 7.3$) and bilinguals just below the mean ($M = 98$, $SD = 9.0$). The difference in the composite score was also significant, $t(98) = 8.85$, $p < .001$, $d = 1.77$, with monolingual scoring higher ($M = 105$, $SD = 7.7$) than bilinguals ($M = 90$, $SD = 8.8$).

WMLS-R Spanish: The SS on the Spanish version (bilinguals only) were 77 ($SD = 7.9$), 90 ($SD = 10.8$), and 81 ($SD = 9.3$) for PV, VA, and OL, respectively. For all three measures, participants performed on average better on the English version than the Spanish version ($ts > 4.88$, $ps < .001$, $ds > 0.77$), showing that as a group, they were dominant in English.

What is the effect of socio-economic status on verbal ability? Monolinguals and bilinguals differed in terms of socio-economic status, measured by mother's education. Therefore, it was of interest to determine the influence of SES on oral language ability. For this purpose, the education levels college, some grad school, and grad school were combined into one

category, college+. The other categories were *less than high school*, *high school*, and *some college*. When both groups were considered in one analysis, including group as a factor, mother's education was not a significant factor. However, when each group was considered on its own, mother's education was a significant predictor of verbal ability for monolinguals ($b = 3.6$, $SE = 1.5$, $p = .017$, $R^2 = .11$) but not bilinguals. When examining the distribution of mother's education level for the bilinguals, 86% of participants reported that their mother's education level was less than high school or high school. Therefore, there may not have been enough variance in the bilingual speakers' SES distribution to find a significant effect. It would likely be necessary to test participants from a wider range of SES levels or to employ a more fine-grained measure of SES to determine how much of the variance can be attributed to language group and how much to SES. Spanish verbal ability was not associated with SES, either ($r(48) = -.05$, $p = .716$).

Factors explaining Spanish and English proficiency in bilinguals: Based on previous research, the differences in English proficiency between monolinguals and bilinguals were expected. A more interesting question is therefore what factors may predict proficiency in English and Spanish in the bilingual group.

The predictors for Spanish proficiency were the number of people who spoke Spanish with the participants, the percentage of Spanish exposure at the five life stages described above (age 0 – 2; 3 – 5; elementary school; middle school; high school). Participants also estimated their parents' use of English and Spanish (in %). These estimates were significantly correlated with percentage of Spanish exposure at age 0-2 and 3-5. The means and standard deviations are shown in Table 11.

Table 11. *Mean number of Spanish speakers and percent exposure to Spanish*

Stage	Number of Speakers of Spanish <i>M (SD)</i>	Percent exposure to Spanish <i>M (SD)</i>
0-2 years	5.8 (3.9)	91.4% (18.3)
3-5 years	6.6 (5.1)	76.3% (22.2)
Elementary school	9.2 (6.7)	45.5% (14.9)
Middle school	9.3 (6.7)	35.0% (13.9)
High school	9.8 (6.8)	33.9% (16.9)
Mean	8.1 (4.4)	56.6% (11.5)

Note. Participants were asked to estimate how many people they interacted with in Spanish regularly and the percentage they were exposed to Spanish at each of the five stages in life shown on the right.

Initial correlation analyses with the outcome variable Oral Language Ability Spanish (standard score) and the predictor variables showed that for percentage exposure to Spanish, the correlation was only significant at age 3-5 ($r(48) = .48, p < .001$) and for the mean percentage exposure ($r(48) = .36, p = .011$). For the number-of-Spanish-speakers variable, the correlation was only significant for age 0-2 ($r(48) = .30, p = .040$) and 3-5 ($r(48) = .32, p = .028$). The mother's use of English in the home while growing up was negatively but not significantly related to the participant's Spanish proficiency ($r(48) = -.18, p = .231$). The correlation with the father's use of English, on the other hand, was significant ($r(46) = -.34, p = .021$)¹³. A regression model using the number of speakers and percent exposure to Spanish at age 3-5 and the father's use of English as predictor variables explained 29% of the variance (adjusted R^2) in Spanish proficiency. In order to show how much additional variance was explained by each variable after accounting for variance explained by the other two variables, stepwise regressions were carried out. Number of Spanish speakers explained 4% additional variance, and percent exposure to Spanish and father's use of Spanish explained 10% each. Adding age of acquisition of English to

¹³ Two participants did not provide an estimate for their father's use of English.

the model did not increase the explained variance. Age of arrival to the US, on the other hand, had a positive effect and the adjusted R^2 increased to 40%. However, only 11 out of 48 bilingual participants were not born in the US, which may make this variable unreliable.

Next, I look at the influence of different variables on English and Spanish proficiency simultaneously. For example, more use of Spanish may be associated with greater Spanish proficiency and lower English proficiency. For this analysis, the data were arranged in the long format with language (English vs. Spanish) as a predictor variable so that each participant contributed two observations. Next to the variables reported above, participants were also asked to estimate the current relative time (in percent) spent listening, speaking, and reading in English and Spanish, respectively. An average was taken for this variable, referred to here as current English use (current English and Spanish use always added up to 100% for each participant as they were not exposed to other languages). The outcome variable in this analysis was the picture vocabulary standard score instead of the oral language score. The reason for this is that oral language is a composite score of VA and PV but PV is more strongly associated with language exposure. Because VA in English and Spanish were correlated, this suggests that verbal reasoning skills transfer from one language to the other and so VA may be less associated with exposure to each language (see below). All model coefficients show the change in the outcome variable (PV on the standard score scale) associated with 1 SD increase in the predictor variable.

Language of test alone explained 25.5% of the variance in the scores on the English and Spanish versions, with participants scoring on average 9.6 points lower on the Spanish version compared to the English version. More current English use was associated with a marginally higher English score ($b = 2.1$, $SE = 1.1$, $p = .061$) and a lower Spanish score ($b = -3.7$, $SE = 1.5$, $p = .020$). More Spanish exposure from birth through high school was associated with lower

English proficiency ($b = -3.3$, $SE = 1.1$, $p = .003$) and higher Spanish proficiency ($b = 5.7$, $SE = 1.5$, $p < .001$). Together, these variables explained 37.6% of the variance.

The next question was what the relationship was between language dominance and PV scores. Language dominance was calculated by subtracting the Spanish picture vocabulary score from the English score. Are more balanced bilinguals less proficient in each of their languages compared to the stronger language of less balanced bilinguals? The mean language dominance score was 9.6 ($SD = 11.7$), showing that most bilingual participants were English dominant. A regression model of picture vocabulary standard scores with Language dominance and Language as predictors explained 63% of the variance (adjusted R^2 ; see Table 12).

Table 12. *Results of the regression analysis predicting picture vocabulary scores*

Variable name	<i>Beta</i>	<i>SE</i>	<i>p</i>
Intercept	81.1	1.1	
Language dominance (LD)	0.5	0.1	< .001
Test Language (baseline = English)	0.0	1.5	1.000
LD*Test Language	-1.0	0.1	< .001

Note. Language dominance was calculated by subtracting Spanish scores from English scores. Thus a positive score means English dominance. Test language was a factor with two levels, English and Spanish.

Because 0 is the score for a perfectly balanced bilingual (an individual who obtained the same score on the English and Spanish version of the test), the intercept of the model shows that the mean PV score in both languages for a balanced bilingual was 81. Every one-point increase in English dominance was associated with a half-point increase on the English version and a one-point decrease on the Spanish version of the test. In other words, participants with higher English scores tended to have lower Spanish scores (see Figure 14). As might be expected when testing bilingual participants who live in a predominantly English environment, there were no participants with very strong dominance in Spanish; 8 out of 48 participants were dominant in Spanish and 3 participants were balanced.

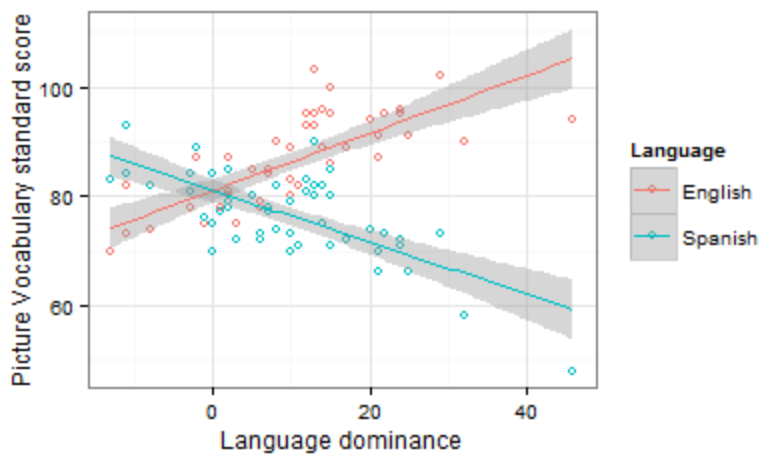


Figure 14. Relationship between language dominance and proficiency in English and Spanish. Language dominance was calculated by subtracting Spanish scores from English scores. Thus a positive score means English dominance and a negative score means Spanish dominance.

As is evident from Figure 14, there was great variance in the data with some participants being fairly balanced and others being clearly dominant in English. To see if some of this variance could be explained by biographical variables related to exposure to English and Spanish, further analyses were run. For these analyses, the bilingual sample was split into balanced bilinguals and English-dominant bilinguals. This split was done on the median, which was 10 (i.e., the English standard score was higher than the Spanish score by ten points). This resulted in 23 balanced and 25 unbalanced bilinguals.

First, it was investigated whether the two groups differed in their use of Spanish from birth through high school. Because all participants started out with more or less the same amount of exposure to Spanish, the question was whether the decline in Spanish exposure was faster for individuals who were later to become English dominant compared to those who remained more balanced. For this, a regression analysis was run with Percent exposure to Spanish as outcome variable and Age and Language dominance (balanced/unbalanced) as predictor variables. Because of the nonlinear decline in Spanish exposure as a function of age, age squared and cubed were also entered. For this analysis, age was treated as a continuous variables although it

technically was a factor with five levels. The results showed that age, age squared, and age cubed were significant predictors (linear term: $b = -322.0$, $SE = 24.3$, $p < .001$; quadratic term: $b = 78.9$, $SE = 24.3$, $p = .001$; cubic term: $b = 60.5$, $SE = 24.3$, $p = .013$). English-dominant bilinguals' exposure to Spanish was, on average, 10% lower than that of balanced bilinguals ($b = -10.6$, $SE = 2.2$, $p < .001$), but Language dominance did not interact with any of the polynomial terms ($|ts| < 1.2$, $ps > .243$), suggesting that the difference between groups remained constant. However, Figure 15 suggests a trend for a steeper decline in Spanish exposure in the English-dominant group. Whereas both groups did not differ significantly until age 5, the two groups started to differ from elementary school onwards (see Table 13). Table 13 also shows that the effect size increases as a function of age from a small effect in infancy to a large effect in middle school and high school. Other biographical variables shown in Table 13 confirm the same trend, although few of the other variables reach statistical significance. The table shows that parents of English dominant participants tended to use more English and were more proficient in English when participants were growing up compared to the balanced bilinguals. Furthermore, English dominant participants interacted with more English speakers during childhood. Balanced bilinguals tended to have participated more in transitional or bilingual programs when entering school compared to English dominant bilinguals, suggesting that these programs aided Spanish language maintenance. A correlation analysis showed that hours (square-rooted to account for outliers) in Spanish immersion programs was positively correlated with oral language ability in Spanish ($r(48) = .30$, $p = .038$) but not with English oral language ability ($r(48) = -.14$, $p = .326$).

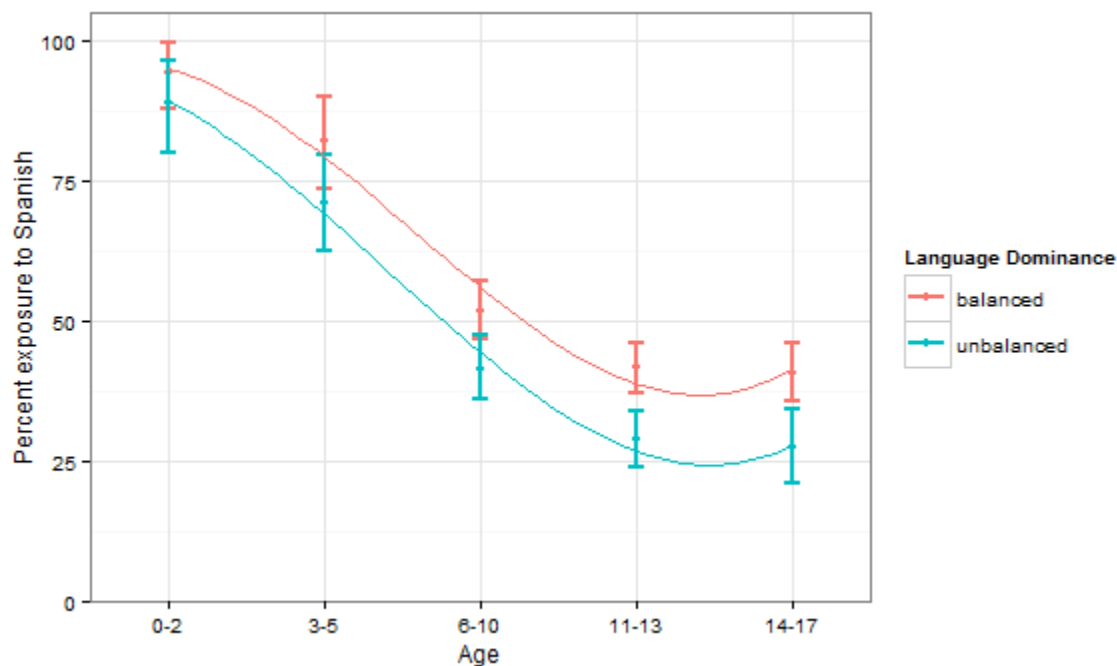


Figure 15. Relationship between percent of exposure to Spanish and age in the bilingual sample. Participants were divided into a balanced and an unbalanced group based on the difference between their Spanish and English score on the WMLS (see text).

Importantly, both groups did not differ in Age of Acquisition and mother’s education level, suggesting that these variables did not determine language dominance. Interestingly, though, groups differed in years of musical training. This may suggest greater integration of the English-dominant bilinguals’ families into the dominant culture but participants were not asked in what language they had received musical training so this explanation is only speculative. The difference may also be indicative of differences between parent characteristics. For example, one study found that length of musical training in 7-9 year olds was correlated with parental income (Corrigall & Schellenberg, 2015). If this was true in the present study, it may also indicate greater integration into the dominant culture.

Table 13. *Differences in background variables between balanced and unbalanced bilingual participants.*

Variable name	Balanced	English dominant	<i>t</i> -value	<i>d</i>
Percent exposure Spanish				
0-2 years	94.1% (15.0)	88.8% (20.9)	1.0	0.29
3-5 years	82.2% (20.9)	70.9% (22.3)	1.8 ⁺	0.52
Elementary school	51.7% (13.0)	41.6% (15.2)	2.5*	0.71
Middle school	41.7% (11.3)	28.8% (13.3)	3.6***	1.05
High school	40.9% (13.4)	27.4% (17.4)	3.0**	0.86
Mother's use English	6.7% (21.4)	8.2% (16.3)	0.3	0.08
Mother's proficiency English (1-10)	2.5 (2.4)	3.5 (2.5)	1.4	0.41
Father's use English	4.6% (12.0)	14.6% (24.7)	1.7	0.51
Father's proficiency English (1-10)	3.6 (2.9)	5.1 (3.1)	1.6	0.51
Number English speakers				
0-2 years	0.4 (1.2)	1.2 (2.1)	1.5	0.43
3-5 years	1.0 (1.4)	4.5 (5.8)	2.9**	0.82
Number Spanish Speakers				
0-2 years	6.6 (4.8)	5.1 (2.8)	-1.3	-0.37
3-5 years	7.5 (6.8)	5.8 (2.7)	-1.2	-0.34
Spanish Immersion program ($\sqrt{h \text{ total}}$)	28.2 (26.3)	15.0 (23.4)	-1.8 ⁺	-0.53
Mother's education level	1.8 (0.9)	1.8 (0.8)	0.2	0.07
Years of musical training	0.3 (0.8)	1.6 (2.8)	2.0*	0.58
Age of English Acquisition	4.7 (2.5)	4.1 (2.5)	-0.8	-0.23

Note. *** $p < .001$; ** $p < .01$; * $p < .05$; ⁺ $p < .1$. See text for an explanation of variables. Spanish Immersion program: Participants were asked how many hours per week of Spanish instruction they had received in bilingual and transitional programs. These were added up to the total number of hours, which were subsequently square-rooted to achieve a normal distribution.

What is the relationship between picture vocabulary and verbal reasoning? The above analyses showed that picture vocabulary in English and Spanish was associated with relative exposure to English and Spanish. On the other hand, verbal reasoning, measured by the verbal analogies subtest of the WMLS, involves higher order thinking skills, which may develop independently of relative language exposure. Several observations support this assumption. Scores on the English VA version were correlated with scores on the Spanish VA version ($r(48)$

= .42, $p = .003$). The PV scores, on the other hand, were not correlated between both languages ($r(48) = -.03$, $p = .822$). Also, current relative exposure and past relative exposure only explained 15.5% of the variance on VA test compared to 37.6% on the PV test (see previous analysis). Therefore, verbal reasoning may provide a better indication of a bilingual participant's actual verbal ability than PV. To test this hypothesis, the relationship between VA and PV was compared between the monolingual and bilingual participants.

The results of the regression analysis show that a bilingual matched with a monolingual participant on their verbal analogies score would, on average, perform 7.8 points lower on the picture vocabulary test compared to the monolingual participant ($b = 7.83$, $SE = 1.59$, $p < .001$). This relationship can best be seen in Figure 16.

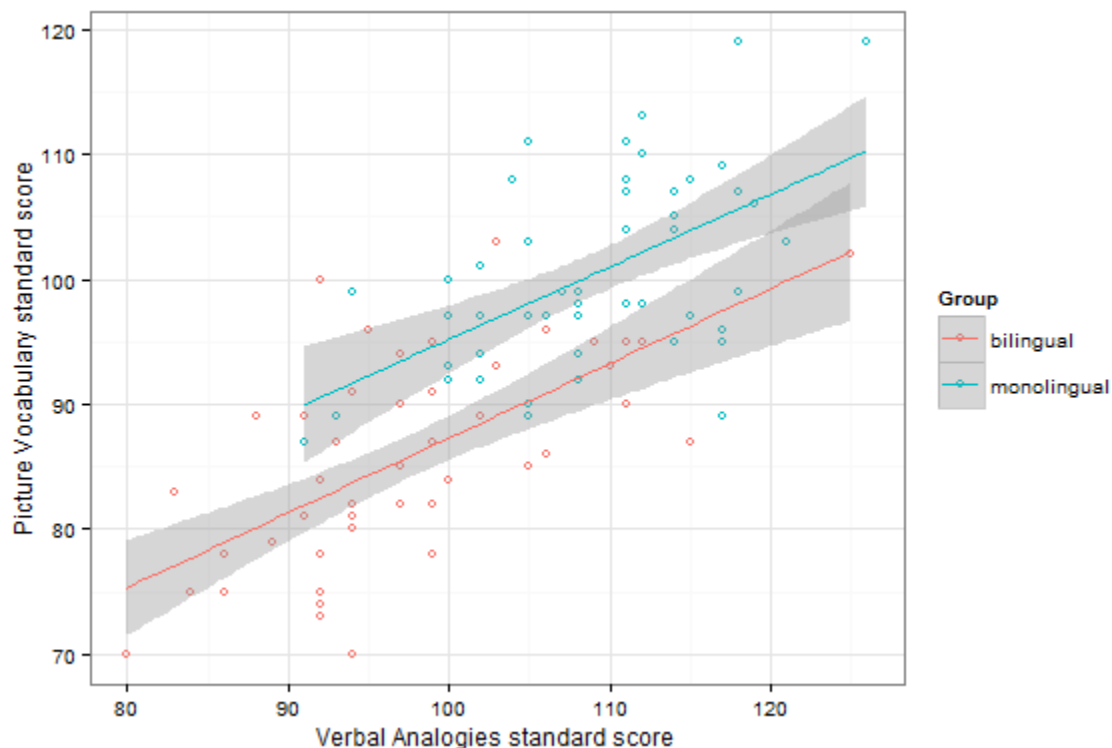


Figure 16. Relationship between the picture vocabulary and the verbal analogies subtests of the WMLS. Compared to the monolingual participants, bilinguals performed lower on the picture vocabulary test as would be expected from the verbal analogies score.

5.2.4 Discussion

The results reported here showed that monolinguals scored higher on both measures of the WMLS-R, PV and VA, compared to the bilinguals. The effect sizes were large, which may be surprising given that all participants were enrolled at a university and were matched on level of education. However, there were significant between-group differences in mother's education level, which is a commonly-used indicator of SES. The SES of the bilinguals was significantly lower than that of the monolinguals. SES has been shown to be associated with vocabulary knowledge (e.g., Farkas & Beron, 2004) and the link between SES and vocabulary knowledge is believed to be reflected in the way mothers from different SES interact with their children (Hoff, 2003).

SES cannot explain all differences between groups, though, because participants talked mostly Spanish at home and learned English at school or kindergarten. This may explain why SES was not a significant predictor of English language ability in the bilinguals. However, because SES was not associated with Spanish language ability, either, a more likely explanation is that the variance in the data did not permit finding an association with only 7 mothers having received any schooling beyond high school. A more nuanced measurement of SES may be necessary to find the association that is usually very robust. For example, information about the parents' occupation and annual income may be collected in addition to education level. Finding a greater range of SES, however, will likely remain difficult in the current population because many Spanish-English bilingual speakers come from immigrant backgrounds and are more likely to have received limited education. For example, Capps et al. (2005) report that in the year 2000 in the US, 32% of children of immigrants had parents with no high school degree compared to

9% of children of natives (parents born in the US). This shows that the distribution of SES in the present study is not uncommon in this population.

Despite the differences in mother's education level between the two groups, SES is unlikely to be the only explanation for the observed differences. The regression analyses showed that proficiency in English and Spanish was closely related to the amount of language exposure in each language. Language exposure, or amount of parental verbal input directed to the child, is a significant predictor of vocabulary growth in children who grow up monolingual (e.g., Huttenlocher & Haight, 1991; Weisleder & Fernald, 2013) and so it is reasonable to assume that the same holds true for bilingual children (Hoff et al., 2012; Hurtado et al., 2013). And because bilingual children are exposed to two languages, they hear each language less often compared to a monolingual child with the same overall amount of language input. Recent evidence suggests that it is not only the amount of language exposure but also the number of speakers a child interacts with that predicts language proficiency (Gollan et al., 2014). In the present study there was also some evidence for this relationship. The number of speakers a participant regularly interacted with at age 3-5 explained variance above and beyond his or her relative exposure to Spanish. The variance explained by this variable was 4%, which is less than in Gollan et al., who reported that frequency of exposure explained 26% and number of speakers an additional 10%. A difference between the Gollan et al. study and the present is that in the present study, the mean number of people a participant interacted with from birth through high school was not a significant predictor but only the number of speakers in childhood (Gollan et al. only asked participants to estimate the number of speakers they regularly spoke to from birth through high school). One reason for this difference in findings may be that some participants in the present study overestimated the number of people they regularly spoke to. Several participants indicated

20 or more once they entered school, which may not be realistic. Another possibility is that the number of speakers a person interacts with in childhood is more important than later in life. But because of the retrospective nature of the data, more evidence would be needed to confirm this hypothesis. The present results also do not preclude the conclusion that more speakers just equaled more input. For example, a child that grows up with two parents and older siblings may receive more input than a child growing up with a single parent. However, Gollan et al. conducted a more controlled experiment with children in which they carefully counted the number of hours of exposure in the heritage language (Hebrew) and the number of speakers through parental report. In their experiment, the number of speakers was still a significant variable (also see Place & Hoff, 2011), suggesting independent contributions from amount of input and the number of interactions with different speakers.

The effect of number of speakers fits well with the broader hypothesis of this dissertation that differences between monolinguals and bilinguals on verbal tasks result from differences in the precision of phonological representations. Frequency of exposure strengthens phonological representations. This is why pictures with high frequency labels are named with greater accuracy than those with low frequency labels (Gollan et al., 2008). Hearing input from more diverse speakers may help children learning a language to form more exact representations of phoneme categories. For example, Maye, Werker, and Gerken (2002) found that infants are sensitive to the statistical distribution of phoneme exemplars. Hearing input from a greater variety of speakers will provide more evidence what the mean and the allowable variance of a phoneme category is (Rost & McMurray, 2009). A different view poses that listeners store exemplars of words every time they encounter a word; phoneme categories emerge from the accumulated evidence of stored exemplars (Pierrehumbert, 2003). A finding from the infant literature is that infants at 14

months of age confuse similar-sounding words such as *bih* and *dih* on a word learning task (Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002; but see Yoshida, Fennell, Swingley, & Werker, 2009). Rost and McMurry (2009) replicated the finding of Werker et al. (2002) in their Experiment 1 with the words /buk/ and /puk/, showing that 14-month-olds failed to discriminate between the two words. However, in Experiment 2 they used the same task with the same words but recorded tokens from 18 different speakers. This time infants were able to distinguish the two words. When measuring VOT of /b/ and /p/ across all exemplars, the authors found considerable variation among speakers and this may have provided infants with information about the category boundary. In contrast, when infants receive input from only one speaker, they may be less confident that /b/ and /p/ are two different phonemes as opposed to two exemplars of the same category. Thus receiving input from multiple speakers may lead to more precise phonological representations of words.

In addition to the infant literature, there is evidence from adult vocabulary acquisition studies that suggest that speaker variability aids in learning new words. Sommers and Barcroft (2011) present evidence for the representation quality hypothesis. This hypothesis states that acoustic variability is beneficial for learning new words because it leads to a more distributed mental representation of the new word. As in a previous study (Barcroft & Sommers, 2005), words were learned with greater accuracy when they were presented by six speakers as opposed to one speaker. In addition, Sommers and Barcroft (2011) found that recognition of words learned from multiple speakers was more robust under adverse listening conditions. These findings suggest that phonological representations of newly learned words became more precise through greater talker variability.

The results also showed that oral language ability was associated with frequency of exposure to each language. Frequency of exposure may act in two ways. Because many words are tied to specific circumstances, bilingual participants may encounter those words in only one of their languages. For example, many bilingual participants were not able to name a picture of a high chair in English. Because most participants only spoke Spanish at home, they may have never heard the word in English. Consistent with this explanation is the finding that while bilingual children know fewer words in each of their languages compared to monolingual children, the total number of words they know is equal to monolingual children (Hoff et al., 2012). Another explanation may be that participants had heard the word for high chair before but they had not encountered the word sufficient times to be able to recall it. This explanation is consistent with the observed bilingual disadvantage in tip-of-the-tongue (TOT) states (Gollan & Acenas, 2004; Gollan & Silverberg, 2001). Gollan and colleagues have shown that bilinguals suffer more TOTs compared to monolinguals. Because TOTs are more common for low frequency words than high frequency words, Gollan and colleagues suggest that the reason for the bilingual disadvantage in lexical retrieval is a frequency effect; that is, all words in each language are less frequent because they are encountered less frequently by someone who speaks two languages (see section 1.4.3). Also consistent with this explanation is the finding that the gap between receptive and productive vocabulary is larger in bilinguals compared to monolinguals (Gibson, Oller, Jarmulowicz, & Ethington, 2012; Gibson, Peña, & Bedore, 2014). Knowledge of a word may be sufficiently precise to recognize a word and match it with a picture but not precise enough to produce it when presented with a picture.

With regard to language dominance, an interesting picture emerged. Language dominance was correlated with language proficiency so that more English dominant participants

were more proficient in English and less proficient in Spanish compared to less English dominant participants (see Figure 14). In fact, only four participants scored within 1 SD of the mean of the normative sample of both the English and Spanish versions of the test. This suggests a trade-off between English and Spanish proficiency. Because proficiency in English and Spanish was closely related to exposure to each language, it may be difficult for bilinguals to achieve and maintain high proficiency in two languages. The results also suggest that language dominance in young adulthood can be predicted relatively early in life. Already in elementary school did balanced and English dominant participants differ in English exposure by 10% points. With the caveat that all biographical data were based on retrospective self-report, the results suggest that increased exposure to the heritage language through immersion programs may be effective for heritage language maintenance but children may also need increased support in the L2 to not stay behind in their language development. At the same time, it may be unrealistic to expect bilinguals to perform equivalently to monolinguals on language tests when language maintenance is the goal of a bilingual speaker.

Lastly, one interesting finding was that verbal reasoning in English and Spanish was correlated while picture vocabulary was not. In addition, picture vocabulary was more strongly associated with language exposure. This suggests that verbal reasoning skills transfer from one language to the other. Furthermore, when compared to monolingual speakers, bilinguals performed lower on the picture vocabulary test than would be expected based on their verbal analogies score. These findings have important practical implications for bilingual language assessment in schools. Because bilingual children usually have less language exposure to each of their languages and thus perform less well on verbal tests, they are more likely to be diagnosed with having a language disorder (Paradis, Genesee, & Crago, 2011). Testing them with a verbal

analogies test may therefore be a better indicator of actual language development that is independent of amount of exposure in each language (although the total amount of language exposure and the quality of interactions remain important, of course).

5.3 Working memory

Previous studies found that verbal working memory (VWM) may be reduced in bilinguals as a function of language proficiency (Dellcenserie & Genesee, 2013; Gutiérrez-Clellen, Calderón, & Ellis Weismer, 2004; Luo et al., 2013; Ratiu & Azuma, 2015). As discussed in Chapter 2, the connection between VWM and language proficiency may be the quality of phonological representations in LTM. For example, high frequency words are remembered better on STM tests than low frequency words (e.g., Hulme et al., 1991). In the same way, more proficient speakers may have overall stronger phonological representations. As a result, they may have to devote fewer attentional resources to retrieving and maintaining those representations on a WM test and can thus devote more resources to the processing part of the WM task.

5.3.1 Materials and procedure

The Working Memory test used for this study comes from the NIH Toolbox. Just as the WIN, it was administered over the internet. In the WM test, participants see pictures and their labels and hear their names (in English). The set-size differs from two to seven pictures. Pictures are either animals or food items. After each set of pictures, participants are asked to repeat what they just saw in size order from smallest to biggest. For example, if they saw a bear, a duck, and an elephant, they would say duck, bear, elephant. To establish the size order, participants have to pay attention to the size of the object on the screen but in most cases, the relative proportions on the screen correspond to real life. The test has two parts. In the first part, sets consist of only animals or only food items. In the second part, sets consist of animals and food and participants

are asked to repeat the food first, from smallest to biggest, and then the animals, from smallest to biggest. Both parts start with two practice sets to ensure that participants understood the directions. If they made a mistake in either practice set, the instructions were repeated and the set was administered again. After the practice items, the test starts with a set size of two. If a participant correctly repeats all picture labels, the set size of the next trial increases by one. If the participant makes an error, another set of the same size but with different items is administered. Testing stopped when a participant could not correctly repeat two sets in a row or when the last set was administered. Responses were recorded on a paper sheet and a score for each participant was calculated by counting the total number of items of all correctly repeated sets. Thus the total score for each part is 27 (2+3+4+5+6+7) and the total possible score is 54. This test was only administered in English.

Recently, the reliability of the test was established (Tulsky et al., 2014). The test-retest intraclass correlation coefficient was .77. The test also correlated with other established WM tests ($r = .57$) and tests of executive function ($r = .43 - .58$) from a standardized cognition battery (see Tulsky et al., 2014). The correlation with a test of receptive vocabulary, on the other hand, was low ($r = .24$). Also interesting with respect to the present study was the finding that Hispanic participants scored, on average, .41 SDs below Caucasian participants.

5.3.2 Results

The monolingual group ($M = 37.6$, $SD = 8.0$) scored higher than the bilingual group ($M = 32.4$, $SD = 7.9$) and this difference was significant ($t(99) = 3.29$, $p < .001$, $d = 0.66$). The next question was whether this difference would still be significant when the picture vocabulary score was included as a covariate. A regression analysis showed that PV was a significant predictor ($b = 0.40$, $SE = 0.13$, $p = .002$), showing that 1 point increase on the PV standard score scale was

associated with an increase in WMC of 0.4 points. The factor Group was no longer significant ($b = 5.75, SE = 17.70, p = .746$) and neither was the interaction between Group and vocabulary ($b = -0.06, SE = 0.19, p = .738$), suggesting that vocabulary knowledge fully accounted for the differences between groups. This is further illustrated in Figure 17. The model explained 22% of the variance and was significant ($F(3, 97) = 9.19, p < .001$).

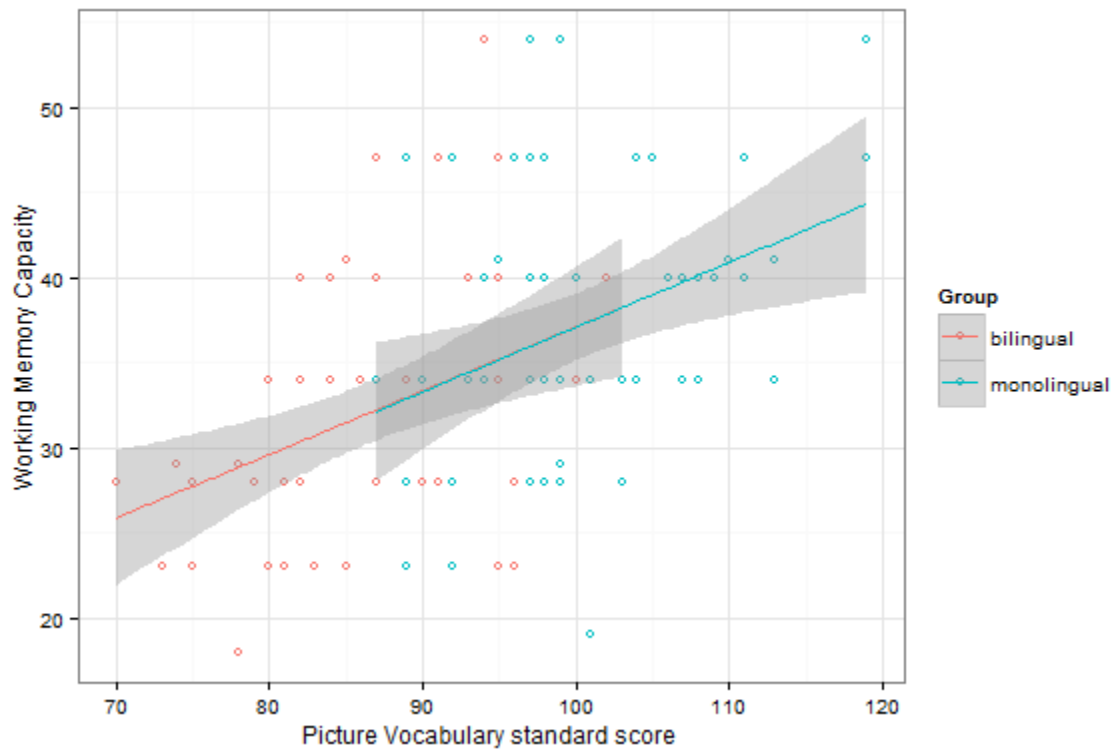


Figure 17. Relationship between working memory capacity and picture vocabulary scores. Grey-shaded area shows the 95% confidence interval of the regression line.

5.3.3 Discussion

The results confirmed the hypothesis that VWM is related to vocabulary knowledge. While the differences between groups were significant, vocabulary knowledge could fully account for these differences. This suggests that bilinguals did not have generally lower WMC but performed more poorly on the WM test as a group because of their lower vocabulary knowledge in English. The direction of this relationship could go in either direction. For one, a

lower WMC may lead to a smaller vocabulary because WMC may be involved in vocabulary acquisition (Baddeley et al., 1998). Conversely, a larger vocabulary may subserve WM via more precise phonological representations in LTM. A third possible explanation is that the relationship may be bidirectional. The first explanation is unlikely because it would suggest that bilinguals had a smaller general WMC than monolinguals. However, general WMC has been shown to be constrained by neural limitations (Vogel & Machizawa, 2004) and is therefore unlikely to be influenced by bilingualism. Indeed, when vocabulary knowledge was regressed on WMC, the residual variance was exactly the same for monolinguals and bilinguals (see Figure 18). This finding contrasts with Luo et al. (2013) who found that monolinguals still scored higher than bilinguals on a VWM test after accounting for differences in vocabulary knowledge. The different results in this study and the present one may be due to the type of vocabulary knowledge tested. Luo et al. tested receptive vocabulary whereas bilinguals in the present study completed a test of productive vocabulary (as mentioned in the Materials section, Tulskey et al. 2014, also did not find a correlation between receptive vocabulary and WM scores in the norming sample). Productive vocabulary may be more indicative of the quality of phonological representations because they can be less precise for recognition memory.

The present results lend further support to the hypothesis that the quality of phonological representations is the main reason for differential performance of monolinguals and bilinguals on verbal tasks. Importantly, the same relationship between vocabulary knowledge and WMC was seen in bilingual and monolingual participants. These findings have implications for studies employing VWM tests to predict performance on other cognitive or perceptual tests. If vocabulary knowledge is not controlled for, it is not clear whether an observed effect is truly caused by WMC or verbal ability. One solution to this problem would be to use more than one

test of WMC measuring different modalities (e.g., visual WM, VWM, spatial WM) and calculate a composite score based on the shared variance between the tests (Conway et al., 2005; Kane et al., 2004). The results also have important implications for teaching second language speakers. Teachers have to bear in mind that English Language Learners with a more limited vocabulary will have greater difficulty following lectures because of a more limited capacity to maintain verbal information in memory.

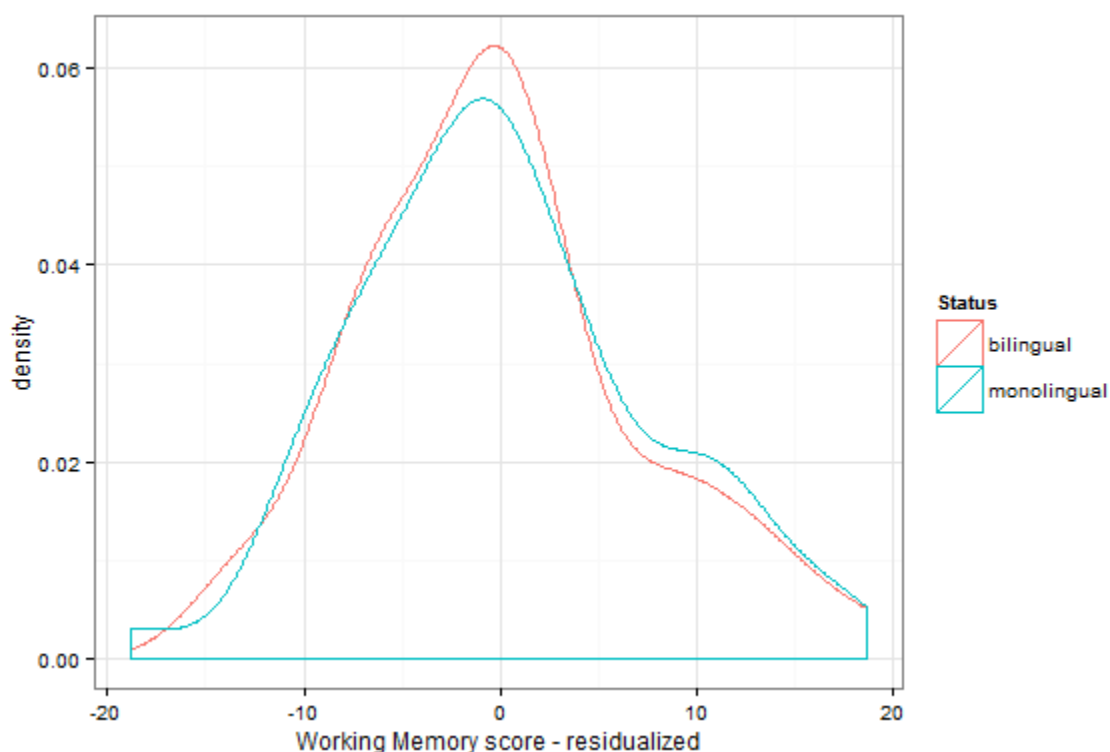


Figure 18. Distribution of working memory scores when the effect of picture vocabulary was partialled out (residual variance).

5.4 Consonant perception in noise

The next test in this test battery was a test of consonant perception. There were two research questions associated with this test: first, do monolinguals and bilinguals differ in the accuracy of consonant perception, and, second, what factors can explain these differences?

Consonant perception in a second language may be influenced by the phoneme inventory of the

first language (Cutler, Garcia Lecumberri, & Cooke, 2008; Cutler, Weber, Smits, & Cooper, 2004; Garcia Lecumberri & Cooke, 2006). Plosives in Spanish and English differ in VOT so that an English /b/ can sound more like a Spanish /p/. Also, Spanish does not have the consonants /f/ and /ɬ/. <v> and represent one phoneme in Spanish with two allophonic realizations, /β/ and /b/. Likewise, /s/ and /z/ are allophones. It was therefore hypothesized that the bilingual participants may experience interference from Spanish, especially since they heard consonants decontextualized, that is, without language cues. For example, English /aba/ may be heard as /apa/. In addition, it was also hypothesized that accuracy would be correlated with vocabulary knowledge in English. Exemplar theory (Pierrehumbert, 2003) proposes that phonetic categories are refined by type statistics in the lexicon, that is, top-down information can influence perception. Thus individuals with a larger lexicon may possess more refined phonetic categories that guide them in perception. For example, /d/ and /b/ differ on many different dimensions such as formant transition, burst amplitude, spectrum, and the ratio of the closure to the voice onset time (Pierrehumbert, 2003, p. 120). Because there is redundant information, representations can be relatively coarse without affecting perception. However, more refined representations may be beneficial under adverse listening conditions, when some of the information such as formant transitions is overshadowed by a competing acoustic signal.

5.4.1 Materials and Procedure

In the consonant perception test (CP), participants heard 16 different consonants in a /VCV/ cluster and were asked to identify them by clicking on one of 16 options on the computer screen. The consonant recordings came from Shannon, Jensvold, Padilla, Robert, and Wang (1999). The original recordings done by Shannon and colleagues included 25 consonants in three different vowel contexts /u/, /a/, and /i/ in medial /VCV/ position and initial /CV/ position.

Following Garcia Lecumberri and Cooke (2006), stimuli were reduced to 16 consonants (/p b t d k g tʃ f v s z ʃ m n l r/) in only one vowel context (aCa) and one consonant position. Two male speakers of standard American English (M2 and M3 from Shannon et al., 1999) were chosen from the original set of 5 male and 5 female speakers and each token was repeated four times for a total of 128 items. The experimental items were mixed with background noise (multi-talker babble) taken from the original SPIN recording. Three different sections from the babble noise track were cut and mixed at a SNR of -4 dB in Praat (Boersma & Weenink, 2014). One of those babble segments was repeated once and the other two were played once. The SNR was chosen based on a pilot study. Participants in the pilot study performed at about 85% accuracy at an SNR of -2 dB. To avoid ceiling effects, the SNR was lowered to -4 dB in the present study. Participants also heard each token in silence at the beginning of the experiment so they could adapt to the pronunciation of each speaker. These trials were only used as practice trials and were not scored. When a participant made a mistake on those practice trials, the same token was repeated until the participant made a correct response.

5.4.2 Results

Mean accuracy for monolinguals was 76.9% ($SD = 5.4$) and for bilinguals 66.9% ($SD = 9.1$). A logistic mixed-effects regression model with subjects and items as random effects showed that this difference was significant, indicating that monolinguals were overall more accurate than bilinguals ($b = 0.65$, $SE = 0.12$, $p < .001$). Additional factors were added to the model to establish whether the two different speakers and the three different babble segments had an effect on recognition accuracy and whether the effect was the same or different for mono- and bilingual participants. Speaker 1 was easier to identify than speaker 2 ($b = -0.78$, $SE = 0.08$, $p < .001$). Speaker interacted with Babble segment ($b = 0.31$, $SE = 0.11$, $p = .005$), showing that

the benefit for Speaker 1 was smaller when paired with babble segment 3 (see Figure 19).

Importantly, Speaker and Babble segment did not interact with Group, suggesting that the effects were the same for both groups.

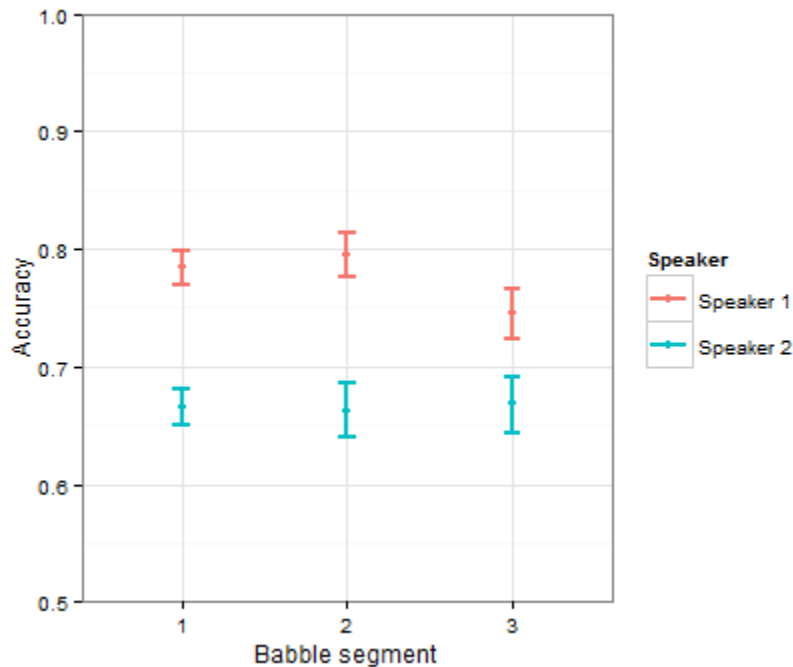


Figure 19. Mean accuracy on the consonant perception test divided by babble segment and speaker. Whiskers show the 95% confidence interval. Note the limited range of the y-axis to highlight the effects.

The next question was whether the monolingual benefit extended over all consonants or was specific to certain consonants only. Figure 20 suggests that performance differed depending on the consonant. First, those consonants that are the same in both languages were recognized with the same accuracy (/tʃ/, /m/, and /n/). In addition, the voiceless plosives /k/, /p/, and /t/ were recognized with the same accuracy by both groups. The largest differences existed for those consonants for which VOTs in English and Spanish overlap (/b/, /d/, and /g/) and those that are allophonic in Spanish (/s/ and /z/, and /v/). Lastly, /f/ was misidentified more often by bilinguals compared to monolinguals, which was not predicted based on native language influence.

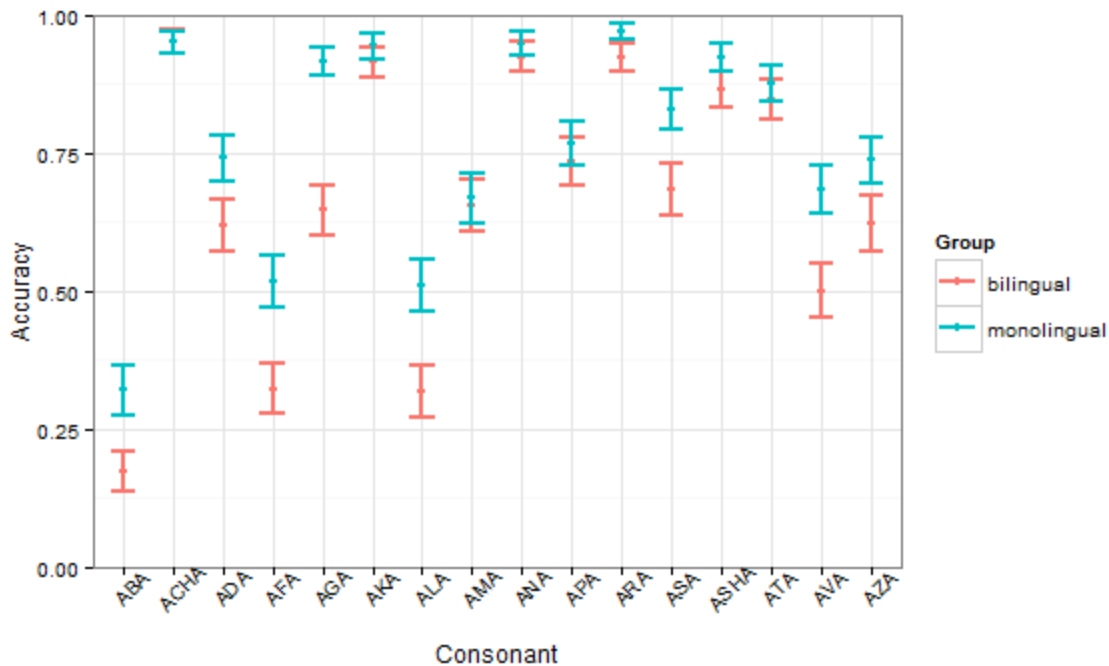


Figure 20. Mean accuracy for each consonant on the consonant perception test. Whiskers show the 95% confidence interval.

The matrices in Table 14 and Table 15 show the average percentage of correct responses (diagonal bolded figures) and which consonant was most often heard when participants did not identify the correct one. If the first language interfered with correct recognition of the English phonemes, then bilinguals should have chosen the consonant that would be predicted based on Spanish phonology more often than monolinguals. For example, the VOT of English /b/ is more similar to a Spanish /p/ so there should be more *apa* responses in the bilingual group compared to the monolingual group. To test whether groups differed in their responses when the target consonant was not correctly identified, a χ^2 analysis was performed. A significant result shows that group differences in the ratio of responses to a certain consonant to the total number of incorrect responses is greater than would be expected by chance. The results for select consonants, those for which we would expect a native-language influence, are shown in Table 16.

Table 14. *Confusion matrix - bilingual participants.*

Consonant		stimulus															
		b	tf	d	f	g	k	l	m	n	p	r	s	ʃ	t	v	z
response	missing	1	1	1		2	1	1		1	1	1	1	1	1	1	1
	b	17		1	8			7	5		1	1				17	
	tf	0	95	2		2	6							11	10		
	d	10		62		1		5	1	2	1				1	6	11
	f	5		1	32			2	1							1	
	g	1	1	12		65		2	1	1	2	1			1	1	3
	k			4		29	91		1		10				1		
	l	16		1	1			32	3	1			1			5	7
	m	2			5			13	66	1						1	
	n	3		2		1		1	7	92					1	1	10
	p	27		1	40	1	1	12	6		73	2				12	
	r	1				1		1				92					
	s	1			1			1					68	1		2	2
	ʃ		2	1									5	86	1		2
	t	1		12							11				85		
	v	16		1	11			24	9			2				50	1
	z	1		1								1	23			2	62

Note. Columns indicate the consonant that was played and rows indicate the response that participants gave. All values are shown as percentages. Values below 1% are not shown, which is why not all columns add up to 100%. Missing = missing response.

Table 15. *Confusion matrix - monolingual participants.*

Consonant		stimulus															
		b	tf	d	f	g	k	l	m	n	p	r	s	ʃ	t	v	z
response	missing	1	1		1	1	1	2	1	1			1	1	2	1	1
	b	32			14	1		4	6		5	1				12	
	tf		95	1			4							4	4		
	d	2		74						1					4		4
	f	6		1	52			1	1		1					1	
	g		3	13		92											2
	k					5	94				9						
	l	23		2	1			51	4		1		1			8	13
	m	1			7			8	67	1						1	
	n	3		2				1	8	95					1		5
	p	3			8	1		1	2		77					1	
	r							1	0			97					
	s												83	1			1
	ʃ		1										2	92			
	t			2							4				88		
	v	27			15			30	11		2					68	1
	z	1		2									11			4	74

Note. Columns indicate the consonant that was played and rows indicate the response that participants gave. All values shown as percentages. Values below 1% are not shown, which is why not all columns add up to 100%. Missing = missing response.

Table 16. *Typical consonant confusions by monolingual and bilingual participants.*

Target consonant	Misidentified Consonant	Misidentified/Total wrong		χ^2	φ
		Bilinguals	Monolinguals		
/b/	/p/	103/318	12/288	78.3***	0.36
/b/	/v/	62/318	113/288	28.7***	0.22
/d/	/t/	46/146	7/110	24.2***	0.31
/g/	/k/	112/136	21/36	9.4**	0.23
/s/	/z/	90/122	45/72	2.7 ⁺	0.12
/v/	/b/	67/192	52/134	0.5	0.04
/v/	/p/	47/192	6/134	23.2***	0.27
/ʃ/	/tʃ/	43/52	19/32	5.6*	0.26
/f/	/p/	154/260	34/205	86.6***	0.43
/f/	/v/	43/260	63/205	13.1***	0.17
/f/	/b/	32/260	60/205	20.8***	0.21
/l/	/v/	91/262	126/208	31.2***	0.26
/l/	/p/	47/262	6/208	26.3***	0.24

*** $p < .001$, ** $p < .01$, * $p < .05$, ⁺ $p < .1$.

Note. The table shows how many times a target consonant was misidentified as another consonant compared to the total number of misidentification. The χ^2 -test tested whether the ratio was significantly different between groups and φ shows the effect size of the difference.

The results suggest that native language influence can explain some of the confusions.

For the voiced consonants /b/, /d/, and /g/, bilinguals were more likely to choose the voiceless counterparts than monolinguals. The influence of the merging of /b/ and /v/ in Spanish can also be observed. Both /v/ and /b/ were confused with /p/. However, bilinguals were less likely to confuse /v/ with /b/ than monolinguals. Also, /s/ and /z/ were not more confusable for bilinguals than monolinguals, contrary to what may be expected based on Spanish phonology.

Monolinguals were more likely to confuse /f/ with /v/ or /b/ and bilinguals were more likely to confuse it with /p/. Because /f/ is produced very similarly in English and Spanish, these results suggest that monolinguals and bilinguals may have attended to different cues in the signal rather than L1 influence. The pattern of these results is strikingly similar to those reported in Garcia

Lecumberri and Cooke (2006), who also tested native speakers of Spanish (albeit European Spanish). However, differences between the present study and theirs were observed for the consonant /f/. The L2 speakers in Garcia Lecumberri and Cooke attained high accuracy for /f/ in noise (92%) and did not typically confuse it with /tʃ/ (2% of responses). This may be because many of their participants also spoke Basque, a language that has the /f/ sound. For other sounds, both monolingual and bilingual speakers were less accurate in the present study. This was true for /l/ and /z/. For example, in Garcia Lecumberri and Cooke, native English speakers reached 97% accuracy for /z/, compared to 74% in the present study. It may be that these differences are attributable to the different noise maskers used in the present study and the fact that Garcia Lecumberri and Cooke used all five male speakers from Shannon et al. (1999) with two repetitions per speaker whereas the present study only used two speakers with four repetitions of each consonant.

The next question was whether English proficiency would be associated with CP test performance. One possibility is that knowing two languages interferes with consonant perception when consonants share overlapping spaces such as Spanish /p/ and English /b/, which may lead to intermediate category boundaries that are unlike those of monolingual speakers of either language. In this case, English proficiency may not correlate with performance. However, some studies have shown that bilinguals are able to shift their category boundaries depending on language mode (Antoniou, Tyler, & Best, 2012; Elman, Diehl, & Buchwald, 1977; Garcia-Sierra, Diehl, & Champlin, 2009). For example, in Elman, Diehl, and Buchwald (1977) bilinguals listened to five tokens on a /b-/p/ continuum, with VOT ranging between -69 to +66 msec. The authors created an English and a Spanish version with the same test syllables but filler words and prompts were either in English or Spanish to put participants in the respective

language mode. The results showed that the same stimulus was identified more often as /b/ in the English context than in the Spanish context for strong (balanced) bilinguals but weak (unbalanced) bilinguals did not show this shift as a result of language mode. Nonetheless, even the performance of strong bilinguals was different from monolinguals in either language, suggesting that bilinguals may be unable to completely turn one language off, as it were, when listening in the other language. Elman and colleagues also assessed proficiency in each language through an oral interview and degree of bilingualism (L1 proficiency/L2 proficiency) was correlated with the size of the category shift ($r = .52$). This suggests that proficiency is related to perception accuracy. The prediction was therefore that higher English proficiency would be associated with more native-like (monolingual) consonant perception.

To address the role of proficiency in consonant perception, mean accuracy was calculated for each participant and the result was used as the outcome variable in a linear regression analysis. Group and English proficiency (oral language ability) were the predictor variables. A visual inspection of the data suggested that the relationship between proficiency and accuracy was not linear (see Figure 21). Rather, the effect on CP was stronger in the lower proficiency range. Therefore, proficiency was entered as a cubic spline with 2 degrees of freedom. Results showed that both terms of the spline function (first term: $b = 0.31$, $SE = 0.06$, $p < .001$; second term: $b = 0.11$, $SE = 0.04$, $p = .010$) and Group ($b = 0.04$, $SE = 0.02$, $p = .017$) were significant predictors. The effect of group shows that after proficiency in English was taken into account, the difference between groups was 4% points. This was smaller than the 10% difference between group that was found above. The model explained 46.3% of the variance (adj. $R^2 = .447$). Group by itself explained 32.0% and proficiency by itself 43.0% of the variance. This suggests that proficiency was a better predictor of performance on the test than Group. Because a spline

function was used, a breaking point was imposed by the function. This point was at the median of 99.5, suggesting that the steepness of the slope differed for individuals below and above this point. This can be seen in Figure 21. Because most participants above the break point were monolinguals, this may suggest that the relationship between CP and proficiency was stronger in bilinguals than monolinguals. Thus separate analyses for each group were run. For monolinguals, the model was not significant ($F(2, 49) = 1.6, p = .216, R^2 = .061$) but for bilinguals it was ($F(2, 45) = 10.1, p < .001, R^2 = .309$). However, some of the variance is lost when aggregating data and so a logistic mixed-effects model was also run on the raw data. The disadvantage is that these models do not provide a R^2 statistic that would allow for model comparisons but the estimates are likely more accurate because error attributable to subject and item variance is taken into account.

Subjects and items and items nested within subjects were entered as random effects to account for the fact that each subject heard each item four times spoken by two speakers and contributed 128 data points. As in the previous analysis, the first and second term of the spline function for language proficiency were significant (first term: $b = 2.32, SE = 0.44, p < .001$; second term: $b = 1.14, SE = 0.30, p < .001$), as was Group ($b = 0.25, SE = 0.12, p = .041$). When the model was run for each group separately, proficiency was a significant predictor in both groups (monolinguals only: first term: $b = 0.90, SE = 0.44, p = .039$; second term: $b = 0.74, SE = 0.39, p = .011$; bilinguals only: first term: $b = 2.17, SE = 0.58, p < .001$; second term: $b = 1.26, SE = 0.44, p = .004$).

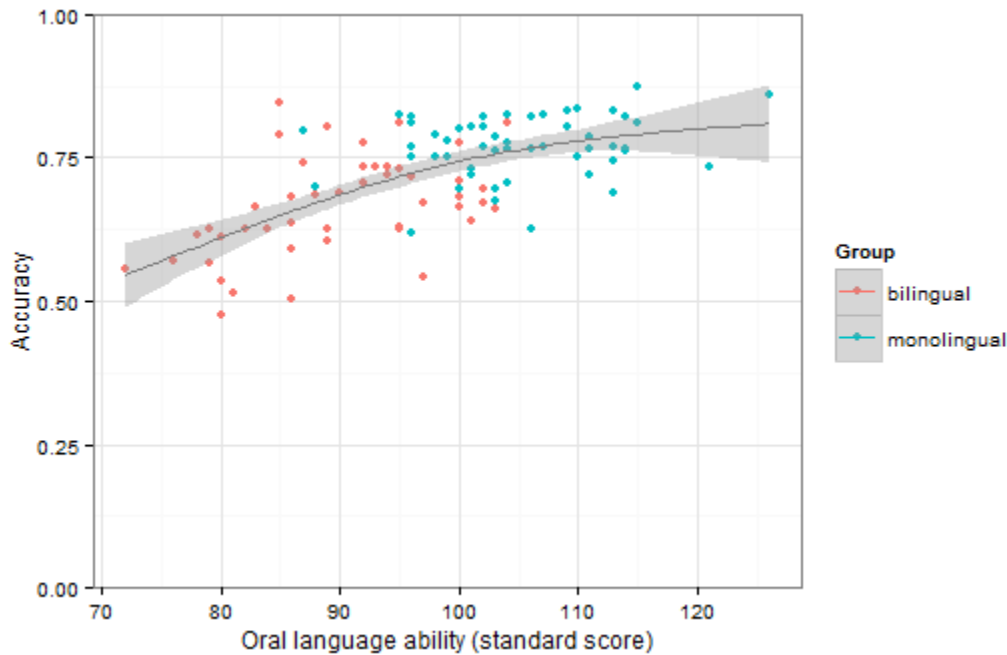


Figure 21. Relationship between accuracy on the consonant perception test and oral language ability. The regression line included one knot at 99.5.

To illustrate the role of proficiency further, each group was divided into high and low proficiency based on a median split of oral language proficiency. A *t*-test showed that the monolingual low and the bilingual high group were not significantly different in language proficiency ($M_{\text{monolingual low}} = 527$ W, $M_{\text{bilingual high}} = 525$ W, $t(51) = 0.97, p = .339, d = 0.27$). Therefore, any differences between those groups are likely not attributable to differences in English proficiency but other factors such as L1 influence. After establishing these four groups, another mixed effect regression analysis was run with group as a predictor variable with four levels (monolingual high/low, bilingual high/low). The results indicated that the bilingual high group was significantly different from the bilingual low group ($b = -0.54, SE = 0.14, p < .001$) and both the monolingual low ($b = 0.34, SE = 0.14, p = .011$) and the monolingual high group ($b = 0.57, SE = 0.15, p < .001$). When the monolingual low group was used as the reference category, they were not significantly different from the monolingual high group ($b = 0.22, SE = 0.14, p = .106$). This suggests that differences in consonant perception still persist even when

groups are matched on proficiency (i.e., monolingual low and bilingual high) but those differences become smaller (see Figure 22).

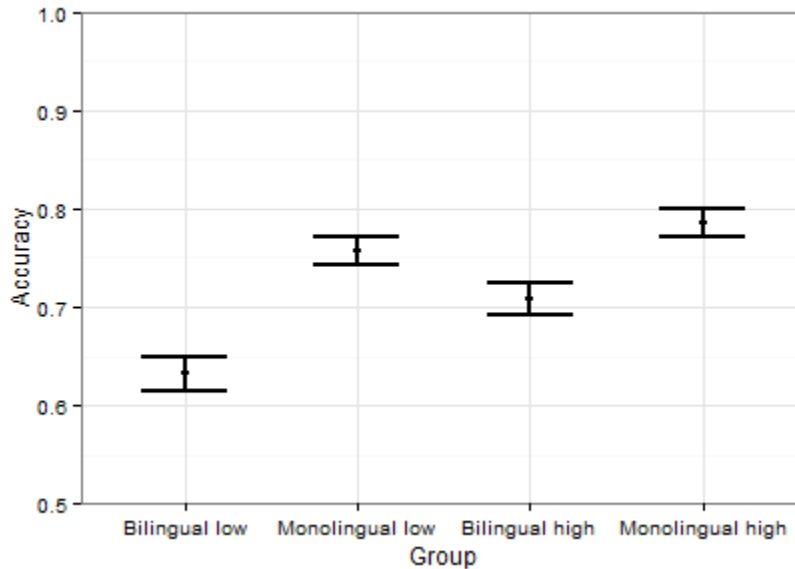


Figure 22. Accuracy on the consonant perception test as a function of group. The monolingual and bilingual groups were each divided into a high and low proficiency group based on a median split of their verbal ability score. Whiskers show the 95% confidence interval.

The results for each consonant are shown in Figure 23. The figure shows that whereas the bilingual low group and the monolingual high group perform significantly differently for most consonants, the bilingual high and the monolingual low group perform more similarly. Differences still exist for some consonants (/g/ and /l/), which may suggest a native language influence for those consonants.

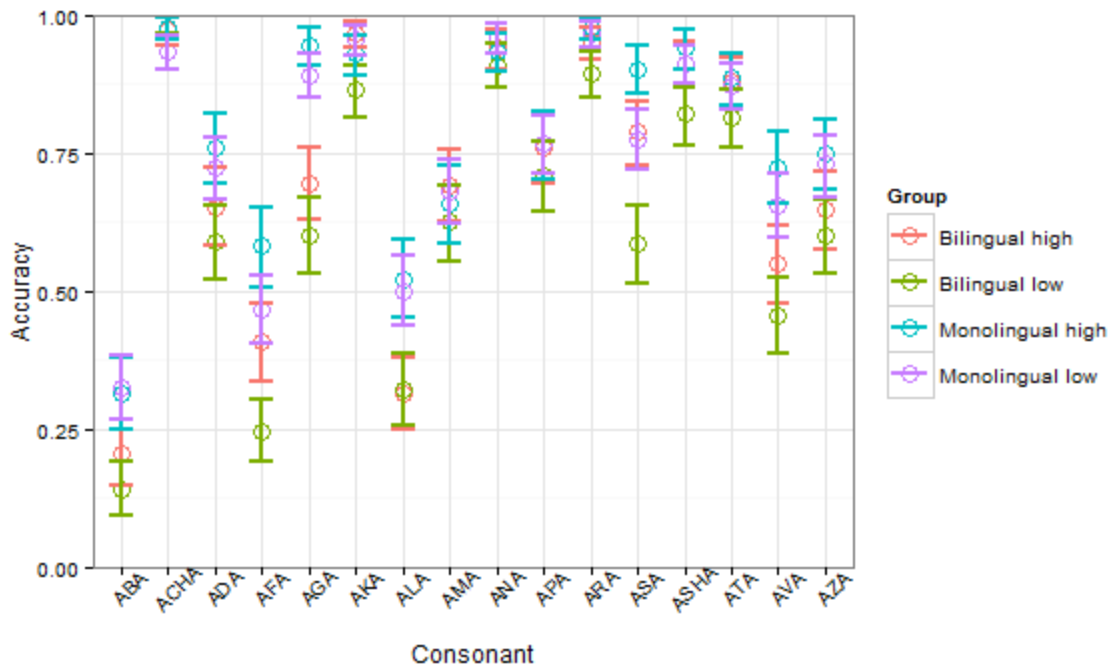


Figure 23. Mean accuracy for each consonant on the consonant perception test. The monolingual and bilingual groups were each divided into a high and low proficiency group based on a median split of their verbal ability score. Whiskers show the 95% confidence interval.

The results so far suggest a relationship between language proficiency and consonant perception in noise. The hypothesis of this dissertation is that a larger vocabulary results in more precise phonological representations in long term memory. Likewise, assuming that phonetic categories are extracted from the phonetic information stored in the entire mental lexicon, a larger vocabulary should result in more precise phonetic categories, which would be more robust to the effect of noise.

To test this hypothesis, the phonotactic probability of each of the 16 consonants (only the probability of the consonant in the VCV cluster was considered) was calculated using the phonotactic probability calculator (Vitevitch & Luce, 2004). The results showed that phonotactic probability was not normally distributed ($M = 0.021$, $SD = 0.025$, $Median = 0.009$). To account for this skew, phonotactic probability was divided into *high* and *low* probability based on a median split. The prediction was that consonants with higher phonotactic probability would be

recognized with greater accuracy. In addition, we may expect individuals with a larger vocabulary would be more sensitive to phonotactic probability and thus be more accurate on VCV clusters with low phonotactic probability. The reason is that the probabilities based on a corpus analysis will only roughly correspond to experienced probabilities. For subjects with less language experience, low probability clusters will be of even lower experienced frequency. As in the case of the frequency effect described in section 1.4.3, we may therefore expect an interaction between phonotactic probability and English proficiency.

A mixed-effects regression model was run with subjects and items and items nested within subjects as random effects. As before, the results showed a main effect of oral language ability ($b = 0.23$, $SE = 0.07$, $p = .002$) and Group ($b = 0.31$, $SE = 0.13$, $p = .014$). Importantly, the interaction between language ability and phonotactic probability was significant ($b = 0.16$, $SE = 0.07$, $p = .015$). Because language ability was centered, the main effect of phonotactic probability shows the estimated effect for a participant with mean language ability, which was not significant ($b = -0.80$, $SE = 0.67$, $p = .231$). These effects can be best interpreted by looking at Figure 24.

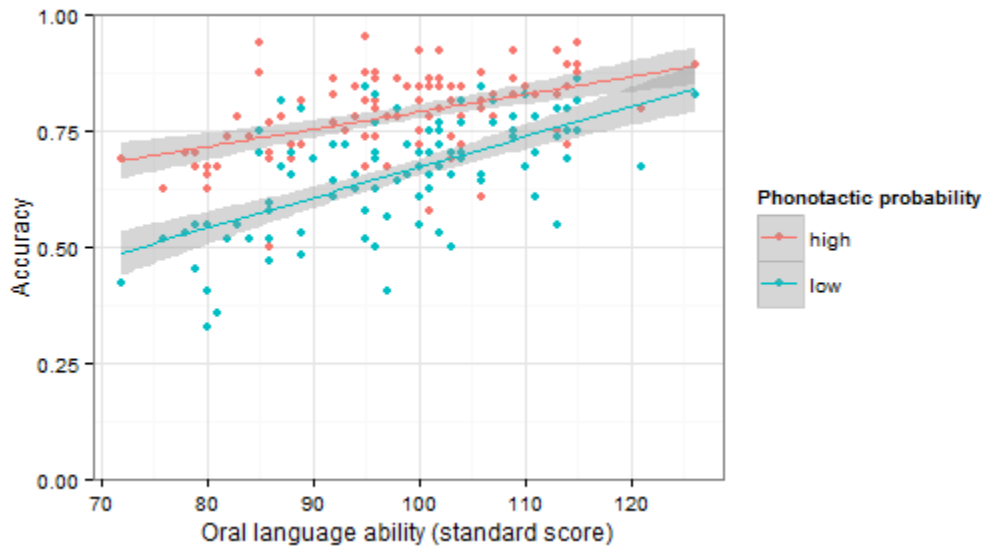


Figure 24. Relationship between mean accuracy on the consonant perception test and oral language ability. Consonant were divided into high and low phonotactic probability based on a median split. The interaction between phonotactic probability and language ability was significant.

5.4.3 Discussion

The results from the consonant perception test showed that bilinguals performed significantly differently from monolinguals with an effect size of about 10% points. The pattern of consonant confusions resemble those reported in Garcia Lecumberri and Cooke (2006) for Spanish native speakers who had learned English as a foreign language. As in their study, bilingual participants in the present study often misperceived the voiced consonants /b/, /g/, /d/, and /v/ as voiceless /p/, /k/, and /t/. This suggests a native language influence on L2 perception even for early bilinguals¹⁴. However, the present study extends the results of Garcia Lecumberri and Cooke by showing that the effect of L1 influence becomes smaller as proficiency in the tested language increases (Figure 23). Importantly, the relationship between proficiency and

¹⁴ Garcia Lecumberri and Cooke did not report detailed information about their participants' age of L2 acquisition and L2 proficiency but the participants lived in Spain, which suggests more limited exposure to English.

accuracy was also found for the monolingual speakers to a certain extent. This suggests that differences between monolinguals and bilinguals cannot solely be attributed to L1 influence.

Two possible explanations for the effect of proficiency come to mind. Higher language proficiency may be associated with more precise phonetic categories and/or individuals with higher language proficiency may be better at attending to those acoustic cues that penetrate the background noise. Both explanations are consistent with a usage-based view of phonetic categories (Pierrehumbert, 2001, 2003). According to this view, mental representations are “gradually built up through experience with speech” (Pierrehumbert, 2001, p. 137). As individuals gain more experience with a language and hear more words in a wider range of contexts, their phonetic categories of those sounds that distinguish meaning in the language become more refined (also see Hardison, 2012). At the same time, individuals may learn to attend to those cues in the speech signal that are most informative, especially when the speech signal is not optimal. For example, aspiration is a good cue in English to distinguish voiced from voiceless plosives (although the main cue is VOT, Flege & Eefting, 1987). However, Spanish does not have aspiration so native speakers of Spanish need to learn to attend to this cue. Not attending to aspiration as a cue may explain why bilinguals often chose /p/ where monolinguals were more likely to hear /v/ or /b/ (see Table 16). At a general level, the effect of language proficiency is also in line with Flege’s speech learning model (Flege, 1995), which states that new, nonnative phonetic categories can be established with increased language experience.

The results provided some evidence that language ability - specifically vocabulary knowledge - is directly related to consonant perception in noise. Individuals with a larger vocabulary were less influenced by phonotactic probability. This effect is interpreted best by an entrenchment account of phonetic categories (Pierrehumbert, 2001). More frequent phonemes

are better entrenched than less frequent phonemes. The effect of proficiency is small for high-probability phonemes because these are well-entrenched for all speakers. However, the low-probability phonemes are less entrenched in speakers with a smaller vocabulary, leading to an interaction between vocabulary size and CP.

Bilinguals showed signs of L1 influence when listening to English consonants although they learned English early in life and were mostly immersed in an English-speaking environment (all participants attended school in the US from first grade). This resembles findings from Sebastián-Gallés and colleagues who found that early Spanish-Catalan bilinguals had difficulty distinguishing between a Catalan vowel contrast nonexistent in Spanish (Sebastián-Gallés, Echeverría, & Bosch, 2005; Sebastián-Gallés & Soto-Faraco, 1999). In the present study, the differences between bilinguals and monolinguals were attenuated when English language proficiency was considered but even a subset of monolinguals and bilinguals matched on proficiency still performed significantly differently from each other. Results from other studies, though, have shown that bilinguals are able to shift phonemic categories depending on the language mode they are in. Antoniou et al. (2012) tested early Greek-English bilinguals on stimuli involving voiced and voiceless consonants as those have a shorter VOT category boundary in Greek. The results showed that the bilinguals were able to shift their category boundaries depending on language context. For example, when in Greek mode they perceived a Greek /p/ most often as /p/ but when in English mode, they were more likely to hear it either as /b/ or as /p/. However, Anoniou et al.'s study employed ideal listening conditions. The results from the present study differ insofar as stimuli were presented in noise. This may reveal more subtle differences in perception, especially in cases where bilinguals and monolinguals rely on different phonetic cues. It should be noted, though, that although only English was used in the

experiment (Spanish was not used until all English tests were completed), the task gave no language context cues. Putting bilinguals into a stronger monolingual mode by providing a context cue such as a carrier sentence for each token might have changed the results. In contexts without strong language cues it may even be beneficial to have more intermediate phonetic boundaries to accommodate language switches. Another reason may be frequent exposure to accented English. One eye-tracking study (Ju & Luce, 2004) found that when listening to Spanish, Spanish-English bilinguals only exhibited cross-language activation (as measured by eye movements to English competitor pictures) when VOT of Spanish words was manipulated to be consistent with English. For example, when VOT was Spanish-like, participant did not look to a picture of pliers more than to a control picture when hearing *playa*. When VOT was English-like, on the other hand, participants looked more to the pliers than to the control picture. Thus Ju and Luce's (2004) study showed that lexical access in bilinguals is constrained by language specific cues such as VOT. Bilinguals who are frequently exposed to accented English may thus treat /b/ and /p/ or /d/ and /t/ as allophonic variants for the purposes of lexical access (cf. Samuel & Larraza, 2015, p. 67). For example, bilinguals might frequently hear /t/ as in /ten/ with a VOT acceptable for Spanish /t/ but more akin to English /d/. The boundary from /t/ to /d/ is around 85 ms in English but around 19 ms in Spanish (Flege & Eefting, 1986). Consequently, the category boundary from /t/ to /d/ may shift to allow shorter VOTs as acceptable for English /p/. Or speakers frequently exposed to Spanish-accented English may ignore VOT as a cue altogether because of its unreliability and may rely more on context. For example, in some r-less New York City dialects, the vowels in the words *source* and *sauce* have nearly merged. Speakers who produce two different vowels in these two words are nevertheless not able to reliably indicate which one they heard, presumably because of the great variability of this vowel distinction in the

speech community (Pierrehumbert, 2003, p. 138). The same may be true for Spanish-English bilingual speakers regarding those consonants whose category boundaries overlap in English and Spanish but further research is necessary to corroborate this hypothesis.

Despite the L1 influence on L2 speech perception, the results showed clearly that differences do not only exist between monolinguals and bilinguals but also within monolinguals. The relationship between vocabulary knowledge and speech perception could be bidirectional given that previous studies have found a relationship between speech discrimination ability and vocabulary development in infants (Tsao, Liu, & Kuhl, 2004). Nevertheless, the present results suggest that differences in speech perception between monolinguals and bilinguals may be less categorical than previously thought. One striking result of the present study is the large difference in vocabulary knowledge between groups, which amounted to 1 SD (see Table 1). Given such differences, monolingual college students may not be a good comparison group. Especially Figure 21 suggests that individual differences in speech perception get smaller as language proficiency increases. Thus differences between monolinguals and bilinguals in speech perception might wrongfully be attributed to bilingual status when in fact differences are in fact attributable to differences in language experience in general.

5.5 Test of Attention in Listening

The main purpose for including the TAIL in this test battery was that previous research has indicated that attentional control, or executive functions, may be recruited when listening under adverse conditions. As outlined in the ELU model (Rönnberg et al., 2013), word recognition is effortless when the speech signal is optimal. However, when the signal is distorted in some way, listening becomes effortful and requires additional attentional resources. A secondary purpose of the study was to test the hypothesis that bilingualism improves attentional

control, often referred to as the bilingual advantage (e.g., Bialystok, Craik, & Luk, 2012; Hilchey & Klein, 2011). This second hypothesis will be explored in this section. At first, it seems unrelated to the topic of this dissertation; however, it has been proposed that there is a relationship between attentional control and language processing in bilinguals (Abutalebi et al., 2013; D. W. Green, 1998; Mercier, Pivneva, & Titone, 2013; Pivneva, Palmer, & Titone, 2012). Because of this literature, it was hypothesized that individual differences in language experience may be associated with individual differences in attentional control.

5.5.1 The bilingual advantage

One of the first studies to report a bilingual advantage was Bialystok, Craik, Klein, and Viswanathan (2004). These authors administered the Simon test, a test that is designed to measure inhibitory control. Inhibitory control is the ability to suppress a prepotent response in the presence of response conflict. In one version of the test, participants press a right or left arrow depending on the direction of an arrow they see on a computer screen. Response conflict arises when a left-pointing arrow appears on the right side of the screen and the other way round. Compared to trials without response conflict, that is, a right-pointing arrow on the right side of the screen, RTs in conflict trials are usually larger, referred to as the Simon effect. Bialystok et al. (2004) found that the Simon effect was much smaller for bilinguals than monolinguals, suggesting that bilingualism may be associated with better inhibitory control.

One explanation for this advantage is the bilingual's need to control access to both languages when speaking in one language and that the constant recruitment of these domain general attentional networks improves nonlinguistic tests of executive function. Costa, Hernández, and Sebastián-Gallés (2008) pointed out that all theories of bilingual lexical access involve some type of control mechanism. In Green's (1998) inhibitory control model, for

example, translation equivalents become active when a bilingual person accesses a word in one language. For instance, when accessing the concept of DOG, the word forms *dog* and *perro* receive activation in a Spanish-English bilingual speaker. In Green's model, these word forms have language tags attached to them and the form with the wrong tag is inhibited. This inhibition mechanism may be the same as the one recruited during tasks used to measure attentional control (Bialystok, Craik, & Luk, 2008).

However, this hypothesis has recently come under criticism. In a review of the bilingual advantage literature, Hilchey and Klein (2011) came to the conclusion that there is no consistent evidence for a bilingual advantage in inhibitory control, but there is a bilingual advantage in general attentional control with bilinguals often being faster on conflict and nonconflict trials. Therefore, Hilchey and Klein (2011) concluded that inhibition of the irrelevant language during bilingual speech production may not be an adequate explanation of the bilingual advantage. Since this review, several studies have been published that did not find any evidence for a bilingual advantage (Antón et al., 2014; Duñabeitia et al., 2014; V. C. M. Gathercole et al., 2014; Paap & Greenberg, 2013), which has some researchers led to question the reliability of the effect (e.g., de Bruin, Treccani, & Della Sala, 2015; Klein, 2015; Paap, 2015). For example, it has been suggested that differences in SES (Morton & Harper, 2007) and immigrant status (Kousaie & Phillips, 2011) can explain purported bilingual advantages (these studies come from Canada where immigrants often have a high SES). One way forward to resolve these conflicting results may be to relate performance on tests of executive function to bilingual experience in a correlational design with a more homogeneous group of bilingual participants in terms of SES and other background variables. For example, one study found that the degree of bilingualism (dominant language proficiency divided by the nondominant language proficiency) was

positively associated with the age of diagnosis of Alzheimer's disease in a sample of low-educated Spanish-English bilinguals (Gollan, Salmon, Montoya, & Galasko, 2011).

5.5.2 Methods

5.5.2.1 Materials

The Test of Attention in Listening (TAIL) was adapted from Zhang, Barry, Moore, and Amitay (2012). In this test, participants have to decide whether two tones were played to the same ear or different ears. What makes this test challenging is that the frequency of the two tones is sometimes the same and sometimes different. Because participants are only supposed to respond based on the location of the tones, response conflict arises on trials on which the location is different but the frequency the same or the location the same but the frequency different. The manipulation of frequency and location results in four conditions, same-frequency same-location (SFSL), same-frequency different-location (SFDL), different-frequency same-location (DFSL), different-frequency different-location (DFDL). The original test also has a second condition where frequency is the task-relevant dimension and location is the irrelevant dimension that has to be ignored. However, only the first condition was used in the present study to reduce the time needed to administer the test.

Three different measures can be derived from the TAIL, baseline RT, involuntary orientation, and conflict resolution. Baseline RT is the mean RT in the SFSL condition. In Zhang et al. (2012), baseline RT correlated with the RTs in a separate test that did not involve response conflict and therefore the authors suggested that this measure reflects information processing speed. Involuntary attention can be calculated by subtracting RTs on trials with the same frequency from those of different frequency ($[DFDL+DFSL] - [SFSL+SFDL]$). Conflict resolution can be calculated by subtracting the mean RTs on trials where location and frequency

were both different or both the same (no response conflict) from those where they were different ([SFSL+DFDL] – [SFDL+DFSL]).

The tones were created in Praat (Boersma & Weenink, 2014) as pure tones with a length of 100 ms. The frequency ranged between 500 and 1400 Hz in 100 Hz intervals, which resulted in ten different sound files. There were a total of 96 experimental trials, 24 trials in each condition. The experiment was programmed in E-Prime.

5.5.2.2 Procedure

Participants were seated in front of a computer and were given written and oral instructions for the experiment. They were told that they would hear two tones and then would decide whether the two tones were played to the same ear or different ears. They were also told to ignore the frequency of the two tones and just pay attention to location. For their responses, participants used the keys Q and P on the keyboard and they were encouraged to respond as fast and as accurately as possible. The experiment started with 16 practice trials for which participants received automated feedback from the computer. If a participant did not get 85% accuracy on these test trials, the instructions were repeated and the participant did another round of 16 practice trials. Most participants reached the accuracy criterion in the first round and everyone else in the second. On each trial, a sound file was randomly chosen. For same-frequency trials, the same sound file was played twice and for the different-frequency condition, the second sound file was randomly chosen so that the difference in frequency was at least 100 Hz.

5.5.3 Analysis

For the accuracy data, a logistic mixed-effects model (Bates et al., 2014) was run with subjects as random effect and Group (monolingual/bilingual), Frequency (same/different), and

Location (same/different) as fixed effects. For the RT data, only correct trials were used and the model included the same random and fixed effects as the previous one. Of particular interest was the interaction between Group and Frequency and Group and Location. One participant from the bilingual group was excluded because of low accuracy (60%).

5.5.4 Results

Is there a bilingual advantage? Accuracy on the test was high ($M = 96.3\%$, $SD = 18.8$, range = 87.5% - 100%). The result of the regression model showed that compared to the SFSL condition, participants were less accurate when Frequency was different ($b = -0.95$, $SE = 0.19$, $p < .001$) but this effect was attenuated when both Frequency and Location were different ($b = 0.66$, $SE = 0.22$, $p = .003$; see Figure 25). The Frequency by Group interaction was also significant, showing that bilinguals were less distracted by a different frequency ($b = 0.45$, $SE = 0.19$, $p = .045$). All other main effects and interactions were not significant ($|z| < 1$). The Frequency by Group interaction is shown in Figure 26. The figure suggests that the interaction arose from the fact that the difference between same and different trials was larger for monolinguals than bilinguals.

Next, RTs were investigated. Compared to the SFSL condition, responses were slower when Frequency was different ($b = 72.0$, $SE = 7.5$, $p < .001$) and when Location was different ($b = 51.7$, $SE = 7.5$, $p < .001$). These effects were attenuated when both Frequency and Location were different ($b = -74.2$, $SE = 8.8$, $p < .001$; see Figure 27). Group interacted with Frequency, showing that the effect of Frequency was smaller in bilinguals ($b = -19.7$, $SE = 8.9$, $p = .026$), as illustrated in Figure 28.

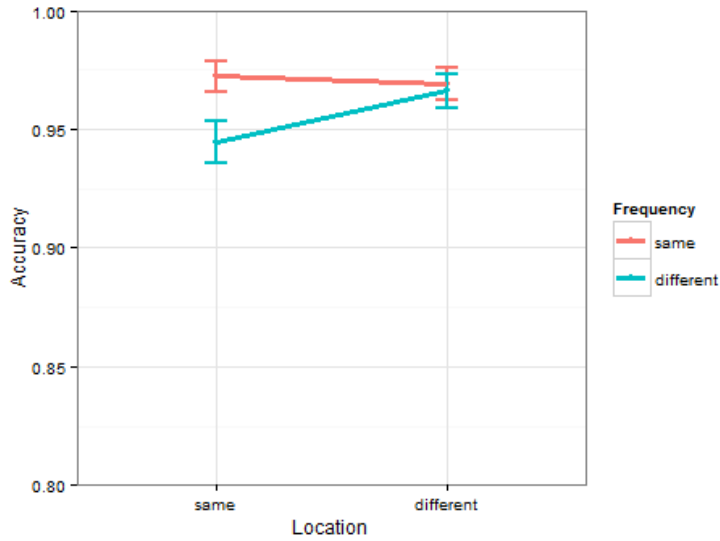


Figure 25. Mean accuracy on the TAIL in each of four conditions. Whiskers show the 95% confidence interval. Note the limited range of the y-axis to highlight the effect.

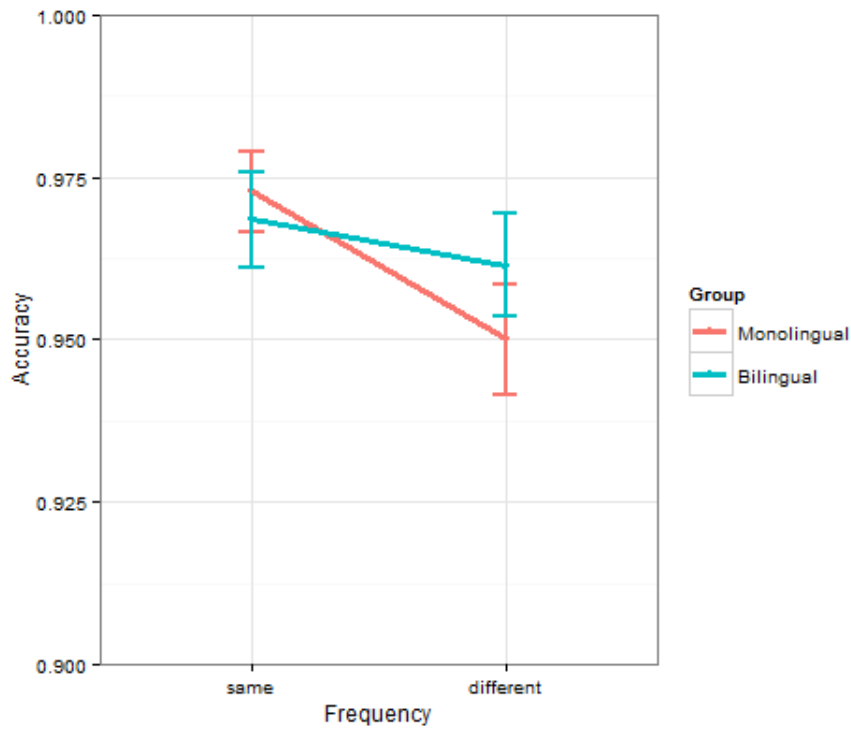


Figure 26. Mean accuracy on the TAIL for monolinguals and bilinguals. The difference between same frequency and different frequency trials was larger for monolinguals than for bilinguals. Whiskers show the 95% confidence interval. Note the limited range of the y-axis to highlight the effect.

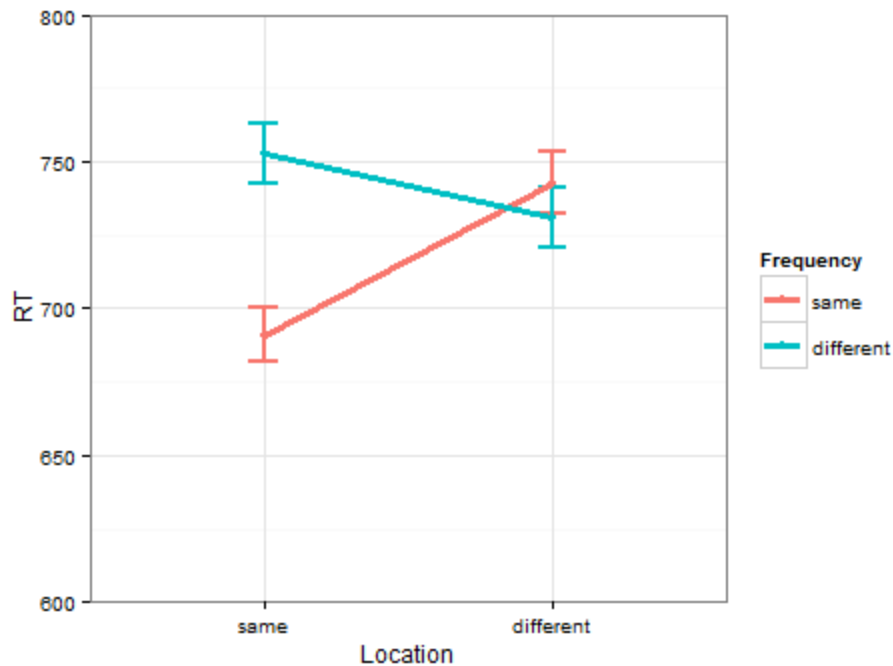


Figure 27. Mean response time (RT) on the TAIL in each of four conditions. Whiskers show the 95% confidence interval.

So far the results seem to show that monolinguals and bilinguals performed differently on some aspect of the test. Monolinguals were faster than bilinguals when the frequency was the same for both tones and slower when the frequency was different. This gave rise to an interaction between Frequency and Group. As described in the Methods sections, there were two versions of the experiment, with one half of each group using Q for *same* responses and P for *different* responses, and the other way around for the other half of each group. To test whether experiment version had an effect on the results, RTs were plotted separately for version 1 and 2. Figure 28 shows that monolinguals were faster in the DFDL and SFDL condition on version 1 than on version 2. All other 95% CIs overlap, which suggests that performance was similar in both versions. When the previous model was rerun including an interaction with test version, the effects changed. Because of the complexity of the results, they are reported in Table format. Table 17 shows that the Location effect was larger for bilinguals compared to monolinguals on

version 1 but smaller on version 2. The Frequency by Group interaction, on the other hand, was only present on version 2, with bilinguals showing a reduced effect.

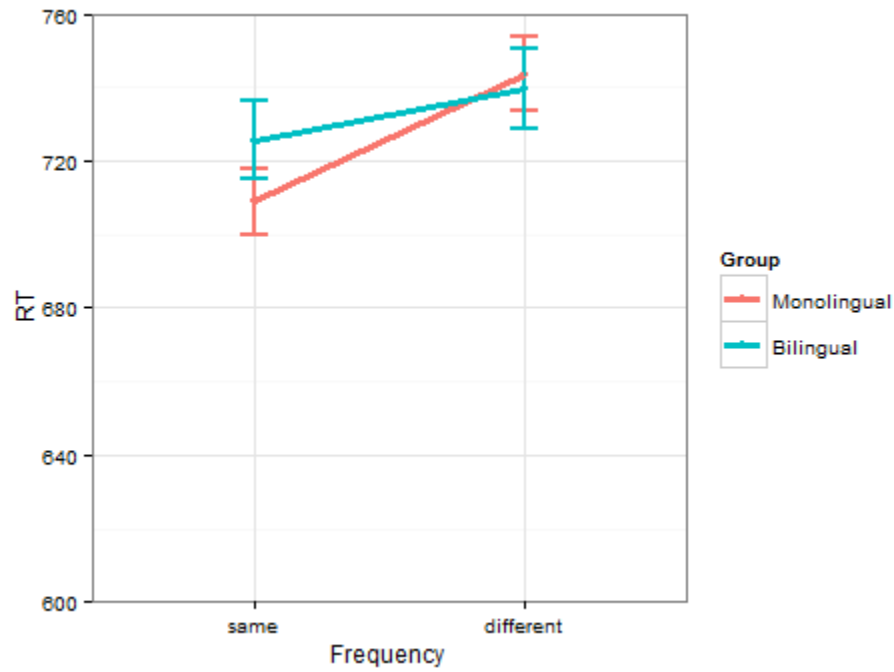


Figure 28. Mean response time (RT) in msec. on same and different frequency trials. Whiskers show the 95% confidence interval. Note the limited range of the y-axis.

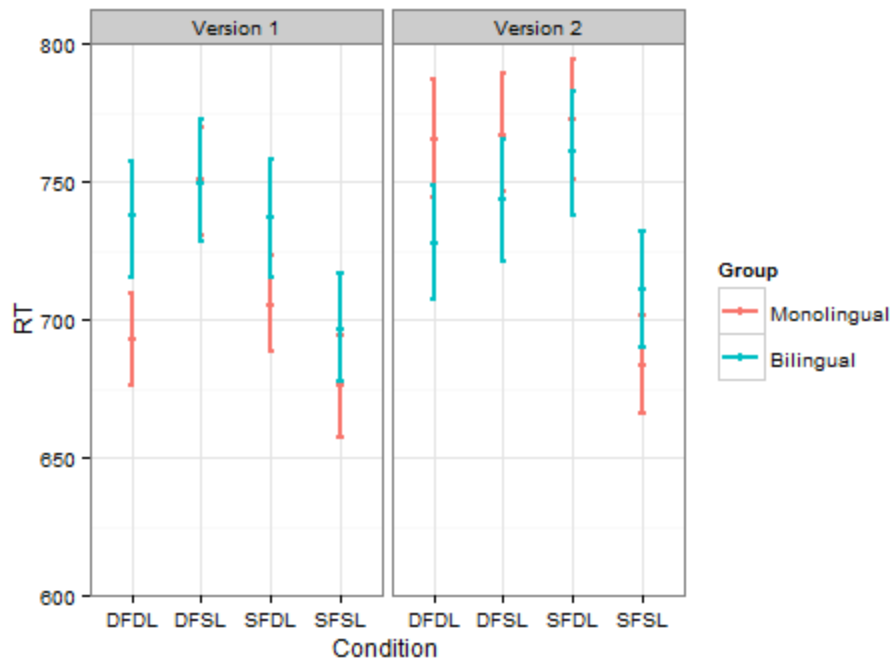


Figure 29. Mean response times (RT) in msec. in each of the four conditions of the TAIL. DF/SF = different/same frequency, DL/SL = different/same location. The difference between Version 1 and 2 was the location of the response keys (see Methods section in text).

Table 17. *Results of the regression analysis of TAIL response times.*

Effect	Beta	SE	p
Intercept (baseline = SFSL condition)	680.1	26.2	< .001
Frequency (baseline = same)	65.1	10.5	< .001
Location (baseline = same)	21.6	10.4	.039
Frequency*Location	- 70.4	12.2	< .001
Group (baseline = monolingual)	10.0	37.5	.791
Test version (baseline = version 1)	5.6	37.4	.882
Frequency*Group	- 1.5	12.2	.903
Location*Group	27.3	12.2	.026
Frequency*Test version	13.9	15.0	.357
Location*Test version	61.5	15.0	< .001
Group*Test version	12.2	54.2	.821
Frequency*Location*Test version	- 7.8	17.7	.661
Frequency*Group*Test version	- 38.5	17.7	.030
Location*Group*Test version	- 53.0	17.7	.003

Note. SFSL = same frequency, same location condition. Frequency and Location were variables with two levels, same and different. Group had two levels, monolingual and bilingual. Test version had two levels, version 1 and version 2.

A question that arises regarding the result is whether the frequency manipulation was successful. When the first and the second tone had a different frequency, the difference could

vary between 100 Hz and 900 Hz. If this manipulation was successful, then a greater difference should have led to a larger frequency effect. To test this, the difference in frequency between the first and second tone was entered as a continuous predictor into a mixed-effects regression model. For this, only the trials for which the frequency was different were included. A visual observation of the data suggested that a third-order polynomial would best fit the data since the relationship between RTs and the difference in frequency was not linear (see Figure 30). The results showed that the difference in frequency was a significant predictor (linear term: $b = 645.2$, $SE = 216.6$, $p = .003$; quadratic term: $b = -396.4$, $SE = 217.1$, $p = .068$; cubic term: $b = -458.8$, $SE = 216.8$, $p = .034$), but the standard errors show that there was quite some uncertainty in these estimates. The interactions with Location, Test version and Group were not significant.

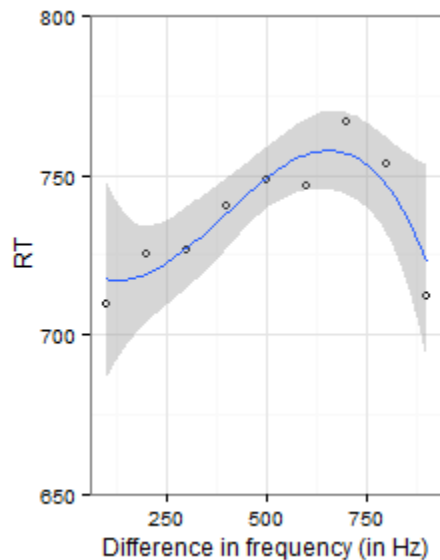


Figure 30. Effect of frequency difference between the first and second tone on response times (RT) in msec. The regression line shows the best fit with a polynomial function with three terms.

Is bilingual experience related to variance in attentional control?

Language dominance was used as a continuous variable in lieu of bilingual experience with the assumption that more balanced bilinguals would have greater language experience in each of their languages. Dominance was calculated by subtracting the English oral language

score from the Spanish oral language score. The resulting variable was normally distributed with a mean of 12.6 and a standard deviation of 14.5. Zero indicates that an individual was equally proficient in English and Spanish, negative values indicate greater proficiency in Spanish and positive values greater proficiency in English. Thus most participants were dominant in English as is typical for bilinguals who live in a mostly English monolingual environment.

For the analysis, a mixed-effects regression model on the RTs was run with Frequency, Location, and Language dominance as main effects and their interaction. The results are summarized in Table 18 and are graphically displayed in Figure 31. There was a main effect for Language dominance, with more English dominant participants being overall faster. In addition, Language dominance interacted with Location with a significantly larger Location effect the less balanced a bilingual was. For the accuracy data, there was a negative effect for Language dominance, with more English dominant participants being less accurate. However, the effect was small. One SD change in Language dominance was associated with a 2.5% decrease in accuracy. Nevertheless, there may have been a trade-off between speed and accuracy. To test for a speed-accuracy trade-off, mean accuracy was correlated with mean RTs in each condition. The correlations were small ($r_{s\ 47} < .24$, $p_s > .100$) so it is unlikely that participants as a group traded accuracy for speed.

Because language dominance was correlated with proficiency in English and Spanish (see section 5.2.3), it is not clear whether dominance is responsible for the results or proficiency in English or Spanish. To answer this question, two separate analyses were run, replacing Language dominance with English and Spanish proficiency, respectively. When English scores were entered into the model, the main effect and the interactions with Frequency and Location were not significant. With Spanish scores, on the other hand, the pattern of results did not change

compared to those in Table 18. Using simple correlations on the mean RTs in the baseline condition (SFSL) with either Language dominance or the Spanish scores as a predictor showed that these correlations were $r_{47} = .39$, 95% CI = [.12, .61] and $r_{47} = .36$, 95% CI = [.09, .59], respectively. The large overlap of the CIs suggests that those correlations were not significantly different.

Table 18. *Results of the regression analysis of response times on the TAIL.*

Variable Name	RTs			Accuracy		
	<i>Beta</i>	<i>SE</i>	<i>p</i>	<i>Beta</i>	<i>SE</i>	<i>p</i>
Intercept (baseline = SFSL)	702.2	19.6		3.66	0.20	
Frequency (different vs. same)	45.0	9.3	.000	-0.53	0.23	.020
Location (different vs. same)	45.6	9.3	.000	-0.06	0.25	.802
Location*Frequency	-59.7	13.2	.000	0.57	0.33	.087
Language dominance (continuous variable)	-58.2	19.6	.004	-0.49	0.16	.003
Language dominance*Location	20.3	6.6	.004	0.28	0.17	.096
Language dominance*Frequency	4.0	6.6	.541	0.33	0.16	.044

Note. Only the data from bilingual participants were analyzed. Language dominance was transformed into a z-score so that the estimate shows the change associated with a 1 SD change in language dominance.

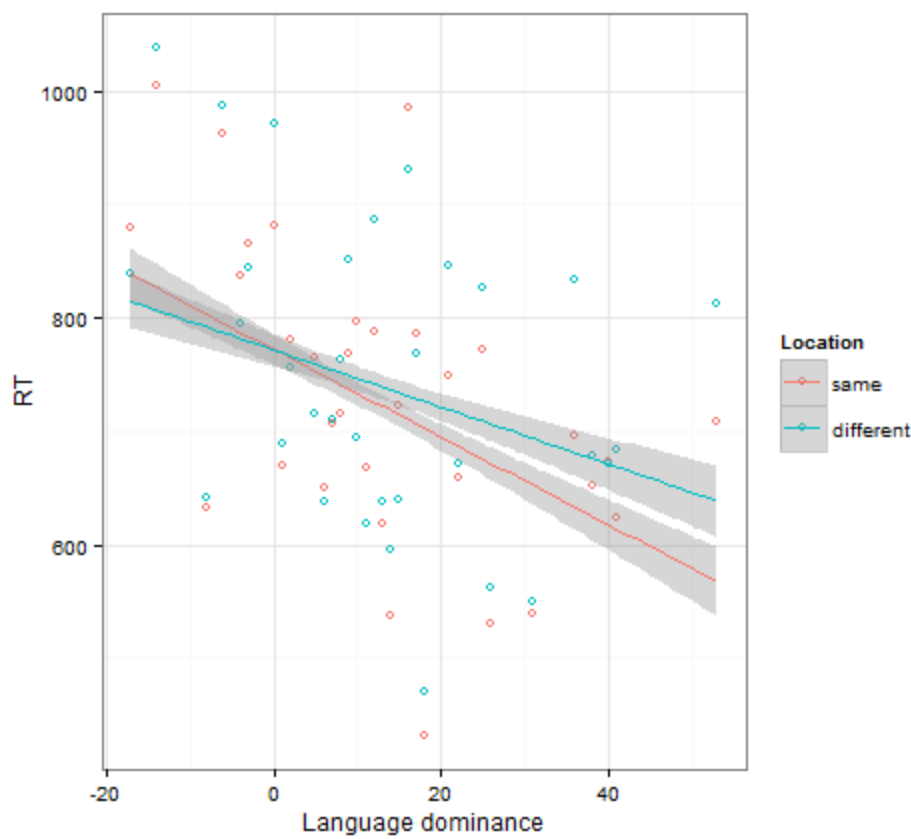


Figure 31. Effect of language dominance on response times (RT, in msec.) and the location effect. Language dominance was calculated by subtracting Spanish proficiency scores from English proficiency scores so that scores above 0 indicate English dominance.

5.5.5 Discussion

In light of the differences between test versions, the results are somewhat difficult to interpret. Bilinguals showed a reduced Frequency effect, that is, whether the frequency of the two tones was the same or different had a smaller effect on them compared to monolinguals. However, this was only true for version 2 of the test when both versions were considered separately. On version 1, on the other hand, bilinguals showed a larger Location effect, that is, they were slower to respond to trials where the location of the two tones was different compared to monolinguals. One possible explanation is that one version was easier than the other for monolinguals but for bilinguals, both versions had the same difficulty. More research would be needed to determine whether these results could be replicated or are idiosyncratic to this study.

A further investigation of the frequency effect showed that the manipulation was successful. A larger difference generally resulted in longer RTs, suggesting that participants were more distracted by this irrelevant dimension when the difference was larger. However, the relationship was not linear. When the difference was large (900 Hz), participants were as fast to respond as when the difference was small (100 Hz). It may be that very large differences were easier to ignore because they were more obvious. It is interesting to note that the effect did not interact with Location, suggesting that frequency was distracting even when there was response congruency (i.e., both Frequency and Location required a *different* response).

A further question that was investigated was whether specific variables relating to bilingualism would be associated with attentional control. The reasoning was that if the bilingual advantage is related to bilingual language use, then more balanced bilinguals may be expected to perform better than less balanced bilinguals. The results were surprising in that bilinguals who were more dominant in English were overall faster. In addition, they displayed a larger Location effect, which was mainly caused by faster responses to *same* trials compared to *different* trials. This suggests that the larger Location effect was due to an advantage for same location trials rather than a disadvantage attributable to greater distraction. The direction of the main effect of Language dominance was unexpected since it was hypothesized that more balanced bilinguals would be faster. Further analyses showed that the Spanish scores were also associated with overall faster RTs on the TAIL test. One possible interpretation of these results is that participants who were more dominant in English were more integrated into the dominant culture. As discussed in section 5.2.4, English dominant participants had more exposure to English, which may be equated to greater influence of the dominant (American) culture. It is well established that sociocultural differences can influence performance of tasks of executive

function (Chasiotis, Kiessling, Hofer, & Campos, 2006; Oh & Lewis, 2008; Sabbagh, Xu, Carlson, Moses, & Lee, 2006). For example, Chasiotis et al. (2006) suggest that cultures that differ in interpersonal distance (separateness – relatedness) and agency (autonomy – heteronomy; see Kagitcibasi, 1996) may differ on tasks of executive function. The researchers found some evidence for this hypothesis by testing children from three cultures that differ on these two dimensions, Germany, Cameroon, and Costa Rica. Differences emerged for the Cameroonian sample compared to the other two samples. Cameroonian children performed less well on conflict-inhibition tasks but better on a delay inhibition task (on this task, the child is told not to take a snack in his view until the experimenter rings a bell). Chasiotis et al. (2006) suggest that this may have to do with parenting style. Parents in Cameroon may favor obedient and inhibited behavior but may disregard impulse behavior (p. 258). Likewise, immigrant families of Mexican descent (the majority of the bilingual sample) may differ in their parenting style from Caucasian nonhispanic American families (the majority of the monolingual sample; Varela et al., 2004). One tentative explanation of the present results may thus be that greater English dominance was associated with greater adaptation to values of the dominant US culture such as independence and autonomy. It should be noted, though, that this is only speculative and that there is no direct evidence for this hypothesis. One way to investigate this hypothesis would be to survey bilingual children's parents about parenting style and cultural values and relate this to their children's executive function development (see, e.g., Bernier, Carlson, & Whipple, 2010). In addition, greater English dominance may not only be associated with parents' cultural values but also the participants' own adaptation to the dominant culture and its values such as autonomy. While Spanish contact is often determined by external forces in the early years (e.g., parents may choose to put their child into a bilingual program), later in life bilinguals may

choose to build social networks with members of the dominant culture and abandon some of their traditional values.

Studies investigating the bilingual advantage have often relied on group comparisons. However, as has been recently pointed out (Valian, 2014), individuals can differ on many dimensions that have been linked to advantages in executive function such as musical training, video gaming, and exercise. Thus we can never be sure that differences between groups are attributable to bilingualism or some other unobserved variable, especially when sample sizes are small. In the present study, one possible confounding factor is SES. SES as measured by mother's education level was significantly lower in the bilingual than the monolingual sample. Because there was almost complete separation, it is impossible to statistically control for this variable. Thus there may have been a bilingual advantage but it may have been obscured by the lower SES of the bilinguals. In light of these difficulties, other researchers have suggested to employ individual differences designs to directly relate aspects of bilingualism to advantages in executive function (Titone, Pivneva, Sheikh, Webb, & Whitford, 2015). In the present study it was hypothesized that more balanced bilinguals would have greater attentional control compared to less balanced bilinguals. The opposite effect was found with English dominant bilinguals being overall faster compared to balanced bilinguals. To explain this unexpected result, a literature search showed that executive functions may be related to cultural values and parenting style. This adds even more variables to the task of singling out bilingualism as a factor of benefits in executive function. What seems clear, though, from the present results is that while bilingualism influences verbal variables, general cognitive function is not negatively affected.

CHAPTER 6: CONCLUSION

Interim discussions of each test can be found after the presentation of the results of each test. In this conclusion I will summarize the main results and relate the different findings to each other. Figure 32 summarizes the results schematically. English proficiency, as measured by the WMLS-R subtests picture vocabulary and verbal analogies, turned out to be the strongest predictor of individual differences in SUN, the topic of this dissertation. The main finding from Experiment 2 (section 0) was that differences between monolinguals and bilinguals may be less categorical as may be thought when looking at group comparisons only. While bilingual participants were overall less accurate on the SPIN than monolinguals (71.8% vs. 80.8%), English proficiency was a mediating factor in both groups. English proficiency, in turn, was related to exposure to English. Bilingual participants with more exposure to English were also more proficient in English. On the other hand, more English exposure was necessarily related with less Spanish exposure. This was further expressed in the finding that higher English proficiency was associated with lower Spanish proficiency. Other predictors of Spanish proficiency were the number of speakers a participant regularly interacted with during childhood. This variable was not negatively correlated with English proficiency, which suggests that a trade-off between a bilingual's languages may be attenuated by certain variables. Another factor that seemed to have positively influenced Spanish proficiency without negatively affecting English proficiency was participation in Spanish immersion programs (not shown in Figure 32). Because these results were based on retrospective reports only they have to be interpreted with caution but the results again suggest that a trade-off between languages may not be inevitable. English proficiency furthermore predicted WMC and consonant perception in noise and a weak association was found between Spanish proficiency and the Spanish WIN test.

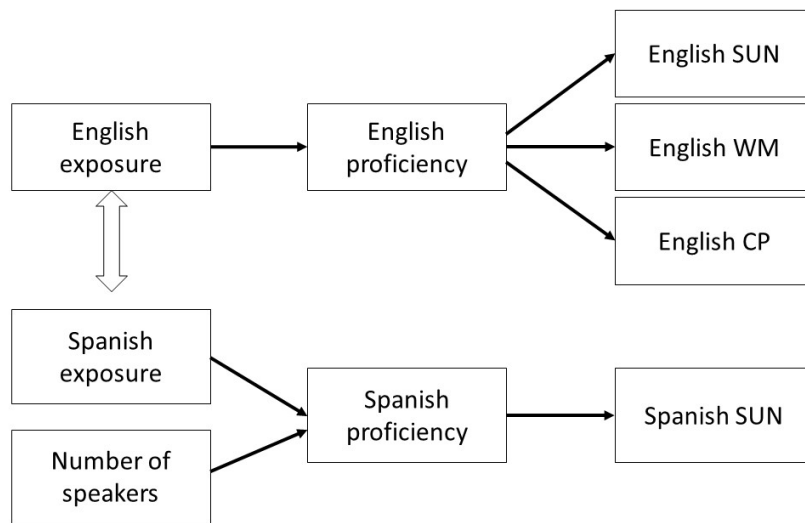


Figure 32. Schematic representation of the results in this study. Arrows indicate significant relationships between variables. The two-way arrow indicates that more exposure to one language is associated with less exposure to the other language. SUN = speech understanding in noise. WM = working memory. CP = consonant perception.

One limitation of this study and any correlational study is that causation cannot be established. The arrows in Figure 32 merely show the hypothesized direction of the relationship between variables. While it is established in the literature that more exposure to the language, for example in the form of mother-child interactions, will lead to vocabulary growth (e.g., Hoff, 2006; Weisleder & Fernald, 2013), a remaining question is whether vocabulary size is causally related to SUN or whether exposure causally predicts both SUN and vocabulary size. The lexical restructuring model (Metsala & Walley, 1998; Walley, 2008), for example, assumes a causal relationship. Furthermore, Pierrehumbert (2001, 2003) suggests that phonetic categories in listeners are fine-tuned by the type statistics computed over the entire lexicon. That is, the mental lexicon provides listeners with feedback about which phonetically dissimilar sounds nevertheless belong to the same phonetic category. Listeners with more refined phonetic categories may be attending to more detailed phonetic information and may thus be less affected some of this information is overshadowed by a competing signal. Some evidence for the relationship between

vocabulary size and the precision of phonetic categories came from the consonant perception test. But again, these results cannot establish causation and it may be that English exposure is the mediating factor, leading to a larger vocabulary and more precise phonetic categories.

As I already mentioned, the results of the SPIN, WIN, CP, and WM tests suggest that differences between monolinguals and bilinguals may be more gradual than previously thought. Especially the fact that language proficiency predicted performance on these tests for monolinguals and bilinguals suggests that the less accurate performance of the bilinguals may be a natural consequence of being exposed to two languages, and, as a consequence, spending less time in each language. The results of the WMLS-R showed that even highly proficient bilinguals who have received all their schooling in English and are studying at a major US university may still have a smaller vocabulary than the general population. Whereas the monolinguals performed above the population mean, the mean standard score of the bilinguals was 2/3 of a standard deviation below the population mean. The context of the present study may be different from studies on bilingualism in other regions of the world such as Catalonia or Montreal where bilingualism is not associated with SES. However, given the relationship between language proficiency and language exposure found across many studies, it seems that balanced bilinguals as a group will always be less proficient in each language compared to monolingual speakers. Therefore, if a goal of a study is to make inferences about bilingualism (i.e., the consequences of speaking more than one language) and not about language proficiency in general, one must test two groups of participants who are matched on language proficiency. Otherwise language proficiency will be a confounding factor and it will not be clear whether differences are attributable to bilingualism or to lower proficiency. Matching bilinguals with monolinguals is not straightforward, however. For example, if we tested a large group of Spanish-English bilingual

speakers in the US, the current sample would probably have performed above the population mean, given that these participants were college students, that is, they would be drawn from the right tail of the bilingual population distribution. If we wanted to find a monolingual sample of the same mean proficiency, they would be drawn from the left tail of the monolingual population distribution. Thus the two samples may differ in other ways and may not be easily comparable.

While differences between groups could be attributed to differences in language proficiency to a large extent, some differences may also be attributable to cross-language influence. Evidence for this assumption comes from the CP test. A comparison of the confusion matrices (*Confusion matrix - bilingual participants*. Table 14 and Table 15) suggested that bilingual speakers tended to misperceive those consonants that have overlapping category boundaries in English and Spanish (i.e., /b/, /d/, and /g/). This result may have been a consequence of the decontextualized nature of the test, which would suggest that bilinguals cannot simply switch their languages on and off and function like a monolingual of the presently relevant language (cf. Grosjean, 2001).

Lastly, the results from the SPIN and the WIN tests also suggest that individual differences in domain-general, cognitive abilities play a role for SUN. A lower baseline RT on the TAIL was associated with higher accuracy on the SPIN and WIN. Baseline RT may reflect processing speed (Zhang et al., 2012). A decrease in processing speed has been proposed as a major contributor to the age-related decline in cognition (Salthouse, 1996). If processing speed is associated with SUN, then the age-related decline in processing speed may also explain why SUN gets harder as a function of age (cf. Wingfield, 1996). In the present study, the effect of processing speed was small, which may be due to the fact that participants were young-adult college students. Larger effects may be found if a more diverse sample was tested.

APPENDIX

APPENDIX

Table 19. *Mean item accuracy on the WIN.*

SNR	Word	Monolingual	Bilingual	SNR	Word	Monolingual	Bilingual
0	back	39.6%	40.4%	12	search	100.0%	100.0%
0	bath	38.5%	38.3%	12	shack	100.0%	100.0%
0	calm	5.7%	0.0%	12	tool	100.0%	97.9%
0	dab	1.9%	0.0%	12	voice	100.0%	100.0%
0	gaze	23.1%	19.1%	12	witch	100.0%	100.0%
0	get	25.0%	36.2%	16	base	100.0%	100.0%
0	kill	11.3%	4.3%	16	date	100.0%	100.0%
0	life	30.8%	25.5%	16	dog	100.0%	100.0%
0	nice	22.6%	23.4%	16	gas	98.1%	83.0%
0	read	7.7%	29.8%	16	have	94.3%	93.6%
4	beg	17.0%	14.9%	16	judge	100.0%	100.0%
4	far	50.9%	29.8%	16	live	100.0%	95.7%
4	learn	88.7%	89.4%	16	red	100.0%	100.0%
4	long	24.5%	17.0%	16	time	84.9%	70.2%
4	mess	58.5%	72.3%	16	wire	100.0%	100.0%
4	mood	22.6%	4.3%	20	chair	100.0%	100.0%
4	mouse	60.4%	59.6%	20	ditch	100.0%	100.0%
4	note	39.6%	40.4%	20	gun	100.0%	100.0%
4	sheep	96.2%	83.0%	20	haze	100.0%	100.0%
4	talk	71.7%	78.7%	20	kick	100.0%	100.0%
8	bite	77.4%	66.0%	20	luck	100.0%	100.0%
8	deep	79.2%	70.2%	20	ring	100.0%	100.0%
8	doll	37.7%	61.7%	20	shawl	96.2%	76.6%
8	half	49.1%	42.6%	20	such	100.0%	100.0%
8	make	98.1%	95.7%	20	tire	100.0%	100.0%
8	pick	96.2%	100.0%	24	cool	100.0%	97.9%
8	soap	67.9%	46.8%	24	dodge	100.0%	100.0%
8	sour	86.8%	78.7%	24	food	100.0%	100.0%
8	turn	92.5%	73.9%	24	hire	100.0%	97.9%
8	young	58.5%	80.4%	24	juice	100.0%	100.0%
12	chief	100.0%	95.7%	24	late	100.0%	100.0%
12	good	100.0%	100.0%	24	pain	100.0%	100.0%
12	hate	100.0%	100.0%	24	road	100.0%	100.0%
12	pass	100.0%	100.0%	24	wheat	100.0%	97.9%
12	rush	100.0%	95.7%	24	youth	100.0%	100.0%

Note. SNR = signal-to-noise-ratio.

Table 20. *Items discrimination index for E-WIN words.*

Word	SNR	Accuracy	Accuracy Bottom 27%	Accuracy Top 27%	Item discrimination	Correlation
far	4	41.0%	25.9%	59.3%	0.33	0.34
gaze	0	21.2%	7.4%	40.7%	0.33	0.32
soap	8	58.0%	33.3%	70.4%	0.37	0.30
turn	8	83.8%	70.4%	92.6%	0.22	0.30
talk	4	75.0%	55.6%	88.9%	0.33	0.29
half	8	46.0%	18.5%	66.7%	0.48	0.29
life	0	28.3%	14.8%	44.4%	0.30	0.29
kill	0	8.0%	0.0%	14.8%	0.15	0.28
get	0	30.3%	22.2%	51.9%	0.30	0.28
shawl	20	87.0%	74.1%	92.6%	0.19	0.28
mood	4	14.0%	3.7%	25.9%	0.22	0.27
live	16	98.0%	92.6%	100.0%	0.07	0.27
long	4	21.0%	3.7%	33.3%	0.30	0.25
calm	0	3.0%	0.0%	11.1%	0.11	0.25
learn	4	89.0%	88.9%	100.0%	0.11	0.24
bite	8	72.0%	55.6%	81.5%	0.26	0.24
mess	4	65.0%	51.9%	81.5%	0.30	0.23
note	4	40.0%	37.0%	55.6%	0.19	0.23
back	0	40.0%	33.3%	59.3%	0.26	0.23
dab	0	1.0%	0.0%	3.7%	0.04	0.21
young	8	68.7%	55.6%	77.8%	0.22	0.21
sheep	4	90.0%	81.5%	92.6%	0.11	0.20
deep	8	75.0%	59.3%	92.6%	0.33	0.20
chief	12	98.0%	92.6%	100.0%	0.07	0.20
read	0	18.2%	7.4%	25.9%	0.19	0.19
time	16	78.0%	66.7%	92.6%	0.26	0.19
have	16	94.0%	92.6%	96.3%	0.04	0.17
make	8	97.0%	92.6%	100.0%	0.07	0.16
bath	0	38.4%	29.6%	48.1%	0.19	0.16
gas	16	91.0%	81.5%	92.6%	0.11	0.14
doll	8	49.0%	48.1%	48.1%	0.00	0.12
rush	12	98.0%	92.6%	100.0%	0.07	0.12
beg	4	16.0%	7.4%	22.2%	0.15	0.09
sour	8	83.0%	77.8%	85.2%	0.07	0.09
tool	12	99.0%	96.3%	100.0%	0.04	0.07
hire	24	99.0%	96.3%	100.0%	0.04	0.07
mouse	4	60.0%	63.0%	66.7%	0.04	0.06
cool	24	99.0%	100.0%	100.0%	0.00	0.00
wheat	24	99.0%	100.0%	100.0%	0.00	0.00

Table 20 (cont'd).

Word	SNR	Accuracy	Accuracy Bottom 27%	Accuracy Top 27%	Item discrimination	Correlation
nice	0	23.0%	18.5%	22.2%	0.04	-0.01
pick	8	98.0%	100.0%	100.0%	0.00	-0.05
good	12	100.0%	100.0%	100.0%	0.00	
hate	12	100.0%	100.0%	100.0%	0.00	
pass	12	100.0%	100.0%	100.0%	0.00	
search	12	100.0%	100.0%	100.0%	0.00	
shack	12	100.0%	100.0%	100.0%	0.00	
voice	12	100.0%	100.0%	100.0%	0.00	
witch	12	100.0%	100.0%	100.0%	0.00	
base	16	100.0%	100.0%	100.0%	0.00	
date	16	100.0%	100.0%	100.0%	0.00	
dog	16	100.0%	100.0%	100.0%	0.00	
judge	16	100.0%	100.0%	100.0%	0.00	
red	16	100.0%	100.0%	100.0%	0.00	
wire	16	100.0%	100.0%	100.0%	0.00	
chair	20	100.0%	100.0%	100.0%	0.00	
ditch	20	100.0%	100.0%	100.0%	0.00	
gun	20	100.0%	100.0%	100.0%	0.00	
haze	20	100.0%	100.0%	100.0%	0.00	
kick	20	100.0%	100.0%	100.0%	0.00	
luck	20	100.0%	100.0%	100.0%	0.00	
ring	20	100.0%	100.0%	100.0%	0.00	
such	20	100.0%	100.0%	100.0%	0.00	
tire	20	100.0%	100.0%	100.0%	0.00	
dodge	24	100.0%	100.0%	100.0%	0.00	
food	24	100.0%	100.0%	100.0%	0.00	
juice	24	100.0%	100.0%	100.0%	0.00	
late	24	100.0%	100.0%	100.0%	0.00	
pain	24	100.0%	100.0%	100.0%	0.00	
road	24	100.0%	100.0%	100.0%	0.00	
youth	24	100.0%	100.0%	100.0%	0.00	

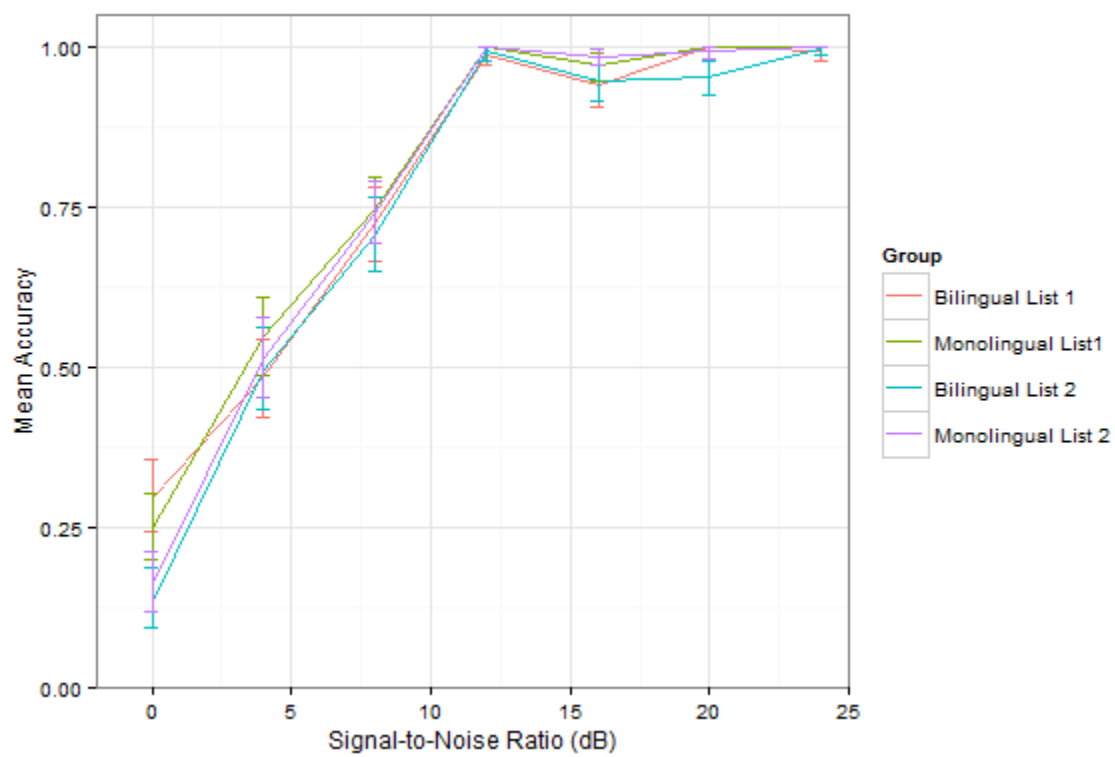


Figure 33. Mean accuracy on the English Words in Noise test for List 1 and 2. Whiskers show the 95% confidence interval.

REFERENCES

REFERENCES

- Abutalebi, J., Della Rosa, P. a, Ding, G., Weekes, B., Costa, A., & Green, D. W. (2013). Language proficiency modulates the engagement of cognitive control areas in multilinguals. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 49(3), 905–11. doi:10.1016/j.cortex.2012.08.018
- Acheson, D., Hamidi, M., Binder, J., & Postle, B. (2011). A common neural substrate for language production and verbal working memory. *Journal of Cognitive Neuroscience*, 23(6), 1358–1367.
- Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, 47(Suppl. 2), S53–S71. doi:10.1080/14992020802301142
- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, 38(4), 419–439. doi:10.1006/jmla.1997.2558
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73(3), 247–64.
- Alvarado, C. G., & Woodcock, R. W. (2005). *Comprehensive manual*. Rolling Meadows, IL: Riverside Publishing.
- Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., ... Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology*, 5(May), 1–12. doi:10.3389/fpsyg.2014.00398
- Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics*, 40(4), 582–594. doi:10.1016/j.wocn.2012.05.005
- Arlinger, S., Lunner, T., Lyxell, B., & Pichora-Fuller, M. K. (2009). The emergence of cognitive hearing science. *Scandinavian Journal of Psychology*, 50(5), 371–84. doi:10.1111/j.1467-9450.2009.00753.x
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. doi:10.1016/j.jml.2007.12.005
- Baddeley, A. D. (1992). Working memory. *Science*, 255(5044), 556–559.

- Baddeley, A. D. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29. doi:10.1146/annurev-psych-120710-100422
- Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105(1), 158–173. doi:10.1037//0033-295X.105.1.158
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *The Psychology of Learning and Motivation*, 8, 47–89.
- Balota, D., Yap, M. J., Cortese, M. J., Hutchison, K. a, Kessler, B., Loftis, B., ... Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3), 445–59.
- Barcroft, J., & Sommers, M. S. (2005). Effects of Acoustic Variability on Second Language Vocabulary Learning. *Studies in Second Language Acquisition*, 27, 387–414. doi:10.1017/S0272263105050175
- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4.
- Bernier, A., Carlson, S. M., & Whipple, N. (2010). From external regulation to self-regulation: Early parenting precursors of young children's executive functioning. *Child Development*, 81(1), 326–339. doi:10.1111/j.1467-8624.2009.01397.x
- Bialystok, E., Craik, F. I. M., Klein, R. M., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and Aging*, 19(2), 290–303. doi:10.1037/0882-7974.19.2.290
- Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences*, 16(4), 240–250. doi:10.1016/j.tics.2012.03.001
- Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 859–73. doi:10.1037/0278-7393.34.4.859
- Bialystok, E., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition*, 15(2), 397–401. doi:10.1017/S136672891100040X
- Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2009). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13(04), 525–531. doi:10.1017/S1366728909990423
- Bilger, R., Nuetzel, J., Rabinowitz, W., & Rzechowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research*, 27, 32–48.

- Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer.
- Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. a. (2008). Context Variation and Definitions in Learning the Meanings of Words: An Instance-Based Learning Approach. *Discourse Processes*, 45(2), 122–159. doi:10.1080/01638530701792826
- 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. doi:10.1121/1.2642103
- Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America*, 106(4), 2074–2085. doi:10.1121/1.427952
- Brännström, K. J., Zunic, E., Borovac, A., & Ibertsson, T. (2012). Acceptance of background noise, working memory capacity, and auditory evoked potentials in subjects with normal hearing. *Journal of the American Academy of Audiology*, 23(7), 542–52. doi:10.3766/jaaa.23.7.6
- Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review*, 18(6), 1189–1196. doi:10.3758/s13423-011-0167-9
- Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–90. doi:10.3758/BRM.41.4.977
- Calandruccio, L., & Zhou, H. (2013). Increase in speech recognition due to linguistic mismatch between target and masker speech: Monolingual and simultaneous bilingual performance. *Journal of Speech, Language, and Hearing Research*, 57, 1089–1097.
- Capps, R., Fix, M., Murray, J., Ost, J., Passel, J., & Herwantoro, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act*. Urban Institute. Washington, D.C.
- Charles-Luce, J., & Luce, P. a. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, 17(01), 205–215. doi:10.1017/S0305000900013180
- Chasiotis, A., Kiessling, F., Hofer, J., & Campos, D. (2006). Theory of mind and inhibitory control in three cultures: Conflict inhibition predicts false belief understanding in Germany, Costa Rica and Cameroon. *International Journal of Behavioral Development*, 30(3), 249–260. doi:10.1177/0165025406066759
- Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition*, 28(1), 143–53. doi:10.3758/BF03211582

- Coady, J. a., & Aslin, R. N. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language*, 30(2), 441–469. doi:10.1017/S0305000903005579
- Conway, A. R. A., & Engle, R. W. (1996). Individual Differences in Working Memory Capacity: More Evidence for a General Capacity Theory. *Memory*, 4(6), 577–590.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769–786.
- Corrigall, K. a., & Schellenberg, E. G. (2015). Predicting who takes music lessons: parent and child characteristics. *Frontiers in Psychology*, 6(282), 1–8. doi:10.3389/fpsyg.2015.00282
- Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: evidence from the ANT task. *Cognition*, 106(1), 59–86. doi:10.1016/j.cognition.2006.12.013
- Cowan, N. (1993). Activation, attention, and short-term memory. *Memory & Cognition*, 21(2), 162–7.
- Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge University Press.
- Crandell, C., & Smaldino, J. (1996). Speech perception in noise by children for whom English is a second language. *American Journal of Audiology*, 5, 47–51.
- Cutler, A. (2012). *Native listening: language experience and the recognition of spoken words*. Cambridge, MA: The MIT Press.
- Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: effects of local context. *The Journal of the Acoustical Society of America*, 124(2), 1264–8. doi:10.1121/1.2946707
- Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, 116(6), 3668. doi:10.1121/1.1810292
- Dahan, D., & Magnuson, J. S. (2006). Spoken Word Recognition. In M. Traxler & M. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 249–283). Amsterdam, NL: Academic Press.
- Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: evidence from eye movements. *Cognitive Psychology*, 42(4), 317–67. doi:10.1006/cogp.2001.0750

- Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 466, 450–466.
- De Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive Advantage in Bilingualism: An Example of Publication Bias? *Psychological Science*, 26(1), 99–107. doi:10.1177/0956797614557866
- Delcenserie, A., & Genesee, F. (2013). Language and memory abilities of internationally adopted children from China: evidence for early age effects. *Journal of Child Language*, 41(6), 1195–1223. doi:10.1017/S030500091300041X
- Delgado, P., Guerrero, G., Goggin, J. P., & Ellis, B. B. (1999). Self-Assessment of Linguistic Skills by Bilingual Hispanics. *Hispanic Journal of Behavioral Sciences*, 21(1), 31–46. doi:10.1177/0739986399211003
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149–79. doi:10.1146/annurev.psych.55.090902.142028
- Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, 66(5), 843–863. doi:10.1080/17470218.2012.720994
- Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, 59(3), 294–311. doi:10.1016/j.jml.2008.06.006
- Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science : A Journal of the American Psychological Society / APS*, 21(11), 1664–1670. doi:10.1177/0956797610384743
- Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: Myth or reality? *Experimental Psychology*, 61, 234–251. doi:10.1027/1618-3169/a000243
- Edwards, J. R., Beckman, M. E., & Munson, B. (2004). The Interaction between Vocabulary Size and Phonotactic Probability Effects on Children's Production Accuracy and Fluency in Nonword Repetition. *Journal of Speech, Language and Hearing Research*, 47, 421–436.
- Elman, J. L., Diehl, R. L., & Buchwald, S. (1977). Perceptual switching in bilinguals. *The Journal of the Acoustical Society of America*, 62(4), 971–974. doi:10.1121/1.381591
- Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39(3), 253–260. doi:10.1016/S0095-4470(11)00055-6

- Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: differences by class and race. *Social Science Research*, 33(3), 464–497. doi:10.1016/j.ssresearch.2003.08.001
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (Vol. 92, pp. 233–277). Timonium, MD: York Press. doi:10.1111/j.1600-0404.1995.tb01710.x
- Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, 43, 155–171.
- Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, 15, 67–83.
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age Constraints on Second-Language Acquisition. *Journal of Memory and Language*, 41(1), 78–104. doi:10.1006/jmla.1999.2638
- Garcia Lecumberri, M. L., & 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. doi:10.1121/1.2180210
- Garcia-Sierra, A., Diehl, R. L., & Champlin, C. (2009). Testing the double phonemic boundary in bilinguals. *Speech Communication*, 51(4), 369–378. doi:10.1016/j.specom.2008.11.005
- Gasquoine, P. G., & Dayanira Gonzales, C. (2012). Using Monolingual Neuropsychological Test Norms with Bilingual Hispanic Americans: Application of an Individual Comparison Standard. *Archives of Clinical Neuropsychology*, 27(3), 268–276. doi:10.1093/arclin/acs004
- Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, 28(2), 200–213. doi:10.1016/0749-596X(89)90044-2
- Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Viñas Guasch, N., ... Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, 5(11), 1–14. doi:10.3389/fpsyg.2014.00011
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.
- Gibson, T. a, Oller, D. K., Jarmulowicz, L., & Ethington, C. a. (2012). The receptive-expressive gap in the vocabulary of young second-language learners: Robustness and possible mechanisms. *Bilingualism: Language and Cognition*, 15(1), 102–116. doi:10.1017/S1366728910000490

- Gibson, T. a., Peña, E. D., & Bedore, L. M. (2014). The relation between language experience and receptive-expressive semantic gaps in bilingual children. *International Journal of Bilingual Education and Bilingualism*, 17(1), 90–110. doi:10.1080/13670050.2012.743960
- Giraud, A.-L., & Poeppel, D. (2012a). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511–7. doi:10.1038/nn.3063
- Giraud, A.-L., & Poeppel, D. (2012b). Introduction: Terminology and concepts. In D. Poeppel, T. Overath, A. N. Popper, & R. R. Fay (Eds.), *The Human Auditory Cortex* (Vol. 43, pp. 225–260). New York, NY: Springer New York. doi:10.1007/978-1-4614-2314-0
- Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1166–83.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279. doi:10.1037/0033-295X.105.2.251
- Gollan, T. H., & Acenas, L.-A. R. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish-English and tagalog-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1), 246–69. doi:10.1037/0278-7393.30.1.246
- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58(3), 787–814. doi:10.1016/j.jml.2007.07.001
- Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, 33(7), 1220–34.
- Gollan, T. H., Montoya, R. I., & Werner, G. a. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, 16(4), 562–576. doi:10.1037//0894-4105.16.4.562
- Gollan, T. H., Salmon, D. P., Montoya, R. I., & Galasko, D. R. (2011). Degree of bilingualism predicts age of diagnosis of Alzheimer’s disease in low-education but not in highly educated Hispanics. *Neuropsychologia*, 49(14), 3826–30. doi:10.1016/j.neuropsychologia.2011.09.041
- Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew English bilinguals. *Bilingualism: Language and Cognition*, 4(1), 63 – 83.
- Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: the frequency-lag hypothesis. *Journal of Experimental Psychology: General*, 140(2), 186–209. doi:10.1037/a0022256

- Gollan, T. H., Starr, J., & Ferreira, V. S. (2014). More than use it or lose it: The number-of-speakers effect on heritage language proficiency. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-014-0649-7
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(02), 67–81.
- Green, K., Kuhl, P. K., Meltzoff, A. N., & Stevens, E. (1991). Integrating speech information across talkers, gender, and sensory modality: female faces and male voices in the McGurk effect. *Perception & Psychophysics*, 50(6), 524–536. doi:10.3758/BF03207536
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28(4), 267–283.
- Grosjean, F. (2001). A bilingual's language modes. In J. L. Nicol (Ed.), *One Mind, Two Languages: Bilingual Language Processing* (pp. 1 – 22). Malden, MA: Blackwell Publishers.
- Grosjean, F. (2008). *Studying bilinguals*. Oxford University Press.
- Gupta, P., & Tisdale, J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language*, 61(4), 481–502. doi:10.1016/j.jml.2009.08.001
- Gutiérrez-Clellen, V. F., Calderón, J., & Ellis Weismer, S. (2004). Verbal working memory in bilingual children. *Journal of Speech, Language, and Hearing Research : JSLHR*, 47(4), 863–76. doi:10.1044/1092-4388(2004/064)
- Hammer, C. S., Komaroff, E., Rodriguez, B. L., Lopez, L. M., Scarpino, S. E., & Goldstein, B. (2012). Predicting Spanish – English Bilingual Children's Language Abilities. *Journal of Speech, Language, and Hearing Research*, 55(October 2012), 1251–1264. doi:10.1044/1092-4388(2012/11-0016)Journal
- Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, 24, 495–522. doi:10.1017/S0142716403000250
- Hardison, D. M. (2012). Second language speech perception: A cross-disciplinary perspective on challenges and accomplishments. In *The routledge handbook of second language acquisition* (pp. 349–363).
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.

- Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review*, 18(4), 625–58. doi:10.3758/s13423-011-0116-7
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, 93(4), 411–428. doi:10.1037/0033-295X.93.4.411
- Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74(5), 1368–1378.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26, 55–88. doi:10.1016/j.dr.2005.11.002
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39(1), 1–27. doi:10.1017/S0305000910000759
- Holt, L. L., & Lotto, A. J. (2008). Speech Perception Within an Auditory Cognitive Science Framework. *Current Directions in Psychological Science*, 17(1), 42–46.
- Howes, D. (1957). On the Relation between the Intelligibility and Frequency of Occurrence of English Words. *The Journal of the Acoustical Society of America*, 29(2), 296–305.
- Hulme, C., Maughan, S., & Brown, G. D. . (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, 30(6), 685–701. doi:10.1016/0749-596X(91)90032-F
- Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. DA, Martin, S., & Stuart, G. (1997). Word-Frequency Effects on Short-Term Memory Tasks: Evidence for a Redintegration Process in Immediate Serial Recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5), 1217–1232.
- Hurtado, N., Grüter, T., Marchman, V. a., & Fernald, A. (2013). Relative language exposure, processing efficiency and vocabulary in Spanish–English bilingual toddlers. *Bilingualism: Language and Cognition*, 1–14. doi:10.1017/S136672891300014X
- Huttenlocher, J., & Haight, W. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27(2), 236–248.
- Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *The Journal of the Acoustical Society of America*, 117(2), 896. doi:10.1121/1.1823291
- Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, 127(2), 277–88. doi:10.1016/j.actpsy.2007.06.003

- Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59(4), 434–446. doi:10.1016/j.jml.2007.11.007
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21(1), 60–99.
- Ju, M., & Luce, P. A. (2004). Falling on sensitive ears: Constraints on Bilingual Lexical Activation. *Psychological Science*, 15(5), 314–318.
- Kağıtçıbaşı, Ç. (1996). The Autonomous-Relational Self: A New Synthesis. *European Psychologist*, 1(3), 180–186.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology. General*, 133(2), 189–217. doi:10.1037/0096-3445.133.2.189
- Kavé, G., Knafo, A., & Gilboa, A. (2010). The rise and fall of word retrieval across the lifespan. *Psychology and Aging*, 25(3), 719–724. doi:10.1037/a0018927
- Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: a lexical decision study on 14,000 dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, 1(174), 1–15. doi:10.3389/fpsyg.2010.00174
- Killion, M., Niquette, P., Gudmundsen, G., Revit, L., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 116(4), 2395–2405.
- Kilman, L., Zekveld, A., Hällgren, M., & Rönnberg, J. (2014). The influence of non-native language proficiency on speech perception performance. *Frontiers in Psychology*, 5(July), 651. doi:10.3389/fpsyg.2014.00651
- Klein, R. M. (2015). Is there a benefit of bilingualism for executive functioning? *Bilingualism: Language and Cognition*, 18(1), 29–31. doi:10.1017/S1366728914000613
- Kousaie, S., & Phillips, N. A. (2011). Ageing and bilingualism: Absence of a “bilingual advantage” in Stroop interference in a nonimmigrant sample. *The Quarterly Journal Of Experimental Psychology*, 65(2), 356 – 369.
- Kucera, H., & Francis, N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown university press.

- Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606–608.
- Kuperman, V., & Van Dyke, J. A. (2013). Reassessing Word Frequency as a Determinant of Word Recognition for Skilled and Unskilled Readers. *Journal of Experimental Psychology: Human Perception and Performance, Advance on*. doi:10.1037/a0030859
- Lewellen, M., Goldinger, S. D., Pisoni, D. B., & Greene, B. (1993). Lexical familiarity and procesing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General*, 122(3), 316–330.
- Liberman, A., & Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1–36.
- Liberman, A., & Mattingly, I. (1989). A specialization for speech perception. *Science*, 243, 489–494.
- Ljung, R., Israelsson, K., & Hygge, S. (2012). Speech Intelligibility and Recall of Spoken Material Heard at Different Signal-to-noise Ratios and the Role Played by Working Memory Capacity. *Applied Cognitive Psychology*, 27, 198–203.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19(1), 1–36.
- Luo, L., Craik, F. I. M., Moreno, S., & Bialystok, E. (2013). Bilingualism interacts with domain in a working memory task: evidence from aging. *Psychology and Aging*, 28(1), 28–34. doi:10.1037/a0030875
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109(1), 35–54. doi:10.1037//0033-295X.109.1.35
- Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, 132(2), 202–227. doi:10.1037/0096-3445.132.2.202
- Majerus, S., Linden, M. Van Der, Mulder, L., Meulemans, T., & Peters, F. (2004). Verbal short-term memory reflects the sublexical organization of the phonological language network: Evidence from an incidental phonotactic learning paradigm. *Journal of Memory and Language*, 51(2), 297–306. doi:10.1016/j.jml.2004.05.002
- Marian, V., Blumenfeld, H. K., & Kaushkanskaya, 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, 940 – 967.

- Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25, 71–102.
- Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, 69(4), 574–588. doi:10.1016/j.jml.2013.08.001
- Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, 59(3), 203–243. doi:10.1016/j.cogpsych.2009.04.001
- Mattys, S. L., Carroll, L. M., Li, C. K. W., & Chan, S. L. Y. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication*, 52(11-12), 887–899. doi:10.1016/j.specom.2010.01.005
- Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, 27(7-8), 953–978.
- Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology. General*, 134(4), 477–500. doi:10.1037/0096-3445.134.4.477
- Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, 65(2), 145–160. doi:10.1016/j.jml.2011.04.004
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101–11.
- Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research: JSLHR*, 40(3), 686–93.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86.
- McQueen, J. M. (2007). Eight questions about spoken-word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 37–53). Oxford: Oxford University Press.
- Meador, D., Flege, J. E., & Mackay, R. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition*, 3, 55–67.
- Mercier, J., Pivneva, I., & Titone, D. (2013). Individual differences in inhibitory control relate to bilingual spoken word processing. *Bilingualism: Language and Cognition*, 1–29. doi:10.1017/S1366728913000084

- Metsala, J. L., & Walley, A. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Mahwah, NJ: Lawrence Erlbaum Associates.
- Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, 21(1), 8–14. doi:10.1177/0963721411429458
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology*, 41, 49–100. doi:10.1006/cogp.1999.0734
- Monsell, S. (1991). The nature and locus of word frequency effects. In *Basic processes in reading: Visual word recognition* (pp. 148–197).
- Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish-English bilinguals. *Cognitive Brain Research*, 22(2), 205–220. doi:10.1016/j.cogbrainres.2004.08.010
- Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, 10(6), 719–26. doi:10.1111/j.1467-7687.2007.00623.x
- Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: the rank hypothesis. *Psychological Review*, 111(3), 721–56. doi:10.1037/0033-295X.111.3.721
- Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. T. (2012). The Influence of Language Proficiency on Lexical Semantic Processing in Native and Late Learners of English. *Journal of Cognitive Neuroscience*, 24(5), 1205–1223. doi:10.1162/jocn_a_00143
- Norris, D. (1994). Shortlist: a connectionist model continuous speech recognition. *Cognition*, 52, 189–234.
- Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–95. doi:10.1037/0033-295X.115.2.357
- Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences*, 13(1), 14–9. doi:10.1016/j.tics.2008.09.005
- Obleser, J., Eisner, F., & Kotz, S. a. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *The Journal of Neuroscience*, 28(32), 8116–23. doi:10.1523/JNEUROSCI.1290-08.2008

- Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common α oscillatory network. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 32(36), 12376–83. doi:10.1523/JNEUROSCI.4908-11.2012
- Oh, S., & Lewis, C. (2008). Korean preschoolers' advanced inhibitory control and its relation to other executive skills and mental state understanding. *Child Development*, 79(1), 80–99. doi:10.1111/j.1467-8624.2007.01112.x
- Oldfield, R., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17(4), 273–281.
- Paap, K. R. (2015). Do many hones dull the bilingual whetstone? *Bilingualism: Language and Cognition*, 18(1), 41–42. doi:10.1017/S1366728914000431
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66(2), 232–58. doi:10.1016/j.cogpsych.2012.12.002
- Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual language development and disorders: a handbook on bilingualism and second language learning*. Baltimore, Md: Paul H. Brookes Pub. Co.
- Parra, M., Hoff, E., & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually developing 2-year-olds. *Journal of Experimental Child Psychology*, 108(1), 113–25. doi:10.1016/j.jecp.2010.07.011
- Perfetti, C. A. (2007). Reading Ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357–383.
- Perfetti, C. A., & Hart, L. (2002). The Lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Peyton, J. K., Ranard, D. A., & McGinnis, S. (2001). *Heritage Languages in America: Preserving a National Resource*. McHenry, IL: Delta Systems and Center for Applied Linguistics.
- Pichora-Fuller, M. K., Schneider, B., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, 97(1), 593–608.
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3), 105–110. doi:10.1016/j.tics.2006.12.002

- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–158). Amsterdam, NL: John Benjamins.
- Pierrehumbert, J. B. (2003). *Phonetic Diversity, Statistical Learning, and Acquisition of Phonology. Language and Speech* (Vol. 46). doi:10.1177/00238309030460020501
- Piquado, T., Cousins, K. a Q., Wingfield, A., & Miller, P. (2010). Effects of degraded sensory input on memory for speech: behavioral data and a test of biologically constrained computational models. *Brain Research, 1365*, 48–65. doi:10.1016/j.brainres.2010.09.070
- Pitt, M. A., Dilley, L., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics, 39*(3), 304–311. doi:10.1016/j.wocn.2010.07.004
- Pivneva, I., Palmer, C., & Titone, D. (2012). Inhibitory control and l2 proficiency modulate bilingual language production: evidence from spontaneous monologue and dialogue speech. *Frontiers in Psychology, 3*(1-18), 57. doi:10.3389/fpsyg.2012.00057
- Place, S., & Hoff, E. (2011). Properties of Dual Language Exposure That Influence 2-Year-Olds' Bilingual Proficiency. *Child Development, 82*(6), 1834–1849. doi:10.1111/j.1467-8624.2011.01660.x
- Portocarrero, J. S., Burright, R. G., & Donovanick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology : The Official Journal of the National Academy of Neuropsychologists, 22*(3), 415–22. doi:10.1016/j.acn.2007.01.015
- R Core Team. (2014). A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Ratnu, I., & Azuma, T. (2015). Working memory capacity: Is there a bilingual advantage? *Journal of Cognitive Psychology, 27*(1), 1–11. doi:10.1080/20445911.2014.976226
- Rimikis, S., Smiljanic, R., & Calandruccio, L. (2013). Nonnative English Speaker Performance on the Basic English Lexicon (BEL) Sentences. *Journal of Speech, Language, and Hearing Research, 56*, 792–805. doi:10.1044/1092-4388(2012/12-0178)materials
- Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics, 27*(03), 465–485. doi:10.1017/S014271640606036X
- Rönnberg, J., Lunner, T., Zekveld, A. A., Sörqvist, P., Danielsson, H., Lyxell, B., ... Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience, 7*, 1–31. doi:10.3389/fnsys.2013.00031

- Roodenrys, S., Hulme, C., Alban, J., Ellis, A. W., & Brown, G. D. a. (1994). Effects of word frequency and age of acquisition on short-term memory span. *Memory & Cognition*, 22(6), 695–701. doi:10.3758/BF03209254
- Rosenhouse, J., Haik, L., & Kishon-Rabin, L. (2006). Speech perception in adverse listening conditions in Arabic-Hebrew bilinguals. *International Journal of Bilingualism*, 10(2), 119–135. doi:10.1177/13670069060100020101
- Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, 12, 339–349. doi:10.1111/j.1467-7687.2008.00786.x
- Rubenstein, H., & Pollack, I. (1963). Word predictability and intelligibility. *Journal of Verbal Learning and Verbal Behavior*, 2(2), 147–158. doi:10.1016/S0022-5371(63)80079-1
- Sabbagh, M. a, Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The Development of Executive Functioning and Theory of Mind. *Psychological Science*, 17(1), 74–81.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103(3), 403–28.
- Samuel, A. G., & Larraza, S. (2015). Does listening to non-native speech impair speech perception? *Journal of Memory and Language*, 81, 51–71. doi:10.1016/j.jml.2015.01.003
- Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: Evidence from pupillometry. *Frontiers in Psychology*, 5(137). doi:10.3389/fpsyg.2014.00137
- Schneider, B. a., Avivi-Reich, M., & Daneman, M. (2014). How age and linguistic competence alter the interplay of perceptual and cognitive factors when listening to conversations in a noisy environment. *Frontiers in Systems Neuroscience*, 8(February), 1–17. doi:10.3389/fnsys.2014.00021
- Schrank, F., & Woodcock, R. W. (2005). WMLS-R scoring and reporting program [Computer software]. In *Woodcock-Muñoz Language Survey-Revised*. Rolling Meadows, IL: Riverside Publishing.
- Sears, C. R., Siakaluk, P. D., Chow, V. C., & Buchanan, L. (2008). Is there an effect of print exposure on the word frequency effect and the neighborhood size effect? *Journal of Psycholinguistic Research*, 37(4), 269–91. doi:10.1007/s10936-008-9071-5
- Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, 52(2), 240–255. doi:10.1016/j.jml.2004.11.001

- Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, 72(2), 111–23.
- Service, E., Simola, M., Metsänheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, 14(3), 383–408.
- Shannon, R. V, Jensvold, a, Padilla, M., Robert, M. E., & Wang, X. (1999). Consonant recordings for speech testing. *The Journal of the Acoustical Society of America*, 106(6), L71–4.
- Shannon, R. V, Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303–4.
- Shi, L.-F. (2009). Normal-hearing English-as-a-second-language listeners' recognition of English words in competing signals. *International Journal of Audiology*, 48, 260–270. doi:10.1080/14992020802607431
- 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), 821–35. doi:10.1044/1092-4388(2010/09-0081)
- Shi, L.-F. (2012). Contribution of Linguistic Variables to Bilingual Listeners' Perception of Degraded English Sentences. *Journal of Speech, Language and Hearing Research*, 55, 219–234. doi:10.1044/1092-4388(2011/10-0240)that
- Shi, L.-F., & Farooq, N. (2012). Linguistic and Attitudinal Factors in Normal-Hearing Bilingual Listeners' Perception of Degraded English Passages. *American Journal of Audiology*, 21, 127–140. doi:10.1044/1059-0889(2012/11-0022)At
- Shi, L.-F., & Sánchez, D. (2010). Spanish/English bilingual listeners on clinical word recognition tests: what to expect and how to predict. *Journal of Speech, Language, and Hearing Research*, 53(5), 1096–1110. doi:10.1044/1092-4388(2010/09-0199)
- Shi, L.-F., & Sánchez, D. (2011). The role of word familiarity in Spanish/English bilingual word recognition. *International Journal of Audiology*, 50(2), 66–76. doi:10.3109/14992027.2010.527862
- Sommers, M. S., & Barcroft, J. (2011). Indexical information, encoding difficulty, and second language vocabulary learning. *Applied Psycholinguistics*, 32, 417–434. doi:10.1017/S0142716410000469
- Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: the interaction of lexical competition and semantic context. *Psychology and Aging*, 14(3), 458–72.

- Sörqvist, P., Hurtig, A., Ljung, R., & Rönnerberg, J. (2014). High second-language proficiency protects against the effects of reverberation on listening comprehension. *Scandinavian Journal of Psychology*, 55(2), 91–6. doi:10.1111/sjop.12115
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388, 381–382. doi:10.1038/41102
- Swingle, D. (2003). Phonetic Detail in the Developing Lexicon. *Language and Speech*, 46(2-3), 265–294. doi:10.1177/00238309030460021001
- Swingle, D., & Aslin, R. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13(5), 480–484.
- Swingle, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76(2), 147–166. doi:10.1016/S0010-0277(00)00081-0
- Swingle, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, 54(2), 99–132. doi:10.1016/j.cogpsych.2006.05.001
- Tabri, D., Smith Abou Chacra, M. K., & Pring, T. (2011). Speech perception in noise by monolingual, bilingual and trilingual listeners. *International Journal of Language Communication Disorders*, 46(4), 411–422. doi:10.3109/13682822.2010.519372
- Tamati, T. N., Gilbert, J. L., & Pisoni, D. B. (2013). Some factors underlying individual differences in speech recognition on PRESTO: a first report. *Journal of the American Academy of Audiology*, 24(7), 616–634. doi:10.3766/jaaa.24.7.10.Some
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632 – 1634.
- Thorn, A. S. C., & Gathercole, S. E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 52(2), 303–24. doi:10.1080/713755823
- Titone, D., Pivneva, I., Sheikh, N. A., Webb, N., & Whitford, V. M. (2015). Doubling down on multifactorial approaches to the study of bilingualism & executive control. *Bilingualism: Language and Cognition*, 18(1), 43–44. doi:10.1017/S1366728914000595
- Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech Perception in Infancy Predicts Language Development in the Second Year of Life: A Longitudinal Study. *Child Development*, 75(4), 1067–1084. doi:10.1111/j.1467-8624.2004.00726.x
- Tulsky, D. S., Carlozzi, N., Chiaravalloti, N. D., Beaumont, J. L., Kisala, P. a, Mungas, D., ... Gershon, R. (2014). NIH Toolbox Cognition Battery (NIHTB-CB): list sorting test to

- measure working memory. *Journal of the International Neuropsychological Society*, 20(6), 599–610. doi:10.1017/S135561771400040X
- Vaden, K. I., Halpin, H., & Hickok, G. S. (2009). Irvine Phonotactic Online Dictionary, Version 2.0 [Data file]. Retrieved from <http://www.iphod.com>
- Valian, V. (2014). Bilingualism and cognition. *Bilingualism: Language and Cognition*, 18(01), 3–24. doi:10.1017/S1366728914000522
- 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. doi:10.1016/j.specom.2010.05.002
- Varela, R. E., Vernberg, E. M., Sanchez-Sosa, J. J., Riveros, A., Mitchell, M., & Mashunkashey, J. (2004). Parenting style of Mexican, Mexican American, and Caucasian-non-Hispanic families: social context and cultural influences. *Journal of Family Psychology : JFP : Journal of the Division of Family Psychology of the American Psychological Association (Division 43)*, 18(4), 651–657. doi:10.1037/0893-3200.18.4.651
- Vitevitch, M. S., & Luce, P. a. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, 36(3), 481–487. doi:10.3758/BF03195594
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428, 748–751.
- Von Hapsburg, D., & Peña, E. D. (2002). Understanding bilingualism and its impact on speech audiometry. *Journal of Speech, Language, and Hearing Research*, 45(1), 202–13. doi:10.1044/1092-4388(2002/015)
- Walley, A. (2008). Speech Perception in Childhood. In D. B. Pisoni & R. Remez (Eds.), *The handbook of speech perception* (pp. 449–468). Blackwell Publishing Ltd.
- Walley, A., Metsala, J. L., & Garlock, V. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing*, 16, 5–20.
- Weisleder, A., & Fernald, A. (2013). Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143–52. doi:10.1177/0956797613488145
- 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. doi:10.3766/jaaa.19.1.2

- Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' Ability to Learn Phonetically Similar Words: Effects of Age and Vocabulary Size. *Infancy*, 3(1), 1–30. doi:10.1207/15250000252828226
- White, K. S., Yee, E., Blumstein, S. E., & Morgan, J. L. (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. *Journal of Memory and Language*, 68(4), 362–378. doi:10.1016/j.jml.2013.01.003
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *The Journal of Neuroscience*, 32(40), 14010–21. doi:10.1523/JNEUROSCI.1528-12.2012
- Wilson, R. H., Abrams, H., & Pillion, A. (2003). A word-recognition task in multitalker babble using a descending presentation mode from 24 dB to 0 dB signal to babble. *Journal of Rehabilitation Research and Development*, 40(4), 321–328.
- Wilson, R. H., Carnell, C. S., & Cleghorn, A. L. (2007). The Words-in-Noise (WIN) test with multitalker babble and speech-spectrum noise maskers. *Journal of the American Academy of Audiology*, 18(6), 522–529.
- Wilson, R. H., McArdle, R., & Smith, S. L. (2007). An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners With Normal Hearing and Listeners With Hearing Loss. *Journal of Speech, Language, and Hearing Research*, 50(4), 844–56. doi:10.1044/1092-4388(2007/059)
- Wingfield, A. (1996). Cognitive factors in auditory performance: context, speed of processing, and constraints of memory. *Journal of the American Academy of Audiology*, 7(3), 175–182.
- Woodcock, R. W., Muñoz Sandoval, A. F., Ruef, M. L., & Alvarado, C. G. (2005). Woodcock-Muñoz Language Survey - Revised. Itasca, IL: Riverside Publishing.
- Yap, M. J., Balota, D., Sibley, D., & Ratcliff, R. (2012). Individual Differences in Visual Word Recognition: Insights From the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance*, 38(1), 53–79. doi:10.1037/a0024177
- Yoshida, K. a, Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science*, 12(3), 412–8. doi:10.1111/j.1467-7687.2008.00789.x
- Zhang, Y.-X., Barry, J. G., Moore, D. R., & Amitay, S. (2012). A new test of attention in listening (TAIL) predicts auditory performance. *PloS One*, 7(12), e53502. doi:10.1371/journal.pone.0053502
- Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F.-X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National*

Academy of Sciences of the United States of America, 102(39), 14110–5.
doi:10.1073/pnas.0504446102

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*, 12(5), 732–45. doi:10.1111/j.1467-7687.2009.00817.x