



The Bilingual Disadvantage in Speech Understanding in Noise Is Likely a Frequency Effect Related to Reduced Language Exposure

Jens Schmidtke*

Michigan State University, East Lansing, MI, USA

OPEN ACCESS

Edited by:

Jerker Rönnerberg,
Linköping University, Sweden

Reviewed by:

Sari Ylinen,
University of Helsinki, Finland
Sandra Campeanu,
Lehman College – City University of
New York, USA

*Correspondence:

Jens Schmidtke
schmi474@msu.edu

Specialty section:

This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Psychology

Received: 29 January 2016

Accepted: 22 April 2016

Published: 13 May 2016

Citation:

Schmidtke J (2016) The Bilingual
Disadvantage in Speech
Understanding in Noise Is Likely
a Frequency Effect Related
to Reduced Language Exposure.
Front. Psychol. 7:678.
doi: 10.3389/fpsyg.2016.00678

The present study sought to explain why bilingual speakers are disadvantaged relative to monolingual speakers when it comes to speech understanding in noise. Exemplar models of the mental lexicon hold that each encounter with a word leaves a memory trace in long-term memory. Words that we encounter frequently will be associated with richer phonetic representations in memory and therefore recognized faster and more accurately than less frequently encountered words. Because bilinguals are exposed to each of their languages less often than monolinguals by virtue of speaking two languages, they encounter all words less frequently and may therefore have poorer phonetic representations of all words compared to monolinguals. In the present study, vocabulary size was taken as an estimate for language exposure and the prediction was made that both vocabulary size and word frequency would be associated with recognition accuracy for words presented in noise. Forty-eight early Spanish–English bilingual and 53 monolingual English young adults were tested on speech understanding in noise (SUN) ability, English oral verbal ability, verbal working memory (WM), and auditory attention. Results showed that, as a group, monolinguals recognized significantly more words than bilinguals. However, this effect was attenuated by language proficiency; higher proficiency was associated with higher accuracy on the SUN test in both groups. This suggests that greater language exposure is associated with better SUN. Word frequency modulated recognition accuracy and the difference between groups was largest for low frequency words, suggesting that the bilinguals' insufficient exposure to these words hampered recognition. The effect of WM was not significant, likely because of its large shared variance with language proficiency. The effect of auditory attention was small but significant. These results are discussed within the Ease of Language Understanding model (Rönnerberg et al., 2013), which provides a framework for explaining individual differences in SUN.

Keywords: speech understanding in noise, bilingual, working memory, frequency effect, spoken word recognition

INTRODUCTION

Spoken language comprehension is a complex process that entails encoding an acoustic signal, matching it to the right phonological representation stored in long-term memory (LTM) out of thousands of such representations, and finally retrieving the semantic information associated with the phonological information and integrate it with the preceding information. Yet understanding

spoken language under optimal listening conditions is usually a seemingly effortless process. Only when it comes to listening to speech under suboptimal conditions do we become conscious of this process and individual differences in people's ability to understand speech become obvious. This is especially true in a second language, as many second language speakers can attest to and has also been shown in many studies (for a review see Garcia Lecumberri et al., 2010). What is surprising is that even speakers who learned their second language early in life and became dominant in that language still show poorer performance on speech understanding in noise (SUN) tests (Mayo et al., 1997; von Hapsburg et al., 2004; Rogers et al., 2006; Shi, 2010). To explain these findings the present study tested the hypothesis that bilinguals are disadvantaged in SUN because of their reduced exposure to each of their languages relative to monolinguals. Its contribution to the current discussion on bilingual SUN is a larger sample size of early bilingual speakers compared to previous studies and the presentation of a framework to explain bilingual disadvantages in auditory language comprehension.

The ease of language understanding (ELU) model (Rönnerberg et al., 2013) provides a framework for explaining the effects of a suboptimal speech signal on listening effort. The model assumes that during listening 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 bottom-up information from the speech signal is insufficient. In such cases where information in RAMBPHO cannot be matched to a LTM representation, explicit processing that involves working memory (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 (e.g., distorted speech or an unfamiliar accent) or speaker-internal reasons (imprecise phonological representations; Rönnerberg et al., 2013, p. 3).

The degree of similarity between the acoustic signal and an internal phonological representation determines the amount of processing that is needed for lexical access to be successful. When the match is optimal, processing is automatic and effortless. The greater the mismatch, the greater is the need for explicit processing of the signal. This explicit processing loop is dependent on WM resources. Thus, according to the model, individual differences in SUN can be attributable to two sources, individual differences in WM capacity and individual differences in the quality of speaker-internal phonological representations of words in LTM.

How can we explain differences in the quality of phonological representations? Exemplar models of the mental lexicon (Klatt, 1979; Goldinger, 1996, 1998; Pierrehumbert, 2000, 2003;

Hawkins, 2003; Johnson, 2005) may be especially useful here. In contrast to models that assume that words in the mental lexicon are stored in an abstract form without any indexical information (e.g., speaker voice characteristics such as gender, age, etc.), exemplar models assume that each encounter with a word token leaves a separate episodic trace in memory. Thus pronunciation variants and reduced forms, for example, are also assumed to be stored (e.g., Pierrehumbert, 2001; Pitt et al., 2011). Phonetic categories are not understood as discrete symbols but as distributions in a multidimensional space that develop through experience. With increased experience, listeners develop selective attention (c.f. Nosofsky, 1986) to those acoustic-phonetic dimensions that are relevant in a given language. In these models, the effect of word frequency arises from the assumption that words that are encountered often are represented with more exemplars on a "cognitive map" than infrequent words (Pierrehumbert, 2003). During retrieval, all exemplars with a certain degree of similarity to an acoustic signal receive activation. Thus frequently reoccurring units of speech (e.g., words) receive more activation since they are associated with more exemplars. This gives high frequency words an advantage over low frequency words in terms of speed of lexical access. Furthermore, the selection of high frequency words will be more robust when information in the acoustic signal is missing or when there is noise in the signal. However, it is not just the mere frequency with which words are encountered that determines the robustness of a representation. For example, research shows that variability in the signal as it occurs through different speakers helps infants extract the distribution of phonetic categories from the signal so that minimal pairs (e.g., *buk* and *puk*) sound less similar, presumably because variability directs infants' attention to the relevant dimension that distinguishes the minimal pairs (in this case voice onset time; Rost and McMurray, 2009). Exemplar models can also be extended to explain second language (L2) speech perception (Hardison, 2003, 2012). Because the acoustic-phonetic space is arbitrarily divided into phonetic categories that differ from language to language, listeners need to create new categories when learning a L2. Proponents of an exemplar-based mental lexicon assume that phonetic differences between a first language (L1) and a L2 can be perceived; however, at first old category labels will continue to be activated by L2 input. Again, acoustic variability in the signal may help the L2 learner create new phonetic categories by directing his attention to those dimensions that may be irrelevant in the L1 but vary systematically in the L2. For example, Japanese listeners need to learn to attend to the third formant (the third resonance peak of the vocal tract) to differentiate between American English /r/ and /l/ because this dimension is not relevant in their first language (Lotto et al., 2004). Perceptual training studies of the /r/-/l/ distinction with native Japanese speakers showed superior identification ability between the two phonetic categories when training stimuli were spoken by multiple speakers compared to a condition with a single speaker (Lively et al., 1993; Hardison, 2003, 2005). Also relevant for the present discussion is a finding from a study on second language vocabulary learning. Native English speakers who learned new Spanish words spoken by six

different speakers showed better retention and faster retrieval of those words compared to those who heard the novel words spoken by one speaker only (Barcroft and Sommers, 2005; also see Sommers and Barcroft, 2011). These findings suggest that token frequency of words in the input determine the quality of mental representations of words. Multiple exemplars associated with one word will make the retrieval of that word more efficient and robust.

Within the account described above, the assumption is made that the quality of phonological representations differs within and between speakers. Within speakers they differ because high frequency words are represented with more phonetic detail than low frequency words, and between speakers because some speakers have more language experience (i.e., more exposure) than others. These assumptions are similar to the lexical quality hypothesis developed by Perfetti and Hart (2002) and Perfetti (2007) to explain individual differences in reading comprehension. A further assumption made here is that bilingual speakers differ in language experience in one language from monolinguals because they speak and hear each of their languages less often compared to someone who only speaks one language. This assumption is expressed in the weaker-links hypothesis developed by (Gollan et al., 2002, 2005, 2008) to explain differences in lexical access between monolinguals and bilinguals (see Ivanova and Costa, 2008; Diependaele et al., 2013; Cop et al., 2015). As a result of reduced language experience, all words in a bilingual's mental lexicon will be of lower experienced frequency compared to a monolingual speaker. Frequency effects in general are pervasive in language processing (Ellis, 2002). Word frequency in particular affects lexical retrieval times (e.g., Oldfield and Wingfield, 1965; Murray and Forster, 2004) and recognition accuracy for words presented in noise (Howes, 1957). Frequency effects are logarithmic in nature, which means that changes in frequency at the low end affect lexical retrieval times and recognition accuracy more than changes at the high end (Murray and Forster, 2004). As a consequence, reduced language exposure will especially affect low frequency words. In one study, Kuperman and Van Dyke (2013) asked subjects with more and less education to rate words for their subjective, or experienced, frequency. When comparing the two groups, subjective ratings for words that are highly frequent in the language (based on a corpus count) were very similar but the lower the objective frequency, the more subjective frequency ratings of both groups diverged. This suggests that frequency estimates that are based on large corpora such as SUBTLEX (Brysbaert and New, 2009) may overestimate the frequency with which certain words are encountered for individuals with less language experience such as bilinguals. Thus the idea behind the weaker-links hypothesis and similar theories is that slower verbal processing in bilinguals is a frequency effect. Bilinguals encounter all words less frequently compared to monolinguals and so they process all words more slowly. Less efficient spoken word recognition has been shown for late and also early bilinguals (e.g., Weber and Cutler, 2004; Schmidtke, 2014).

While occurrence counts in large corpora of language can give us an idea of the relative quality of representations of

words in memory (the less frequent a word the less precise its representation), it is more difficult to estimate the overall language experience of individuals. Different means of data collection are possible such as asking participants to keep a diary of daily interactions for a week or similar techniques. However, these measures are based on self-report and do not capture language experience over longer periods of time. In this paper, the assumption is made that vocabulary knowledge, and more precisely productive vocabulary knowledge, closely resembles language experience and thus the quality of phonological representations. Individuals who are able to recall infrequent words must have been exposed to these words more often than someone who is not able to recall low frequency words. Someone with a weaker phonological representation of a word may be able to recall the first sound or a similar sounding word but lexical retrieval may not be successful. This phenomenon is usually referred to as a tip-of-the-tongue state in the literature (Brown and McNeill, 1966). A second reason for not knowing a word is that the participant may have never encountered the word before. This would also suggest reduced language experience because the more someone is exposed to language, the more likely they are to encounter an infrequent word (Kuperman and Van Dyke, 2013). The prediction is then that individuals with a higher score on a vocabulary test will be overall more accurate on a word-recognition-in-noise test, and the difference compared to someone with a lower vocabulary score will be most pronounced for low frequency words. The frequency effect might thus explain why SUN in a L2 is usually more difficult compared to one's first language and why this effect is modulated by experience in the L2 (e.g., Mayo et al., 1997; Rogers et al., 2006; Shi, 2009, 2010; Shi and Sánchez, 2010). At the same time, the frequency effect could also explain individual differences between monolingual speakers that have been shown to exist in normal hearing subjects (see Tamati et al., 2013).

As mentioned before, in the ELU there are two sources for speaker-internal individual differences in SUN. One source is the quality of internal phonological representations, as described above. The other source is differences in WM capacity. When mismatches between the acoustic signal and phonological representations occur, speech processing relies more on explicit processes, which presumably are more susceptible to individual differences in processing resources than implicit processes. Examples of such explicit processes include "inference making, semantic integration, switching of attention, storing of information, and inhibiting irrelevant information" (Rönnberg et al., 2013, p. 3). Individuals with greater WM capacity have more resources available for such processes and are thus better able to make up for missing information from the speech signal. In support of this hypothesis, studies have established a link between the quality of sensory information and maintenance of such information in WM. For instance, hearing verbal stimuli under suboptimal listening conditions leads to reduced recall accuracy of such stimuli even when intelligibility is not impaired (Rabbitt, 1966; Pichora-Fuller et al., 1995; Amichetti et al., 2013). Other studies have used brain imaging and found that alpha power, an indication of WM load, increased as a function of speech intelligibility (Obleser et al.,

2012) and degree of hearing loss of listeners (Petersen et al., 2015). Importantly, these studies established increased power in alpha oscillations during the retention phase of a memory test, suggesting that retaining degraded speech in WM is more effortful than clear speech, even when overall intelligibility is high.

Several studies have established a correlation between tests of verbal WM, typically assessed through the reading-span test (see Daneman and Carpenter, 1980), and performance on SUN tests. The problem with such studies is that no direct causation can be established as performance on both tests may be influenced by a third variable. Specifically, it has been shown that short-term memory (STM) for words is not independent of LTM representations of those words; both word frequency and phonotactic probability influence serial recall of words (Hulme et al., 1991; Hulme et al., 1997; Gathercole et al., 1999). At the same time, SUN is dependent on these factors as described above. Thus the quality of phonological representations in LTM, which is dependent on language experience, may influence both verbal WM and SUN. Therefore, studies that assess the correlation between verbal WM and SUN need to control for language experience to ensure that the correlation is not confounded by this third factor. Two recent studies found that verbal WM was no longer a significant predictor of SUN in a second language when proficiency in that language was controlled for (Kilman et al., 2014; Sörqvist et al., 2014).

Other executive functions next to WM may be recruited during SUN. When individuals follow a conversation in background noise, they have to selectively attend to one speaker and ignore other sounds or speakers (e.g., Mesgarani and Chang, 2012; Wild et al., 2012). In addition, during word recognition, words that are semantically and acoustically related to the target words also become active and inhibiting these competitors may require executive functions (Sommers and Danielson, 1999; Lash et al., 2013). Two recent studies assessed the relationship between individual differences in attention and SUN. Anderson et al. (2013) used structural equation modeling and found that a latent variable consisting of auditory attention, auditory STM, and auditory WM explained a large amount of variance in SUN. However, the contribution of auditory attention was only small compared to the memory measures, which suggests that in this specific study the role of auditory attention was limited. The second study comes from Tamati et al. (2013), who found that individuals who performed high and low on a SUN test did not differ in their performance on a color-Stroop test. This last finding might suggest that auditory attention is more important in SUN than more general attention that is measured by the Stroop test.

The purpose of the present study was to find individual differences that would predict SUN. Based on the ELU model, it was hypothesized that language experience, measured through vocabulary knowledge, verbal WM, and auditory attention would predict SUN. It was further hypothesized that differences between monolingual and bilingual participants would mostly be attributable to differences in language experience. To test this hypothesis, word frequency of to be recognized words was manipulated.

MATERIALS AND METHODS

Participants

The study included 53 monolingual and 48 bilingual participants. The inclusion criteria for monolinguals were that they did not learn a second language before the age of 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 8¹. In addition, participants had to be between 18 and 35 years old. Six additional monolinguals and five additional bilinguals were tested 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 study was approved by the local Institutional Review Board and all subjects gave their written informed consent to participate.

Experimental Design

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 6–10 min to administer.

Speech Understanding in Noise

Materials for the SUN test were taken from the revised Speech Perception in Noise test (SPIN; Bilger et al., 1984), which was obtained as a digitized recording. 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 from the Department of Speech and Hearing Science from the University of Illinois Urbana-Champaign. The sound file was edited so that each sentence was saved in a separate file. For the background babble, a short sequence from the original babble track (12-talker babble) was chosen and mixed with each sentence in Praat (Boersma and Weenink, 2014) at two different speech-to-noise ratios (SNRs; –2 dB and 3 dB). These SNRs were chosen based on a pilot experiment. The sound intensity of the sentence was held constant and so the intensity of the babble differed for the two SNRs.

¹Four 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 the present study, 128 sentences from the test were chosen and divided into four lists of 32 words². Words in each list were matched on word frequency, phonotactic probability, and on neighborhood density. 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 log₁₀ word frequency of the stimuli used in the present study was 2.70 ($SD = 0.44$) and the mean frequency per million was 15.92 ($SD = 16.46$). 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.

²The word *fun* was later dropped from all analyses because its frequency per million of 235 was several SDs away from the mean of 15.9. Thus there were 127 unique items.

TABLE 1 | Participant characteristics divided by language group.

	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
Verbal ability <i>W</i> -score	533.2 (8.9)	515.6 (11.4)
Verbal ability standard score	105 (7.7)	90 (8.8)
Picture vocabulary <i>W</i> -score	537.1 (11.0)	516.1 (13.5)
Picture vocabulary standard score	101 (7.6)	86 (8.4)
Verbal analogies <i>W</i> -score	529.5 (9.2)	515.3 (11.8)
Verbal analogies standard score	109 (7.3)	98 (9.0)
Verbal ability <i>W</i> -score – Spanish	–	503.0 (11.9)
Verbal ability standard-score – Spanish	–	81 (9.3)
Picture vocabulary <i>W</i> -score – Spanish	–	500.8 (11.8)
Picture vocabulary standard score – Spanish	–	77 (7.9)
Verbal analogies <i>W</i> -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)
Current amount of time spent		
Listening to English	–	64.6% (18.4)
Speaking English	–	65.5% (17.4)
Reading English	–	81.3% (16.7)

Years of musical experience: participants were asked if they have ever played a musical instrument or sung in a choir or band and for how many years. See text for an explanation of the different language measures. W-scores are based on an arbitrary equal-interval scale. Standard scores have a population mean of 100 and a standard deviation of 15. Values in parentheses are standard deviations.

The number of neighbors of a word were calculated based on the English Lexicon Project (Balota et al., 2007). The correlation between biphone probability and log-frequency was $r = 0.16$ and the correlation between log-frequency and neighborhood density was $r = 0.16$.

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 predictive 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, USA).

Working Memory

The WM test used for this study comes from the US National Health Institute's (NIH) so called Toolbox³. 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 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, and 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 make a mistake in either practice set, the instructions are repeated and the set is 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 stops when a participant cannot correctly repeat two sets in a row or when the last set is 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 0.77. The test also correlated with other established WM tests

³www.nihtoolbox.org

($r = 0.57$) and tests of executive function ($r = 0.43 - 0.58$). The correlation with a test of receptive vocabulary, on the other hand, was low ($r = 0.24$).

Verbal Ability

Verbal ability was assessed with the Woodcock-Muñoz Language Survey – Revised (WMLS-R; Woodcock et al., 2005), which 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 and 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 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” while pointing at a picture. The second test administered is called Verbal Analogies. 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 (henceforth verbal ability). This score correlates highly with the cluster score that is based on all tests of the WMLS-R ($r = 0.9$). The standard error of the mean for all tests is between 5.55 and 5.93 and the internal consistency reliability coefficients were around $r_{11} = 0.9$ (Alvarado and Woodcock, 2005).

Auditory Attention

The auditory attention test was adapted from Zhang et al. (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 in 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 part 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 test, 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)]. Preliminary correlational analyses (see Supplementary Materials) with each of these three measures and overall accuracy on the SUN test showed that only processing speed correlated significantly with SUN accuracy and so only this variable was used in the analyses reported below.

The tones for this test were created in Praat (Boersma and 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. For different-frequency trials, the second tone was randomly chosen. There were a total of 96 experimental trials, 24 trials in each condition. The experiment was programmed in E-Prime.

Procedure

All participants completed all tests in the following order: consent form, background questionnaire, attention test, Words-in-Noise test (Wilson et al., 2007, not reported here), SUN test, verbal ability, WM test, and a consonant perception test (not reported here). Bilingual participants then also completed the verbal ability and the Words-in-Noise test in Spanish.

Analysis

Incorrect responses on the SUN test were manually checked for any spelling mistakes. A misspelled word was counted as correct in the following cases: letter transposing (e.g., thief 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 in this way, which is comparable to 2.5% in Luce and Pisoni (1998) who used a similar procedure.

For the analysis, mixed-effects regression models were run in R (R Core Team, 2015) using the *lme4* package (Bates et al., 2015). *P*-values were calculated using the *Anova* function in the *car* package (Fox and Weisberg, 2011) using the type II sums of squares method. Subjects and items were entered as random effects.

RESULTS

First a model was run with four predictor variables to analyze group-level effects, language group (bilingual/monolingual), predictability (low/high), noise level (low/high), and word frequency (see the Supplementary Materials for model specifications). The results showed that words in low noise ($M = 85.5\%$, $SD = 35.2$) were recognized with higher accuracy than words in high noise [$M = 67.6\%$, $SD = 46.8$; $\chi^2(1) = 712.4$, $p < 0.001$], and words in a predictive context ($M = 88.7\%$,

$SD = 31.6$) better than words in an unpredictable context [$M = 64.4\%$, $SD = 47.9$; $\chi^2(1) = 1059.3$, $p < 0.001$]. The difference between a low and a highly predictive context was 28.2% when noise was high and 20.5% when noise was low and this interaction was significant [$\chi^2(1) = 30.7$, $p < 0.001$]. Monolinguals recognized words more accurately ($M = 80.8\%$, $SD = 39.4$) than bilinguals [$M = 71.8\%$, $SD = 45.0$; $\chi^2(1) = 76.7$, $p < 0.001$]. When noise was low, the difference between monolinguals and bilinguals was smaller [$M_{\Delta} = 7.1$ percentage points (pp)] than when noise was high ($M_{\Delta} = 10.9$ pp) but this interaction did not reach significance [$\chi^2(1) = 3.19$, $p = 0.074$]. The effect of predictability was only slightly larger for monolinguals ($M_{\Delta} = 24.8$ pp) than bilinguals ($M_{\Delta} = 23.8$ pp). Nevertheless, the interaction between predictability and language group was significant [$\chi^2(1) = 47.56$, $p < 0.001$]. As can be seen in **Figure 1**, this interaction was likely caused by the fact that monolinguals benefitted more from a predictive context compared to bilinguals when noise was high. In the high noise condition, the benefit for monolinguals was $M_{\Delta} = 31.06$ pp and $M_{\Delta} = 24.87$ pp for bilinguals. The main effect of frequency was significant [$\chi^2(1) = 6.00$, $p = 0.014$], showing that high frequency words were recognized with greater accuracy than low frequency words. The interaction between frequency and language group was also significant [$\chi^2(1) = 5.65$, $p = 0.017$]. **Figure 2** suggests that this interaction was driven by the steeper slope of the frequency effect in the bilingual group compared to the monolingual group.

The following variables were added to the analysis to investigate the effect of individual differences: verbal ability, WM, and processing speed. All continuous variables were centered around the mean. The mean values for each variable can be seen in **Table 2**. WM and verbal ability were highly correlated [$r(99) = 0.527$, $p < 0.001$] and WM and processing speed were moderately correlated [$r(99) = 0.229$, $p = 0.021$]. Processing speed and verbal ability were not correlated [$r(99) = 0.034$, $p = 0.737$; see the Supplementary Materials for a detailed correlation matrix].

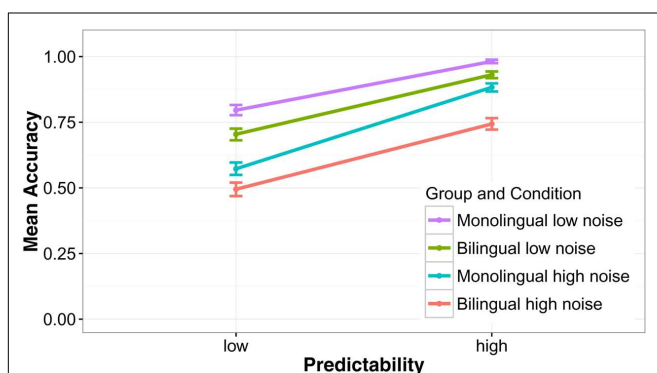


FIGURE 1 | Mean accuracy on the speech understanding in noise test. Sentences were presented at two noise levels (low/high) and the predictability of the target word was either high or low. Error bars show the 95% confidence interval.

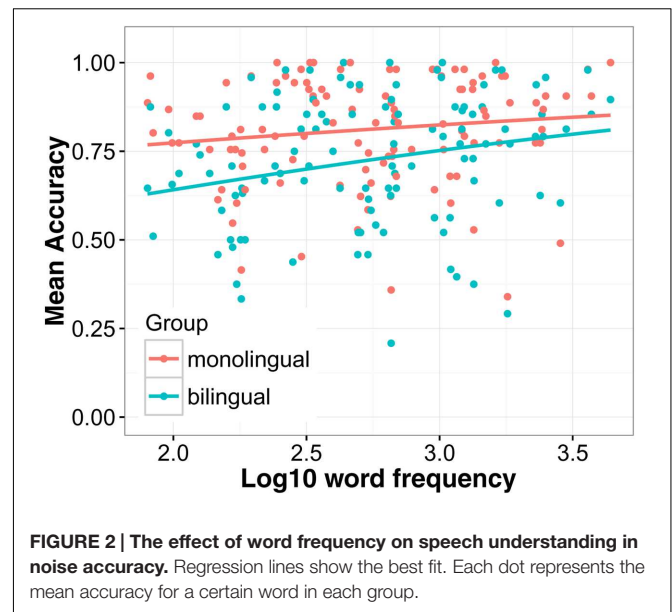


FIGURE 2 | The effect of word frequency on speech understanding in noise accuracy. Regression lines show the best fit. Each dot represents the mean accuracy for a certain word in each group.

A model was built with the same variables as above, that is, language group, word frequency, noise level, and predictability, plus the individual difference variables. Besides the main effects, only the significant interactions are reported here. The full model can be seen in the Supplementary Materials.

The main effects of language group, noise level, and predictability were highly significant as before (all $\chi^2 > 10$, $ps < 0.001$). Furthermore, main effects of verbal ability [$\chi^2(1) = 44.51$, $p < 0.001$] and processing speed [$\chi^2(1) = 5.87$, $p = 0.015$] were significant, showing that higher verbal ability and faster processing speed (lower RTs) were associated with higher accuracy on the SUN test. This can be seen in **Figures 3** and **4** respectively. The interaction between verbal ability and predictability was significant [$\chi^2(1) = 53.10$, $p < 0.001$]. As **Figure 3** shows, participants with higher verbal ability benefitted more from a predictive context compared to those with lower verbal ability. The interaction between word frequency and verbal ability was also significant [$\chi^2(1) = 5.13$, $p = 0.024$]. This interaction can best be interpreted using **Figure 5**. The difference in accuracy between listeners with high and low verbal ability was most pronounced for low frequency words. WM was not a significant predictor of SUN accuracy [$\chi^2(1) < 0.01$, $p = 0.978$], likely because of its high correlation with verbal ability (when verbal ability was taken out of the model, WM became a significant predictor; see Supplementary Materials). These analyses show that verbal ability was a powerful predictor of SUN accuracy. Expressed as an odds-ratio, compared to someone with average verbal ability, someone with verbal ability 1 SD above the mean was 2.14 times more likely to recognize a target word. Compared to verbal ability, the effect of processing speed was much smaller. Compared to someone with mean processing speed, someone 1 SD below the mean was 1.09 times more likely to recognize a target word.

To check whether the effect of verbal ability was true for both groups or was simply driven by group differences, follow-up

TABLE 2 | Mean values for the individual differences variables.

	Verbal ability	Working memory	Processing speed
Monolingual	533.2 W (8.9)	37.6 (8.0)	680 ms (125)
Bilingual	515.6 W (11.4)	32.4 (7.9)	702 ms (139)
Total sample	524.8 W (13.4)	35.2 (8.3)	690 ms (132)

W-scores are based on an arbitrary equal-interval scale. See text for an explanation of working memory scores. Processing speed is the baseline mean response time on the attention test. Standard deviations are shown in parentheses.

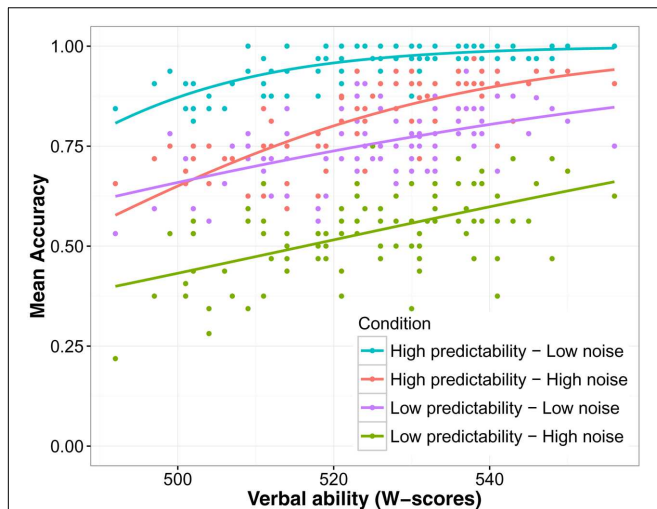


FIGURE 3 | The effect of verbal ability on speech understanding in noise accuracy. *W*-scores are on an arbitrary equal-interval scale. Each dot in each condition represents one participant.

analyses were run for each group separately. The main effect of verbal ability and the interaction with predictability were highly significant in both groups (all $\chi^2 > 15$, $ps < 0.001$) but the interaction with frequency was no longer significant (both $\chi^2 < 1$). The main effect of frequency was significant in the bilingual group [$\chi^2(1) = 8.61$, $p = 0.003$] but not in the monolingual group [$\chi^2(1) = 3.27$, $p = 0.071$]. Furthermore, the effect of processing speed did not reach significance in either group ($ps = 0.058$ and 0.129 for the bilingual and monolingual group, respectively). This may have been due to insufficient power in these smaller samples.

The analyses so far suggest that verbal ability had an effect on SUN in both the monolingual and the bilingual group. Yet, even when verbal ability was controlled for, language group was still a significant predictor. To investigate further what the added difficulty for bilinguals might be, two subgroups were formed from each group, respectively, that were closely matched on their vocabulary score⁴ by randomly selecting participants from each group with a similar score (see Table 3). A *t*-test confirmed that the difference in vocabulary scores between these subgroups was not significantly different [$t(44) = 0.63$, $p = 0.534$]. The

⁴For matching, only the vocabulary score (i.e., Picture Vocabulary) was compared because in the bilingual group, English verbal reasoning was correlated with Spanish verbal reasoning and so vocabulary is likely a better indicator of English exposure.

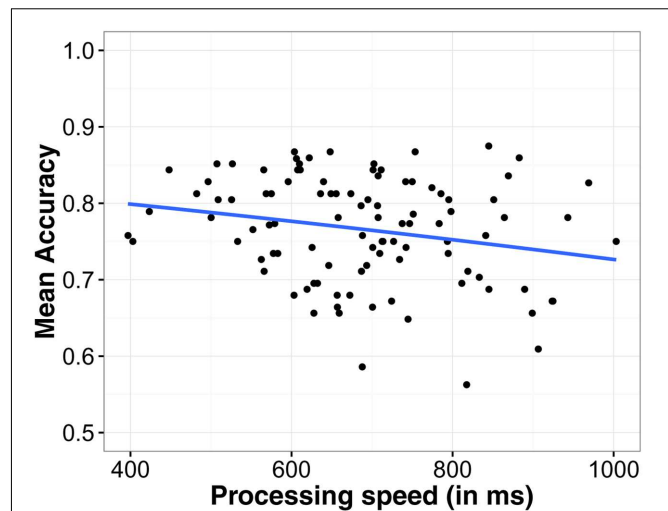


FIGURE 4 | The effect of processing speed on speech understanding in noise accuracy. Processing speed was the baseline measure on the attention test. Note the reduced range of the *y*-axis to highlight the effect.

mean group difference in SUN accuracy in this subsample was $M_{\Delta} = 5.1$ pp, which is smaller than in the total sample ($M_{\Delta} = 9.0$ pp). Yet this difference was still statistically significant [$\chi^2(1) = 15.35$, $p < 0.001$]. The interaction between word frequency and language group was not significant [$\chi^2(1) = 2.02$, $p = 0.155$] but Figure 6 suggests that it was especially the low frequency words that were more difficult for bilinguals. Also the language group by predictability interaction was still significant in this subsample [$\chi^2(1) = 4.07$, $p = 0.044$], suggesting that differences in language proficiency alone cannot explain this interaction.

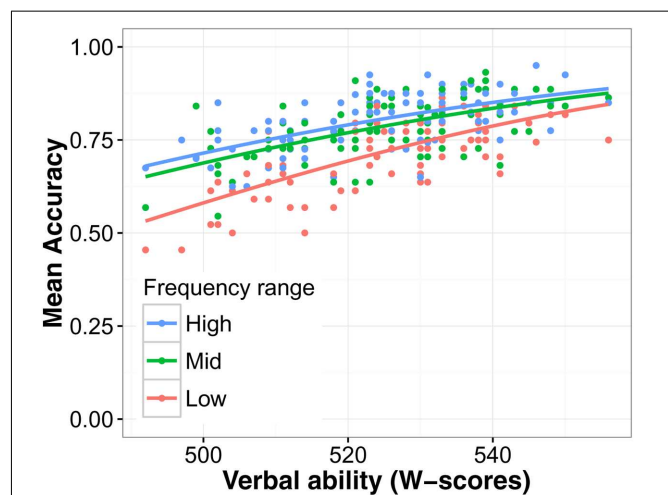


FIGURE 5 | The effects of oral language ability and word frequency on speech understanding in noise accuracy. To show the effect of word frequency, the continuous variable was split into a factor with three levels, high, mid, and low frequency.

TABLE 3 | A subsample from each group matched on language proficiency.

	<i>n</i>	PV score	VA score	SUN accuracy
Monolingual	24	529.8 W (7.8)	521.3 W (11.5)	80.5% (39.7)
Bilingual	22	528.3 W (8.1)	526.8 W (9.5)	75.4% (43.1)

A subsample from each group was randomly chosen and matched on their picture vocabulary score. PV = picture vocabulary. VA = verbal ability. SUN = speech understanding in noise. Standard deviations are shown in parentheses.

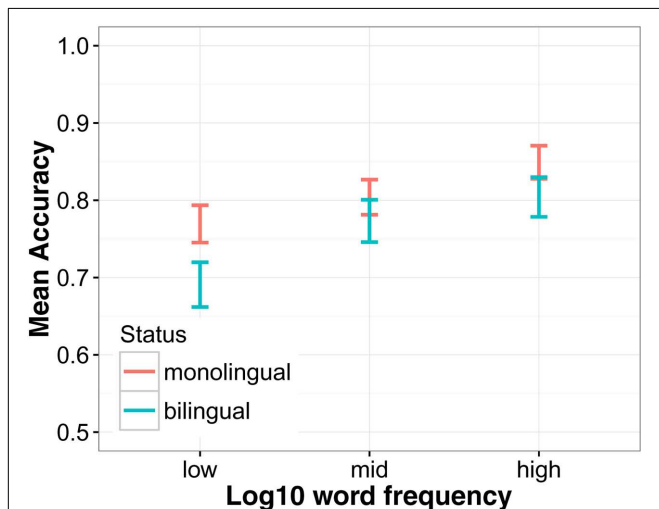


FIGURE 6 | The effect of word frequency in a subsample of monolinguals and bilinguals matched on vocabulary knowledge. Word frequency was split into a factor with three levels, high, mid, and low frequency, to highlight the effect. Also note the limited range of the y-axis.

DISCUSSION

The results confirmed previous studies by showing that noise had a disruptive effect on speech understanding whereas a predictive context was facilitative. The effect of a predictive context was stronger when noise was high compared to when it was low and monolinguals benefitted more from a predictive context than bilinguals. Word frequency had an effect on recognition accuracy, high frequency words were recognized with greater accuracy than low frequency words. However, in follow-up analyses, this effect was only marginally significant in the monolingual group, while it remained significant in the bilingual group. Next, an analysis of the effect of individual differences in verbal ability, WM, and attention was conducted. The effect of verbal ability was highly significant in both groups, as was the interaction between verbal ability and predictive context, showing that individuals with higher verbal ability recognized more words in general and also benefitted more from a predictive context. The effect of WM was not significant, likely because of its shared variance with verbal ability. The effect of processing speed was significant when both groups were analyzed together but did not reach significance when each group was analyzed separately. Finally, two subsamples from each group that were matched on their vocabulary scores were compared. This analysis showed that

group differences were reduced when subjects were matched on verbal ability but the differences were still statistically significant, suggesting that differences in verbal ability cannot completely explain the bilingual disadvantage in SUN.

As in previous studies (e.g., Mayo et al., 1997; Meador et al., 2000; Rogers et al., 2006; Bradlow and Alexander, 2007; Shi, 2009, 2010; Van Engen, 2010), the bilingual speakers recognized fewer words on average than the monolingual speakers. However, the effect was additive rather multiplicative, meaning there was no interaction between noise level and group. This is in line with Rogers et al. (2006). Yet, other studies found an interaction (Mayo et al., 1997; Shi, 2010; Tabri et al., 2011). The reason may be that in the present study only two noise levels were tested. Another reason may be that the bilinguals in the present study learned English early and had grown up in an English-speaking environment. They were thus more proficient than many of the second language speakers tested in previous research.

An improvement to many previous studies that compared monolingual to early bilingual listeners (e.g., Mayo et al., 1997; Shi and Sánchez, 2010) was the inclusion of a larger sample. Thus there is more robust evidence that even early bilinguals have greater difficulties recognizing words in noise. Previous research also established that more exposure, a younger age of acquisition, and greater proficiency in the target language is positively associated with SUN (Meador et al., 2000; Shi, 2009, 2012; Rimikis et al., 2013; Kilman et al., 2014). The present study sought to contribute to the current discussion of bilingual SUN by not only showing the existence of a so-called *bilingual disadvantage* and which factors contribute to it but also to find possible explanations for this disadvantage. In this respect, an improvement to previous research was that monolingual and bilingual participants were tested with the same standardized language test. A standardized test is not only important to make results comparable across studies but also to be able to compare the samples of monolinguals and bilinguals within a study. This is important to note because the present study found that verbal ability was associated with SUN in both groups. Since bilinguals often have a smaller vocabulary in each of their languages compared to monolinguals (e.g., Portocarrero et al., 2007; Bialystok and Luk, 2012; Gasquoine and Dayanira Gonzales, 2012), one reason for the bilingual disadvantage for SUN in previous studies may be that groups simply differed in verbal ability. This assumption was confirmed when two subsamples were compared that were matched on vocabulary size. Compared to the total sample, the difference went down from 9.0 to 5.1 pp, which is a decrease of 43%. At the same time, differences in language proficiency cannot be the only explanation because even these two subsamples matched on proficiency were still significantly different in SUN accuracy.

Word frequency may be a second, albeit related, explanation for the bilingual disadvantage in SUN. All participants recognized high frequency words with higher accuracy than low frequency words. As described in the introduction, exemplar models of speech perception assume that each encounter with a word leaves a trace in memory and that words that are encountered frequently are represented in memory with more phonetic detail. The more a word is encountered in different contexts,

spoken by different speakers, the more robust its recognition will be under suboptimal listening conditions. Because on average bilingual speakers have not had as much exposure to each of their languages compared to a monolingual speaker, all words are encountered less often (c.f. Gollan et al., 2008) and disproportionately less so low frequency words (Kuperman and Van Dyke, 2013). This can explain the interaction between group and word frequency, which showed that the bilinguals as a group recognized especially low frequency words with lower accuracy than monolinguals. This explanation suggests that the bilingual disadvantage stems from their reduced exposure to each of their languages. Thus we would expect the same to be true for monolinguals who, for various reasons, have not had as much exposure to lower frequency words. For example, Tamati et al. (2013) tested a large sample of native English speakers on a SUN test and also had subjects rate their subjective familiarity with certain words. They found that those who performed well on the test reported to be more familiar with low frequency words than those who performed not so well on the test. Assuming that familiarity is closely related to the frequency of encounter with a word, their study and the present one suggest that subjective word frequency is an important factor influencing individual differences in SUN.

Both explanations for individual differences in SUN, verbal ability and word frequency, are related because both depend on language experience. Someone who is exposed to language in many different contexts is more likely to learn the meaning of more words compared to someone with more limited exposure and, at the same time, they will encounter words of lower frequency more often. How, then, can we explain that the two subsamples that were matched on verbal ability still performed significantly different on the SUN test? It may be that for the bilinguals, vocabulary knowledge overestimated their actual exposure to English. Even though they knew the meaning of a less common word, they may not have encountered that word as often as a monolingual speaker. Also, assuming that a bilingual speaker hears English in school and Spanish outside of school, they will hear each language not only less often but also from a more limited number of speakers. These may be factors that determine the quality of phonological representations (Gollan et al., 2014; Schmidtke, 2015) and thus SUN. Suggestive of this explanation is that, as in the whole sample, the largest difference between these two subsamples was in the low frequency range (see **Figure 6**), although the interaction between group and frequency did not reach significance. In this respect it is interesting that the size of the frequency effect changed as a function of proficiency. The effect was most pronounced for participants at the lower end of the proficiency range. In the matched subsamples, however, the proficiency range was smaller and this may be why the interaction was no longer significant.

While the present hypothesis for the bilingual disadvantage was based on exemplar models, the data do not necessarily contradict the predictions of models that assume an abstract level of representation of words. For example, TRACE (McClelland and Elman, 1986) assumes three levels of representation, a feature level, a phoneme level, and a word level, with each level of representation being more abstract. Frequency effects

can be modeled by adjusting the resting-activation levels of words so that words with high resting levels require less activation from the speech signal, which results in earlier selection compared to words with low resting-activation levels (Dahan et al., 2001). A noisy signal could result in fewer features that receive activation so that words with a low resting-activation level do not receive sufficient activation to pass the threshold necessary for selection. Proponents of a mental lexicon with abstract representations of words can explain differences between native and non-native speech perception by assuming differences at a perceptual level. Because categorical speech perception develops very early in life (Kuhl, 2004), even an early learned second language will be perceived through the phonemic inventory of the first language (e.g., Sebastián-Gallés and Soto-Faraco, 1999), which will result in non-native-like phonological representations in the mental lexicon (Pallier et al., 2001). However, the two models do not have to stand in opposition to each other and more recently researchers have developed hybrid models that include aspects of exemplar and abstract models to be able to explain the whole range of phenomena (e.g., Goldinger, 2007; Ernestus, 2014; Kleinschmidt and Jaeger, 2015; Pierrehumbert, 2016). This being said, exemplar models provide a more elegant solution to explain the present results. Differences in the quality of mental representations of words between and within speakers are a fundamental part of exemplar-based models and so they can readily explain individual differences in word recognition. Abstract models, on the other hand, have to assume additional mechanisms to be able to explain individual differences.

Exemplar-based models may also be useful to explain the finding that individual differences in WM capacity were not a significant predictor of SUN when controlling for language ability. A verbal WM test was included in the current study because of the ELU's prediction that individuals with a larger WM capacity would recognize words in noise with less effort and thus be more accurate. The test required individuals to remember items in different set sizes and to mentally manipulate the order of the items according to their size. Because of these storage and processing components, the test is believed to tap into WM. Individuals who can correctly recall more sets are assumed to have a larger WM capacity. The items were common animals and food items such as *mouse*, *pig*, and *banana* that all participants were likely very familiar with. It was therefore surprising that the test correlated highly with the language test ($r = 0.5$). Exemplar-based models can explain this finding because they assume that not only one representation is activated at the time of encoding but all exemplars of a word. If a word is represented by many exemplars then it is more likely that a memory trace is still active in LTM at the time of retrieval. Related to this explanation is also the finding that items stored in WM are not independent from LTM representations (e.g., Hulme et al., 1997; Acheson et al., 2011). Additionally, in individuals with larger mental lexicons the phonological representations of words may be overall more precise, which may reduce the spread of activation to similar sounding words and therefore prevent interference during rehearsal (cf. Cowan et al., 2005). However,

although WM was not a significant predictor of SUN accuracy in the present sample, this does not necessarily imply that individual differences in WM are not important for SUN. The participants here were all young adults and a more diverse sample in terms of age may be needed to find an effect of verbal WM above and beyond verbal ability. For example, Parbery-Clark et al. (2011) found a correlation between auditory WM and SUN ability even when controlling for vocabulary knowledge in a sample of older listeners. But the present results may further inform the ELU in that the quality of lexical representations in LTM and capacity limits of WM are not independent constructs. This view would be more akin to the model of WM developed by Cowan et al. (2005) and Cowan (2008) rather than to a limited capacity system for temporary storage of items as it is currently defined in the ELU (Rönnberg et al., 2013, p. 2). The present results also have implications for future research. Researchers interested in the relationship between SUN and cognition should always also include a proficiency test that measures vocabulary knowledge in their test batteries when they administer a verbal WM test. Otherwise correlations may be attributed to WM (or some other covariate) when in fact language experience is the underlying factor. However, the type of verbal ability test used may also lead to differing results, since an effect of verbal ability is not always found (e.g., Benichov et al., 2012). In the same way, in the norming study of the WM used here the authors found a much weaker correlation between WM and receptive vocabulary ($r = 0.24$; Tulsy et al., 2014).

The next finding that merits discussion is the effect of a predictive context. Previous research found that bilingual and second language speakers do not benefit as much from a predictive context as monolinguals under certain circumstances (Mayo et al., 1997; Bradlow and Alexander, 2007; Shi, 2010). However, the present results suggest that individual differences in the effective use of context also exist between monolinguals and that verbal ability is the mediating factor. This would again suggest that differences between monolinguals and bilinguals might emerge because of differences in verbal ability (see above). As a result, the less effective use of context cues attributed to bilingualism is not a bilingual disadvantage *per se* but may be a result of reduced language experience (cf. Newman et al., 2012). But what is the relationship between verbal ability and the effective use of context cues? One explanation is that individuals with lower verbal ability generally understood fewer words and so if they missed words in the preceding context of the target words, they were not able to form any predictions. Another explanation may be the relationship between verbal ability and WM. In order to make predictions about the target word, subjects need to maintain preceding words in WM. This process might take up more resources depending on the ease with which phonological representations are retrieved and maintained in WM. A third explanation may be the association strength between words (Spence and Owens, 1990). One example sentence from the SUN test is *the ship sailed along the coast*. Here, *ship* and *sailed* may be used to predict the target word *coast*. If individuals with larger vocabularies have more language experience overall, then they have likely heard words such as *ship* and *coast* more often in the same context and thus

there is a stronger association of *ship* and *coast* compared to an individual with less language experience (c.f. Nation and Snowling, 1999).

Given the findings discussed so far, a frequency-based explanation of differences between monolinguals and bilinguals seems to be the most powerful because it cannot only account for group differences but also differences between individual participants. Furthermore, a frequency-based account can give a united explanation of the language-related effects such as language proficiency, word frequency, predictive context, and the null-effect of verbal WM. The last variable to be discussed, attention, stood out in this respect because it was not language related. The attention test was included in the study to give a more complete picture of individual differences in SUN, as recent studies have pointed to the potential role of non-linguistic factors in language comprehension and especially SUN (e.g., Anderson et al., 2013; Fedorenko, 2014).

The attention test based on Zhang et al. (2012) provided three different variables but no prediction was made as to which variable would be associated with SUN. In the analysis, only processing speed was used because it provided the most robust correlations with the SUN test of the three variables. The results showed a small but significant effect of processing speed on SUN accuracy. The reason why this effect was small might be that there was not enough variance in the data for a stronger effect to emerge. As with WM, processing speed may become more important as a factor in older populations. The general speed of information processing slows down in older adults (Salthouse, 1996), which may explain why cognitive factors are sometimes a better predictor of SUN than hearing acuity (Wingfield, 1996; Benichov et al., 2012). However, further studies are needed to confirm or disconfirm that processing speed is indeed a better predictor of SUN than the conflict resolution or involuntary attention components of the test.

One practical implication of the study for hearing testing is that word frequency needs to be taken into account. One possibility is to only use high frequency words when testing patients to avoid a possible confound. On the other hand, it may be useful to test high and low frequency words and to have norms for each set. If a patient fares especially poor for the low frequency words then this might be an indication for the practitioner that part of the patient's hearing difficulties may stem from factors unrelated to hearing acuity.

Some limitations of the present study that qualify the results should be addressed. Inherent to the design of the study, no inferences about causation can be made. The results suggest that a larger vocabulary is associated with better SUN but the nature of this relationship requires further investigations. Here the assumption was made that exposure frequency is the mediating variable but vocabulary size could also have a direct influence on word recognition. Alternatively, though less likely, people with better SUN ability may be better able to pick up new words through listening and therefore have larger vocabularies. Another limitation is that only one WM and one attention test were used. Future studies would benefit from the use of multiple tests for each construct, which, along with a larger sample size, would allow more sophisticated statistics such as

structural equation modeling. Finally, the two samples did not only differ in language status (monolingual vs. bilingual) but also in the age of acquisition (AoA) of the tested language and socio-economic status (SES; assessed by maternal education level). In the present study, additional tests showed that neither variable was a significant predictor of SUN once language proficiency was accounted for but these results may be different in a sample where AoA and SES are not correlated with verbal ability.

CONCLUSION

The purpose of the present study was to find factors that would explain individual differences in SUN between listeners, especially between monolingual and bilingual listeners. Previous research had established that bilinguals often performed below monolinguals on SUN tests, even when the bilinguals had learned the second language early in life. The present study confirmed these results but the general conclusion was that differences between groups could largely be explained by frequency effects, which suggests that differences between groups are less categorical than might be assumed based on previous research. Based on the ELU model (Rönnberg et al., 2013), it was hypothesized that listening difficulty arises from mismatches between the speech signal and internal phonological representations. Mismatches can occur because of a poor signal and because of poor phonological representations in LTM. In the current ELU model, the definition of what poor phonological representations are is underspecified and so the ELU was extended to exemplar models of the mental lexicon (e.g., Goldinger, 1996, 1998). These models assume that each encounter with a word leaves an episodic trace in memory. The present study showed that recognition of high frequency words was more robust to noise compared to low frequency words. Exemplar models can explain this finding in that high frequency words are represented in memory with more exemplars and more highly activated exemplars than low frequency words (Pierrehumbert, 2001). Word retrieval of high frequency words is more robust because a new exemplar will more likely be similar to an already stored exemplar when more exemplars of a word exist in memory. Following these assumptions, the premise of the study was that the bilingual disadvantage in SUN is a frequency effect (c.f. Gollan et al., 2008). Because bilinguals are exposed to each of their languages less often than monolinguals, they encounter all words less frequently. Consequently, bilinguals will have fewer stored exemplars in LTM for all words. This will especially affect the recognition of low

frequency words as bilinguals will encounter these even more rarely than monolinguals and consequently recognition of these words under noise is expected to be more fragile. In support of this hypothesis, the present study found that differences in SUN between groups were largest for low frequency words. Another consequence of reduced exposure to each language is a smaller vocabulary. As in previous research (Portocarrero et al., 2007; Bialystok and Luk, 2012), bilinguals scored on average below monolinguals on verbal ability test, and higher verbal ability was associated with better performance on the SUN test. Importantly, however, there was a relationship between verbal ability and SUN for both groups, suggesting that some of the group differences might be explained by the overall lower English proficiency of the bilinguals. When two subgroups that were matched on language proficiency were compared, the difference in performance on the SUN test was much smaller (5.1% compared to 9.0%). These results support the hypothesis that differences in SUN between monolinguals and bilinguals are a result of the bilinguals' reduced exposure to each of their languages as a consequence of being bilingual.

FUNDING

The research was supported by a Doctoral Dissertation Improvement Grant (NSF-DDIG 1349125) from the US National Science Foundation.

ACKNOWLEDGMENTS

The research presented here was part of the author's doctoral work (Schmidtke, 2015) that he conducted while at Michigan State University. He is currently at the China University of Petroleum East, Qingdao, China. I am grateful to my dissertation committee, Aline Godfroid, Laura Dilley, Debra Hardison, and Paula Winke, and the two reviewers for their comments, which greatly improved the present article. I thank Kara Morgan-Short for providing research space at UIC and Karthik Durvasula for assistance with Praat.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00678>

REFERENCES

- Acheson, D., Hamidi, M., Binder, J., and Postle, B. (2011). A common neural substrate for language production and verbal working memory. *J. Cogn. Neurosci.* 23, 1358–1367. doi: 10.1162/jocn.2010.21519
- Alvarado, C. G., and Woodcock, R. W. (2005). *Comprehensive Manual*. Rolling Meadows, IL: Riverside Publishing.
- Amichetti, N. M., Stanley, R. S., White, A. G., and Wingfield, A. (2013). Monitoring the capacity of working memory: executive control and effects of listening effort. *Mem. Cogn.* 41, 839–849. doi: 10.3758/s13421-013-0302-0
- Anderson, S., White-Schwoch, T., Parbery-Clark, A., and Kraus, N. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hear. Res.* 300, 18–32. doi: 10.1016/j.heares.2013.03.006
- Balota, D., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behav. Res. Methods* 39, 445–459. doi: 10.3758/BF03193014
- Barcroft, J., and Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Stud. Second Lang. Acquisit.* 27, 387–414. doi: 10.1017/S0272263105050175

- Bates, D. M., Maechler, M., Bolker, B., and Walker, S. (2015). lme4: linear mixed-effects models using Eigen and Eigen. *J. Statist. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Benichov, J., Cox, L. C., Tun, P. A., and Wingfield, A. (2012). Word recognition within a linguistic context: effects of age, hearing acuity, verbal ability, and cognitive function. *Ear. Hear.* 33, 250–256. doi: 10.1097/AUD.0b013e31822f680f
- Bialystok, E., and Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism* 15, 397–401. doi: 10.1017/S136672891100040X
- Bilger, R., Nuetzel, J., Rabinowitz, W., and Rzezczkowski, C. (1984). Standardization of a test of speech perception in noise. *J. Speech Hear. Res.* 27, 32–48. doi: 10.1044/jshr.2701.32
- Boersma, P., and Weenink, D. (2014). *Praat: Doing Phonetics by Computer (Version 5.3.51) [Computer Program]*. Available at: <http://www.praat.org/>
- Bradlow, A. R., and Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J. Acoust. Soc. Am.* 121, 2339–2349. doi: 10.1121/1.2642103
- Brown, R., and McNeill, D. (1966). The “tip of the tongue” phenomenon. *J. Verbal Learn. Verbal Behav.* 5, 325–337. doi: 10.1016/S0022-5371(66)80040-3
- Brybaert, M., and 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. *Behav. Res. Methods* 41, 977–990. doi: 10.3758/BRM.41.4.977
- Cop, U., Keuleers, E., Drieghe, D., and Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychon. Bull. Rev.* 22, 1216–1234. doi: 10.3758/s13423-015-0819-2
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Prog. Brain Res.* 169, 323–338. doi: 10.1016/S0079-6123(07)00020-9
- Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., et al. (2005). On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cogn. Psychol.* 51, 42–100. doi: 10.1016/j.cogpsych.2004.12.001
- Dahan, D., Magnuson, J. S., and Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: evidence from eye movements. *Cogn. Psychol.* 42, 317–367. doi: 10.1006/cogp.2001.0750
- Daneman, M., and Carpenter, P. (1980). Individual differences in working memory and reading. *J. Verbal Learn. Verbal Behav.* 466, 450–466. doi: 10.1016/S0022-5371(80)90312-6
- Diependaele, K., Lemhöfer, K., and Brybaert, M. (2013). The word frequency effect in first- and second-language word recognition: a lexical entrenchment account. *Q. J. Exp. Psychol.* 66, 843–863. doi: 10.1080/17470218.2012.720994
- Ellis, N. C. (2002). Reflections on frequency effects in language processing. *Stud. Second Lang. Acquisit.* 24, 297–339. doi: 10.1017/S0272263102002140
- Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* 142, 27–41. doi: 10.1016/j.lingua.2012.12.006
- Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Front. Psychol.* 5:335. doi: 10.3389/fpsyg.2014.00335
- Fox, J., and Weisberg, S. (2011). *An R Companion to Applied Regression*, 2nd Edn. Thousand Oaks, CA: Sage Publication.
- Garcia Lecumberri, M. L., Cooke, M., and Cutler, A. (2010). Non-native speech perception in adverse conditions: a review. *Speech Commun.* 52, 864–886. doi: 10.1016/j.specom.2010.08.014
- Gasquoine, P. G., and Dayanira Gonzales, C. (2012). Using monolingual neuropsychological test norms with bilingual hispanic americans: application of an individual comparison standard. *Arch. Clin. Neuropsychol.* 27, 268–276. doi: 10.1093/arclin/acs004
- Gathercole, S. E., Frankish, C. R., Pickering, S. J., and Peaker, S. (1999). Phonotactic influences on short-term memory. *J. Exp. Psychol.* 25, 84–95. doi: 10.1037/0278-7393.25.1.84
- Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *J. Exp. Psychol.* 22, 1166–1183. doi: 10.1037/0278-7393.22.5.1166
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 105, 251–279. doi: 10.1037/0033-295X.105.2.251
- Goldinger, S. D. (2007). “A complementary-systems approach to abstract and episodic speech perception,” in *Proceeding of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 49–54.
- Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: aging, bilingualism, and the weaker links hypothesis. *J. Mem. Lang.* 58, 787–814. doi: 10.1016/j.jml.2007.07.001
- Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., and Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Mem. Cogn.* 33, 1220–1234. doi: 10.3758/BF03193224
- Gollan, T. H., Montoya, R. I., and Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology* 16, 562–576. doi: 10.1037//0894-4105.16.4.562
- Gollan, T. H., Starr, J., and Ferreira, V. S. (2014). More than use it or lose it: the number-of-speakers effect on heritage language proficiency. *Psychon. Bull. Rev.* 22, 147–155. doi: 10.3758/s13423-014-0649-7
- Hardison, D. M. (2003). Acquisition of second-language speech: effects of visual cues, context, and talker variability. *Appl. Psycholinguist.* 24, 495–522. doi: 10.1017/S0142716403000250
- Hardison, D. M. (2005). Second-language spoken word identification: effects of perceptual training, visual cues, and phonetic environment. *Appl. Psycholinguist.* 26, 579–596. doi: 10.1017/S0142716405050319
- Hardison, D. M. (2012). “Second language speech perception: a cross-disciplinary perspective on challenges and accomplishments,” in *The Routledge Handbook of Second Language Acquisition*, eds S. M. Gass and A. Mackey (Abingdon: Routledge Taylor & Francis), 349–363.
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *J. Phonet.* 31, 373–405. doi: 10.1016/j.wocn.2003.09.006
- Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of english words. *J. Acoust. Soc. Am.* 29, 296–305. doi: 10.1121/1.1908862
- Hulme, C., Maughan, S., and Brown, G. D. (1991). Memory for familiar and unfamiliar words: evidence for a long-term memory contribution to short-term memory span. *J. Mem. Lang.* 30, 685–701. doi: 10.1016/0749-596X(91)90032-F
- Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., and Stuart, G. (1997). Word-frequency effects on short-term memory tasks: evidence for a reintegration process in immediate serial recall. *J. Exp. Psychol.* 23, 1217–1232.
- Ivanova, I., and Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychol.* 127, 277–288. doi: 10.1016/j.actpsy.2007.06.003
- Johnson, K. A. (2005). “Speaker Normalization in Speech Perception,” in *The Handbook of Speech Perception*, eds D. B. Pisoni and R. Remez (Malden, MA: Blackwell Publishing Ltd), 363–389. doi: 10.1002/9780470757024.ch15
- Kilman, L., Zekveld, A. A., Hällgren, M., and Rönnberg, J. (2014). The influence of non-native language proficiency on speech perception performance. *Front. Psychol.* 5:651. doi: 10.3389/fpsyg.2014.00651
- Klatt, D. H. (1979). Speech perception: a model of acoustic-phonetic analysis and lexical access. *J. Phonet.* 4, 75–95. doi: 10.1162/jocn.2011.21620
- Kleinschmidt, D. F., and Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychol. Rev.* 122, 148–203. doi: 10.1037/a0038695
- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843. doi: 10.1038/nrn1533
- Kuperman, V., and Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *J. Exp. Psychol.* 39, 802–823. doi: 10.1037/a0030859
- Lash, A., Rogers, C. S., Zoller, A., and Wingfield, A. (2013). Expectation and entropy in spoken word recognition: effects of age and hearing acuity. *Exp. Aging Res.* 39, 235–253. doi: 10.1080/0361073X.2013.779175
- Lively, S. E., Logan, G. D., and Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: the roles of phonetic environment and talker variability in learning new perceptual categories. *J. Acoust. Soc. Am.* 94, 1242–1255. doi: 10.1121/1.408177
- Lotto, A. J., Sato, M., and Diehl, R. L. (2004). “Mapping the task for the second language learner: the case of Japanese acquisition of /r/ and /l/,” in *From Sound*

- to Sense: 50+ Years of Discoveries in Speech Communication, eds J. Slifka, S. Manuel, and M. Matthies (Cambridge, MA: MIT Press), 181–186.
- Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear Hear.* 19, 1–36.
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007/067)
- Mayo, L. H., Florentine, M., and Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *J. Speech Lang. Hear. Res.* 40, 686–693. doi: 10.1044/jslhr.4003.686
- McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0
- Meador, D., Flege, J. E., and Mackay, R. (2000). Factors affecting the recognition of words in a second language. *Bilingualism* 3, 55–67. doi: 10.1017/S1366728900000134
- Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi: 10.1038/nature11020
- Murray, W. S., and Forster, K. I. (2004). Serial mechanisms in lexical access: the rank hypothesis. *Psychol. Rev.* 111, 721–756. doi: 10.1037/0033-295X.111.3.721
- Nation, K., and Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: evidence from semantic priming. *Cognition* 70, B1–B13. doi: 10.1016/S0010-0277(99)00004-9
- Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., and Ullman, M. T. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of English. *J. Cogn. Neurosci.* 24, 1205–1223. doi: 10.1162/jocn_a_00143
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol. Gen.* 115, 39–61. doi: 10.1037/0096-3445.115.1.39
- Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., and Maess, B. (2012). Adverse listening conditions and memory load drive a common α oscillatory network. *J. Neurosci.* 32, 12376–12383. doi: 10.1523/JNEUROSCI.4908-11.2012
- Oldfield, R., and Wingfield, A. (1965). Response latencies in naming objects. *Q. J. Exp. Psychol.* 17, 273–281. doi: 10.1080/17470216508416445
- Pallier, C., Colomé, A., and Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: exemplar-based versus abstract lexical entries. *Psychol. Sci.* 12, 445–449. doi: 10.1111/1467-9280.00383
- Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., and Kraus, N. (2011). Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. *PLoS ONE* 6:e18082. doi: 10.1371/journal.pone.0018082
- Perfetti, C. A. (2007). Reading ability: lexical quality to comprehension. *Sci. Stud. Read.* 11, 357–383. doi: 10.1080/10888430701530730
- Perfetti, C. A., and Hart, L. (2002). “The lexical quality hypothesis,” in *Precursors of Functional Literacy*, eds L. Verhoeven, C. Elbro, and P. Reitsma (Amsterdam: John Benjamins Publishing Company), 189–213.
- Petersen, E. B., Wöstmann, M., Obleser, J., Stenfelt, S., and Lunner, T. (2015). Hearing loss impacts neural alpha oscillations under adverse listening conditions. *Front. Psychol.* 6:177. doi: 10.3389/fpsyg.2015.00177
- Pichora-Fuller, M. K., Schneider, B., and Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J. Acoust. Soc. Am.* 97, 593–608. doi: 10.1121/1.412282
- Pierrehumbert, J. B. (2000). What people know about sounds of language. *Stud. Linguist. Sci.* 29, 111–120.
- Pierrehumbert, J. B. (2001). “Exemplar dynamics: word frequency, lenition and contrast,” in *Frequency Effects and Emergent Grammar*, eds J. Bybee and P. Hopper (Amsterdam, NL: John Benjamins), 137–158.
- Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Lang. Speech* 46(Pt 2–3), 115–154. doi: 10.1177/00238309030460020501
- Pierrehumbert, J. B. (2016). Phonological representation: beyond abstract versus episodic. *Annu. Rev. Linguist.* 2, 33–52. doi: 10.1146/annurev-linguist-030514-125050
- Pitt, M. A., Dille, L., and Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *J. Phonet.* 39, 304–311. doi: 10.1016/j.wocn.2010.07.004
- Portocarrero, J. S., Burrell, R. G., and Donovick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Arch. Clin. Neuropsychol.* 22, 415–422. doi: 10.1016/j.acn.2007.01.015
- R Core Team (2015). *A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rabbitt, P. (1966). Recognition: memory for words correctly heard in noise. *Psychon. Sci.* 6, 383–384. doi: 10.3758/BF03330948
- Rimikis, S., Smiljanic, R., and Calandruccio, L. (2013). Nonnative english speaker performance on the basic english lexicon (BEL) sentences. *J. Speech Lang. Hear. Res.* 56, 792–805. doi: 10.1044/1092-4388(2012/12-0178)materials
- Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Appl. Psycholinguist.* 27, 465–485. doi: 10.1017/S014271640606036X
- Rönning, J., Lunner, T., Zekveld, A. A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The ease of language understanding (ELU) model: theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* 7:31. doi: 10.3389/fnsys.2013.00031
- Rost, G. C., and McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Dev. Sci.* 12, 339–349. doi: 10.1111/j.1467-7687.2008.00786.x
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychol. Rev.* 103, 403–428. doi: 10.1037/0033-295X.103.3.403
- Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Front. Psychol.* 5:137. doi: 10.3389/fpsyg.2014.00137
- Schmidtke, J. (2015). *Consequences of Bilingualism for Speech Understanding in Noise*. Ph.D. Dissertation (Publication No. 3716774.), Michigan State University. Ann Arbor, MI: ProQuest/UMI.
- Schrank, F., and Woodcock, R. W. (2005). *WMLS-R Scoring and Reporting Program [Computer Software]*. Rolling Meadows, IL: Riverside Publishing.
- Sebastián-Gallés, N., and Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition* 72, 111–123. doi: 10.1016/S0010-0277(99)00024-4
- Shi, L.-F. (2009). Normal-hearing English-as-a-second-language listeners’ recognition of English words in competing signals. *Int. J. Audiol.* 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. *J. Speech Lang. Hear. Res.* 53, 821–835. doi: 10.1044/1092-4388(2010/09-0081)
- Shi, L.-F. (2012). Contribution of linguistic variables to bilingual listeners’ perception of degraded english sentences. *J. Speech Lang. Hear. Res.* 55, 219–234. doi: 10.1044/1092-4388(2011/10-0240)that
- Shi, L.-F., and Sánchez, D. (2010). Spanish/English bilingual listeners on clinical word recognition tests: what to expect and how to predict. *J. Speech Lang. Hear. Res.* 53, 1096–1110. doi: 10.1044/1092-4388(2010/09-0199)
- Sommers, M. S., and Barcroft, J. (2011). Indexical information, encoding difficulty, and second language vocabulary learning. *Appl. Psycholinguist.* 32, 417–434. doi: 10.1017/S0142716410000469
- Sommers, M. S., and Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: the interaction of lexical competition and semantic context. *Psychol. Aging* 14, 458–472. doi: 10.1037/0882-7974.14.3.458
- Sörqvist, P., Hurtig, A., Ljung, R., and Rönning, J. (2014). High second-language proficiency protects against the effects of reverberation on listening comprehension. *Scand. J. Psychol.* 55, 91–96. doi: 10.1111/sjop.12115
- Spence, D. P., and Owens, K. C. (1990). Lexical co-occurrence and association strength. *J. Psycholinguist. Res.* 19, 317–330. doi: 10.1007/BF01074363
- Tabri, D., Chacra, M. K., and Pring, T. (2011). Speech perception in noise by monolingual, bilingual and trilingual listeners. *Int. J. Lang. Commun. Disord.* 46, 411–422. doi: 10.3109/13682822.2010.519372
- Tamati, T. N., Gilbert, J. L., and Pisoni, D. B. (2013). Some factors underlying individual differences in speech recognition on PRESTO: a first report. *J. Am. Acad. Audiol.* 24, 616–634. doi: 10.3766/jaaa.24.7.10.Some
- Tulsky, D. S., Carullo, N., Chiaravalloti, N. D., Beaumont, J. L., Kisala, P. A., Mungas, D., et al. (2014). NIH Toolbox Cognition Battery (NIHTB-CB): list sorting test to measure working memory. *J. Int. Neuropsychol. Soc.* 20, 599–610. doi: 10.1017/S135561771400040X

- Van Engen, K. J. (2010). Similarity and familiarity: second language sentence recognition in first- and second-language multi-talker babble. *Speech Commun.* 52, 943–953. doi: 10.1016/j.specom.2010.05.002
- Vitevitch, M. S., and Luce, P. A. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behav. Res. Methods Instruments Comput.* 36, 481–487. doi: 10.3758/BF03195594
- von Hapsburg, D., Champlin, C. A., and Shetty, S. R. (2004). Reception thresholds for sentences in bilingual (Spanish/English) and monolingual (English) listeners. *J. Am. Acad. Audiol.* 15, 88–98. doi: 10.3766/jaaa.15.1.9
- Weber, A., and Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *J. Mem. Lang.* 50, 1–25. doi: 10.1016/S0749-596X(03)00105-0
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., and Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *J. Neurosci.* 32, 14010–14021. doi: 10.1523/JNEUROSCI.1528-12.2012
- Wilson, R. H., Carnell, C. S., and Cleghorn, A. L. (2007). The Words-in-Noise (WIN) test with multitalker babble and speech-spectrum noise maskers. *J. Am. Acad. Audiol.* 18, 522–529.
- Wingfield, A. (1996). Cognitive factors in auditory performance: context, speed of processing, and constraints of memory. *J. Am. Acad. Audiol.* 7, 175–182.
- Woodcock, R. W., Muñoz Sandoval, A. F., Rief, M. L., and Alvarado, C. G. (2005). *Woodcock-Muñoz Language Survey - Revised*. Itasca, IL: Riverside Publishing.
- Zhang, Y.-X., Barry, J. G., Moore, D. R., and Amitay, S. (2012). A new test of attention in listening (TAIL) predicts auditory performance. *PLoS ONE* 7:e53502. doi: 10.1371/journal.pone.0053502

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Schmidtke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.