

The Effect of Background Language and Speaker Differences on Serial Recall of Aurally

Presented Single-Talker Stimuli

By

Copyright 2011

Breanna I. Steidley

Submitted to the graduate degree program Linguistics and the Graduate Faculty of the University
of Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

Chairperson Allard Jongman

Joan Sereno

Utako Minai

Date Defended: March 28, 2011

The Thesis Committee for Breanna I. Steidley
certifies that this is the approved version of the following thesis:

The Effect of Background Language and Speaker Differences on Serial Recall of Aurally
Presented Single-Talker Stimuli

Chairperson Allard Jongman

Date approved: March 28, 2011

Abstract

Previous studies suggest that hearing a sound during the completion of a working memory task negatively affects recall performance. This is true for a wide variety of competing background sounds including multitalker babble, unfamiliar-language multitalker babble, white noise, and reversed speech. It is also suggested that the closer the background speaker is to the target speaker physically, the more difficult it is to separate the two speech streams, which results in lowered recall performance. The present study investigates serial recall performance of English words in single-talker background speech when the background language is either the listeners' native language (English) or an unknown language (Korean). Moreover, the effect of speaker differences is tested by using the same (bilingual) speaker for both foreground and background in one condition and introducing a novel foreground speaker in a second condition. An effect of background language did not emerge overall, but on some lists, Korean background speech is less distracting than English background speech. The comparison of recall performance between conditions revealed that there is not a significant difference overall when the background speaker is either the same as or different from the foreground speaker. On individual lists, when the background speaker is physically different from the foreground speaker, listeners recall significantly more words than when the background and foreground speaker are the same. These results support previous work and provide evidence for theories on target-masker separation.

Acknowledgements

I would like to thank those who have supported me throughout the duration of this project. My gratitude is extended to Ji-Yeon Lee for connecting me with a bilingual speaker for this study. I would also like to thank my colleagues and friends for providing me with a sounding board for ideas and for providing me with their support. They always managed to keep me focused and optimistic.

Perhaps my best motivation has been my family, and a debt of gratitude is owed them. My husband, Brian, and my son, Gracen, inspired me to work diligently, but to remember to have fun each and every day. I am honored to thank my parents, Roy and Richelle, and my siblings, Katie and Joe, for their guidance as well as for encouraging me to work hard to achieve my goals—whatever they may be.

Thank you to my committee members, Dr. Joan Sereno and Dr. Utako Minai for your comments and questions throughout. I am grateful to have had your input.

Finally, I would like to thank my advisor, Dr. Allard Jongman, for helping me to become a critical thinker and a researcher. These skills are priceless and I am ever grateful to have obtained them.

Table of Contents

Title Page	i
Acceptance Page	ii
Abstract	iii
Acknowledgements	iv
I. Introduction	1
II. Experiment 1: Presentation Rate and Sound-to-Noise Ratio	13
II.A. Participants	13
II.B. Stimuli	14
II.B.1. Target Words	14
II.B.2. Background Speech	15
II.B.3. Mixing	16
II.C. Procedure	17
II.D. Results	17
III: Experiment 2: Independent Ratings of the Bilingual Speaker	19
III.A. Participants	19
III.A.1. Speakers	19
III.A.2. Listeners	20
III.B. Stimuli	20
III.C. Procedure	20
III.D. Results	21
IV: Experiment 3: Individual Word Intelligibility Test	23
IV.A. Participants	23
IV.A.1. Speakers	23
IV.A.2. Listeners	23
IV.B. Stimuli	24
IV.C. Procedure	24
IV.D. Results	24
IV.E. Discussion	26
V. Experiment 4: The Effect of Background Language on Serial Recall	27
V.A. Participants	28
V.A.1. Speaker	28
V.A.2. Listeners	28
V.B. Stimuli	28
V.C. Procedure	28
V.D. Results	29
V.E. Discussion	32
VI. Experiment 5: Differing language and talker	34
VI.A. Participants	34
VI.A.1. Speakers	34
VI.A.2. Listeners	34
VI.B. Stimuli	34
VI.C. Procedure	35
VI.D. Results	35

VI.E. Discussion	38
VII. Experiment 4 and 5 Combined Results.....	39
VII.A. Results	39
VII.B. Discussion.....	41
VIII. Experiment 4 and 5 Performance by Gender: Within-Groups Analyses.....	42
VIII.A. Results	43
VIII.B. Discussion	44
IX. Conclusion.....	46
Works Cited.....	50
Appendix I.....	52
Appendix II.....	53

I. Introduction

Each day, we are required to focus on tasks regardless of other distractions around us. Studies on working memory and immediate serial recall have shown that hearing a sound that is unrelated to the completion of the task lowers performance as a result of auditory distraction. This has been shown to be true for both speech (Drullman et al., 1999; Rhebergen et al., 2005; Tremblay et al., 2001; Van Engen & Bradlow, 2006) and non-speech sounds (Beaman et al., 2005; Bradlow et al., 2007; Sperry et al. 1997) used as competitors. More specifically, this is true for recall tasks where the to-be-remembered stimuli are embedded in a competing speech signal. Previous studies have found that participants perform worse when the intensity of the background speech is the same or higher than that of the target speech (Van Engen & Bradlow, 2006). It has also been found that the number of talkers in multitalker babble (Van Engen & Bradlow, 2006), native and non-native single-talker speech (Rhebergen et al., 2005) and reverse speech (Rhebergen et al., 2005; Sperry et al., 1997) all are contributors to different levels of distraction of working memory.

The Baddeley-Hitch model (1974) of working memory proposes that recall of aurally presented stimuli is more difficult because the “phonological loop” is occupied when competing sounds are played and that this inhibits the listener’s ability to rehearse the stimuli for recall because it is being occupied by both the background speech and the target speech (Baddeley, 1998). In addition to deficits in working memory, the competing speech also may affect perception of sounds due to interference and confusion. Van Engen and Bradlow (2006) found that correct identification of sounds was least accurate when the background speech was very dense. In the 2-talker babble background speech conditions, where the speech signals of two different speakers are mixed into one, listeners were more accurate than in the denser 6-talker babble background speech condition (Van Engen & Bradlow, 2006). Reduced perception coupled

with a taxing on the phonological loop in working memory then should result in low performance in serial recall tasks. The present study will investigate this competing sound effect further to determine if differences in the language of the background speech affect listeners in similar or different ways in recall tasks.

Baddeley and Hitch first proposed their working memory model in 1974 and it described working memory as a sort of working space that could be filled with information for a short time. The information is then processed and either moved to long-term memory storage or forgotten (Baddeley, 2003; Baddeley & Hitch, 1974). The model also includes a phonological loop, which is of key importance to the present study. The phonological loop allows for rehearsal of material placed in working memory in order to lengthen its time on the working-memory slate (Baddeley, 1998; Baddeley, 2003). In order for the phonological loop to be useful for this purpose, it must be unoccupied. If there are competing sounds present, particularly speech sounds, then those occupy the phonological loop and must be processed, which yields lower performance in tasks that require working memory such as serial recall tasks.

The working memory model describes the process of how speech affects recall performance from a wide scope, but does not explicitly describe how different types of speech may affect recall. There are many different types of background speech that can be used and in turn, there are many different tasks that can be applied. Combining background speech and tasks in different ways may yield widely conflicting results because some types of competitor sounds may affect the phonological loop more than others; language, intelligibility, frequency of words and predictability are just a few of the factors that may influence the ability of listeners to disregard background competing speech (Bradlow & Alexander, 2007; Oswald et al., 2000; Poirier & Saint-Aubin, 1996; Rhebergen et al., 2005; Van Engen & Bradlow, 2006). In addition

to the competing sound, some tasks may be more difficult than others without competing speech. There are several studies that investigate these issues using permutations of these variables: type of background speech and task.

Two types of recall tasks are typically used in working memory tests: free recall and serial recall. In free recall, listeners are given a list of words, digits or consonants and are allowed to recall the words in any order they choose. Beaman & Jones (1998) found that even when free recall is allowed, listeners tend to recall words serially or by semantic category, if the presented lists contain words that can be grouped together. Free recall may also include recall of sentences (Rhebergen et al., 2005; Van Engen & Bradlow, 2006). In many cases, researchers choose keywords from each sentence to quantify correct recall. Free recall of sentences can be quite useful in tests of competing speech because the task is simple, yet shows to what extent linguistic knowledge and working memory are taxed when competing sounds are present. This method yields more holistic results than simply recalling digits or consonants because listeners are required to access different semantic and morphological classes in order to complete the task, which is most similar to background speech conditions in day-to-day hearing.

Serial recall removes the aides that are sometimes found in free recall, like grouping by semantic category or by phonetic structure. In free recall, order is not required, but in serial recall it is. In order for a word to be scored as correct, it must be the correct word and in the correct serial position. This method requires two simultaneous responses from the listeners: correct serial order and correct word. Fewer tokens are used per list in serial recall due to the difficulty in recalling both the word and the serial position of the word on the list as a whole (Beaman, 2005; Tremblay et al., 2001).

In addition to various types of recall tasks, the presentation of the target stimuli must be

considered. Visual presentation of the stimuli requires participants to view the presentation of stimuli on a screen or to read information on paper and then recall or repeat it. During the reading of the target stimuli, background speech is played. In aural presentation, both the target stimuli and the background speech are played simultaneously and typically in a single sound stream. With visual presentation, there is little chance that the participant's perception of the targets is impaired because perception of the visual targets is not disrupted by the background speech, only recall ability is impaired.

In aural presentation of both target and masker, the listener must discern which stream is the focus and then focus on that target while disregarding the background masker. Van Engen and Bradlow (2006) support this with their comment that in their 2-talker English babble condition, listeners often transcribed words from the babble instead of the target speech stream. This suggests that when target and masker are presented aurally, the listener has difficulty with the perception of the target speech. The present study will use auditory presentation of targets and maskers.

One of the main studies of speech perception in differing background language speech conditions is Van Engen and Bradlow (2006). The study investigated whether background speech in the listener's native language is more distracting than background speech in a language unfamiliar to the listener. The two languages used were English and Mandarin. The researchers also varied signal-to-noise ratio (SNR) and the number of background talkers. By implementing these variables, the researchers determined if the language of the background speech is the sole cause of the detriment to performance or if it is a result of other factors like SNR or number of talkers because they could analyze performance both between participants and across groups. The groups were organized by number of talkers heard. Table 1 shows the group organization for

this study.

	No. of Talkers in Background Babble	Background Language			
		Mandarin (Block 1)	English (Block 2)	Mandarin (Block 3)	English (Block 4)
<i>Condition 1</i>	6	SNR: +5	SNR: +5	SNR: 0	SNR: 0
<i>Condition 2</i>	6	SNR: 0	SNR: 0	SNR: -5	SNR: -5
<i>Condition 3</i>	2	SNR: +5	SNR: +5	SNR: 0	SNR: 0
<i>Condition 4</i>	2	SNR: 0	SNR: 0	SNR: -5	SNR: -5

Adapted from Van Engen and Bradlow 2006: Table 1 shows the counterbalanced design of Van Engen and Bradlow (2006) with respect to SNR as well as the organization of number of talkers in babble and background language.

The target stimuli for this study are simple, meaningful sentences recorded by a female native speaker of English that listeners must recall. *The children dropped the bag* and *Five men are working* are two of the target sentences used. Four lists of 16 sentences were played in the varying conditions. 50 keywords per list were chosen for a total of 200 words to be recalled by participants. The sentences produced for the background stimuli were semantically anomalous (e.g. *Your tedious beacon lifted our cab*). These were produced in English by 6 monolingual English speakers (3 M, 3 F) and in Mandarin by 6 native Mandarin Chinese speakers (3 M, 3 F). To create the babble, the researchers added a multiple of 100ms to the beginning of each speaker's sentences varying between 0-500ms, then added and cut at the point where all recordings had begun. The same procedure was used for 2-talker babble. Foreground speech and background babble were then mixed at the three SNRs for each group: -5 dB SNR, 0 dB SNR, +5 dB SNR. Participants were played the stimuli over headphones in a quiet room and were asked to write down what they heard after the presentation of each sentence. The listeners were encouraged to guess and completed all trials in one session.

The results of this study show that listeners recalling sentences heard in varying speech competitor situations perform better in some conditions than in others. The researchers converted the percent correct to a rationalized arcsine unit scale (RAU) where -23 RAU corresponds to 0% correct and 123 RAU corresponds with 100% correct. The results show that higher SNRs yielded better results overall than lower SNRs. In the +5 dB SNR condition (overall across talker and language conditions), participants scored an average of 109.75 RAU, while in the -5 dB SNR condition participants scored 30 RAU. This is to be expected because at higher SNRs, the foreground speech is at a higher volume and therefore, easier to attend to. In addition, performance was poorer overall in the six-talker condition (63.25 RAU mean correct across language and SNR conditions) compared to the 2-talker conditions (84.5 RAU mean correct across language and SNR) because the babble was highly disruptive and listeners could not rely on strategies that they could use in the two-talker babble condition. Van Engen and Bradlow (2006) suggest that in the two-talker babble condition, listeners are able to take advantage of the temporal gaps in the background signal. The researchers noted that the listeners reported awareness of background language changes from block to block in the 2-talker babble conditions (see Table 1), but not in the 6-talker babble; this suggests that the background speech in the 6-talker condition is very dense and no linguistic aspects can be deciphered. This is perhaps why the effect of language only emerged in 2-talker conditions.

Van Engen and Bradlow (2006) found language effects in what they call the “hard” 2-talker conditions where the SNR was 0 dB in Condition 3 (Mandarin background = 99 RAU; English background = 89 RAU) and -5 dB in Condition 4 (Mandarin background = 64 RAU; English = 37 RAU). In these conditions, recall performance was significantly better in the Mandarin background speech condition. This suggests that listeners are more readily able to

disregard babble when it is in a language that is unknown. It also suggests that, in easier conditions, listeners are able to take advantage of the easier aspects of the task and language does not become a significant distractor. When the task is difficult, as in the “hard” condition, listeners are distracted by language because they do not have other advantages like a high SNR or a low number of talkers.

This study shows that language effects do exist in sentence recall tasks where both the targets and the maskers are presented aurally and simultaneously. It also shows that performance is at its best when the SNR is high. This is because the targets are at a higher intensity to the background speech and thus the targets are easier to discern. In addition to SNR, the number of talkers had an effect. When 6 talkers are used, no language effect is shown and all participants do poorly in these conditions—likely demonstrating a floor effect. When 2 talkers are used, however, the task seems to become easier and the effect of language emerges where Mandarin background speech is less distracting to native English speakers than English background speech. What is not known from this study is if the effect can be generalized to other types of recall tasks and to different background speech conditions, like single-talker speech. It also does not address the issue of the differences between the target speaker and the background speaker. In the 2-talker babble conditions, both the target speaker and the maskers were female and in the 6-talker babble conditions the target speaker was female while the maskers were males and females. This mismatch between the two conditions introduces a potential variable that may affect performance.

The second study that is relevant to the current study is Rhebergen et al. (2005). This study attempts to separate the semantic information from the prosodic information of Dutch and Swedish, testing Dutch listeners. In order to tease these two variables apart, the researchers

played native (Dutch) and unknown (Swedish) speech forward and backward for native Dutch listeners. The researchers note that Dutch and Swedish vary widely in prosodic factors such as intonation, rhythm and mean pitch and that these factors may be responsible for distraction in speech perception and recall, and not the informational masking found in the forward native background speech (Rhebergen et al., 2005). The researchers required free recall of the target sentences, similar to Van Engen and Bradlow's (2006) methods. Instead of scoring keywords from each recalled sentence, Rhebergen et al. (2005) tested perceptual accuracy through the use of a speech reception threshold (SRT) test. The SRT test determines the minimally required SNR to ensure 100% accuracy in perception.

Rhebergen et al. (2005) tested the SRT for native speakers of Dutch listening to Dutch sentences embedded in Dutch and Swedish background speech as well as the same Dutch and Swedish speech reversed. The background speakers for this study were both female (one native speaker of Dutch, one of Swedish). The background stimuli were concatenated sentences edited to have no pauses. The target speaker was a male native speaker of Dutch. The target stimuli were short common sentences. The SRT test requires listeners to repeat the sentence that was heard back to the researcher. The researcher then scores how many words were correct from the sentence that was actually presented. If it was less than 100%, the SNR was systematically raised in 4 dB intervals after each presentation of the sentence. Once the 100% recall level was reached for the first target sentence, the following 12 sentences were presented only once. If the sentence was not recalled 100% correctly, the SNR was raised in steps of 2 dB and the next novel sentence was presented. Once listeners were able to recall all of the words in at least 50% of the 12 test sentences, the SRT in dB was reached and recorded. A low SRT indicates that listeners do not require a large difference in intensity between target and background speech.

Through the use of forward and reversed single talker background speech, the researchers intended to show that the prosodic and semantic features that are familiar and intelligible to native speakers of Dutch will result in a higher SRT. The SNR must be very high due to the similarities between target speech and masking speech, according to their hypothesis. If there is an effect for semantic aspects of language, but not prosodic elements, then we would expect a detrimental effect for Dutch forward speech but not reversed Dutch. In addition, if informational masking is the cause of the detriment, we would expect performance on reversed Dutch and reversed Swedish to be similar because the informational element has been removed.

As expected, Dutch played forward was most difficult for Dutch listeners. The Dutch listeners on average required an SNR of -11 dB in order to recall at least 50% of the sentences correctly. The listeners performed better with Swedish played forward because it required an SNR of -14 dB. These findings support Van Engen and Bradlow's (2006) conclusion that when the target and background speech are in the same intelligible language, it is much more difficult to separate the speech streams than when the background language is different from that of the language of the target speech. Rhebergen et al.'s (2005) use of different genders of talkers for target and maskers likely aided this process as well. It seems that there needs to be some familiarity in the background speech stream, but if there is too much that can be perceived, like entire words or phrases, then the task becomes much more difficult. This sensitivity is also reflected in Van Engen and Bradlow (2006) when they realize that language effects only emerge in the "hard" conditions.

Sperry et al. (1997) found similar results to Rhebergen et al. (2005) and Van Engen and Bradlow (2006). Their study required listeners to recall words heard in forward multitalker speech, reversed multitalker speech and amplitude-modulated speech-spectrum noise (SSN).

SSN is synthesized speech-like sound. The researchers generated this by analyzing the long-term average speech spectra of the multitalker babble they had generated and set a spectral range for the SSN. Then they created a pink-noise signal that falls within this spectral range and smoothed it according to the amplitude peaks from the multitalker speech analysis. The aim of this is to create a sound that reflects the prosodic fluctuations in the speech stream without retaining the informational masking. This intent is similar to what Rhebergen et al. (2005) did with their reversed speech, but provides a different view of perception because it does not introduce additional energetic masking found in reversed speech. This study takes the multitalker aspect of Van Engen and Bradlow (2006) and includes the reversed speech of Rhebergen et al. (2005) and includes the speech-like noise as a background competitor. Listeners were allowed to recall words freely and were required to give a written response. The researchers also varied SNR five ways: -8 dB, -4 dB, 0 dB, 4 dB and 8 dB. Like the prior studies, Sperry et al. (1997) found that higher SNRs made it easier for listeners to filter out the competing sounds and focus on the recall of the word lists. They also found that at the most difficult SNR (-8 dB), listeners recalled the most words in the SSN background condition and the least in the forward multitalker speech condition. As SNR increases, so does accuracy in all three background masking conditions—and the pattern between background conditions remains the same even at +8 dB SNR (Sperry et al., 1997).

Few studies of speech perception in competitor speech have considered possible listener gender effects in their results. Prior research shows that females and males perform differently in tasks of working memory (Kramer et al., 1988; Speck et al., 2000). Kramer et al. (1988) tested males and females on their immediate and delayed free recall ability as well as recognition of words, learning and error recognition. Presentation was auditory with an examiner reading the

to-be-recalled stimuli aloud. For the free recall tasks, the researchers found that females perform better on both recall tasks overall than males and attribute this to “better retrieval” and therefore better recall ability. This study was not performed with background speech or noise present. Speck et al. (2000) used fMRI to test females and males in 4 tasks of working memory and found that males and females differ in the functional organization of the brain and may differ with respect to problem solving. These differences suggest that females and males may perform differently on tasks of working memory. Since the current study uses a working memory task, serial recall, to test speech perception, it is possible that gender has an effect on this performance. Therefore, in each analysis gender will be included as an independent variable.

The previous work on background competitors allows for some generalizations. First, if listeners are able to interpret meaning of words from the background masker, it is more difficult for them to ignore the masker and focus on the target (Rhebergen et al., 2005; Sperry et al., 1997; Van Engen & Bradlow, 2006). Second, if the background competitor is prosodically different from the native language of the listener, greater interference may occur; this extends to reversed speech as well. Despite these generalizations, there are still some unresolved issues. These issues will be the target of the present study.

The present study will focus on the following questions:

- 1) Is there more or less interference with serial recall when the foreground and background talkers share the same language as compared to when they speak different languages?
- 2) Does the physical difference between foreground and background talker affect serial recall?

The above questions will be answered through a series of experiments. The first three

experiments will take baseline measures to ensure that there are no extraneous factors at work other than background language, while Experiments 4 and 5 will test the effect of speaker differences and language effects in serial recall of words heard in two background speech conditions: one in which the background and target language is shared and the other with a different background language from the targets.

The present study will test listeners' ability to serially recall English target words mixed with background speech spoken in either English or Korean by one bilingual speaker. Serial recall will be implemented as a task because it is simple to explain to participants and very simple to score, but it also requires listeners to engage working memory. The task is simple to understand, but difficult to do because of the stress it places on working memory along with the occupation of the phonological loop through the simultaneous presentation of background speech and target speech. The results of this task performed when background speech is present will reveal if listeners respond more or less accurately in two background language conditions: English or Korean. English is the native language of the listeners while Korean is unknown to the listeners.

By using one bilingual speaker, we control for speaker difference effects and ensure that any effects that are seen are indeed from the background language. Using some methods from Van Engen and Bradlow (2006), two word lists and one practice list containing 9 words each are played at a +5 dB sound-to-noise ratio (SNR). Words will be scored as correct if they are the correct words and if they are in the correct serial order.

Previous studies indicate that in babble and reverse-speech conditions, the native language is most difficult for listeners. We can feasibly expect a similar result in a single-talker environment. It may be even more difficult for listeners because the background speech is easily

comprehensible and is processed by listeners at a semantic level (Beaman et al. 2005, Van Engen and Bradlow 2006). On the other hand, the single-talker speech may yield more accurate results in the recall task because listeners can take advantage of temporal gaps in the background speech stream (Van Engen and Bradlow 2006).

II. Experiment 1: Presentation Rate and Sound-to-Noise Ratio

In order to ensure that the experimental design does not yield a floor or ceiling effect, a pilot study was implemented to ensure that both presentation rate and SNR were optimal. These baselines must be established to ensure that the SNR is high enough for listeners to successfully separate the background speech from the target speech. This is especially crucial for this design because listeners cannot rely on physical differences between background talker and target talker since the present study uses one bilingual talker for both target and masker speech.

A. Participants

Ten native English speakers participated in this study, five females and five males. None reported hearing loss. They were divided into four independent groups as shown in Table 2:

Background Speech		SNR	ISI
Condition 1	English (N=2)	+5 dB	500 ms
	Korean (N=3)		
Condition 2	English (N=2)	+5 dB	750 ms
	Korean (N=3)		

Table 2: Grouping design for participants in Experiment 1

One female bilingual speaker of Korean and English was recorded in an anechoic chamber with an Electrovoice RE 20 microphone and a Marantz PMD 671 solid state recorder at

a sampling rate of 22.05 kHz. This speaker was 19 years old and had grown up in the Midwest and spoke Korean and English in the home and was taking a university level course in Korean as a heritage learner. The speaker exhibited literacy in both languages. All participants were required to sign participant consent forms and complete a language background questionnaire.

B. Stimuli

Selection of the stimuli was guided by previous research on phonological similarity and frequency effects in serial and free recall tasks. These two factors were controlled in order to prevent recall effects due to frequency and phonological variables instead of language variables. Poirier and Saint-Aubin (1996) conducted a study of French native-speakers' serial recall performance on lists containing high (occurrence > 100 per million), medium (between 14 and 25 per million) and low frequency words (1 per million). In addition to frequency, the researchers varied phonological similarity for each frequency level; there was no background speech or noise presented in this study and presentation was visual. The researchers found that participants were most accurate when words were highly frequent and phonologically dissimilar (Poirier and Saint-Aubin, 1996). Due to these results, the chosen words are highly frequent to control for any frequency effects that would otherwise present themselves.

1. Target Words

Twenty-seven monosyllabic frequent words were selected from the Brown Corpus (Francis & Kucera, 1964). For complete word lists, see Appendix I. The practice list, had a mean frequency of 227.89 per million words. The mean frequency of the tested words is 217.89 occurrences per million words. Word list one had a mean frequency of 245.11 per million words

and word list two had a mean frequency of 190.67 per million. A two-tailed independent samples t-test reveals that the difference between List A and List B is not significant ($p=0.151$). These words were also controlled for phonetic structure as well as phonetic diversity per list (Poirier & Saint-Aubin, 1996). Words were all CVCC, CCVC or CVC words. The words were assigned to lists such that there were no rhyming words, minimal pairs, or semantically similar words on either list.

Using Praat, each word was extracted from the list file and set to an average peak intensity of 60 dB, following the method of Van Engen and Bradlow (2006). Since the words that will be combined in lists may have come from different recording files, each word must be set to the same average peak intensity when the words are then combined into lists. After all of the words were individually selected and edited, they were then concatenated with a fixed interval (500 ms for Condition 1 and 750 ms for Condition 2). Each list was then set to an average peak intensity of 60 dB for Condition 1 and 62dB for Condition 2 to create a +5 dB and +7 dB SNR, respectively, to the 55 dB background speech.

2. Background Speech

Background recording passages were selected from two bilingual Korean-English books (Fulton & Fulton, 1997; Voorhees & Mueller, 1990). The passages were Korean folk tales in both Korean and English. Proper names were changed to pronouns or common nouns to maintain neutrality of the passage. In addition, passages containing dialog were excluded to control for unexpected intonation fluctuations by the speaker. The speaker read each passage twice in Korean and twice in English.

Each passage was set to an average peak intensity of 55 dB for Condition 1 and 53 dB for

Condition 2. When combined with the word lists that are set at 60 dB, this will yield a SNR of +5 dB for Condition 1 and +7 dB for Condition 2. Pauses in the background speech signal longer than 500ms were shortened to 70ms, based on the natural pauses of the speaker, to prevent listeners from taking advantage of these gaps in the speech signal and to provide a continuous steady presentation of background speech. Three unique passages were used, but six total recordings were used in the study because each passage was read in English and in Korean by the bilingual speaker so that each condition would have the same passages combined with the same word lists across background language conditions. For example, one unique passage was used as background speech for the practice word list, but the language of the passage differed depending on if the condition was Korean or English. The duration of each word list was noted, and a selection from the passages was taken to match this duration using Praat. The total duration of Word List A is 11.84s and Word List B is 11.85s.

3. Mixing

The word lists were mixed with their background speech signal. A 1-second pause was inserted at the beginning of each list and 500 ms inserted at the end of the list to allow for a controlled lead-in of the background speech before presentation of the word lists. For both conditions, the word lists and the background passages were exactly the same. The only differing element was the SNR and presentation speed. Condition 1 has a +5 dB SNR and 500ms between words, while Condition 2 has a +7 dB SNR and 750 ms between words.

C. Procedure

Participants were tested in a quiet room on PC computers. As many as four participants were tested at the same time. All participants listened to the stimuli over headphones. The experiment was setup using Paradigm software on a PC computer. Participants signed consent forms before beginning.

The listeners were told to listen to the word lists, ignoring the background speech. They were ensured that they would not be tested on the content of the background speech in any way. Participants were then asked to write their responses on response sheets in the correct serial order. Guessing was encouraged. Participants began the trial after reading instructions on the computer screen by pressing the spacebar. After a 3 second interval, the presentation of the practice list began. When prompted on the screen, participants began to write their responses. This prompt appeared immediately after each word list ceased. The following two test lists were started by the spacebar by the participants by an onscreen prompt at their own pace. When participants finished all word lists, they were asked to complete a language questionnaire detailing their gender, age and proficiency in other languages.

D. Results

The two conditions in this experiment tested the SNR and presentation speed of the word lists to ensure that floor effects were not occurring due to task difficulty. Table 3 suggests that participants did better in Condition 1, where the SNR was +5 dB and the word list was presented with a 500ms pause between words.

	Language Background	Mean Words Correct
Condition 1		
N=3	Korean	6.67
N=2	English	5.50
Condition 2		
N=3	Korean	5.00
N=2	English	5.50

Table 3: The overall mean words correct (out of 18) for each condition as well as for each background speech condition. Condition 1 is +5 dB SNR and 500 ms ISI. Condition 2 is +7 dB SNR and 750 ms ISI.

In Table 3, the Condition 1 results are numerically higher than Condition 2, but a one-way ANOVA for differences between groups on overall recall performance reveals that these are not significantly different [$F(1,9)= 1.136$ $p=0.318$]. Despite this nonsignificance, the final design of the study is that of Condition 1 with a SNR of +5 dB and a pause of 500ms between words by virtue of the higher numeric mean of these very preliminary results. It is likely that the most important factor in determining performance in this study is the presentation speed. It is not likely that an increased SNR would lead to a more deficient performance. As Sperry et al. (1997) and Van Engen and Bradlow (2006) observed, the higher the SNR, the better the performance regardless of background language. The results of the preliminary measures seem to contradict this, but it must be considered that the presentation of the word list was slower (750 ms pause between words) for the +7 dB SNR condition, and this may be why participants were less accurate in this condition. They may forget words due to the longer pause between words and a longer word list overall than Condition 1. Items in working memory may begin to decay if the listener is unable to rehearse; since listeners in this study are unable to rehearse due to the interference from the background speech, a shorter list duration is expected to yield higher recall (Baddeley, 2003).

One-way ANOVAs were run for each condition for language effects. In Condition 1 where the SNR was +5 dB and ISI 500 ms, the test revealed that while participants performed better in the Korean background speech condition, this difference was not significant [$F(1,4)=0.535$ $p=0.518$]. For Condition 2, the result of the ANOVA is [$F(1,4)=0.138$ $p=0.735$]. Neither condition shows significant language effects, but it does seem that Condition 1 yields more expected results when prior research is considered. Participants do better in the Korean condition and English seems to be more distracting. This, coupled with the overall higher mean performance of Condition 1 suggests that this condition is more likely to yield an accurate representation of performance on serial recall tasks in background speech conditions.

III: Experiment 2: Independent Ratings of the Bilingual Speaker

In addition to ensuring the presentation rate and SNR are at optimal levels for speech perception, the speaker used in the study received independent ratings to ensure that she did not speak accented English due to bilingualism. This control lowers the possibility that any language effects observed are due to accentedness of the speaker and ensures that any effects observed are indeed due to the change in background language.

A. Participants

1. Speakers

The speaker was the same Korean/English bilingual female speaker from Experiment 1. Another female bilingual speaker of Tuvan/English was used as a filler to mask the purpose of the study. The recordings for the Tuvan/English bilingual speaker were from an earlier pilot study and will be described in detail in the stimuli section below.

2. Listeners

Five monolingual native English speakers, 3 M and 2 F, evaluated the intelligibility and accentedness of the two bilingual speakers. These five participants did not participate in any other experiments.

B. Stimuli

The stimuli were all recorded in an anechoic chamber with an Electrovoice RE 20 microphone and a Marantz PMD 671 solid state recorder. The two speakers each read a unique passage. The passage read by the Korean/English bilingual speaker was an English passage from Experiment 1, but was not used for that study. The Tuvan/English bilingual speaker read a passage in English from *Through the looking-glass and what Alice found there* (Carroll, 1872). From each of these passages, 12 seconds from the onset of the passage were selected for presentation.

C. Procedure

This experiment was set up using the Paradigm software package on a PC. Listeners read instructions on a computer screen and began presentation of the first passage by pressing the spacebar. Then they were again prompted on-screen to press the spacebar when they were ready to listen to the second passage. Each listener heard one passage from the Tuvan/English bilingual and one passage from the Korean/English bilingual. Again, the two passages were from different books.

Listeners first heard the Tuvan/English bilingual and were asked to answer three questions:

- 1) Where do you think this speaker is from?
- 2) How understandable was this speaker?
- 3) Explain your answer for number (2)

Participants would then proceed to the Korean/English speaker's passage and answer the same questions.

The choice of words for Question 1 is important because it does not imply that the speaker may speak a language other than English. Participants could answer a state, country, continent, or any other geographical location. This provides information about which language the listener believes is the speaker's native language. Question 2 was followed by a 1-5 scale with 1 being "very difficult to understand" and 5 being "very easy to understand."

D. Results

All listeners rated the target speaker (Korean/English bilingual) but not the Tuvan/English speaker as being from locations within the United States. The listeners are aware of the accentedness of the speaker when the ratings and locations are considered. Responses to question 1 and question 2 are shown in Table 4:

Participant # and (Gender)	1(M)	2(F)	3(M)	4(M)	5(F)	Mean
Korean/English Bilingual						
Where do you think this speaker is from?	Colorado	USA-Midwest	Midwest (US)	USA	Virginia	
How understandable was the speaker's speech?	5	5	5	5	5	5
Tuvan/English Bilingual						
Where do you think this speaker is from?	Utah	China	N/A*	England	Mexico	
How understandable was the speaker's speech?	2	4	3	5	5	3.8

**Participant believed speech was generated by computer*

Table 4: The accentedness ratings of two bilingual speakers by native English speakers and their intuition about where the speakers are from.

These results tell us that the Korean/English bilingual speaker was perceived as being from the United States, was highly intelligible and easy to understand as agreed upon by five native English speakers. A one-way ANOVA testing the answer to Question 2 between speakers shows that the difference is not significant [$F(1,4)=4.235$ $p=0.074$], but is trending toward significance with the Korean/English bilingual being significantly more “understandable” with an average rating of 5. Table 4 also suggests that the listeners were sensitive to accentedness and rated the Tuvan/English bilingual speaker as not from the United States and much less intelligible. Therefore, Experiments 4 and 5 will use the target and background speech from the Korean/English bilingual.

IV: Experiment 3: Individual Word Intelligibility Test

This experiment explored if the individual target words are intelligible. It is important to establish that the recordings are comprehensible both on their own and when background speech is present. This test will ask listeners to listen to each word one-by-one and write down what they hear. By doing this, the listeners do not have to overload the working memory by trying to remember multiple words in order. They simply listen to one word, write it down and press the spacebar to move on to the next word. This procedure will be carried out with three types of background stimuli. The first is “in the clear.” Words are heard with no background speech; this will determine if the words without background speech are indeed intelligible when no other sound is present. The next two conditions are Korean background and English background. The Korean background will add Korean speech at a -5dB SNR to the English word list. The English background condition will add English speech at a -5 dB SNR to the English word list. The results of this experiment will ultimately decide the choice of words of the lists for Experiment 4 and Experiment 5.

A. Participants

1. Speaker

The female Korean/English bilingual speaker from Experiments 1 and 2 was used.

2. Listeners

Seven native English speakers (6 M; 1 F) were assigned to three background conditions for the word-by-word trial. These conditions were: Clear--no background speech is used (N=2); Korean background speech (N=3); and English background speech (N=2). The final participant tested in the Korean background condition was excluded so each group in the final analysis contained two

participants; the excluded participant was female. Each listener participated in only one of these conditions.

B. Stimuli

The target words were the same as those in Experiment 1. The list contained a total of 27 words. Nine words were from the practice test, nine from the first word list and nine from the second word list (see Appendix 1 for complete lists). For Korean and English background word-by-word conditions, the words were excised from their respective background speech conditions from Experiment 1 so that they contained the same stretch of background speech as what would be tested in the final study. Each word was played individually.

C. Procedure

The experiment was also set up using Paradigm software for PC. Three conditions were used. One condition was the words without any background speech, called “Clear” (N=2). The other two conditions contained background speech in Korean (N=2) and English (N=2). Listeners had a response sheet with 27 blanks. When they were ready to hear the first word, they pressed the spacebar on the keyboard to advance to each trial; the task was self-paced. Each individual word out of the 27 was presented in this way. The words were then scored as correct only if they were heard as intended. Obvious orthographical errors and homophones were counted as correct.

D. Results

Words were scored correct if they were the word intended by the speaker. Table 5 shows

the word-by-word errors across all conditions. Seven out of 27 words were misheard across all conditions, and many of those were agreed upon by more than one listener. This suggests that these words are not ideal for the final experiments. The small number of errors further supports our evidence from Experiment 2 that the speaker is indeed easy to understand and not accented.

Word	Background Speech			Errors by Word
	None	Korean	English	
talk	1	0	0	1
heart	0	2	0	2
month	0	0	0	0
trade	0	0	0	0
power	0	0	1	1
food	0	0	0	0
hand	0	0	0	0
cold	0	0	0	0
trial	0	0	0	0
club	2	0	0	2
fall	0	0	0	0
night	0	2	0	2
car	0	0	0	0
law	0	1	0	1
girl	0	0	0	0
fire	0	0	0	0
line	0	0	0	0
step	0	0	0	0
plan	0	1	0	1
space	0	0	0	0
book	0	0	0	0
name	0	0	0	0
door	0	0	0	0
land	0	0	0	0
call	0	0	0	0
top	0	0	0	0
wife	0	0	0	0
Totals by Condition	3	6	1	10

Table 5: Number of perception errors by condition and by word. “English” means that the participants heard each word mixed -5 dB SNR with English background speech. “Korean” is the word list mixed with -5 SNR Korean background speech. “None” indicates the words were heard “in the clear,” without background speech.

Table 5 provides a view of what listeners were hearing in three background speech conditions. A one-way ANOVA shows that these values are not significantly different from each other [$F(2,4)=1.28$ $p=0.283$]. Listeners correctly recalled the most words in the English background speech condition, with 1 misperceived token, and the least in the Korean background condition with 6 misperceived tokens total. Finally, the Clear condition results show that in the absence of background speech, some tokens are difficult to perceive. These results show that some of the words are difficult to perceive correctly in a variety of background speech situations.

E. Discussion

The results of Experiment 3 led to a modification of the word lists. All tokens that were misperceived were moved to the practice list so that no perceptually ambiguous words were presented in the test lists. These results work together to help ensure that the results found in Experiments 4 and 5 are not due to any of these variables, but to the manipulation of the independent variables being tested. Given these results coupled with those of Experiment 1 and Experiment 2, decisions about the final study may be made.

Experiment 1 showed that more accurate results are found in the condition with +5 dB SNR and a 500 ms interstimulus interval. Although the difference between conditions was not significant, it is likely that the shorter interstimulus interval is influencing the result because the interval between the start of the word list presentation and the onscreen recall cue is shorter. This means that participants are able to recall more quickly and the words recalled are less likely to deteriorate in working memory. Van Engen and Bradlow (2006) found that at higher SNRs, overall performance was better, but effects of language in these conditions did not emerge, likely due to a ceiling effect. Therefore, for Experiments 4 and 5 the SNR is +5 dB and the

interstimulus interval is 500 ms.

Experiment 2 showed that the bilingual speaker was perceived as a highly intelligible speaker from the United States. The purpose of this study was to ensure that any effects found in Experiments 4 and 5 are not due to difficulty understanding the speaker or due to accentedness. Therefore, this speaker is used as the target and background speaker in Experiment 4 and the background speaker in Experiment 5.

V. Experiment 4: The Effect of Background Language on Serial Recall

Experiment 4 will test listeners' ability to perform a serial recall task when presented with background speech and target words in the auditory domain. Target words spoken in English will help to determine if background languages have an effect on perception and working memory capacity. Experiment 1 found that listeners on average had more correct words in the Korean background speech condition than in the English background speech condition. However, when the words were presented in isolation with the serial recall task removed (Experiment 3), we saw that listeners hearing English in the background were the most accurate. Experiment 4 will determine if this is a result of perceptual variability on the part of the listeners or if words are easier to recall in the English as compared to the Korean background speech condition.

It is also important to note that the same Korean/English female bilingual speaker is used for the English and Korean productions in order to control for potential effects of speaker differences. Previous studies that used talkers that varied from background speech to target speech may show different results because the difference in acoustic attributes between talker and masker may aid listeners in perceiving the target speech and disregarding the background speech. Experiment 4 uses the same speaker for both the target speech and the background

speech in order to control for any effects of speaker differences that may occur.

A. Participants

1. Speaker

The speaker is the same female Korean/English bilingual used in the previous experiments.

2. Listeners

Forty native English speakers were recruited from an introductory linguistics course at the University of Kansas. They were given course credit in exchange for their participation. Speakers either heard English or Korean as the background language. In each language condition, there are two groups because the two word lists in the task are counterbalanced to avoid a speaker familiarity effect (Van Engen & Bradlow, 2006). The participants in the four groups were balanced for gender and age.

B. Stimuli

The stimuli were the same recordings of the Korean/English bilingual from Experiment 1. The only differences for the current experiment are that the SNR is fixed at +5 dB for all groups and the words are presented with a fixed interstimulus interval of 500ms between words. The background passages and words are the same as in Experiment 1.

C. Procedure

The procedures are the same as those from Experiment 1. The experiment was set up using the Paradigm software for PC. Participants read instructions on the screen, and then progressed from list to list by pressing the spacebar at their own pace. For each list, participants would write their responses on a unique response sheet numbered from 1 to 9 when they received

the on-screen signal to begin writing. Guessing was encouraged and emphasis was placed on getting the correct serial order. The task and paperwork completion took approximately 15 minutes.

The first and second word lists were counterbalanced within each condition to avoid a speaker familiarity effect (Van Engen and Bradlow 2006) or a task familiarity effect. It is possible that the longer listeners hear the voice of the speaker, the more they will adapt to the speech and become better at perceiving words as the lists progress. Each background condition contains counterbalanced word lists in order to observe this potential assimilation. By counterbalancing the lists, the presentation order can be examined to determine if that was a contributing factor in performance. See Table 6 for presentation for each condition.

Background Condition	Practice	1 st Test List	2 nd Test List
English	Practice List	Word List A	Word List B
English	Practice List	Word List B	Word List A
Korean	Practice List	Word List A	Word List B
Korean	Practice List	Word List B	Word List B

Table 6: Presentations of the lists by background language group. The lists were counterbalanced to counteract any familiarity or learning effects that may have occurred during presentation.

D. Results

Responses were scored two ways. First, they were scored strictly so that the correct word had to be in the correct serial location to be counted as “correct.” Homophones of any of the words (like “night” and “knight”) were scored as correct, but words that were not phonetically the same were counted as incorrect, e.g. “plant” for the correct word “plan” would be counted as incorrect. Obvious orthographic errors of the correct word were scored as correct, e.g. “fal” for “fall” or “powr” for “power” (both of which were observed). Next, the responses were scored

without regard to serial order. In other words, all correct responses that were on the respective lists were counted as correct, even if they were not in the correct serial order. Repetitions of the same word on one list were not counted as correct, nor were morphological endings allowed in either of these scorings. If any of these occurred, the words were scored as incorrect.

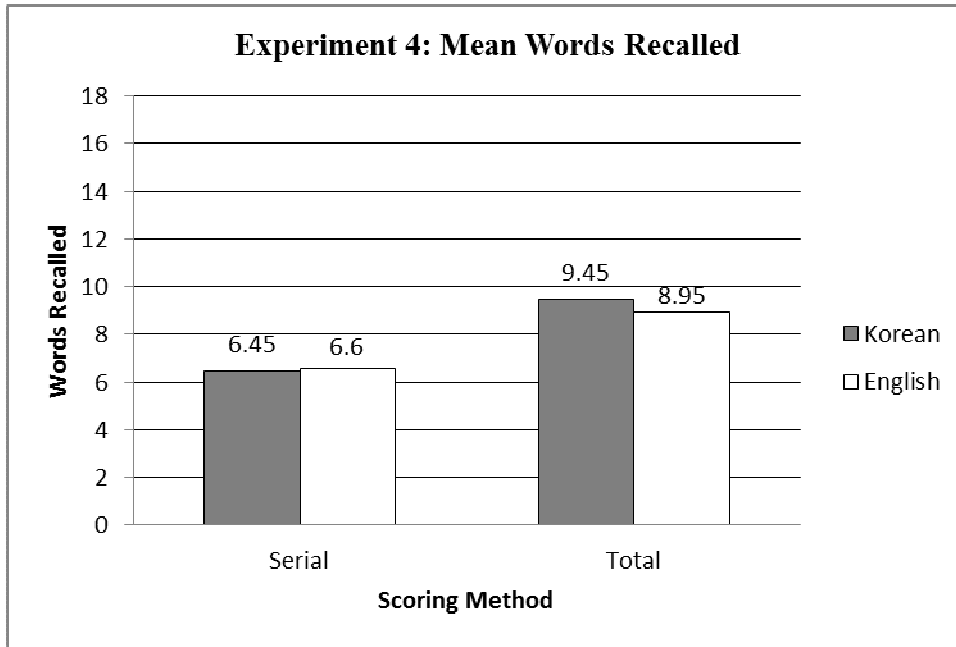


Figure 1: Total number of words correctly identified by participants per condition. Total Words (Total) indicates the mean number of correct words. Serial Order (Serial) indicates the mean number of words written in the correct serial order. Mean values are listed. (English N=20, Korean N=20).

Figure 1 shows the mean scores for English and Korean background speech conditions. When looking between language conditions within the same scoring method, it is apparent that there is little difference between the background speech conditions.

A series of two-way ANOVAs with Language and Gender as independent factors compare the mean scores between groups in several ways. No significant effects for Language emerged for the serial recall condition [$F(1,39)=.008$ $p=0.927$] (Korean $\bar{x}=6.45$; English $\bar{x}=6.6$), nor the total recall condition [$F(1,39)=0.326$ $p=0.572$] (Korean $\bar{x}=9.45$; English $\bar{x}=8.95$). No

significant effect for Gender emerged in serial recall [$F(1,39)=0.041$ $p=0.841$] (Males $\bar{x}=6.64$; Females $\bar{x}=6.46$) or total words correct [$F(1,39)=0.042$ $p=0.838$] (Males $\bar{x}=9.07$; Females $\bar{x}=9.27$). No significant interactions emerged.

Serial Recall Scoring					Total Recall Scoring				
		Language					Language		
		Korean	English	Total			Korean	English	Total
List A	Females	4.08	3.31	3.69	List A	Females	5.62	4.54	5.08
	Males	4.43	3.57	4.00		Males	5.57	4.57	5.07
	Total	4.20	3.40	3.80		Total	5.60	4.55	5.07
List B	Females	2.23	3.31	2.77	List B	Females	3.85	4.54	4.19
	Males	2.29	3.00	2.64		Males	3.86	4.14	4.00
	Total	2.25	3.20	2.72		Total	3.85	4.40	4.13

Table 7: The means of males and females for each language condition and the totals overall by list and by scoring. Rows that are boxed show a near significant effect that is discussed below.

Next, a two-way ANOVA for each list for both serial order and free order scorings was conducted. Means for all conditions are shown in Table 7 above. Language and Gender are the independent variables. This analysis compares the means of each word list disregarding the block position of the lists. Because the lists are counterbalanced to observe potential learning or familiarity effects, List A and List B appear in both first and second block positions. Some participants heard List A first and List B second, while others heard List B first in the block and List A second. This analysis only considers performance on List A or List B regardless of the position in which it was presented. For serial recall of Word List A, no main effects or interactions are significant (see Appendix II). In serial recall of Word List B, Language trends toward significance [$F(1,39)=3.727$ $p=0.061$] with English background speech being slightly less distracting than Korean speech (Korean $\bar{x}=2.25$, English $\bar{x}=3.20$). Analyses for each list were also conducted for the total recall scoring which disregards serial order and counts the total

number of correct words on each list. Total recall on Word List A neared significance for Language [$F(1,39)=3.615$ $p=0.065$] with Korean $\bar{x}=5.60$ and English $\bar{x}=4.55$. No main effects or interactions for total words on Word List B were observed.

Van Engen and Bradlow (2006) found assimilation effects where the listeners performed significantly better on the second list presented. To test if speaker or task familiarity occurred, a series of ANOVAs were run to determine if there was an effect for presentation order of the lists and if there was more accurate performance in the first block or second block.

A series of two-way ANOVAs were conducted with Gender and Language as between-subjects variables. This analysis examined serial recall performance on the list presented first (Word List A or B depending on the block) yielded no significant effects for Language [$F(1,39)=0.023$ $p=0.881$]. Likewise, serial recall performance on the list presented second was not significant for Language [$F(1,39)=0.082$ $p=0.776$]. For total recall of the words that were heard first, the result was not significant [$F(1,39)=1.436$ $p=0.239$]; total recall of the words heard second was also not significant [$F(1,39)=0.009$ $p=0.926$]. Gender was not a significant factor in any of these analyses (for a table of all means from this analysis, see Appendix II).

E. Discussion

The results of this study show no significant effects overall for background language, but some trends toward significance for serial recall of Word List B and total word recall of Word List A. It is possible that this nonsignificant overall result is due to the similarity between the target speaker and the masker. They are, in fact, the same speaker. This was a very specific choice in order to control for speaker difference effects and to ensure that any differences observed are indeed due to the change in background language and not because of some talkers

being more intelligible than others. While it is true that the current study sees near-significant differences in means in the expected direction--Korean background condition participants performed better than English background condition participants for total recall of List A , but this difference is not statistically significant. The results do not show that background language has an overwhelming effect on perception of words and serial recall task performance.

Although some lists show trends for significance in language we do not see any significant effects or interactions emerge for Experiment 4. One reason for the lack of significance overall may be the fact that the target speaker and masker are one and the same.

Rhebergen et al. (2005) suggest:

“...Intelligibility appears to vary greatly between conditions, which in part might be due to the degree of similarity between the target speaker and the interfering speakers. The more similar the target and masker are, the more the listener is confused or distracted, which in turn results in poorer performance.”

Rhebergen et al. (2005) did not specifically test for this in their study, nor did Van Engen and Bradlow (2006), but based on the difference between their 2-talker and 6-talker conditions, Van Engen and Bradlow conclude that:

“...greater similarity between target and masker in the linguistic domain creates great interference in target intelligibility and must be taken into consideration in a principled account of speech perception in noise.”

This suggests that perhaps using the same talker for target and masking speech yields very poor results because the listeners have difficulty choosing which signal to ignore. It seems that a controlled word list played over a fluidly spoken passage would be fairly simple to distinguish, but this may not be the case when the speaker is the same for target and masker. While Rhebergen et al. (2005) and Van Engen and Bradlow (2006) suggest that this similarity may yield poorer performance, neither tested it. In order to test if this is the reason for the current lack of a language background effect, a final experiment was performed.

VI. Experiment 5: Differing language and talker

The results of Experiment 4 were different from what previous researchers found. Because Van Engen and Bradlow (2006) and Rhebergen (2005) both suggest that greater similarity between talker and masker may result in more difficulty in separating the target speech from the masking speech in multi-talker perception, a different speaker was added to determine if language effects emerge when target speaker and masker are physically different and to determine if the results of Experiment 4 were caused by a floor effect.

A. Participants

1. Speakers

The same female Korean/English bilingual speaker was used for the background speech. A novel female monolingual English speaker recorded the same word lists from Experiment 4.

2. Listeners

Thirty-nine native English speakers were recruited from an introductory linguistics course at the University of Kansas. These participants did not participate in Experiment 4 or any of the preliminary experiments. They were granted course credit in exchange for their participation.

B. Stimuli

The new female speaker was recorded in an anechoic chamber with an Electrovoice RE 20 microphone and a Marantz PMD 671 solid state recorder. The words were then mixed with background speech from the Korean/English bilingual female speaker from the first study. The

word lists were exactly the same as in the first experiment and the SNR and scale peak intensity were manipulated in the same way. The SNR was +5 dB; the scale peak intensity for the background speech was 55 dB while the word lists were set to 60 dB. The background speech passages were the same for each word list as in the previous study as well. The only difference in stimuli is the different speaker for the word lists. All else is identical to Experiment 4.

C. Procedure

Testing procedures are the same as those in Experiment 4.

D. Results

Results were scored the same way as in Experiment 4. In Figure 2, Serial Order correct indicates that the participant placed the correct word in the correct serial position on the response form. Total Words (Total) indicates the total number of words correct per list, disregarding serial position. Figure 2 shows the means for this data.

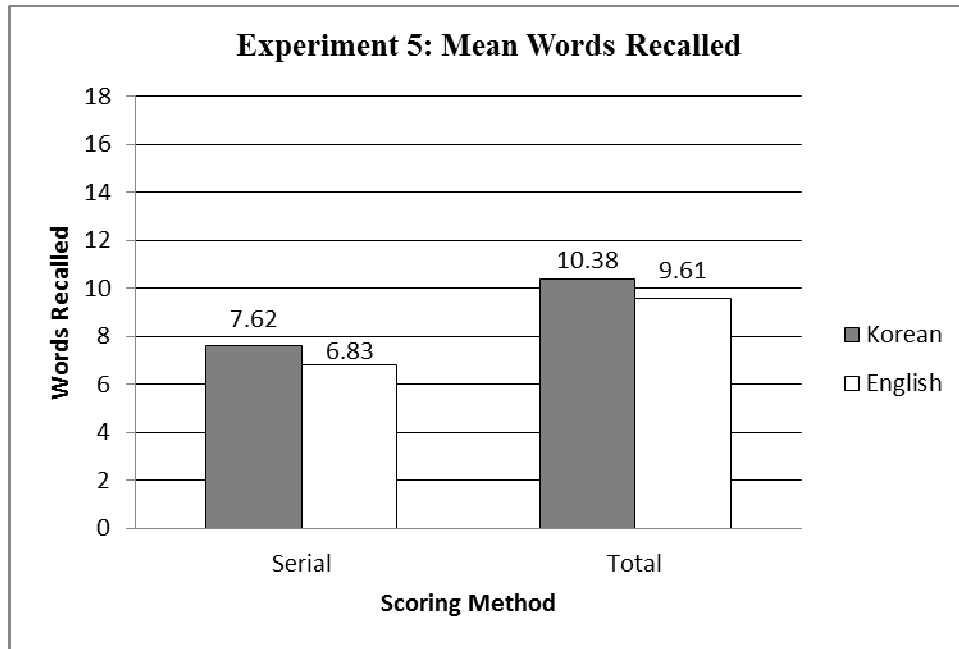


Figure 2: Mean scores for Korean and English. This experiment used a different word list speaker than in Figure 1. Total Words Correct (Total) is how many words are correct in all disregarding serial order. Serial Order Correct (Serial) indicates the number of words correct out of 18 that were the correct word in the correct serial order.

For serial order overall, a two-way ANOVA with Language and Gender as the independent variables showed no significant effect for Language [$F(1,38)=1.310$ $p=0.260$]. For total words recalled, there was also no effect for Language [$F=(1,38)=2.710$ $p=0.109$]. No effect of Gender emerged overall for serial recall [$F(1,38)=0.065$ $p=0.800$] or total recall [$F(1,38)=0.959$ $p=0.334$].

As in Experiment 4, an analysis of each word list was conducted by serial order and by total words recalled.

		Serial Recall Scoring					Total Recall Scoring		
		Language					Language		
		Korean	English	Total			Korean	English	Total
List A	Females	3.92	4.00	3.96	List A	Females	5.31	5.33	5.32
	Males	3.75	3.17	3.50		Males	5.50	4.50	5.07
	Total	3.86	3.72	3.79		Total	5.38	5.06	5.23
List B	Females	3.23	3.08	3.16	List B	Females	5.00	4.75	4.88
	Males	4.63	3.17	4.00		Males	5.00	4.17	4.64
	Total	3.76	3.11	3.46		Total	5.00	4.56	4.79

Table 8: Mean recall performance of males and females overall in Korean and English by list.

For serial recall of Word List A, there was no significant effect for Language [$F(1,38)=0.239$ $p=0.628$] nor in total recall of Word List A [$F(1,38)=1.636$ $p=0.209$]. Serial recall of Word List B was also nonsignificant for Language [$F(1,38)=1.850$ $p=0.182$] and total recall of Word List B is also nonsignificant for Language [$F(1,38)=1.488$ $p=0.231$]. For Gender performance on List A, results of serial recall [$F(1,38)=0.942$ $p=0.338$] and total words [$F(1,38)=0.708$ $p=0.406$] revealed no significant effects. In Gender performance on List B, serial recall [$F(1,38)=1.567$ $p=0.219$] and total words recalled also revealed no significant effect [$F(1,38)=0.431$ $p=0.516$]. No interactions between Language and Gender were significant.

The block position of words presented was analyzed in the same way as Experiment 4, with a series of two-way ANOVAS with Gender and Language as between-subjects variables. Serial recall on the list heard first (either List A or List B) showed no significant effect for Language or Gender (see Appendix II for table of means). Serial recall of the list heard second also showed no significant effects or interactions for Language [$F(1,38)=2.378$ $p=0.132$] or Gender. Total recall of the list heard first showed no significant effect for Language [$F(1,38)=2.396$ $p=0.131$] but there is a Gender * Language interaction that nears significance [$F(1,38)=3.876$ $p=0.057$]. Males in the Korean background condition ($\bar{x}=5.50$) did better than females ($\bar{x}=5.15$). In contrast, males in the English background condition ($\bar{x}=4.00$) did worse

than females ($\bar{x}=5.33$).

E. Discussion

As in Experiment 4, Experiment 5 showed no overall effects for language. However, the results show a marginally significant interaction between Gender and Language for total recall of the words heard first. No other significant effects by list were observed. Experiment 4 did provide near-significant effects for language by some lists, but this was not found to be true in Experiment 5. The significant result for the words heard first could be due to the observation that by the second and final block, listeners were very taxed from the previous blocks and were therefore much less accurate by the time they reached the final block. It was noted that some listeners recalled words from the previous blocks (Practice or Block 1) on their Block 2 response sheets.

No previous work has used the same speaker for the talker and masker and previous studies always seemed to find an effect for language. These studies (Rhebergen et al., 2005; Sperry et al., 1997; Van Engen & Bradlow, 2006) do not account for speaker differences, and according to the findings of the present study, the physical difference between target and masking speaker does not seem to be an overall contributor to language effects in serial recall. Listeners in Experiment 5 are numerically more accurate than in Experiment 4 for both serial and total recall and for both English and Korean as background speech competitors. This could suggest that listeners in Experiment 5 were better able to separate the target speech from the background speech and therefore better able to attend to the targets and disregard the background speech.

VII. Experiment 4 and 5 Combined Results

A series of two-way ANOVAs with Language and Experiment as independent variables was conducted to determine if listeners in Experiment 4 were significantly more or less accurate than those listeners in Experiment 5. This analysis allows for a comparison between those who heard the same talker for target and masker and those who heard a target speaker that was different than the masking speaker. This will further support that having different speakers for the background and target speech may not be a significant aid in separating the target and background speech signals. Moreover, the data from Experiments 4 and 5 were run together through a series of ANOVAs to determine if significant effects emerge overall for Language or Gender. An analysis of each individual list for serial and total recall was conducted as well since statistical trends were observed when individual lists and scorings were considered in each experiment.

A. Results

	Serial Recall Scoring				Total Recall Scoring			
	Experiment 4		Experiment 5		Experiment 4		Experiment 5	
	Korean	English	Korean	English	Korean	English	Korean	English
Males	6.71	6.57	8.38	6.33	9.43	8.71	10.5	8.67
Females	6.31	6.62	7.15	7.08	9.46	9.08	10.31	10.08

Table 9: Comparison of Experiment 4 and Experiment 5, Gender and Background Language results across both Word List A and Word List B

A comparison of the results of Experiments 4 and 5 allows us to investigate the role of physical differences between the foreground and background speaker. Between Experiment 4 and 5, a three-way ANOVA of Experiment, Language and Gender was run to determine if the means of Experiment 5 were significantly higher than in Experiment 4. For serial recall of all test words, there was no effect for Experiment [$F(1,78)=1.128$ $p=0.292$], Language [$F(1,78)=0.571$

$p=0.452$], or Gender [$F(1,78)=0.105$ $p=0.747$] and no interaction for Experiment*Language [$F(1,78)=0.781$ $p=0.380$] nor between Gender and Language [$F(1,78)=0.883$ $p=0.350$] or Experiment and Gender [$F(1,78)=0.002$ $p=0.967$]. For total words recalled, again there was no significant effect for Experiment [$F(1,78)=1.553$ $p=0.217$], no effect for Language [$F(1,78)=1.870$ $p=0.176$] or Gender [$F(1,78)=0.493$ $p=0.485$] and no interaction between Experiment and Language [$F(1,78)=0.173$ $p=0.679$], Gender and Language [$F(1,78)=0.705$ $p=0.404$] or Experiment and Gender [$F(1,78)=0.129$ $p=0.721$]. These results reveal that listeners are not sensitive to the physical differences between the target and masker. Listeners overall perform the same regardless of how different or similar the target speaker is to the masking speaker. The lack of effect here is not surprising, since there were no effects shown for each individual experiment, but effects for list were shown in the previous experiment. Therefore, an overall analysis of each list was conducted.

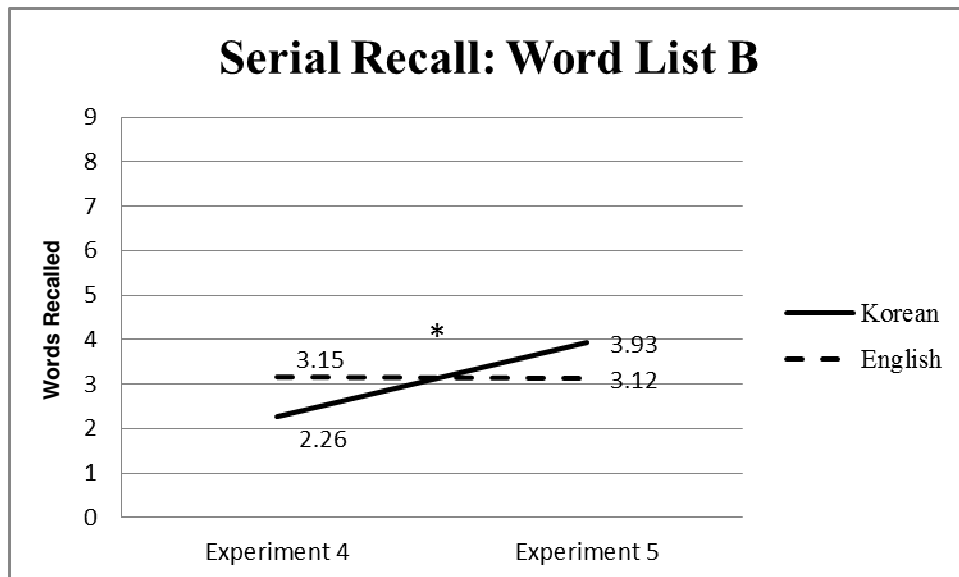


Figure 3: Experiment*Language interaction for serial recall of Word List B

Three-way ANOVAs were run with Gender, Experiment and Language as independent

variables for each scoring of each list to observe effects found in each individual list (for mean values, see Appendix II). For serial recall of Word List A, there was no effect for Gender [$F(1,78)=0.057$ $p=0.811$], no significant effect for Experiment [$F(1,78)=0.111$ $p=0.740$] and no effect for Language [$F(1,78)=1.706$ $p=0.196$]. For total recall of Word List A, however, Language was a significant factor [$F(1,78)=5.197$ $p=0.026$], with Korean performance ($\bar{x}=5.49$) being significantly better than English ($\bar{x}=4.79$). Performance between genders was not significantly different [$F(1,78)=0.237$ $p=0.628$], and Experiment was not a significant factor [$F(1,78)=0.066$ $p=0.798$],

For serial recall of Word List B, there was no significant effect for Gender [$F(1,78)=0.670$ $p=0.416$], but a significant effect for Experiment [$F(1,78)=4.806$ $p=0.032$] with Experiment 5 performance ($\bar{x}=3.46$) being significantly better than Experiment 4 performance ($\bar{x}=2.72$). There was no observed significance for Language [$F(1,78)=0.015$ $p=0.902$]. There was a significant interaction between Experiment and Language [$F(1,78)=5.150$ $p=0.026$] (see Figure 3). All other interactions were nonsignificant. For total recall of Word List B, no significant effect for Experiment was observed [$F(1,78)=3.197$ $p=0.078$], and no effect for Language [$F(1,78)=.006$ $p=0.941$] or Gender [$F(1,78)=0.467$ $p=0.497$] was observed. No interactions were found to be significant.

B. Discussion

The findings of the analysis of Experiment 4 and 5 together reflect what was found in each individual experiment analysis. This supports the idea that having a different speaker for target and masker is not necessary for language effects to emerge. Since effects were only found in individual lists, it suggests that there are factors about each list that are contributing to the

effects observed. In total recall of Word List A, a main effect of Language was found with Korean conditions being significantly better than English. This result again is interesting because it was found in Experiment 4. It is possible that effects emerge for this list because it is slightly more frequent than Word List B, however, the frequency difference between these lists was not found to be significant and the mean frequencies of the list were well within the “high” category created by Poirier and Saint-Aubin (1996).

In serial recall of Word List B, a significant effect for Experiment was observed. Participants recalled more words in the correct order in Experiment 5 than Experiment 4. This result is likely due to the ease of separation of target and masker in Experiment 5. In addition, a significant interaction between Experiment and Language was observed with participants in Experiment 4 showing more sensitivity to background language and a larger separation between Korean and English condition means. Van Engen and Bradlow (2006) found language effects in their “hard” conditions and this could be what is occurring here with speaker perception difficulty in Experiment 4, serial recall being a difficult scoring and Word List B containing slightly less frequent words. Because all other variables are difficult, it is likely that listeners can no longer rely on other methods to aid in their recall and background language sensitivity is heightened.

VIII. Experiment 4 and 5 Performance by Gender: Within-Groups Analyses

As previously discussed, Gender was used as an independent variable for the above analyses, providing a means for measure if males perform differently than females. What the above analyses do not provide is a measurement of how males perform compared to other males and how females compare with each other. This analysis shows us within groups how language

and speaker differences affect recall performance in each gender group. The following analysis provides these results and a discussion of their implications.

A. Results

	N	Total Order Mean	Total Word Mean	Serial Order Condition t-test (p-value)	Total Word Condition t-test (p-value)
Experiment 4 Females					
Korean	13	6.31	9.46	0.761	0.731
English	13	6.62	9.08		
Experiment 4 Males					
Korean	7	6.71	9.43	0.931	0.671
English	7	6.57	8.71		
Experiment 5 Females					
Korean	13	7.15	10.31	0.955	0.796
English	12	7.08	10.08		
Experiment 5 Males					
Korean	8	8.38	10.5	0.081	0.011*
English	6	6.33	8.67		

Table 10: Results of independent samples t-tests when genders are separated with language as the between groups factor.

A series of independent samples t-tests were conducted to compare males' and females' results overall between language conditions for each Experiment. The results in Table 10 show that there is no effect of language for female speakers for either experiment, but for males, there is a significant effect within Experiment 5 for Total Words and a trend effect for language for the serial order scoring. This means that the males had significantly more difficulty with the English background speech condition than with the Korean background speech in Experiment 5 and that the novel female speaker used in Experiment 5 was easier to distinguish from the background speech than in Experiment 4. This may also suggest that Experiment 4 was too difficult for a language effect to emerge because foreground and background speaker were the same. In

addition, females may be better at this task overall and therefore, language effects do not emerge.

A series of ANOVAs was conducted to determine performance accuracy within each gender by list across Experiment 4 and 5. Two-way ANOVAs testing Experiment and Language as independent variables in males' serial recall of Word List A revealed no significant effects. Males' serial recall performance of Word List B across Experiment 4 and 5 show a significant effect for Experiment [$F(1,27)=8.123$ $p=0.009$]. Experiment 4 results were significantly lower ($\bar{x}=2.64$) than Experiment 5 results ($\bar{x}=4.00$). A significant effect for Experiment*Language also was observed [$F(1,27)=6.105$ $p=0.021$] with Experiment 4 performance in Korean conditions at 2.29 words and in English at 3.00 words, while in Experiment 5, males' recall in Korean was 4.63 words on average and 3.17 words in English. Total recall of Word List A showed no significant effects or interactions nor did total recall of Word List B.

Two-way ANOVAs testing Experiment and Language as independent variables for female performance do not show the effects observed in males' performance. For serial recall of Word List A, no significant effects or interactions emerged. There is no significant effect for Experiment [$F(1,50)=0.615$ $p=0.437$], for Language [$F(1,50)=0.884$ $p=0.352$] and there is no interaction between Experiment and Language [$F(1,50)=1.533$ $p=0.222$] for serial recall of Word List B. No effects or interactions for total recall of either Word List A or B were observed.

B. Discussion

In Experiment 4, a significant effect of language did not emerge for males or females. This suggests that the two groups performed similarly across language conditions when speaker variability is absent. What is not immediately clear is why males showed effects of Language in the presence of speaker differences in Experiment 5 when females did not.

As Kramer et al. (1988) and Speck et al. (2000) found, males perform differently on tasks of working memory and specifically on recall tasks than females. The findings of the present study support these findings because males, in the present study, are showing significant sensitivity to the language of the background speech where females are not. This lack of effect shown in females' performance suggests that males are more sensitive to background language than females and their recall performance suffers as a result in the English background speech condition. Females, on the other hand, tend to recall the same amount of words despite background language. These results may be due to differences in performance in each gender as previous studies suggest (Kramer et al., 1988; Speck et al., 2000). It is possible that females are performing at ceiling, which would account for the stability across language and experimental conditions, and that males are not and this is why effects emerge.

Males were also found to be significantly affected by speaker differences. Males perform significantly better in Experiment 5 than in Experiment 4 and the interaction between Experiment and Language is significant. The presence of speaker variability between target and masker in Experiment 5 must provide an aide in separating the target speech and disregarding the background speech. As previous studies suggest, males are less accurate in recall and working memory tasks than females, so the presence of speaker differences significantly helps males to recall more words in Experiment 5 (Kramer et al., 1988; Speck et al., 2000). Females do not seem to require this variability because significant differences are not found between experiments.

IX. Conclusion

The present study tested the effect of background language on serial recall performance and controlled for speaker differences to ensure that any background language effects observed in serial recall tasks with auditory presentation of target stimuli are indeed due to language and not speaker differences. The effects of background language were only found to be significant in specific conditions and experiments, but when language effects emerged listeners in the Korean background speech conditions recalled more words than listeners in the English conditions. This is seen especially when analyzing performance by list in Experiment 4 and across experiments. Along with what previous studies conclude, a shared language between background and target speaker is more difficult for listeners than when the background and target are produced in different languages (Rhebergen et al., 2005; Sperry et al., 1997; Van Engen & Bradlow, 2006). What these previous studies did not test empirically was the potential effect of speaker differences between target and masker on the perception of the to-be-remembered stimuli. Van Engen and Bradlow (2006) and Rhebergen et al. (2005) both suggest that the closer the target is to the masker, the more difficult the task of separating the target and background speech streams will be. The present study set out to test this potential variable and showed that using the same speaker for target and masker indeed elicits a lower recall score than using different speakers, but this difference is only found in specific lists and scorings and is not an overall contributor to recall performance. Therefore, the importance of speaker differences between target and masker does not seem to be a significant contributing factor in recall ability.

The role of physical differences between speakers was tested in the present study through the use of the same bilingual speaker for both targets and background speech in Experiment 4 and then introducing a novel female speaker in Experiment 5 for the target speech to compare

performance between these groups. Although listeners in these two experiments did not perform significantly different from each other overall, when individual list performance was analyzed between experiments, listeners in Experiment 5 performed significantly better than those in Experiment 4 on serial recall of List B. This suggests that the presence of speaker variability does aid listeners in separating the speech signals to an extent. It seems that listeners were able to definitively focus on the target speech better in Experiment 5 than in Experiment 4 where no speaker variability was present. Where we see this effect of Experiment, the means of Experiment 5 are higher than those of Experiment 4.

Results between genders revealed that males and females respond differently to competing speech conditions, and, as previous research indicates, perform differently in working memory tasks. Males are significantly better at disregarding Korean background speech than English background speech, but the background speech must be produced by a different speaker than the target in order for the effect to emerge. Males did not show a significant effect for language in Experiment 4 when the target was the same as the masker. Females also did not show an effect, but perform similarly across language conditions. Typically, females recalled more words than males, which is consistent with previous research (Kramer et al., 1988).

The research questions that the present study addressed are as follows:

- 1) Is there more or less interference with serial recall when the foreground and background talkers share the same language as compared to when they speak different languages?
- 2) Does the physical difference between foreground and background talker affect serial recall?

Question 1 was addressed in the present study in both Experiment 4 and 5; when results

of both Experiments 4 and 5 are considered together, for total recall of words in List A, a significant effect of Language emerges. Where these language effects emerge, Korean background speech shows significantly more words recalled than in English. This suggests that because the listeners are able to understand the background speech, they therefore must process both the background speech and the target words simultaneously in the auditory domain. This yields less accurate results than in the non-native competing speech condition where listeners cannot comprehend the background speech and seem to have an easier time disregarding it.

Question 2 cannot be definitively answered with a simple “yes” or “no.” Experiment 4 and Experiment 5 were intended to test this question through a comparison of results between experiments. Recall accuracy in Experiment 5 was not significantly better than in Experiment 4 overall, but a significant effect for Experiment emerges in serial recall of List B which suggests that listeners are significantly better when differences between target and masker exist.

The effects shown in Word List B support Van Engen and Bradlow’s (2006) findings that language effects tend to emerge when all other helpful variables are difficult. In addition, the effects emerge when the list is presented first in the block, Finally, experiment effects emerge for serial recall of List B, the experiment that lacks speaker variability and numerically yields lower recall of words overall.

The results of the current study support previous findings by both Van Engen and Bradlow (2006) and Rhebergen et al. (2005) that language of background speech affects the performance on working memory tasks. If the background speech stream is as intelligible as the target speech, as in the single-talker background speech condition, listeners have great difficulty in separating the talker from the masker and performing the serial recall task. In addition, the present findings support the working memory model and especially the theory of the

phonological loop (Baddely and Hitch, 1974; Baddeley, 2003). When the phonological loop is occupied by the background speech, listeners are less able to rehearse the target words, and therefore recall is less accurate. A broader implication of the present study is that all speech must be processed and if it contains material that the listener comprehends, it is more difficult to focus on the target speech—especially if the target and masker are physically very similar.

WORKS CITED

- Baddeley, A.D. (1998). Recent developments in working memory. *Current opinion on neurobiology*, 8(2), 234-238.
- Baddeley, A.D. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36(3), 189-208.
- Baddeley, A.D., & Hitch, G.J. (1974). Working memory. *Psychology of Learning and Motivation*, 8, 47-89.
- Banbury, S.P., Macken, W.J., Tremblay, S., & Jones, D.M. (2001). Auditory distraction and short-term memory: Phenomena and practical implications. *Human Factors*, 43(1), 12-29.
- Beaman, C.P., & Jones, D.M. (1998). Irrelevant sound disrupts order information in free recall as in serial recall. *The Quarterly Journal of Experimental Psychology*, 51A(3), 615-636.
- Beaman, C.P. (2005). Irrelevant sound effects amongst younger and older adults: objective findings and subjective insights. *European Journal of Cognitive Psychology*, 17(2), 241-265.
- Bradlow, A.R., & Alexander, J.A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, 124(4), 2339-2349.
- Carroll, L. (1872). *Through the looking-glass and what Alice found there* (2003 ed.) Penguin Group Publishing: New York.
- Francis, W.N. & Kucera, H. (1982). *A standard corpus of present-day edited American English, for use with digital computers*. Brown University: Providence, Rhode Island.
- Fulton B. & Fulton, J. (1997). *Korean Folktale* Seoul, Korea: Green-Stick BD.
- Kramer, J.H., Delis, D.C., & Daniel, M. (1988). Sex differences in verbal learning. *Journal of Clinical Psychology*, 44(6), 907-915.
- Lewis, C.S. (1950). *The Lion, the witch and the wardrobe*. New York: Macmillan.
- Oswald, C.J.P., Tremblay, S., & Jones, D.M. (2000). Disruption of comprehension by the meaning of irrelevant sound. *Memory*, 8(5).
- Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology*, 50(4), 408-412.

- Rhebergen, K.S., Versfeld, N.J., & Dreschler, W.A. (2005). Release from informational masking by time reversal of native and non-native interfering speech (1). *Journal of the Acoustical Society of America*, 118(3), 1274-1277.
- Speck, E., Ernst, T., Braun, J., Koch, C., Miller, E. & Chang, L. (2000). Gender differences in the functional organization of the brain for working memory. *NeuroReport*, 11(11), 2581-2585
- Sperry, J.L., Wiley, T.L., & Chial, M.R. (1997). Word recognition performance in various background competitors. *Journal of the American Academy of Audiology*, 8, 71-80.
- Tremblay, S., Macken, W.J., & Jones, D.M. (2001). The Impact of broadband noise on serial memory: changes in band-pass frequency increase disruption. *Memory*, 9, 323-331.
- Van Engen, K.J., & Bradlow, A.R. (2006). Sentence recognition in native-and foreign-language multi-talker background noise. *Journal of the Acoustical Society of America*, 121(1), 519-526.
- Vorhees, D. & Mueller, M. (1990). *Snail lady; the magic vase* (5th ed.). Elizabeth, New Jersey: Hollym International Corp.

APPENDIX I
Word lists used in Experiments 4 and 5

Note: Words are organized by order of presentation to listeners by list

Practice Word List

talk
heart
club
law
power
plan
night
fire
trial

Mean Frequency: 227.89 per million

Word List A

hand
line
car
space
trade
girl
door
fall
book

Mean Frequency: 245.11 per million

Word List B

food
month
call
land
cold
name
step
wife
top

Mean Frequency: 190.67 per million

APPENDIX II
Tables of Means

Experiment 4: Words recalled in each block, regardless of List A or B

Serial Recall

	Block 1			Block 2		
	Korean	English	Total	Korean	English	Total
Males	3.71	3.00	3.36	3.00	3.57	3.28
Females	3.30	3.84	3.58	3.00	2.77	2.88
Total	3.45	3.55	3.50	3.00	3.05	3.03

Table 11a: Serial recall mean performance of males and females in Korean and English background conditions by block. Because the lists were counterbalanced, both List A and List B may appear in each block.

Total Recall Scoring

	Block 1			Block 2		
	Korean	English	Total	Korean	English	Total
Males	5.14	4.00	4.57	4.28	4.71	4.50
Females	4.85	4.77	4.81	4.61	4.31	4.46
Total	4.95	4.50	4.72	4.50	4.45	4.47

Table 11b: Total recall mean performance of males and females in Korean and English background conditions by block.

Experiment 5: Words recalled in each block, regardless of List A or B

Serial Recall

	Block 1			Block 2		
	Korean	English	Total	Korean	English	Total
Males	4.12	3.00	3.64	4.25	3.33	3.86
Females	3.69	4.25	3.96	3.46	2.83	3.16
Total	3.86	3.83	3.84	3.76	3.00	3.41

Table 12a: Serial recall mean performance of males and females by block and background condition for Experiment 5

Total Recall Scoring

	Block 1			Block 2		
	Korean	English	Total	Korean	English	Total
Males	5.50	4.00	4.85	5.00	4.66	4.85
Females	5.15	5.33	5.24	5.15	4.75	4.96
Total	5.28	4.88	5.10	5.09	4.72	4.92

Table 12b: Total words correct on average by males and females by block and background condition.

Experiment 4 and 5: Performance by List

		Serial Recall Scoring				
		Experiment 4		Experiment 5		
		Males	Females	Males	Females	Total
List A	Korean	4.43	4.08	3.75	3.92	4.02
	English	3.57	3.31	3.17	4.00	3.55
	Total	4.00	3.69	3.50	3.96	3.8
List B	Korean	2.29	2.23	4.63	3.23	3.02
	English	3.00	3.31	3.17	3.08	3.16
	Total	2.64	2.77	4.00	3.16	3.09

Table 13a: Performance of males and females by list and background language condition for serial recall scoring.

		Total Recall Scoring				
		Experiment 4		Experiment 5		
		Males	Females	Males	Females	Total
List A	Korean	5.57	5.62	5.5	5.31	5.49
	English	4.57	4.54	4.5	5.33	4.79
	Total	5.07	5.08	5.07	5.32	5.15
List B	Korean	3.86	3.85	5.00	5.00	4.44
	English	4.14	4.54	4.64	4.75	4.47
	Total	4.00	4.19	4.88	4.64	4.46

Table 13b: Mean performance of males and females by list and background condition for total recall scoring.