

**The Effect of L1 - L2 Vowel Category Mapping on L2 Word learning:  
English Speakers Learning Arabic Pseudo-words**

A thesis submitted to  
The faculty of Graduate Studies and Research in  
partial fulfillment of the requirement for the degree  
Masters of Cognitive Science

by

Lubna Safaa

Carleton University

Ottawa, Ontario

© 2014

Lubna Safaa

## Abstract

Adult language learners often face challenges perceiving unfamiliar speech sounds. Models of second language (L2) speech perception suggest that adults learn L2 through the “filter” of their first language (L1), frequently resulting in misperception and the production of accented speech. The present study had two goals: 1) to examine the learning of a demanding L2 speech contrast by English listeners, and 2) to investigate the role of cognitive resources when learning novel phoneme categories. The first goal was achieved by asking English listeners to learn Arabic vowels embedded in word-like contexts. Unlike English, Arabic uses vowel length as a primary acoustic cue to vowels. English speakers who had no experience with Arabic were presented 36 pseudo-words containing Arabic vowels in a word-learning experiment. The stimuli contained three Arabic short vowels /i,u,a/ and their long counterparts. The task required listeners to learn to associate the pseudo-words with complex images of non-existent objects. The second goal of the study was achieved by looking at the relationship between learning performance and two types of cognitive resources, working memory and attention. Working memory was measured using a span task and the attention was measured using the Attention Network Task. Results yielded no statistically significant effect of vowel type or length on word learnability, nor did participants’ performance on the word-learning task improve significantly over five learning blocks. A small, nonsignificant positive relationship between attentional capacity and overall performance was observed. The results indicate that although listeners may have been able to perceive the difference between the Arabic vowels, learning to associate the novel phonemes to novel concepts may have been too difficult for participants. Suggestions for future work are discussed.

**To my parents**

## **Acknowledgment**

I would like to express my appreciation to my superiors: professors John Logan and Randall Gess for their help and support. Thanks for giving me the opportunity to work and learn from you. Special thanks to Professor John Logan for accepting to supervise me the last minute, and for his input, patience and understanding.

Secondly, I would like to thank my colleague, Elizabeth Rousseau, who was always there to listen and show support.

Lastly, I would also like to express my sincere thanks to my family and friends who always believed in me and surrounded me with their love and warmth.

## **Table of Contents**

<b>Abstract.....</b>	<b>II</b>
<b>Acknowledgment.....</b>	<b>IV</b>
Major models of cross-language speech perception and production .....	4
Key aspects of English and Arabic vowel systems.....	9
English vowel system. ....	11
Standard Arabic vowel system.....	12
Working memory and attention .....	15
<b>Method .....</b>	<b>19</b>
Participants.....	19
Stimuli.....	19
Procedure .....	21
<b>Results and Discussion.....</b>	<b>26</b>
<b>Conclusion .....</b>	<b>31</b>
<b>References .....</b>	<b>33</b>
<b>Appendix A. Stimuli pictures.....</b>	<b>40</b>
<b>Appendix B. List of pseudo-words .....</b>	<b>41</b>
<b>Appendix C. Tables for Mean scores .....</b>	<b>42</b>

## List of figures

<i>Figure 1.</i> A schematic representation of Canadian-English vowels .....	18
<i>Figure 2.</i> A schematic representation of Arabic vowels.....	19
<i>Figure 3.</i> A revised model of working memory (Baddeley, 2000) .....	22
<i>Figure 4.</i> The order of the sessions for the learning and test phases.....	27
<i>Figure 5.</i> The order of the sessions for the learning and test phases.....	27
<i>Figure 6.</i> Illustration of the ANT.....	29
<i>Figure 7.</i> Mean percent correct accuracy in the test phase across the five blocks .....	31
<i>Figure 8.</i> Mean percent correct accuracy in the test phase across vowel type. ....	32
<i>Figure 9.</i> Mean percent correct accuracy in the test phase as a function of vowel type and length. ....	32
<i>Figure 10.</i> Conflict management and word learning.....	36

The effect of L1-L2 vowel category mapping on L2 word learning:  
English speakers learning Arabic pseudo words

Language is a uniquely human capacity that seems to be learned effortlessly in childhood. Adult learners, however, do not learn second languages (henceforth L2) as adroitly as children do, and their speech is usually described as being different from that of native speakers, especially in accent. Bradlow and Pisoni (1999) argued that accented L2 is caused by factors such as a reduction in adults' neural plasticity, which makes it difficult for them to modify their sensorimotor systems to produce the distinctive sounds in the L2 that are not present in their first language (henceforth L1). Flege (1995) claimed that age-related decline in L2 learners' recognition of certain auditorily detectable differences between L1 and L2 sounds is also considered to be a factor that hinders the formation of new phonemic categories, which may lead to developing a foreign accent. Flege, Munro, and MacKay (1995), examined the relation between the perceived foreign accent and age of learning in native Italian subjects speaking Canadian English. The native Italian subjects were asked to read English sentences that were later evaluated by native English speakers on the basis of whether the speakers sounded like a native English speaker. All subjects who began learning English after the age of 15 showed a varying degree of accented speech regardless of how long they had been living in Canada. Flege et al. found a strong positive correlation between age of arrival and perceived foreign accent in speech. In another study, Flege, Mackay, and Meador (1999) tested a number of Italian speakers on their perception and production of English vowels to determine whether accuracy increased as the age of arrival of the subjects decreased. Vowel perception was assessed through using a categorical discrimination test, in which participants had to discriminate between sets of Italian and English vowel contrasts by identifying the odd vowel out on a continuum. They found that

there was a strong negative correlation between age of arrival and vowel discrimination scores; the older the subjects were when they learned English the worse their results were. Both of Flege's studies made a strong claim that the success of early bilinguals in acquiring foreign sounds might be attributed to the establishment of long-term memory representations of the foreign sounds in question, which enabled them to create new phonemic categories. Several other studies have also supported the claim that inaccurate perception and categorization of foreign sounds is correlated with foreign accent (Baker & Trofimovich, 2005; Flege & Hillenbrand, 1984). Sound categorization here refers to subjects' ability to perceive and organize L2 into unique categories to facilitate discrimination and learnability.

Adult learners face more challenges in L2 learning because they have already created lexical, semantic, and phonological representations of their L1 which interfere with their ability to establish new representations for L2 phonemes. For example, Aoyama, Flege, Guion, Akahane-Yamada, and Yamada (2004) showed that Japanese adults were unable to discriminate between English minimal pairs that contain /r/ and /l/ because these phonemes are not contrastive in Japanese and because the Japanese liquid consonant is phonetically closer to English /l/ than it is to English /r/. L2 learners usually face perceptual and categorical problems when acquiring a new language, which arise due to the overlapping phonological units between their L1 and L2.

The processes of learning unfamiliar vocabulary in L2 rely not only on the degree of similarity between the individual sounds between L1 and L2 but also on cognitive factors such as memory and attention. Recent work in cognitive psychology suggests that subjects' ability to repeat or learn novel phonological patterns that rely predominantly on working memory is associated with their long-term vocabulary acquisition. Baddeley, Papagno, and Vallar, (1988) tested a patient with a deficit in short-term phonological memory to investigate what



consequences this might have on long-term phonological learning. They tried to teach the patient either pairs of meaningful words in her native Italian or vocabulary from Russian, which was an unfamiliar language to her. Her performance on pairs of familiar words was quite normal, while her capacity to acquire new vocabulary in a foreign language was grossly impaired, suggesting that a deficit in working/short-term memory is detrimental to the acquisition of foreign language.

In this thesis, my first objective was to focus on the effect of a novel vowel-length distinction on the learnability of non-native words. This objective was achieved by conducting a word-learning experiment in which participants listened to Arabic pseudo-words associated with non-object pictures (adapted from Kroll & Potter, 1981). Participants were evaluated on their ability to learn the difference between long and short Arabic vowels embedded in the pseudo-words. Vowels are segments that form the core of syllables. Every language has a different number of vowels but typical vowel segments that exist in almost all known languages are (/i/, /u/, /a/) (Ashby & Maidment, 2005). Arabic and English have different phonetic and phonological vowel systems that differ not only in the number of vowels but also in how the two languages contrast vowel pairs. Arabic phonology includes a distinctive contrast of vowel length as cued by durational difference, while such a temporal cue is considered secondary in English (Muller, 2012). My second objective in this project was to determine the extent to which general cognitive skills underlie learning unfamiliar sounds by testing the relationship between performance on basic attention and memory tasks and performance on the L2 word-learning task. To obtain this end, participants were asked to complete both a reading span task (Turner & Engle, 1989) and an attention network task (Fan, McCandliss, Sommer, Raz, and Posner, 2002) (more details on these tests are found in the methods section). I hypothesized that first, participants would demonstrate learning through improved scores on the word-learning task

across testing blocks; and second, a positive correlation would be observed between general cognitive task abilities and word learning.

### **Major models of cross-language speech perception and production**

Listening to a novel L2 draws on the same underlying cognitive architecture used to perceive the listener's L1, creating several levels of difficulties for listeners. Having to map and identify the new phonemes appears to be a challenging task for L2 learners especially if the sounds are acoustically similar to each other. Three major models of speech perception and production have been proposed to explain how L1 and L2 interact. These models are Flege's Speech Learning Model (SLM) (Flege & Hillenbrand 1986; Flege, 1995), Kuhl's Native Language Magnet model (NLM) (Kuhl, 1991; 1993), and Best's Perceptual Assimilation Model (PAM) (Best, 1995).

Flege's speech learning model (SLM) focused on adult bilingual speakers' acquisition of speech sounds. This model assumed that the phonetic system established in childhood is dynamic and can undergo change when one language interacts with another. Unsuccessful interactions between the language systems may result in learners failing to learn the phonetic differences between pairs of sounds in L2, or between similar sounds in L1 and L2. Flege attributed this to either a "category assimilation" mechanism whereby L2 sounds become associated with the closest L1 sounds and learners ignore the differences between the two sounds in question, or to a "category dissimilation" mechanism whereby the differences between the pairs of sounds are enhanced and a new L2 category is created. In general, the SLM claimed that L2 sounds are classified according to how "similar" they are to existing L1 sounds. The more similar the sounds are, the harder it is to establish a new category for them; therefore, perception becomes more challenging. In one of several experimental studies testing this model, Flege (1987) showed

that learners failed to produce L2 vowels authentically when they have close counterparts in their L1 vowel space. In a different study, Flege (1980) tested the phonetic contrast between /p,t,k/ and /b,d,g/ produced by native speakers of American English and Saudi speakers of Arabic. Because stop voicing is treated differently in both languages, Flege suggested that this would provide an opportunity to determine how well adult L2 learners modify the phonetic specification of a phonological contrast when that contrast is phonetically implemented in different ways in L1 and L2. Flege concluded that the Arabic-speaking learners of English tried to use different phonetic strategies to mimic the English voicing contrast between the stops, specifically the /p-b/ contrast, which does not exist in Arabic. As a result, they exaggerated the contrast between the sound pairs in the L2 to maximize distinction. The L2 learners were trying to modify their pre-existing phonetic patterns to produce new or different sounds in the L2.

Kuhl (1993), on the other hand, proposed in her native language magnet (NLM) model that children grow up learning to perceive and discriminate sets of sound units, developing acoustic prototypes for those sounds. Speech prototypes refer to the sounds that are identified as ideal representatives of a phonetic category by adult speakers of a given language. When learning a new language, these prototypes function as perceptual magnets that attract “similar” non-native sounds towards them. Some members of a sound category are perceived as better exemplars than others, and having a prototype in mind makes listeners categorize any similar foreign sounds as belonging to that prototype. Kuhl claimed that one can only create these prototypes of the sounds they know from L1, because only L1 provides sufficient relevant acoustic cues. This model suggested that the L1 stands between the learner and his or her ability to perceive the new foreign sounds. This makes it difficult to perceive acoustic characteristics

and cues that are not essential in L1. In other words, language experience constrains the perception of a speech contrast that is phonologically deviant from those of an individual's L1.

Finally, Best (1995) proposed the perception assimilation model (PAM), originally developed to explain why American English speakers were able to discriminate Zulu clicks that they had never heard before. Best argued that this unexpected finding was due to the fact that the listeners did not perceive the clicks as speech sounds (Best, 1988). The PAM mainly attributed the learning of L2 sounds to the perception of either similarity or discrepancy in the L2 sound compared to the closest L1 segment. The PAM emphasized that even though there is a great array of diversity between languages, there is also a good amount of overlap. This is because all sounds are drawn upon gestural possibilities of the human vocal tract, at least at the segmental level.

The PAM discussed primarily how learners incorporate L2 speech sounds into the phonetic system of their L1, and it proposed six ways to treat non-L1 contrast. These ways are 1) two category assimilation (TC type), in which two L2 phones are assimilated to two L1 categories; 2) single category assimilation (SC type), in which two L2 phones are assimilated to a single L1 category; 3) category goodness difference (CG type), in which two L2 sounds assimilate to one L1 category but one of them sounds better suited than the other; 4) uncategorized versus categorized (UC type), in which one sound assimilates to a native L1 category and the other falls outside the phonetic category; 5) both uncategorizable (UU type), in which both L2 sounds fall outside any L1 phonetic category; and finally 6) not assimilable (NA type), in which L2 sounds are perceived as belonging to a system outside the phonetic domain and thus not classified as speech sounds. For both TC and UC, L1 phonology will have a positive effect and make it easier for listeners to perceive the contrast between the speech sounds.

However, it will have a negative effect on SC and UU types, and thus discrimination is expected to be very poor. In addition, discrimination is expected to be moderate for CG and good for NA. According to the PAM, adults' discrimination of a L2 contrast is predicted to depend on how each of the contrasting phones is perceptually assimilated to the most articulatorily similar L1 phoneme. Several studies (Best, 1993, 1994a, 1994b) support the predictions made by the PAM. For example, Best (1994) investigated again the ability of American English speakers to discriminate the click consonants of Zulu. Since the consonants are very deviant from any English phonetic sound (i.e., a NA type), the discrimination level was expected to be very good. The results showed that the prediction was upheld and listeners were able to discriminate those clicks even though they had no prior exposure to the language.

The three major models of speech perception focus on the idea that L1 phonology plays a major and essential role in the perception of L2 phonemes. All three models focus on the role of L1 in the perception of L2 sounds and they all contribute to our understanding of the perception of nonnative speech sounds. Despite their similarities, the models are different in their explanation of the acquisition of L2 sounds. The PAM, for example, emphasizes the role of L1 phonology and suggests that it should aid discrimination when two L2 phones are separated by native phonological boundaries, but should hinder it when both phones assimilate to the same L1 phoneme. However, discrimination of L2 elements that are heard as non-speech sounds is neither helped nor hindered by the native phonology. In addition, both the PAM and the NLM make implicit predictions about L2 speech perception, while the SLM is unique in making explicit predictions regarding both the perception and the production of foreign speech sounds. For example, Flege et al. (1999) examined the perception and production of early and late Italian bilinguals of Canadian English vowels and found that their results supported the predictions the

SLM makes regarding the level of accuracy. They found that early bilinguals did not differ from native English speakers in their vowel production, which suggested that the early bilinguals established new phonemic categories for L2 vowels that did not exist in their L1. On the other hand, the PAM seems unique in its constant reference to pairs of L2 sounds (rather than individual sounds) compared to the L1 inventory, as stressed in its proposed six assimilation patterns.

I particularly focused on the PAM, and considered it to be the most relevant model because it is more concerned with naïve than experienced L2 listeners, as was the case in this study. In addition, the PAM offers several possible patterns for interpretations.

Based on the PAM, the following patterns of results were predicted for the perception of Arabic vowels by English speakers. English speakers were expected to show difficulty categorizing Arabic vowel pairs that contrast mainly in length, because length is a secondary cue in English (see English vowel system below). This difficulty was expected to have a negative effect on their perception and therefore hinders their learning.

Based on the PAM model, several possibilities exist for how L2 listeners would perceive the difference between the vowels. First, if the listeners fail to perceive the difference between the vowel pairs of Arabic they would probably assimilate them to a single counterpart from the English vowel inventory, which would correspond to single category assimilation (SC), and in this case, discrimination between the pairs of words would be hindered. Second, Arabic vowel pairs could assimilate to one English vowel but one of them is considered a better match than the other, which could correspond to category goodness. In this case, listeners might detect that there is a difference between the pairs but fail to realize it.

The models predict that L2 sounds should assimilate to the closest native sound, which means that the results of how participants categorize Arabic long-short vowel pairs would depend on whether they perceive the difference in vowel lengths.

It is worth mentioning that all three models are similar regarding the fact of perceptual assimilation of L2 to L1, and, ideally, they could help interpreting the results in the research presented in this thesis. However, in reality, the models are more of a heuristic because they lack explicit criteria defining what is meant by ‘similar’ and ‘new’ phonemes when describing their hypothesis. In short, they do not clearly specify how ‘close’ an L2 phoneme should be to an L1 phoneme in order to be classified as ‘similar’ to the L1 phoneme.

### **Key aspects of English and Arabic vowel systems**

Languages of the world have different vowel systems, some with as few as three vowels, and others with up to twelve vowels. Vowels are produced when vocal fold vibration excites the resonance of the vocal tract resulting in peaks in the spectrum called formants (Ashby & Maidment, 2005). Some of the basic descriptors usually employed to describe vowels, in terms of their articulation, are the height of the tongue body (high, mid, low); tongue advancement (tongue position in the mouth: front, mid, or back), and lip rounding (rounded and unrounded). The traditional way of organizing vowels uses a two-dimensional grid, where the horizontal dimension indicates tongue advancement and the vertical dimension indicates tongue height (see Figures 1 and 2 for examples). Acoustically, vowels are typically described in terms of their formants, the frequency values of which are expressed in Hertz (formants are concentrated regions of resonant frequencies caused by the different shapes of the vocal tract associated with each vowel). The first two formants (F1 and F2) play a prominent role in describing vowel

quality. F1 correlates with vowel height; for example, high vowels such as /i/ have a low F1 and low vowels such as /a/ have a high F1. F2, on the other hand, correlates with vowel frontness or backness; for example, back vowels such as /u/ have a low F2 and a front vowel such as /i/ has a high F2. Vowel quality is associated with different features, among which is the position of the velum (which determines whether the vowel is nasalized or not), and the position of the root of the tongue. When the tongue root is advanced, the resulting vowels are described as being “tense”, referring to the muscular tension produced when pronouncing them, as opposed to their “lax” counterparts. Languages like English use this distinction to categorize their vowels as explained in the section below. Another parameter, among others, used to describe vowels is vowel quantity. Vowel quantity is used to describe a difference in phonological length. Length is a phonological correlate of the durational difference between vowels in languages that have a “long” and “short” vowel distinction such as Japanese and Arabic (Gussenhoven & Jacobs, 2005).

The research in this thesis is concerned with one of the many differences in the vowel inventories of Arabic and English. It is concerned with the use of different primary acoustic cues in categorizing vowels in each of these two languages. Below is a non-exhaustive summary of the main characteristics in question between the two vowel systems, focusing only on vowel tenseness and length. It is important to take into consideration that only vowel features that are considered relevant to the present study are summarized in the following section.



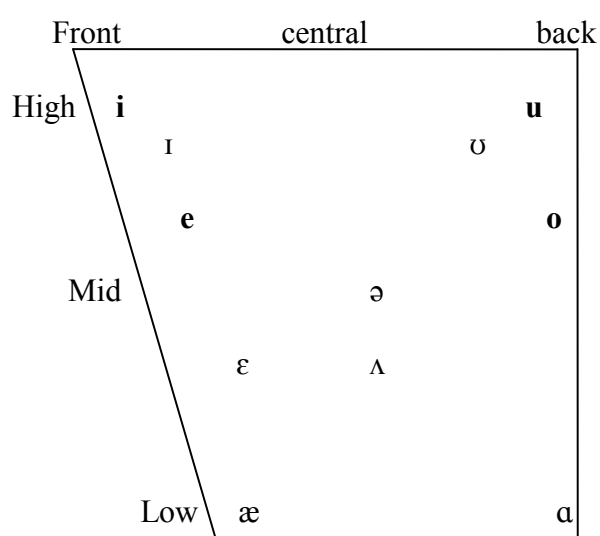
## **English vowel system.**

As mentioned above, vowels are described in terms of three main articulatory properties: tongue height, tongue advancement and lip rounding. Standard North American English has a 12-vowel system, distributed into an unusually rich set of front vowels /i, ɪ, e, ε, æ/, central vowels /ə, ʌ/ and back vowels /u, ʊ, o, ɔ/ (the slashes (/) indicate a broad phonemic transcription). Another feature that imposes additional contrast upon these vowels is tenseness where vowels are further divided into tense long and lax short vowels. English has six tense vowels /i, u, e, a, o, ɔ/ and six lax vowels /ɪ, ʊ, ε, ə, ʌ, æ/ distributed across the vowel space as depicted in Figure 1. Several linguists further describe English tense vowels as having a diphthongal quality, as for the vowels /e, o/ which are also transcribed as /eɪ, oʊ/. Similarly, the tense high front vowels /i, u/ are often transcribed as /ij/ and /uw/ (as in the words ‘bee’ /bij/ and ‘music’ /miuwɪk/ respectively) (Jensen 1993, Carr & Durand 2004).

Tense vowels correlate with phonetic length, being produced with longer duration than the lax vowels. However, contrastiveness in English is based primarily on the quality of the vowels while quantity, i.e. vowel length, is considered an accompanying property that does not create a contrastive feature on its own (Muller, 2012). Kondaurova and Francis (2008) found that native English speakers differentiate vowels based on two independent acoustic dimensions: vowel quality (spectrum) and duration, relying predominantly on the former as the main acoustic cue. Durational difference is not considered phonemic in English. For example if one were to pronounce the word ‘bat’ with a prolonged vowel duration while preserving its quality, it would still be perceived as the word ‘bat’ for the native English hearer.

Several studies have investigated the role of vowel duration in the recognition of English vowel quality (Ainsworth, 1972; Zahorian & Jagharghi 1993; Stevens 1959; Hillenbrand, Clark,

and Houde, 2000). For example, Hillenband et al. (2000) synthesized versions of 300 utterances of a syllable of the type /hVd/, so that in each version the vowels were lengthened, shortened, or kept as original duration while keeping the spectral properties constant across the vowels. They found that participants were able to identify correctly the vowels /i, ɪ, u, ʊ, e, ε / at their original and altered durations; therefore, it was concluded that duration had only a modest overall effect on vowel perception.



*Figure 1.* A schematic representation of Canadian-English vowels (Tense vowels have been bolded)

### **Standard Arabic vowel system.**

As a Semitic language, Arabic is noteworthy for its limited vocalic system but rich consonantal system. Arabic is a 6-vowel system, consisting of /i, u, a/ and their corresponding long vowels /i:, u:, a:/ (the colon (:)) indicates a longer duration) (see Figure 2). others claim that

the vowel inventory of modern standard Arabic does not contain any diphthongs. However, some argue that there exist two diphthongs /aw, ay/, raising the total number of vowels to eight (Alotaibi & Husain, 2009). Vowel length is phonemic in Arabic whereby duration is the primary acoustic correlate for the contrast among certain vowel pairs. In Arabic, for example, the words *sharika* and *shari:ka* which differ only in the length of the middle vowel, have different meanings (*company* and *female partner*, respectively). Although it is often difficult for nonnative speakers of languages with phonemic vowel length to learn to use these contrasts phonemically, native speakers have a stable ability to categorize vowels based on length alone. Regardless of dialect, Arabic vowels show less diphthongization than English vowels.

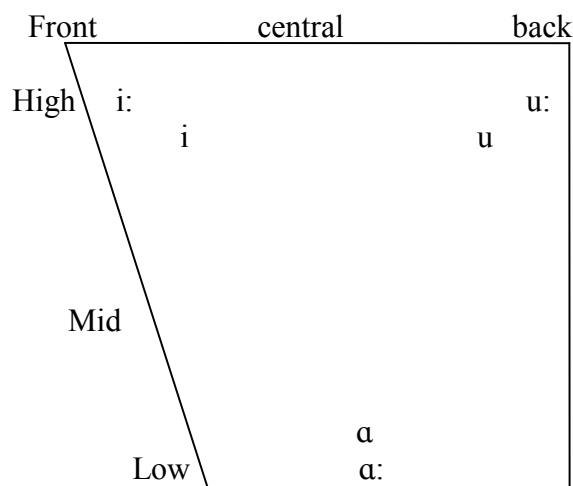


Figure 2. A schematic representation of Arabic vowels.

To conclude, the vowels of Arabic and English may share many similarities but they have distinct characteristics, and it is clear that each language relies on a different primary acoustic cue in discriminating between its vowels. Moreover, based on work done by Saadah (2011), the

English vowel system is classified as ‘a centripetal system’ in which vowels are more at the center of the acoustic space as opposed to languages such as Russian and Spanish, where vowels are located at the periphery of the acoustic space. On the other hand, Arabic has a vowel system that falls in between the two types of systems. As a result, these differences between English and Arabic show that there seem to be no simple one-to-one mapping between the vowels of English and Arabic. Whereas English native speakers rely predominantly on vowel quality in making judgments about vowel types and where quantity is used to enhance the quality difference of the tense/lax pairs, Arabic native speakers use both quality and quantity as main acoustic cues when categorizing vowels.

A study by Munro (1993) shows that in a perceptual task, native Arabic speakers usually categorize English tense/lax vowels in terms of the Arabic vowel inventory and perceive them as long and short. Consequently, they produce the English vowels with Arabic-like properties. Nevertheless, the tense-lax pairs produced by Arabic speakers show a greater duration difference than the English speakers’ tense/lax pairs, which suggests that Arabic L2 English speakers have perceived the English /i/ as a variant of the Arabic /i:/ and the English /ɪ/ as the Arabic/i/. An L2 effect was expected to be present when testing English native speakers’ perception of Arabic vowels and it was expected that they would try to categorize the Arabic vowel pairs in terms of their knowledge of English phonology and rely predominantly on spectrum as the main acoustic cue in distinguishing vowels. Participants were expected to conflate the vowels at the early stages of testing, but would learn that Arabic relies on length in distinguishing its vowels at the later stages of testing.

The focus of the present study is on the fact that duration in English is associated with specific qualities but it is not contrastive on its own as in Arabic. All other characteristics of the

vowel inventories of both English and Arabic are ignored focusing only on what is relevant to what the experiment seeks to test.

### **Working memory and attention**

Learning new words is a cognitive process that requires an interaction between different components of the brain. It requires the learner to select information (attention) and to retain information in an accessible state (working memory). Attention as used here refers to the selection of some information at the expense of other information (Pashler,1998).

Both processes are connected and cannot be separated from each other. Engle (2002) concluded that individuals with high performance on different working memory tasks have higher attention abilities than those who score low on the same working memory tasks.

The initial concept of memory was based on a unitary temporary storage called the short-term memory (STM), but because of the active role of memory in several cognitive tasks, this idea was abandoned, and a multi-component model was proposed by Baddeley and Hitch (1974). The term ‘working memory’ was presented then to emphasize that the memory system is not merely storage and that it is very much involved in different cognitive tasks. Baddeley’s working memory model is a very influential model that is designed to account for how information is stored and manipulated during thinking and reasoning.

Working memory was defined as “the temporary storage of information in connection with the performance of other cognitive tasks such as reading or problem solving or learning” (Baddeley 1983, p. 311). Baddeley and Hitch (1974) originally proposed a three- component working memory model that consists of an attentional control system, the central executive, and

two temporary storage subsystems, the phonological loop and the visuospatial pad. Baddeley (2000) added a fourth subsystem called “the episodic buffer” to the model (see Figure 3).

The central executive plays a crucial role in working memory as it works as an attention control system that coordinates and regulates cognitive processes and decides which information is attended to, so that our attention is directed to a particular activity such as stopping at a red light. The phonological loop is assumed to comprise a temporary phonological store that keeps auditory information and that is specialized for the retention of verbal information such as phonological forms, for short periods of time. The second component of the phonological loop is the articulatory loop, which helps to maintain decaying phonological representations in the phonological store. The phonological loop plays an important role in learning new words (Baddeley & Gathercole, 1998). The visuospatial sketchpad is claimed to be responsible for temporarily maintaining visuospatial information like the shapes, colors, and locations of the objects around us. The episodic buffer links information through the different domains and long-term memory by creating a direct link between the phonological loop and long-term memory. The episodic buffer regulates time sequences in a chronological order and provides an interface between the different codes (visual, verbal, perceptual). It is important to understand that working memory is limited in its resources and there are individual differences in how much information can be temporarily stored in it.

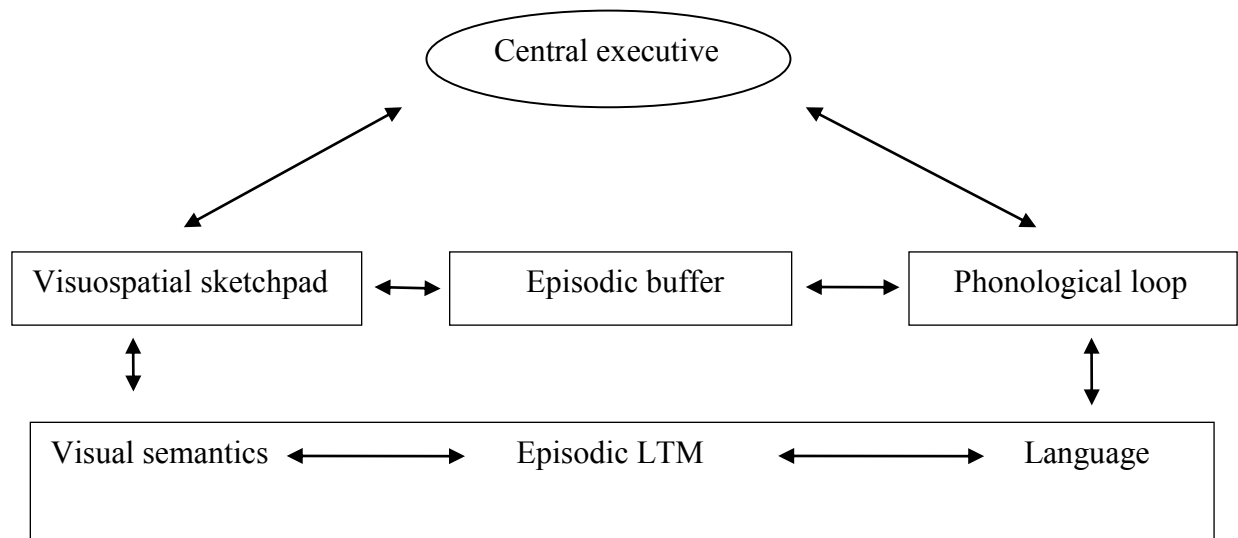


Figure 3. A revised model of working memory (Baddeley, 2000)

The phonological loop is of special importance to this study since there is evidence in the literature that supports its role in L1 word learning in general and in L2 in particular because it is responsible for the retention of verbal material. Baddeley et al. (1998) suggested that the phonological loop is used not only to remember familiar words but also to learn new ones (p.158). Many tests on children have investigated the association between phonological memory and word learning (Gathercole, Hitch, Service, and Martin, 1997; Gathercole & Baddeley, 1990). Children were tested on their ability to remember pairs of familiar words or non-words, and the results showed that there was a strong correlation between phonological loop ability (which was tested by running memory span tests, i.e., the digit span test and the phonological span test) and the rate of learning words. According to Baddeley's model, learning new words depends on the function of the phonological loop through both storing novel phonological forms in short-term memory and transporting traces of the short-term memory to long-term memory by rehearsal, which is a process responsible for repeating information to help maintain decaying

representations in the phonological loop, and the central executive. Several studies have shown that foreign language vocabulary learning is linked with phonological memory capacity by demonstrating that higher scores on different phonological measures such as non-word repetition tasks are highly associated with foreign vocabulary retention (Gathercole & Baddeley, 1989; Gathercole & Baddeley 1990; Baddeley et al., 1998; Service, 1992).

Working memory capacity is an important indicator of the general intellectual ability of an individual. A wide range of research areas use different working memory tasks such as the operation span task or the reading span task to obtain information on how working memory correlates with word learning (Conway et al., 2002, Conway et al., 2003, Engle et al., 1999). In this thesis, I investigated the relation between working memory and L2 word learning by comparing the results obtain from the reading span task that the participants performed and their word learning task scores.



## Method

### Participants

Thirty-five undergraduate (22 females, 13 males) Carleton university students participated in the experiment for course credit. The mean age of the participants was 21.1 years. All participants spoke English fluently and they had no knowledge of Arabic. Participants reported normal or corrected vision, and no hearing or neurological problems. All participants gave a written informed consent before the experiment started.

### Stimuli

Pseudo-word-learning experiment stimuli were made up of 36 sound files and 36 pictures. The pictures were black-and-white drawings adopted from Kroll and Potter (1984). The size of each picture on the computer screen was  $8 \times 8$  cm/300 pixels  $\times$  300 pixels (height  $\times$  width). The pictures depicted non-object shapes that were unfamiliar to participants (see appendix A) (for more information, see Kroll & Potter, 1984). The 36 pictures were divided into two sets of 18 pictures each. The sets did not differ significantly in terms of the object-familiarity ratings provided by Kroll and Potter ( $z = 1.04$ ,  $p = .30$ ). The sound files paired with the non-object pictures used in the experiment were 36 three-syllable Arabic pseudo-words (e.g., /ri.su.ba/, /ka.ti:.mu/) (a period (.) represents syllable boundaries, slashes (/) indicates a broad phonemic transcription, (: ) indicates a long vowel). Each syllable was composed of a consonant and one of the three Arabic vowels /i, u, a/. No identical vowels appeared within the same word boundary. The consonants were selected so that all Arabic emphatic sounds were avoided because they present special difficulties for L2 learners. Any consonant combinations that might sound similar to English words were also excluded to avoid confusion.

The 36 pseudo-words were divided into two conditions (short and long, see appendix B). In the short, vowels in the pseudo-words were all short e.g., /ri.su.ba/ and in the long condition, the second vowel was always long e.g., /ri.**su:**.ba/ (the difference has been bolded for clarity).

The pseudo-words were spoken and recorded by a female native Arabic speaker, and digitized in mono at a 16-bit/44.1KHz sampling rate, normalized with an intensity level of 70 dB. Furthermore, acoustic measurements and analyses using Pratt (Boersma & Weanik, 2005) were carried out on the stimuli to ensure that the two conditions differed significantly only in the length of the second vowel and to avoid having the participants rely on irrelevant acoustic cues. The other acoustic parameters crucial to the experiment such as the duration of the first vowel, the third vowel, and F0 maxima of vowels were kept constant (see Table 1;  $p > .05$  in pair wise comparisons).

Table 1. *Acoustic analysis for the sound stimuli tested in the experiment. The table presents the mean duration (ms) and F0 maxima (Hz) for the three vowels tested in each condition (short & long) in the experiment. Standard deviations for the mean (SD) are provided in parenthesis.*

Condition	Mean duration			F0 maxima		
	1 <sup>st</sup> V	2 <sup>nd</sup> V	3 <sup>ed</sup> V	1 <sup>st</sup> V	2 <sup>nd</sup> V	3 <sup>ed</sup> V
short	.147(.0)	.152(.0)	.148(.0)	265.8(11.5)	244.2(11.3)	206.3(9.8)
long	.143(.0)	.241(.0)	.143(.0)	266.3(16.1)	244.6(13.1)	208.1(8.5)

*Note: N=18 in each condition; 1<sup>st</sup> V = first vowel in the word, 2<sup>nd</sup> V = second vowel in the word, 3<sup>ed</sup> V = third vowel in the word.*

## Procedure

**Pseudo-word-learning task.** Participants were seated in front of a 14" laptop screen in a sound attenuated booth, wearing headphones. Prior to beginning the experimental sessions, participants were familiarized with the experimental procedures in a practice session. The experiment had five blocks consisted of learning and test phases (as depicted in Figure 4). During the learning phase, participants were asked to learn new “words” by listening to audio stimuli played through headphones and associate them with pictures presented simultaneously with the sound. No Feedback was provided during the learning phase.

In the test phase, participants’ ability to associate the pseudo-words with the images was assessed by asking them to choose the image that corresponded to the pseudo-word that they heard in the learning phase. Each learning phase consisted of 36 trials in which participants heard 36 pseudo-words and saw 36 corresponding images. The 36 pseudo-words were randomly ordered so that participants could not predict whether the words were from the short or long condition. There was no intentional testing of minimal pairs, although they might have occurred within the same trial. Participants learned one pseudo-word per trial. Each trial began with a fixation cross “+” presented for 100 ms, then a blank screen was presented for 200 ms. Next, the image was presented on the screen, followed by a 100 ms delay, after which the pseudo-word was presented. The image remained on the screen for 2000 ms. A fixation cross “+” appeared at the end of each trial to indicate the beginning of the next trial. At the end of the 36 trials, a break was provided before the beginning of the test phase. In the test phase, participants were tested on their ability to match the words they heard in the learning phase to the images that they were associated with. In each test phase, the participant first looked at the fixation cross “+” presented for 800 ms, then a blank screen presented for 300 ms. A pseudo-word was then presented, while

the screen remained blank. After the pseudo-word had been presented, two images from the learning phase appeared, one situated to the left side of the screen and the other one to the right side. The participants were instructed to indicate which image was associated with the pseudo-word presented in the learning phase. (Figure 5 depicts what participants saw in the test phase). The order of the pictures was randomized for each trial. The images remained on the screen until the participant responded by pressing the left or right arrow on the computer keyboard. In this manner, all 36 sounds were tested. Both accuracy and response time (RT) were recorded for each trial. Participants repeated the two phases in a series of five blocks as depicted in Figure 4.

Block 1		Block 2		Block 3		Block 4		Block 5	
<b>Learn ng phase</b>	<b>Test phase</b>	Learn ng phase	Test phase	<b>Learn ng phase</b>	<b>Test phase</b>	Learn ng phase	Test phase	<b>Learn ng phase</b>	<b>Test phase</b>




Figure 4. The order of the sessions for the learning and test phases.

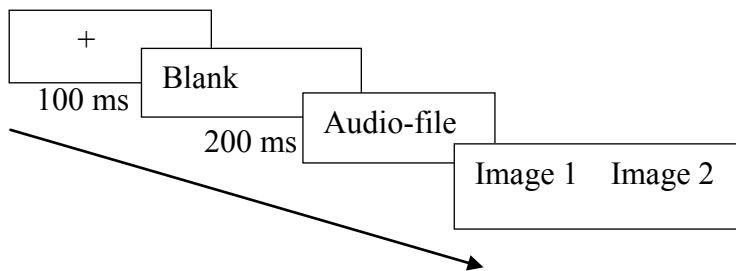


Figure 5. The order of the sessions for the learning and test phases.

**Attention and memory tasks.** Because language learning is associated with basic cognitive skill learning, participants were asked to complete two cognitive tasks, the attention network task, and the reading span task, as part of the main word-learning experiment. Because memory and attention are implicated in language acquisition, these tasks were used to test whether individuals with higher working memory and attention capacities would outperform those with lower capacities in L2 word learning.

**Attention network task (ANT).** This task measured the three functions of attention (alertness, orienting, and executive function). Alertness refers to the ability maintain an alert state. Orienting refers to the ability to select information from sensory input for further processing, and the executive functions refers to conflict management abilities that arises when participants encounter competing stimuli.

Following the procedure used by Fan et al. (2002), participants were asked to monitor the direction of an arrow in the middle of a screen by pressing the 'left' key on a computer screen if the arrow was pointing left, or pressing the 'right' key when the arrow is pointing to the right. In addition to this target arrow, there were two arrows on each side, referred to as 'flankers'. These flankers either pointed to the same direction as the target arrow (congruent), or pointed to the opposite direction (incongruent condition). Some trials did not have any flankers (neutral condition). In addition, participants were asked to attend to four cues, appearing with the fixation cross that preceded each trial: a central cue, a double cue, a spatial cue, or no cue. In the center cue condition, an asterisk was presented at the location of the fixation cross. In the double cue condition, an asterisk appeared at the location of the target above or below the fixation cross. In the spatial cue, an asterisk appeared in the position of the upcoming target. The task was administrated in four blocks, the first block was for practice and it took about two minutes. The

other three blocks were experimental blocks and each took about five minutes. The task lasted for approximately 20 minutes (see Figure 6).

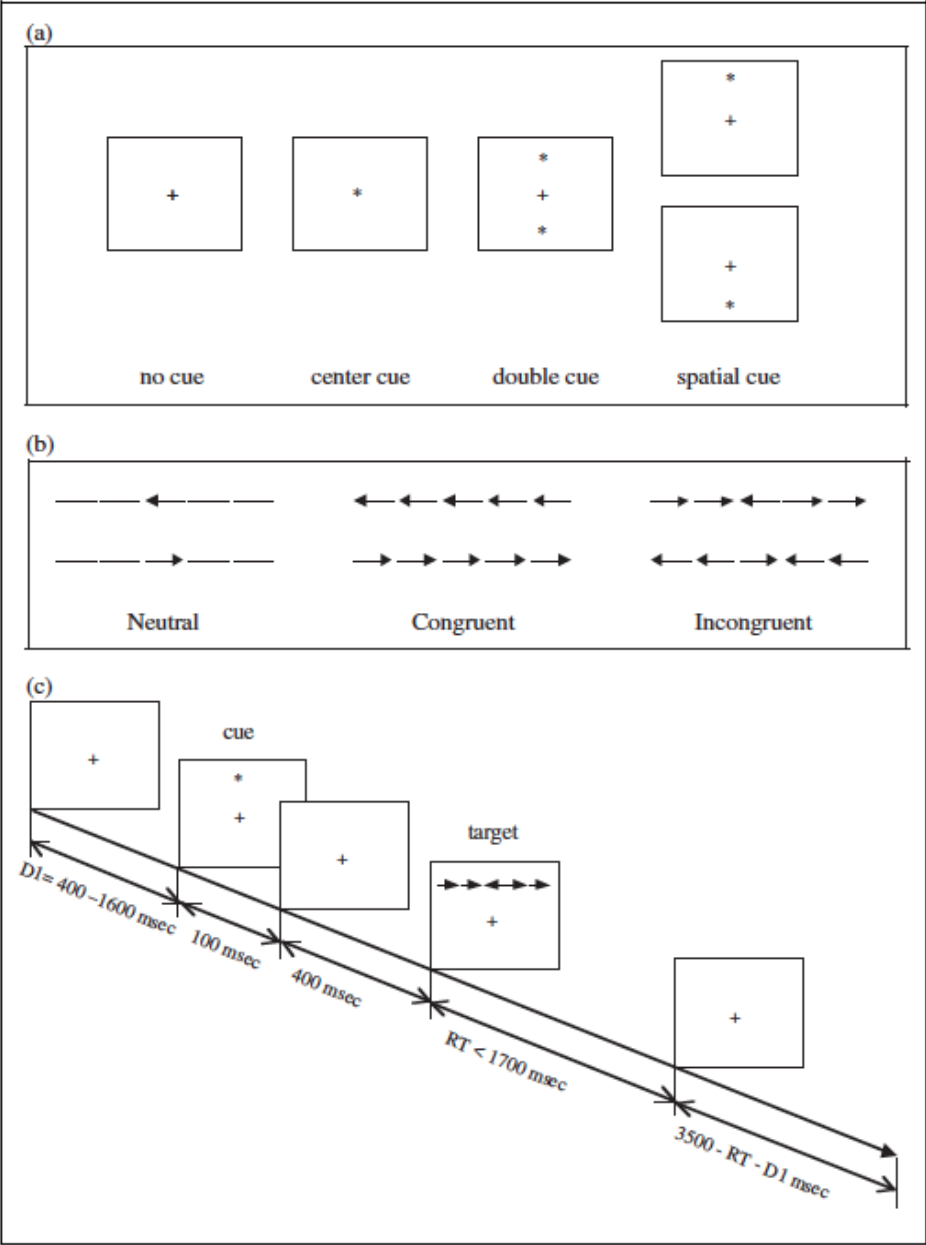


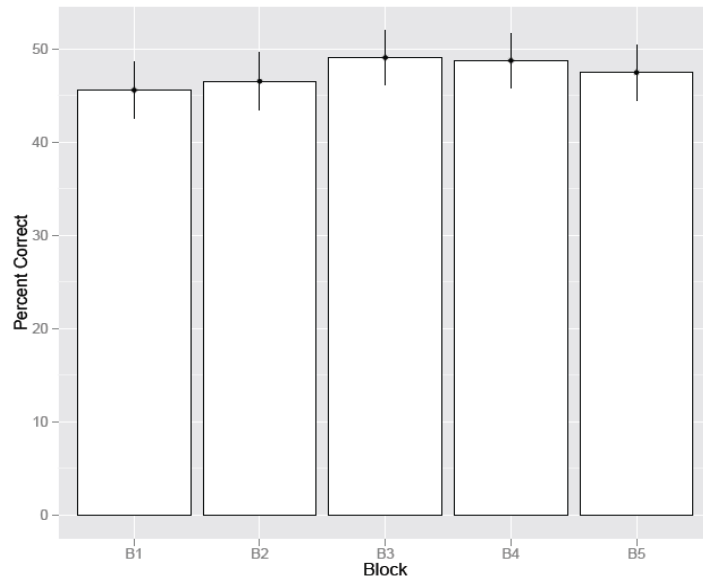
Figure 6. Illustration of the ANT

(This figure is adapted from Fan et.al (2002). It illustrates the four cue conditions (a), the six stimuli presented (b) and an example of the procedure (c)).

**The reading span task (Rspan).** This is a widely used measure of working memory capacity (WMC). This measure focuses on the ability to keep relevant information active in memory while conducting complex cognitive tasks such as sentence comprehension and judgment making (Daneman & Carpenter, 1980). Turner & Engle (1989) created a new version of the reading span task in which the sentences were presented auditorally and visually and where they had to make semantic judgements. Participants were asked to read the sentences aloud simultaneously as they heard them, each of which had a random letter presented right after it (e.g. *Andy was stopped by the policeman because he crossed the yellow heaven. R*). The participants were asked to make judgments on whether the sentences were semantically correct by pressing the letter 'C' on the keyboard, or semantically incorrect by pressing the letter 'M' on the keyboard. After a certain span of sentences, the participants were required to write down on a piece of paper provided to them all the letters they saw in order. There were 12 items (each group of sentences was referred to as one item; each item consisted of a different number of sentences starting from two up to five sentences, referred to as spans).

## Results and Discussion

Participants' performance in the word-learning experiment were analyzed as follows. For each participant the mean percentage of correct answers was calculated for block, vowel, and length. Overall accuracy for learning the pseudo-words was near chance across the five blocks, with the lowest score seen on the first block ( $M = 45.57$ ), and the highest score seen on the third block ( $M = 49.05$ ) as shown in Figure 7. Performance did not vary as a function of the two different conditions (short and long) nor across the three different vowels /i, u, a/ (as depicted in Figure 8 and Figure 9).



*Figure 7.* Mean percent correct accuracy in the test phase across the five blocks ( Error bars represent 95% confidence intervals)



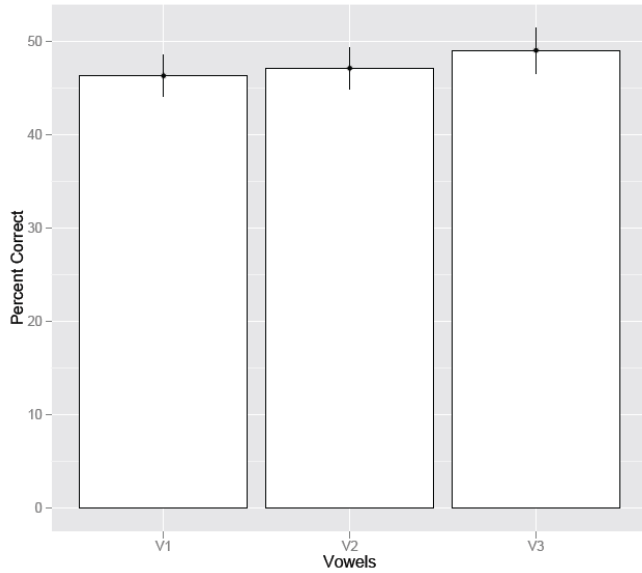


Figure 8. Mean percent correct accuracy in the test phase across vowel type.  
 (Error bars represent 95% confidence intervals. V1 = /i/, V2 = /u/, and V3 = /a/)

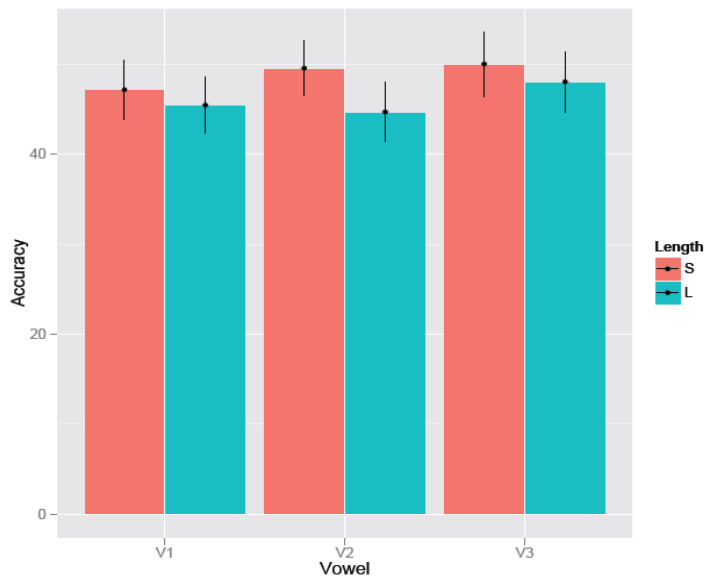


Figure 9. Mean percent correct accuracy in the test phase as a function of vowel type and length.  
 (Error bars represent 95% confidence intervals.)

To evaluate the reliability of effects, the data were subjected to a factorial 5 (block)  $\times$  2 (condition: short and long)  $\times$  3 (vowel type: i,u,a) analysis of variance. Consistent with the patterns of results shown in figures 7, 8 & 9, no statistically significant effects were observed (Block,  $F(4, 136) = 0.89, p = 4.71^{-1}, \eta^2 = .0036$ ; vowel type,  $F(2, 68) = 2.10, p = 1.3^{-1}, \eta^2 = .0027$ ; (vowel length,  $F(1, 34) = 2.26, p = 1.4^{-1}, \eta^2 = .0042$ ). The results obtained did not support my hypothesis that after being exposed to the non-native vowels repeatedly, participants would learn to perceive the vowels and that a positive correlation would be observed between learning and the general cognitive tasks.

The lack of learning observed over the five blocks is consistent with two possibilities. First, it could mean that the participants were unable to identify the differences between the pairs of vowels, perceiving them as identical, which in turn made it too challenging for them to remember the associated pictures. The second possibility is that the participants had a large memory load as a consequence of trying to memorize unfamiliar sounds, and then matching them to previously unseen line drawings may have hindered their ability to perceive an acoustic contrast that is not essential in their L1.

Although the small improvement from the first to the third block is not statistically significant, it could suggest that some learning occurred. Nevertheless, the results do not contradict what would be predicted by the PAM or the SLM. One of the PAM's claims is that participants might fail to perceive the difference between non-native sounds, assimilating them to a single category, and thus perform poorly on the task. The SLM also is clear that non-native sounds are usually assimilated to the closest similar L1 sound, and the more similar the sounds, the harder it is to perceive them accurately. Therefore, one of the interpretations of why performance did not improve could be attributed to a perceptual difficulty of distinguishing the

difference between the short and long vowels and thus confusing the different words and fail to learn them.

Learning to associate the pseudo-words and novel images may also have been problematic, even if participants were able to perceive the vowel and length differences. Perception is just one aspect of word learning. Learning in the context of the present experiment also required the additional step of linking the pseudo-words to the images. Because learning phases did not include any feedback, making this link may have been beyond the capabilities of the participants. This view of the results is reinforced by comparing the results from the present experiment with previous research that attempted to train listeners to perceive a duration contrast. Pisoni, Aslin, Perey, and Hennessey (1982) were able to train monolingual English listeners to perceive the prevoiced voicing category used in Thai and other languages with no feedback. Pisoni et al. used synthesized CV stimuli varying in voice-onset time (VOT), an acoustic cue in which duration is the underlying dimension. While acknowledging the difference between consonant and vowel stimuli used in Pisoni et al. and the present experiment, respectively, both utilize acoustic cues in which time-varying information signals different phoneme categories. One could argue that if listeners can perceive the time-varying information associated with voicing categories in consonants that listeners should be able to perceive time-varying information associated with vowel categories. Unfortunately, this argument by analogy is without direct empirical evidence because there was no evaluation of listeners' perception of the pseudo-words independent of the learning task, which integrates perception and learning together.

Other methodological considerations also may account for the lack of learning observed in the present study. Completing all three tasks (the word-learning task, the working memory

task, and the attention network task) took approximately an hour and a half, during which participants were sitting in front of a computer screen. Boredom and/or fatigue may have been factors that caused their attention to the task to deteriorate. In addition, motivation for the task may have been limited. The participants were all undergraduate students who participated in the research for course credit, and who generally showed little interest in the research, other than obtaining their credit.

The links between the working memory task, attention network task and the word-learning task were investigated by computing the correlation between performance on each of these tasks and performance the word-learning task measured by averaging performance across all five test blocks. No significant correlation was found between the working memory task and the word-learning task. However, a marginally significant positive correlation was found between the overall accuracy on the word-learning task and the conflict network in the attention network task, as illustrated in Figure 7 ( $r = .28$ ,  $p = .09$ ). The effect was not large but it consistent with previous findings showing executive attention performance is positively correlated with performance on language learning tasks. The correlation suggests that the two tasks are related by virtue of both requiring attention to be directed towards one stimulus and not the other. The size of the effect is modest, in part, because of the limited range of learning performance by participants; a truncated range on one variable will generally result in a reduced correlation coefficient. Participants with the highest mean performance on the learning task achieved only ~55%, with the lowest performance ~45% (one exception is the participant who had 25% correct, who could be considered an outlier). If learning performance had not been subject to a floor effect, the correlations with the WM and attention measures would likely have been larger.

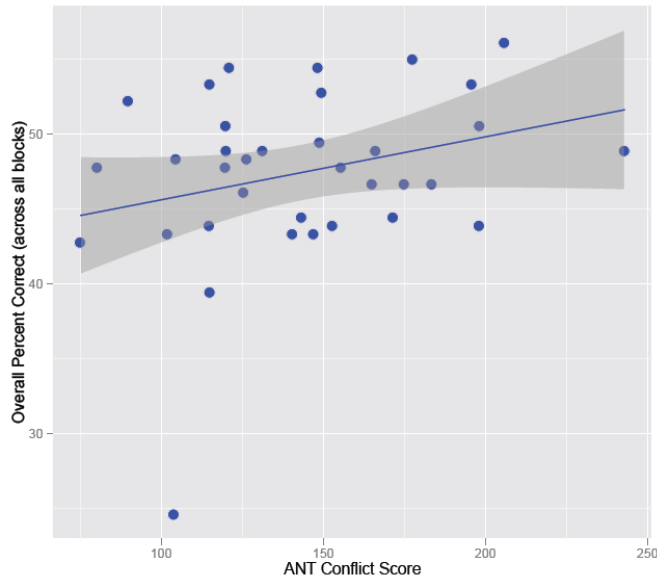


Figure 10. Conflict management and word learning.

## Conclusion

The present study found that English-speaking participants were not able to successfully learn the pseudo-words containing long and short Arabic vowels, either due to mapping Arabic vowels to the closest phonetic category in English or due to the large memory load associated with learning to associate the L2 pseudo-words to novel images. Although there are possible mappings, such as to English stressed vowels, no participants appear to have made this connection. Furthermore, as none of the participants successfully learned the words, there was also no relationship found between working memory and word learning. What this also suggests is that working memory has limited effectiveness when important linguistic elements of the L2 are lost at the level of perception. For future research, it would be useful to present the minimal pairs containing the long and short vowels side by side to participants to directly test whether they were able to differentiate the contrast. In addition, providing feedback could be an important way to facilitate learning. Not only would it be useful in promoting the cognitive aspect of

learning, feedback could also help keep participants alert and motivated. Most studies on error correction in L2 classes have provided evidence that students who receive error feedback improve in accuracy (Swain, & Lapkin, 1995; Ellis, Loewen, & Erlam, 2006; Bitchener, 2008). In addition, a longer training method could be used to expose participants to the target stimuli for longer periods. For example, participants could be asked to perform the task several times a week over a month where a pre-test and a post-test could be conducted to assess their improvement over longer training periods because intensive training can improve learners' performance. An independent test to assess perception outside of the learning task would be useful for localizing the source of performance problems. It would allow the researcher to determine whether problems in learning are due to a problem perceiving the pseudo-words or whether it is learning the link between the pseudo-words and their respective images. Finally, research on joint attention (e.g., Tomasello & Farrar, 1986) suggests that word learning in infants and children is facilitated when the learner is engaged with another. Mutual gaze towards an object enhances the learning experience, possibly due to the involvement of more brain regions than if the learner were simply hearing a word and seeing an object. This might be a profitable avenue to explore given the lack of success in the present experiment.

Language learning is a cognitive activity that depends on a myriad of underlying processes and strategies. The learner's L1 and the L2 interact in complex ways that are only beginning to be understood. The null results obtained in the present study should not deter researchers from working towards a greater understanding of the mechanisms required for L2 learning.

## References

- Ainsworth, W. (1972) Duration as a cue in the recognition of synthetic vowels, *Journal of the Acoustical Society of America*, 51, 648-651
- Altmann, H. (2006) The Perception and Production of Second Language Stress: A Cross-Linguistic Experimental Study. PhD dissertation, University of Delaware.
- Altmann, H., & Vogel, I. (2002). L2 Acquisition of Stress: the role of L1. Paper presented at the DGfS Annual Meeting “Multilingualism Today” in Mannheim, Germany, March 2002.
- Ashby, M., & Maidment, J. (2005). Introducing phonetic science. New York: Cambridge University Press.
- Alotaibi, Y., & Hussain, A. (2009). Formant based analysis of spoken Arabic vowels. *Lecture-notes Computer Science*, 5707, 162-169.
- Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese/r/and English/l/and/r/. *Journal of Phonetics*, 32(2), 233-250.
- Baddeley, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 302(1110), 311-324.
- Baddeley, A., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, 27(5), 586-595.
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105(1), 158.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science*, 4(11), 417-423.

- Best, C. McRoberts, G. & Sithole, N. (1988). The phonological basis of perceptual loss for non-native contrast: Maintenance of discrimination among Zulu clicks by English speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 345-360.
- Best, C. (1994a). The emergence of native-language phonological influence in infants: A perceptual assimilation model. In Goodman J.C. & Nusbaum, H,C (eds). *The development of Speech Perception: The transition from Speech Sound to Spoken Words*, 167-224.
- Best, C. (1994b). Learning to perceive the sound pattern of English. *Advances in Infancy Research*, 8, 271-304.
- Best, C. T. (1995). A direct realist view of cross-language speech perception: New directions in research and theory". In Winifred Strange. *Speech perception and linguistic experience: Theoretical and methodological issues*. Baltimore: York Press, 171-204.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17(2), 102-118.
- Boersma, P. & Weenink, D. (2005). Praat: doing phonetics by computer (Version 4.3.01) [Computer program]. Retrieved from <http://www.praat.org/>
- Bohn, O., & Flege, J. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second language Acquisition*, 14, 131-158.
- Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America*, 106, 2074- 2085.
- Broselow, E. Chen, I. & Huffman,M. (1997). Syllable weight: Convergence of phonology and



- phonetics. *Phonology*, 14(1), 47-82.
- Carr, P., & Durand, J. (2004). General American and New York City English. *La Tribune Internationale des Langues Vivantes*, 36, 56-69.
- Cebrian, J. (2002). Phonetic similarity, syllabification and phonotactic constraints in the acquisition of a second language contrast. *Toronto Working Papers in Linguistics*.
- Clark, E. V. (2002). First language acquisition. Cambridge: Cambridge University Press
- Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, 7(12), 547-552.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, 19(4), 450-466.
- Eid, M. Broselow, E. & MacCarthy, J. (1992). Perspectives on Arabic linguistics 6. Amsterdam, Philadelphia: John Benjamin.
- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in second language acquisition*, 28(02), 339-368.
- Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *The Journal of Educational Research*.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1), 19-23.
- Fan, J., McCandliss, B., Sommer, T., Raz, A., & Posner, M. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, 14(3), 340-347.
- Flege, J. E. (1980). Phonetic approximation in second language acquisition. *Language Learning*, 30, 117-134.

- Flege, J. (1981). The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly*, 15, 443-455.
- Flege, J., & Davidian, R. (1984). Transfer and developmental processes in adult foreign language speech production. *Applied Psycholinguistics*, 4, 323-347.
- Flege, J., & Hillenbrand, J. (1986). Differential use of temporal cues to the /s/- /z/ contrast by native and non-native speakers of English. *Acoustical Society of America*, 79(2), 508 – 517.
- Flege, J. (1987). The production of new and similar phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47-65.
- Flege, J. (1995). Second language speech learning: Theory, findings and problems. In Winifred Strange (ed). *Speech perception and linguistic experience: Theoretical and methodological issues*. Baltimore: York Press. 233–277.
- Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication*, 16(1), 1-26.
- Flege, J. Bohn, O., & Jang, S. (1997). The effect of experience on non-native speakers' perception and production of English vowels. *Journal of Phonetics*, 24,437-470
- Flege, J. Mackay, I., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Acoustical Society of America*, 106(5), 2973-2986.
- Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, 28(2), 200-213.

- Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, 29(3), 336-360.
- Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, 23(1) 83-94.
- Gathercole, S. E., Hitch, G. J., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology*, 33(6), 966.
- Gussenhoven, C., & Jacobs, H. (2005). Understanding phonology. Routledge: Hodder Education.
- Hillenbrand, JM. Clark, MJ., & Houde, RA. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America*, 108, 3013–22.
- Hayes, B. (1995). Metrical stress theory: Principles and case studies. Chicago: The University of Chicago Press.
- Hayes, B. (2009). Introductory phonology. Malden, MA: Wiley-Blackwell.
- Jensen, J. (1993). English phonology. Philadelphia: John Benjamin.
- Kenstowicz, M. (1970). On the notation of vowel length in Lithuanian. *Papers in Linguistics*, 3, 73-110
- Kingston, J. (2003). Learning foreign vowels. *Language and Speech*, 46(2-3), 295-349
- Kondaurova, M., & Francis, A. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *Journal of the Acoustical Society of America*, 124(6), 3959-3971.
- Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, and reality decisions. *Journal of Verbal Learning and Verbal Behavior*, 23(1),39-66

- Kuhl, P. K. (1991). Human adults and human infants show a perceptual magnet effect and the prototype of speech categories monkeys do not. *Perception and Psychophysics*, 50, 93-107.
- Kuhl, P., & Iverson, P. (1995). Linguistics experience and the perceptual magnet effect. In Winifred Strange (ed.). *Speech perception and linguistic experience: Issues in cross-language research*, 121-154. Baltimore: York Press.
- Munro, M. (1993). Production of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, 36(1), 39-66.
- Muller, E. (2012). *Standard vowels systems of English, German and Dutch: Variation in norm*. Oxford: Peter Lang.
- Odlin, T. (1989). *Language transfer*. Cambridge: Cambridge University Press.
- Otake, T., & Cutler, A. (1999). Perception of supersegmental structure in non-native dialect. *Journal of Phonetics*, 27, 229-253
- Pashler, H. E. (1998). *The psychology of attention* (Vol. 15). Cambridge, MA: MIT press.
- Pisoni, D. B., Aslin, R. N., Percy, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 297-314.
- Plag, I. Kunter, G., & Schramm, M. (2011). Acoustic correlates of primary and secondary stress in north American English. *Journal of Phonetics*, 39, 362-374.
- Posner, M., & Rothbart, M. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*, 58, 1-23.
- Saadah, I. (2011). *The production of Arabic vowels by English L2 learners and heritage speakers of Arabic. Ph.D dissertation*, University of Illinois, 1-160.

- Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology*, 45(1), 21-50.
- Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied linguistics*, 16(3), 371-391.
- Stevens, K. N. (1959). The role of duration in vowel identification. *Quarterly progress report*, 52, laboratory of electronics, MIT.
- Tiffany, W. (1953). Vowel recognition as a function of duration, frequency modulation, and phonetic context. *Speech hear. dis.*, 18, 289-301
- Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 57(6), 1454–1463. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/3802971>
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127-154.
- Wang, Y., Spence, M. Jongman, A., & Sereno, J. (1999). Training American listeners to perceive Mandarin tone. *Acoustical Society of America*, 106(6), 3649-3658.
- Zahorian, S., & Jagharghi, A. (1993). Spectra-shape features verses formants as acoustic correlates for vowels. *Acoustic Society of America*, 94, 1966-1982.

Appendix A. Stimuli pictures



### Appendix B. List of pseudo-words

Short condition	Long condition
rusiba	rusi:ba
rasibu	rasu:bu
kutima	kuti:ma
katimu	katimu
zuhida	zuhi:da
zahidu	zahi:du
risuba	risu:ba
rasubi	rasu:bi
kituma	kitu:ma
katumi	katu:mi
zihuda	zihu:da
zahudi	zahu:di
risabu	kisa:bu
rusabi	kusa:bi
kitamu	kita:mu
kutami	kuta:mi
zihadu	ziha:du
zuhadi	zuha:di

**Appendix C. Tables for Mean scores**

*Table 1. Means for correct answers across blocks*

	Mean	Median	SD
B1	45.57	50	22.17
B2	46.51	50	22.89
B3	49.05	50	22.89
B4	48.73	50	21.59
B5	47.46	50	21.78

*Note: 'B' = 'Block', 'SD' = standard deviation*

*Table 2. Means for correct answers across vowel type*

	Mean	Median	SD
V1	46.28	50	21.41
V2	47.09	50	21.44
V3	49.00	50	23.17

*Note: V1= vowel /i/, V2 = vowel /u/, V3 = vowel /a/*

*Table 3. Means for correct answers across conditions*

	Mean	Median	SD
Short V	48.88	50	2.23
Long V	46.04	50	1.76

*Note: V = vowel*