August 2016

The L2 Perceptual Mapping of Arabic and English Consonants By American English Learners

Zafer Lababidi
University of Wisconsin-Milwaukee

THE L2 PERCEPTUAL MAPPING OF ARABIC AND ENGLISH CONSONANTS BY AMERICAN ENGLISH LEARNERS

by

Zafer Lababidi

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy
in Linguistics

at
The University of Wisconsin-Milwaukee
August 2016

ABSTRACT

THE L2 PERCEPTUAL MAPPING OF ARABIC AND ENGLISH CONSONANTS BY AMERICAN ENGLISH LEARNERS

by

Zafer Lababidi

The University of Wisconsin-Milwaukee, 2016
Under the Supervision of Professor Hanyong Park

There has been rapid growth in Arabic learning in the United States. With that increase, many learners of Arabic often experience difficulties in learning some Arabic sounds. Among these are the Arabic plain sounds /t, d, ð, s/ and emphatic sounds /tˤ, dˤ, ðˤ, sˤ/. Many studies have proposed that these difficulties are related to the relationship between sounds in learners' first language (L1) and those in the target language (L2) (Flege, 1987 and 1995; Best, 1995 and 1999; Best, McRoberts, & Sithole, 1988; Best & Tyler, 2007). Previous studies have examined the perceptual patterns of these sounds by relying solely on the articulatory differences between them and the English categories (Al-Mahmoud, 2013).

This study first investigates the categorical representations of the plain and emphatic Arabic sounds in the minds of monolingual native speakers of American English in order to establish accurate category mappings between the two languages following Guion *et al.* (2000). Sixteen listeners participated in an identification task and a goodness-of-fit rating task. Following the sound categorization of the Speech Learning Model (Flege, 1995), the results show that the Arabic consonants /t, d, ð, s, sˤ/ are considered "similar" sounds to the English categories, Arabic /tˤ, ðˤ, dˤ/ are considered "new" sounds.

The study also examined whether the L1-to-L2 mappings found earlier differ between naïve monolingual English listeners and L2 Arabic learners. In addition, the study investigates the perceptual development of the Arabic sounds over time with more L2 exposure. Fifty L2 Arabic learners participated in an L1 labeling task, an L2 labeling task, and goodness-of-fit rating tasks. The results show L2 perceptual development for the emphatic sounds but not for the plain sounds. The results show no difference between L1-to-L2 mappings between the naïve monolinguals and experienced L2 learners at the labeling level. However, the results show differences at the goodness-rating level, suggesting subtle L2 perceptual development.

The study also investigates the degree of reliance on L1 in order to predict the accuracy of L2 identification by following Park and de Jong's (2008) quantitative analysis. The results show that it is not clear to what extent L2 learners are using and facilitating their L1 categories in order to perceive L2 sounds. However, the observed accuracy results are successful at showing how L2 exposure affects the overall learnability of L2 emphatic sounds. The study concludes that the perceptual developmental pattern of the emphatic sounds matches the description of SLM's "new" categories, while the pattern of the plain sounds matches the description of the "similar" sounds.

To

my parents,

my wife,

and our growing family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

بسم الله الرحمن الرحيم
"رب أوزعني أن أشكر نعمتك التي أنعمت علي و على والدي وأن أعمل صالحا ترضاه"
النمل (19)

"My Lord, enable me to be grateful for your favor which you have bestowed upon me and upon my parents and to do righteousness of which you approve."

I would like to express my deepest gratitude to a number of people whom have helped me throughout my academic journey in the past years. First and foremost, I am greatly indebted to Professor Hanyong Park, my major advisor and mentor. Professor Park did not only spark my interest in phonetics, but he also inspired me to think critically about everything in life and not to take anything for granted. He was very patient with me and was always ready to answer all my questions and late night emails. I am very thankful for the countless meetings we had together to discuss my projects, papers and presentations. I will always think of him as the person who shaped my research skills and taught me how to be an objective researcher.

I would also like to extend a sincere appreciation to Professor Hamid Ouali. Professor Ouali was a friend and a brother on top of being an excellent source of knowledge throughout my academic life. From my first day in Milwaukee, Professor Ouali helped me settle down and eased my transition into the academic life at UWM. His encouraging words and valuable advice will always echo in my mind.

I am also very thankful to my committee members Professors Anne Pycha and Jae Yung Song. My conversations with them were very helpful and I definitely enjoyed learning from their expertise in the field. I thank Professor Pycha for providing me with the opportunity to use her lab and get a hands-on experience with the equipment that she has.

I would like to also thank all the faculty members in the linguistics department for their encouragement and for giving me the best training that I could ask for. In addition, I would like to thank the department of Linguistics and the department of Foreign Languages and Literatures for their endless support.

My gratitude also goes to all the friends and colleagues whom I met in Milwaukee and made my journey a memorable and an unforgettable one. My thanks goes to every one of them and all their families for listening to me and sharing my happiest moments with me. You were all very supportive and I am truly honored to have met you all.

Above all, I am greatly thankful to my parents and siblings for their unconditional love and support. Your emotional support and prayers kept me going despite the long distance between us. Thank you for being very understanding and for believing in me and for the endless love that you showed me during this phase in my life.

Finally, no words can describe or express the amount of gratitude that I have in my heart for my lovely wife, Yasmin Soudani. Thank you for the countless times you heard me complain about random stuff due to my frustration phases which were more than I can count or remember. Thank you for proofreading my documents and papers and I apologize for the headaches that I gave you because of the linguistic jargon. Thank you for sticking by my side during this long journey and for encouraging me and pushing me towards my goals. Thank you for taking care of our little son while I had to stay up late and work on my dissertation. You are truly an amazing wife, friend, and mother and I am blessed for having you in my life. I love you and baby soos and I ask god to reward you for all your good deeds. As for my baby son, thank you for staying away from my laptop, for the majority of the time, and not ruining my work. I love you and I cannot wait to hear your first words.

Chapter 1 Introduction

1.1 Purpose and Significance

Since 2001, there has been a steady increase in Arabic language learning in the United States. Furman, Goldberg, and Lusin (2010) report a 46.3% growth in the fall of 2009, based on the findings of the Modern Language Association of America. This rapid increase has illuminated the difficulties which learners of Arabic often experience in learning some Arabic sounds, which, in turn, constitutes one of the obstacles that stands in the way of achieving high second language (L2) proficiency. The Arabic sound system is particularly difficult for L2 learners because it contains sounds with small phonemic differences which do not exist in the English sound system. Nevertheless, being able to fully understand and acquire these fine distinctions is crucial for learners because it allows them to differentiate minimal pairs in Arabic.

Well-cited models in the field of L2 phonology, such as Flege's Speech Learning Model (SLM) (Flege, 1987 and 1995), and the Perceptual Assimilation Model (PAM) and the PAM-L2 by Best and colleagues (Best, 1995 and 1999; Best, McRoberts, & Sithole, 1988; Best & Tyler, 2007), propose that such difficulties are related to the relationship between sounds in a learner's first language (L1) and those in the target language. It has long been established that L2 speakers actually perceive non-native sounds under the influence of the L1 sound system. Polivanov (1931) and Trubetzkoy (1939) were pioneers in pointing out that the native language phonological system of an L2 learner acts as a filter through which L2 sounds are perceived and produced. This filtration process was then believed to consist simply of learners making a contrastive comparison between L1 and L2 sounds, in order to understand the differences between them. Lado's (1957) Contrastive Analysis Hypothesis was the main tenet behind that proposal. However, as time progressed, researchers found this simple contrastive comparison between L1 and L2 sounds to be insufficient, due to its lack of explanation as to why specific differences between L1 and L2 sounds are easy to learn and why specific similarities between L1 and L2 sounds are difficult to learn (Eckman, 1987; Gass & Selinker, 2001; Towel & Hawkins, 1994). Consequently, researchers proposed learning models which attempted to reveal and predict areas of difficulty in learning L2 sounds arising from perceptual similarities between L1 and L2 sounds. The degree of similarity between sounds in L1 and L2, therefore, seems to affect the overall success of learning L2 categories (Flege, 1987 and 1995; Best, 1995; Best *et al.*, 1988).

In addition to attempting to understand how an L1 sound system may affect L2 perception, researchers have attempted to understand and classify the developmental stages through which L2 learners progress. The studies have shown that the L2 learner's ability to successfully perceive phonetic information also relies on his or her L2 experience. In fact, the level of experience with L2 is one crucial element that influences the stages of perceptual development in addition to the effects of the L1 sound system. This L2 experience has been widely linked to the amount of exposure an L2 learner receives throughout his or her L2 learning journey (Flege, 1984; MacKain, Best, & Strange, 1981; Best & Strange, 1992). Previous studies have found that the amount of exposure plays a dynamic role in the developmental stages of L2 learners. For example, Flege, Bohn, and Jang (1997) assess how varying experience with English affects non-native speakers' perceptions. German, Spanish, Mandarin, and Korean speakers, with varying English exposure backgrounds, participated in an identification task of English synthetic vowels /i, ɪ, æ, ɛ/. The participants were placed in different groups, depending on each individual's level of exposure to English. The results indicate that the more experienced participants (more L2 exposure) perceive English vowels more accurately than the less experienced participants (with less L2 exposure). Similarly, Flege and Liu (2001) make comparisons between groups of Chinese speakers with different exposure levels to L2 English. The researchers administered an identification task of word-final English consonants, a grammaticality judgment task, and a listening comprehension task. The participants were divided into four different groups, depending on each participant's amount of L2 exposure. According to the results, participants with more exposure to L2 English scored higher in all three tasks than those with less exposure. These findings align with what previous studies have reported on the effect of L2 exposure, or input, in determining the developmental stages of L2 learners.

A number of models were used extensively, in order to analyze results obtained from perception tasks from an array of different languages (e.g. L1-Japanese learners of L2-English in Guion *et al.*, 2000; L1-Korean learners of L2-English in Park & de Jong, 2008; L1-English learners of L2-Arabic in Al Mahmoud, 2013; and L1-German, L1-Spanish, L1-Mandarin, and L1-Korean learners of L2-English in Flege, Bohn, & Jang, 1997). However, there seems to be a lack of research examining the perceptual similarities between Arabic and English sounds and the overall perception development of Arabic L2 sounds. Studies that have attempted to establish a basic scheme to the matter adopt a model that is solely based on descriptive realizations of sounds.

1.2 Aim of the Study

Based on what has been mentioned in Section 1.1 above, I aim to establish perceptual category mapping between Arabic and English consonants in the minds of native speakers of

American English, with no prior experience with Arabic. More specifically, I seek to create a mental perceptual representation between the Arabic plain consonants /t, d, ð, s/ and Arabic emphatic consonants /tˤ, dˤ, ðˤ, sˤ/ and the English categories, in order to understand how these non-native Arabic sounds are represented in the minds of naïve American English monolinguals. In addition, I aim to examine the relationship between the L1 and L2 sound systems and explore to what degree L1 sound system interference affects the perception of L2 Arabic plain and emphatic consonants. In addition, this study examines the perceptual category mapping of the aforementioned L2 Arabic plain and emphatic consonants but with L2 Arabic learners who possess varying degrees of L2 Arabic experience. In doing so, this study investigates whether L2 experience influences or affects the perceptual category mapping of naïve English monolinguals, who have no prior experience with Arabic.

The study will also investigate the perceptual development of the Arabic plain and emphatic consonants in L2 Arabic learners of varying L2 proficiency levels. In other words, I will examine if and how L2 Arabic experience affects the overall perceptual development of L2 Arabic plain and emphatic consonants. Finally, using the results obtained, the study seeks to shed light on the current perceptual models and examine the findings, in light of their predictions and conclusions. More specifically, I will examine the results in the context of the Speech Learning Model (SLM) — my primary model, and indirectly with the Perceptual Assimilation Model (PAM).

This study aims to answer to the following research questions:

I. *What are the perceptual category mappings between Arabic plain and emphatic consonants with English consonants in the minds of native speakers of American English?*

II. *Does more L2 exposure over time lead to similar learning development with Arabic L2 plain and emphatic sounds?*

III. *Do L2 learners' L1-to-L2 mappings change over time?*

IV. *How can the current perceptual models be evaluated based on the results?*

Answering these research questions will help scholars and educators better understand the degree to which Arabic is perceived by non-native listeners — whether naïve monolinguals or L2 language learners with varying degrees of L2 Arabic proficiency.

1.3 Speech Production and Perception

In this section, I will discuss speech production and perception theories that are directly related and applicable to this study. I will present the SLM that was proposed by Flege (1995) as the speech production model and will explain how Park and de Jong (2008) extend the SLM, in order to make it account for speech perception as well. I will also present the PAM that was proposed by Best and colleagues (Best, McRoberts, & Goodell, 2001; Best et al., 1988; Best, 1995 and 1999) and the Perceptual Assimilation Model-L2 (PAM-L2) that was proposed by Best and Tyler (2007).

1.3.1 Speech Learning Model (SLM)

Flege's (1987, 1992, and 1995) SLM makes a unique prediction regarding the learning paths of two types of L2 sounds: "similar" and "new" sounds. Under this model, Flege explains that some L2 sounds will be "similar" to learners' native categories, while others will be "new" to their L1 sounds. The SLM predicts that learners will demonstrate better abilities at learning the so-called “similar” L2 sounds than the “new” ones at the earlier stage of learning. However, as the acquisition process continues, the “new,” sounds will be acquired with more accuracy, surpassing the accuracy of those “similar” sounds. The reasoning behind this advancement in the acquisition process is explained by SLM as follows: In the early acquisition stages, L2 learners will, in fact, rely on their L1 categories and recycle them for the “similar” L2 sounds to facilitate L2 learning. This practice is due to the interference from L1 categories. As a result, there will be no learning of these sounds, even with more L2 experience; learners simply tend not to put any effort into learning what they already think they know. However, as L2 experience increases, L2 learners must pay more attention to the sound details in order to learn the “new” L2 sounds. As a result, with more L2 experience, learners will be able to acquire these sounds better than the "similar" sounds. Importantly, this model considers the phonetic similarities and dissimilarities between L1 and L2 sounds as the main source of category formation or learning of L2 sounds. The SLM hypothesizes that, if L2 learners are able to detect dissimilarities between L1 and L2 sounds, they will be able to form L2 categories. In return, this formation of L2 categories (i.e., the so-called “new” L2 sounds) will allow L2 learners to suppress the interference of L1 categories and will enable L2 learners to successfully perceive the target L2 sounds in a native-like manner.

Flege’s model is supported by many different studies, examining either L2 vowels or consonants. Flege (1987), for example, demonstrated that native English speakers who are L2 learners of French were able to produce French /y/ in an authentic manner after 12 years of being residents of Paris. On the other hand, the same French learners were not able to produce the French /u/ in an authentic manner; the French learners’ /u/ production was different from the /u/

production of native speakers of French. Flege explained that the French /y/ does not have a phonological counterpart in the English sound system, while the French /u/ is similar to the English /u/, but with some differences. According to the SLM, not having a counterpart for the French /y/ in English allowed French learners to establish a new phonological category for that sound. On the other hand, French learners assumed that the French /u/ is similar to their English /u/ and consequently did not put much effort into trying to learn that sound. The result was their inability to produce the sound accurately, even after 12 years of residing in France.

With regards to consonants, Flege and Hillenbrand (1984) showed that native French speakers who are L2 learners of English were not able to produce the English /t/ in an authentic manner, like native speakers of English. They explained that French speakers consider English /t/ as a “similar” sound to their French /t/. However, French, unlike English, is a language that has short-lag /p t k/ in its system. This means that native French speakers will produce English /t/ with a short-lag. Flege and Hillenbrand explain that L2 learners who are native speakers of such languages will have a tendency to produce English /p t k/ with short-lag VOT values.

Based on these findings, Flege developed what is now known as the SLM — a model which aims to provide an explanation for the ability to produce L2 vowels and consonants in a native-like manner. The model proposed four postulates, which are as follows:

1. The mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span and can be applied to L2 learning.
2. Language-specific aspects of speech sounds are specified in long-term memory representations called *phonetic categories*.

3. Phonetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 or L2 phonemes identified as a realization of each category.
4. Bilinguals strive to maintain contrast between L1 and L2 phonetic categories, which exist in a common phonological space. (Flege, 1995, p. 239).

These four postulates are followed by seven hypotheses (see Flege, 1995, p. 239). The following hypotheses pertain to this study: First, Flege hypothesizes that, "Sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level." Second, he hypothesizes that, "A new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds." Finally, he hypothesizes that, "The greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned."

It is obvious that the SLM emphasizes a certain level of continuity between the practices used for learning L1 speech and the practices used for learning L2 speech; however, the SLM stresses that, in the case of L2 speech learning, this continuity is contingent upon L2 learners' language experience. "As L2 learners gain experience in the L2," Flege contends, "they may gradually discern the phonetic differences between certain L2 sounds and the closest L1 sound(s)" (Flege, 1995: p. 263).

Since the earlier version of the SLM was primarily interested in L2 speech production, and it relies on somewhat crude and subjective criteria to determine which L2 categories are considered "new" and which ones are considered "similar," researchers have attempted to address such drawbacks. Park and de Jong (2008) extends the SLM to make it account for L2 speech perception, using a quantitative approach rather than subjective criteria. In the next chapter I present a detailed explanation of Park and de Jong (2008).

1.3.2 Perceptual Assimilation Model (PAM/PAM-L2)

As another well-cited model, the PAM by Best and colleagues (Best, McRoberts, & Goodell, 2001; Best et al., 1988; Best, 1995 and 1999) is a cross-language perception model, rather than an L2 model. This is because the main focus of the PAM is naïve monolinguals' perceptual patterns of non-native sounds, rather than those of L2 learners. Nevertheless, the PAM is widely cited among L2 researchers, who apply the basic idea of the model to the perception of L2 sounds among learners. In this model, the human capacity to discriminate non-native sounds emanates from the assimilation patterns of non-native segments to categories in the L1 phonological system. The main principle of the model is that perceptual limitations of L2 learners determine what kind of difficulties the L2 learners may face when learning L2 sounds (Almbark, 2012). The PAM is very appealing to many researchers because it proposes that the process of listening to L2 sounds is not limited to the act of deciding which L2 sounds are similar to the L1 sounds and which are not similar (Pilus, 2005). Listening to an L2 sound, according to the PAM, is a complex process which requires discriminating between two L2 sounds, as well as distinguishing L2 sounds from L1 sounds (Pilus, 2005). So, the PAM is, at its core, a perceptual model, not a production model.

In the model, Best explains that non-native segments are segments "whose gestural elements […] do not match precisely any native constellations" (1995, p. 193). The PAM has gained notoriety for judging the similarities between non-native and native segments by articulatory and gestural characteristics. As a result, non-native phonemes are more likely to be assimilated to native ones, when they are perceived as good exemplars of their native equivalent (Al-Mahmoud, 2013). Thus, according to the PAM, non-native phonemes can be assimilated in one of the following ways:

1. Assimilated to an existing native category (sound), which could be heard either as being an identical, acceptable, or deviant exemplar of the native category.
2. Assimilated as an uncategorizable sound within the phonological space; however it is identified as having characteristics of speech, just not representative of any particular native category.
3. Not assimilated to any native category and therefore it exists outside of the native phonological space (Best, 1995, p. 195).

So, the PAM categorizes six pairwise assimilation types between the non-native and the native categories by relying on the combinations of 1–3 above, each with a specific level of discriminatory accuracy (Best, 1995, p. 195):

1. Two-Category Assimilation (TC type): two non-native sounds are assimilated to two different native categories. The discrimination of these two non-native sounds will be excellent.
2. Category Goodness Difference (CG type): two non-native sounds are assimilated to a single native category with different goodness of fits (CG: e.g., one is a good fit and the other is poor), the discrimination of these two sounds will vary from being very good to poor depending on the degree of goodness fit difference.

3. Single-Category Assimilation (SC type): two non-native sounds are assimilated to the same native category (SC); the discrimination of these two sounds will be poor.
4. Both Uncategorizable (UU type): both non-native sounds fall inside a phonetic space but outside any particular native category. The discrimination of these two sounds will vary from poor to very good depending on their proximity to a native category within a native space.
5. Uncategorized versus Categorized (UC type): One non-native sound is categorized as a native category, but the other is not. The other, however, falls in a phonetic space but outside any particular native category. The discrimination of these two sounds will be very good. Guion, Flege, Yamada, and Pruitt (2000) explain that the English /r/-/w/ distinction for Japanese listeners aligns with this assimilation type; the English /w/ is assimilated as Japanese /w/, but the English /r/ is not assimilable to any Japanese category.
6. Nonassimilable (NA type): Both non-native segments are non-speech segments. The discrimination will vary from good to very good. Best *et al.*, (1988) explains that Zulu clicks fit this assimilation type since, for the English listener, the click segments are not speech sounds.

In an attempt to extend and make the PAM account for L2 learners, Guion *et al.* (2000) tests whether the PAM predictions would be applicable to L2 learners. In their study, they examine perceptions of English and Japanese consonants in the context of Japanese categories. They had participants perform an identification task and a goodness-of-fit task (more details will be presented in Chapters 2 and 3). The obtained results were used in order to generate a "fit index," showing numerically how successful the selected L2-English categories were, in relation to native Japanese consonants. From these fit index results, the researchers categorized the selected categories into "good", "fair", or "poor" L2 categories to test the PAM's predictions according to its assimilation types. In order to account for L2 perception development, Guion *et al.* tested three Japanese listener groups of varying English proficiency. The results show that the perceived phonetic distance of L2 consonants from their closest L1 consonant successfully predicted the discrimination of L2 sounds. Additionally, the results indicate the positive role that L2 experience plays in the discrimination of sounds; experienced Japanese groups scored significantly higher than the inexperienced Japanese groups. Even though Guion *et al.* (2000) extended the original PAM to more experienced L2 learners, other researchers have concluded that their results do not help us test a "time-course of development" from cross-language mapping (Park & de Jong, 2008).

The PAM-L2 by Best and Tyler (2007) is the modified and more recent version of the PAM — proposed in order to address L2 perception patterns among language learners. The PAM-L2 is interested in two distinct groups: late L2 learners and naïve monolinguals. Examining these two groups enables the PAM-L2 to account for the amount of L2 exposure in learning non-native (L2) sounds. Similar to the original PAM, naïve monolinguals' perceptions are affected by the "similarities" and "dissimilarities" between L1 and L2 sounds. The PAM-L2 suggests that perception is not restricted to differences pertinent to native phonological contrasts; adult monolinguals demonstrate systematic perceptual sensitivities to non-contrastive phonetic variation, in both native and non-native speech. For example, phonotactic biases, coarticulatory patterns, and allophonic or other phonetic variations influence the perception of L2 contrasts among adult L2 learners. Therefore, the PAM-L2 indicates that categorization and discrimination performance levels differ across L2 contrasts and across L1s through a systematic relation to both the contrastive phonological and gradient phonetic properties of the L1s (Best & Tyler, 2007, p. 10). The model uses the same assimilation types discussed above in the PAM.

Al-Mahmoud (2013), for instance, examines L2 perception of non-native contrasts and tests the PAM's prediction by examining the ability of American learners to discriminate Arabic contrasts. Twenty-two L1-American English speakers were asked to complete a forced choice AXB discrimination task. The results provided partial evidence for the PAM. Two-category (TC) contrast results followed the PAM's prediction. However, the results of category-goodness difference (CG) and uncategorizable contrasts (UU) demonstrated partial support of the PAM, and the results of uncategorized-categorized contrast (UC) discrimination provided counter-evidence to the PAM. In his study, the author relies on descriptive phonetic realizations of the tested phonemes in order to compose the contrasts to be tested. In other words, the author depends on articulatory and acoustic measurements of Arabic and English consonants that were provided by other research studies to determine L2 Arabic contrasts and their assimilation to English. These criteria included the place and manner of articulation of the Arabic and English consonants, burst intensity and duration, and first and second formant frequencies. By adopting those criteria, however, the study still leaves uncertainty about the standards that have been used to govern the relationships between the tested consonant segments. Relying solely on descriptive and acoustic similarities between L1 and L2 sounds may not be sufficient for cross-language comparisons. For example, the perceiver's linguistic background, in addition to a variety of different factors other than the acoustic similarity(s), might influence the perceived similarity of sounds drawn from two languages (Bohn & Flege, 1997).

As indicated earlier, since the PAM claims that different assimilation types are determined based on *gestural* similarity or dissimilarity under articulatory phonology (Browman & Goldstein, 1986), and this criterion is used due to Best and colleagues' theoretical position (i.e., direct realist [ecological] position), some researchers (e.g. Guion et al., 2000) proposed to determine assimilation types by relying on quantitative approaches to be theory-objective. I will discuss how that was done in the next two chapters.

Finally, Table 1.1 below is taken from Almbark (2012) and presents a comparison between SLM and PAM, in addition to another model which is not applicable to this study. The table shows that SLM is primarily interested in advanced L2 learners, whereas the PAM is interested in naïve listeners.

Table 1.1. A Comparison between L2 Speech Perception and Production Models from Almbark (2012).

	L2 Model		
	NLM	PAM	SLM
The kind of L2 learners	L1 cognitive development (with implications for L2 learning)	L2 learners with little or no L2 experience	advanced L2 learners or bilinguals
Objects of speech	phonetic/acoustic properties of native language prototypes	actual articulatory gestures	phonetic/acoustic properties of speech signal
Level of perception	phonetic perception of L1 instances leading to phonological representations of prototypes	phonetic perception of L2 by naive listeners, and phonological perception of L2 by naturalistic L2 learners.	phonological perception of L2 by naturalistic L2 learners.
Perception-production link	perceptual model	perceptual model	links perception and production

The table also demonstrates that the SLM differs from the PAM, in the type of segment characteristics that they are interested in. The SLM, on the one hand, is interested in the phonetic/acoustic properties of the speech signal. On the other hand, the PAM is primarily interested in the actual articulatory gestures as I explained earlier. It is important to note that the PAM-L2, which is not listed in the figure above, unlike the PAM, is interested in naïve listeners and advanced L2 listeners since it extends the original PAM and attempts to account for L2 development.

1.4 Background on Arabic Vowels and Consonants

In this section of the chapter I will present a general outline of the Arabic sound system including the Arabic vowels and Arabic plain and emphatic consonants.

1.4.1 Arabic Vowels

Modern Standard Arabic (MSA) is characterized as being a quantity language that distinguishes between short and long vowels. The vowel system of Arabic consists of a 6-vowel system that has three short vowels (/a/, /u/, /i/) and three corresponding long vowels (/a:/, /u:/, /i:/); vowel length is contrastive and phonemic (Ryding, 2005; Most, Levin, & Sarsour, 2008). Native speakers of Arabic are native speakers of their own regional dialects, characterized by their morphology, phonology, syntax, and lexical selection. For example, some regional variations are described as having more than three distinct vowels, like Gulf Arabic (spoken in the Arabian Peninsula), Egyptian Arabic, and Levantine Arabic (Syrian, Jordanian, Palestinian, and Lebanese) (Lababidi & Park, 2016). As a result, even when dealing with MSA, the aforementioned study contends that it is, “difficult to tease apart MSA and examine it, without making reference to the influence of the dialect of the native speaker” (Lababidi & Park, 2016, p. 68). In fact, some researchers believe that MSA “refers to the Standard Arabic which shows the norms of the speaker’s dialect” (Watson, 2002, p. 8).

The vowel system of MSA has been characterized as a three-way quality vowel system that consists of the following vowels (Newman & Verhoeven, 2002):

1. The high front vowel /i/ or /i:/
2. The high back vowel /u/ or /u:/
3. The low vowel /a/ or /a:/

Researchers have argued that the three vowel qualities of Arabic are the most common three qualities (Newman, 2002).

Research has also shown that Arabic vowels have pharyngealized allophones when they are adjacent to one of the four pharyngealized consonants /$t^ʕ$, $d^ʕ$, $ð^ʕ$, $s^ʕ$/ (Almbark, 2012). When they do occur, these pharyngealized vowels are characterized by being more retracted than their non-pharyngealized cognates — a mechanism that is associated with the production of Arabic emphatics (Saadah, 2011; Al-Ani & El-Dalee, 1984; Khattab et al., 2006). When this happens, pharyngealized vowels show a significant downward shift of the onset of their F2 value and a slight upward shift of the onset of their F1 value (Al Masri & Jongman 2003; Khattab *et al*., 2006). In articulatory terms, Almbark (2012, p. 65) explains that moving the tongue from the retracted position into the vowel position will take time, and this is why the quality and the quantity of the following vowel is argued to determine the degree of the effect of the pharyngealized consonant.

Results from recent research have shown that short and long Arabic vowels have similar F1 and F2 frequencies and that they only differ in duration (Abdel Salam, et al., 2011). For instance, Table 1.2 (below) is taken from Al-Ani (1970) and illustrates that each Arabic vowel pair has similar F1 and F2 frequencies. Other researchers, however, do not believe that the short vowels have the same quality as their long counterparts (Lababidi and Park, 2016). Watson (2002) describes the articulation of /i:/ and /u:/ as being closer than that of their short cognate. In addition, she believes that the articulation of /a:/ is fronter than its short cognate.

Table 1.3. Durations and F1 and F2 Values of MSA Arabic Vowels in Isolation Al-Ani (1970).

Vowel	*Duration*	F_1	F_2	F_3
i	300	290	2200	2700
ii	600	285	2200	2700
u	300	290	800	2150
uu	600	285	775	2050
a	300	600	1500	2100
aa	600	675	1200	2150

In terms of vowel quantity, many researchers have argued that the long vowels are approximately twice as long as the short vowels (Khattab & Al-Tamimi, 2008; Saadah, 2011). Table 1.3 (below) is taken from Lababidi and Park (2016). The figure shows comparisons of short and long vowel durations for four native speakers of Jordanian Arabic, who produced MSA vowels. One can see that the vowel durations of the long vowels are approximately twice as long as vowel durations in the short vowels.

Table 1.5. Vowel Durations (top row) for Four Native Speakers of Arabic from Lababidi and Park (2016).

Groups	Mean normalized vowel duration					
	a	a:	u	u:	i	i:
Control (N = 4)	.271 (.063)	.506 (.094)	.249 (.018)	.443 (.122)	.263 (.029)	.447 (.113)

Researchers have identified a number of linguistic factors that may affect vowel duration in Arabic, such as stress, focus, and voicing of the preceding and following consonants. Another of the factors that was found to have an effect on vowel duration is speech rate. Almbarak (2012) explains that Allatif (2008) tested the effects of speech rate and sentence type (declarative and interrogative) on Syrian Arabic vowel durations. He reports in his findings that rapid speech rate reduced the duration of sort vowels by 20% and long vowels by 19%. He also found that vowels were longer in the interrogatives than in the declaratives.

1.4.2 Arabic Consonants

MSA is characterized by having a set of complex sounds that do not exist in English. Table 1.4 (below) shows the Arabic consonant phonemic inventory for MSA. Among these sounds, this study is interested in the coronal plain obstruents /t, d, ð, s/ and emphatics /$t^ʕ$, $d^ʕ$, $ð^ʕ$, $s^ʕ$/, which are the counterparts of the coronal plain obstruents. The reason for choosing these Arabic consonants for this study is related to general classroom observations and feedback from students and L2 instructors. The coronal plain obstruents and emphatics are often misperceived, since the fine acoustic details that sets them apart are nonexistent in English. The first impression that these sounds give us is that they are not distinguished from each other (e.g., /t/ and /$t^ʕ$/).

From an historical point of view, has Arabic interested many grammarians, who attempted to present a detailed description of its sound system. Bin-Muqbil (2006) explains that the most famous Arabic grammarian Sibawayh (d. 796 A.D.) indicated that the Arabic emphatics are similar in articulation to their plain counterparts, except in the emphatics, the tongue covers the area extending from the main lace of articulation to parts of the palate opposite the tongue

(*Kitab Sibawayh*, vol. 2, p. 406). Bin-Muqbil (2006, p. 31) also explains that another Arab scholar, Ibn Sina (d. 1037 A.D.) elaborated on the emphatics by indicating that they are articulated with “a depressed tongue surface behind the main articulation.”

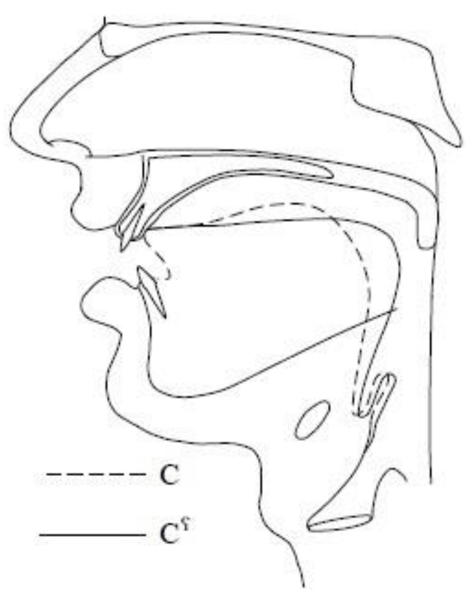
More recent studies have shown that Arabic emphatics /tˤ, dˤ, ðˤ, sˤ/ are produced with the tongue retracted back, resulting in a secondary place of articulation (i.e., pharyngealization), which is not used in producing the plain /t, d, ð, s/ (Al-Ani, 1970). Al-Ani’s x-ray imaging demonstrates how the tongue body is pulled towards the upper back of the mouth during the production of emphatics. Although plain coronal obstruents also exist in English, their phonetic realizations are slightly different between these two languages. In Arabic, the voiceless /t/ and voiced /d/, for instance, are stops produced with the tongue tip touching the inner part of the front teeth (Al-Ani, 1970), whereas the English /t/ is produced with either the tongue tip or the tongue blade touching the alveolar ridge (Ladefoged & Johnson, 2011). Figure 1.1 below is taken from Bin-Muqbil (2006), and it demonstrates the above articulatory configuration based on three distinct cineflurographic studies (Al-Ani, 1970; Ali & Daniloff, 1972; and Ghazeli, 1977).

As for the differences in duration between emphatics and plain sounds, Al-Ani (1970) indicates that there are none such differences between them. However, other studies have found some durational difference patterns; Giannini and Pettorino (1982) indicate that the emphatic sounds tend to occupy a longer duration when they occur before /a/. On the other hand, they show that the plain sounds have a longer duration when they occur before /i/.

Table 1.7. Arabic Consonant Phonemic Inventory.

	Bilabial	Labiodental	Dental	Alveolar	Palato-alveolar	Palatal	Velar	Uvular	Pharyngeal	Glottal
Stops	b			t d tˤ dˤ			k	q		ʔ
Nasals	m			n						
Trills				r						
Fricatives		f	θ ð ðˤ	s z sˤ		ʃ		x ɣ	ħ ʕ	H
Affricates					dʒ					
Glides	(w)					j	w			
Liquids				l						

Figure 1.1. Articulation of an Arabic Emphatic Coronal and Its Non-Emphatic Counterpart Based On Description in Al-Ani (1970), Ali and Daniloff (1972), and Ghazeli (1977). Taken from Bin-Muqbil (2006).



Finally, with regard to the coarticulatory effects of emphatics on their adjacent vowels, studies have reported a lowered F2 and a raised F1, as I explained earlier (Al-Ani, 1970 and Ghazeli, 1977).

To sum up the differences, the four emphatic sounds /tˤ, dˤ, ðˤ, sˤ/ are characterized as being a set of coronal obstruents with a secondary articulation manner (tongue retraction). These sounds are known for lowering F2 and raising F1 in the adjacent vowels. Their plain counterparts do not have that effect on their adjacent vowels.

Chapter 2 Research Questions and General Methods

In the introductory chapter, I presented my research questions, in addition to a number of theoretical models — all of which directly pertain to my research. In this chapter, I revisit my research questions and discuss how my methods and analysis will help address these questions, in light of the presented theoretical models.

To reiterate, my proposed research questions are:

I. *What are the perceptual category mappings between Arabic plain and emphatic consonants with English consonants in the mind of native speakers of American English?*

II. *Does more L2 exposure over time lead to similar learning development with Arabic L2 plain and emphatic sounds?*

III. *Do L2 learners' L1-to-L2 mappings change over time?*

IV. *How can the current perceptual models be evaluated based on the results?*

By seeking an answer to the first research question in my study, I first attempt to examine and identify the perceptual category mappings between Arabic plain and emphatic sounds and English sounds in the minds of native speakers of American English. In other words, I aim to describe and determine the perceptual classification of these L2 Arabic sounds in the minds of native speakers of American English. Chapter 3 will present results from an experiment on how the tested Arabic sounds are mentally represented and categorized by native English speakers. This experiment follows the steps of research conducted by Guion, Flege, Akahane-Yamada, and Pruitt (2000), in their attempt to extend PAM and make it account for L2 learners. In their study, Guion et *al.* (2000) examine the perception of English and Japanese consonants in terms of

Japanese categories. More specifically, the study examines the English consonants /b v w θ t s ɹ l/ and Japanese consonants /b ɰ t d s ɾ h/ in a /Ca/ syllable structure, in order to assess the perceived relation between English and Japanese consonants, in the context of the Japanese categories. Native Japanese listeners with no experience with English listened to the English tokens and categorized them in terms of provided Japanese categories. The provided Japanese category list was based on previous research. After the native Japanese listener selected what they believed to be a matching Japanese category to the English stimulus, they had to determine how good they thought their selections were, by completing a goodness-of-fit rating task. In this task, the Japanese listeners chose one of seven provided numbers on a Likert scale, reflecting the level of category goodness (more details are found in Chapter 3).

The study then analyzes the results obtained by creating matrices that show the mean identifications and goodness ratings for each one of the English stimuli. Table 2.1 below is a sample matrix taken from Guion *et al.* (2000). In that figure, you can see the English stimuli listed in the left column, while the Japanese categories are listed in the top row. The numbers in the table represent the mean of the given identifications for each Japanese category, and the numbers in parentheses represent the mean of the given goodness rating for each one of the selected Japanese categories. The bolded numbers represent the modal (highest) identification response. For example, when the English consonant /b/ was presented to the native Japanese listeners, the Japanese category /ba/ was selected 84% of the time, with a 5.3 goodness rating, and the Japanese category /va/ was selected 16% of the time, with a 4.8 goodness rating.

Table 2.1. Table Showing Mean Percent Identification and Goodness Rating (in parentheses) of English Consonant Stimuli in Terms of Japanese Categories. From Guion *et al*., (2000).

Consonant Stimuli	Percent Identification and Rating																
	ワ ɰa	ラ ɾa	ウラ ɯɾa	リャ ɾʲa	ヤ ja	ハ ha	パ pa	バ ba	ヴァ va	ファ ɸa	タ ta	ダ da	サ sa	ザ za	シャ ʃa	ツァ tsa	ア a
/b/								**84** **(5.3)**	16 (4.8)								
/v/	2 (4.5)						1 (6)	17 (4.2)	**80** **(4.4)**								
/w/	**79** **(3.5)**	10 (3.0)	6 (2.2)			2 (3.0)			3 (2.0)								
/θ/								2 (2.0)	2 (4.5)	38 (3.4)			**39** **(3.8)**	7 (2.7)		12 (3.3)	
/t/											**91** **(3.9)**					8 (2.4)	
/s/													**87** **(4.5)**			13 (3.6)	
/ɹ/		46 (3.4)	**50** **(3.3)**	4 (1.2)													
/l/	7 (3.5)	**50** **(3.2)**	37 (3.0)	1 (1.0)					5 (3.6)								

Up until this point in the analysis, matrices like the above allowed the researchers to determine numerically to what degree English consonants are connected to Japanese categories. Next, in order to determine the degree of fit of these selected categories and test theoretical models, like PAM and SLM, the researchers used a fit index analysis, using the obtained proportions of classifications (mean values of the selected category) and the mean goodness ratings; using these, they classified English sounds as "good", "fair", and "poor" fits with the Japanese categories. With this analysis, the proportion of classification is multiplied by the mean goodness. From the /ba/ example above, 0.84 x 5.3 = 4.45. Then, the researchers used a standard deviation (SD) criterion to determine the classification of the English consonant. The mean fit index of the Japanese-to-Japanese labeling task was used in order to calculate the SD. The researchers determined that if the fit index for an English consonant fell within 1.0 SD of the mean fit index for the Japanese consonants, this would be considered a "good" fit or a good instance of Japanese category. If the fit index fell between 1 and 2 SDs, they classified it as a "fair" fit to the Japanese category. If the fit index was lower than 3 SD, they marked it as a "poor" fit. In this case, 4.45 is labeled as a "good" fit, which means that the Japanese category /ba/ is a good fit for the English consonant /b/. In other words, fit is an indication to how strong the mental representation is between the English consonant and the Japanese category that represents it in the minds of native Japanese listeners. Table 2.2 below is a sample fit index table taken from Guion *et al.* (2000). In the figure, one can see the degrees of fit assigned for each English consonant on the furthest right column of the table.

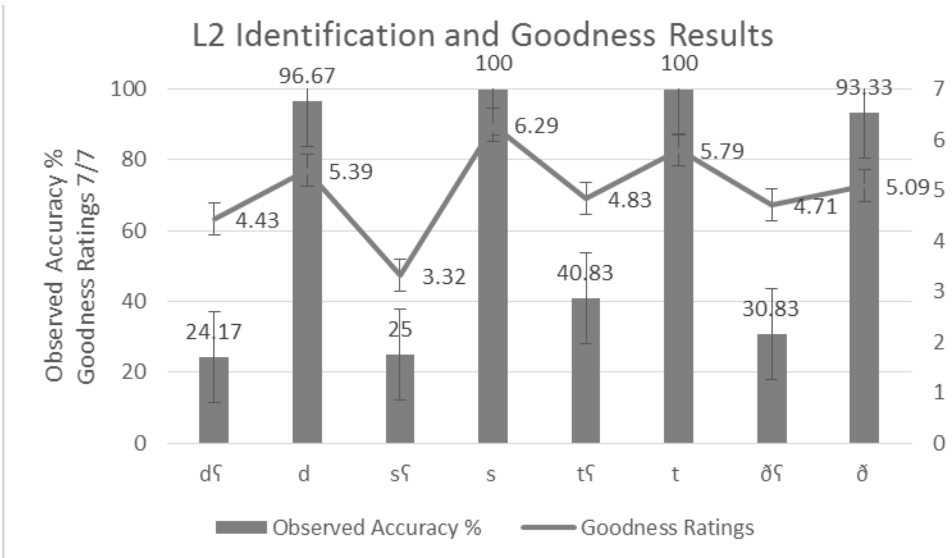
Table 2.2. Fit Indexes Derived for English Consonants in Terms of Japanese Categories. From Guion *et al*., (2000).

English consonant	Most common identification	Proportion of identifications	Goodness rating	Fit index	
/b/	/b/	0.84	5.3	4.5	good /b/
/s/	/s/	0.87	4.5	3.9	good /s/
/t/	/t/	0.91	3.9	3.5	good /t/
/v/	/v/	0.80	4.4	3.5	good /v/
/w/	/ɰ/	0.79	3.5	2.8	fair /ɰ/
/ɹ/	/ɯɾ/	0.50	3.3	1.7	poor /ɯɾ/
	/ɾ/	0.46	3.4	1.6	poor /ɾ/
/l/	/ɾ/	0.50	3.2	1.6	poor /ɾ/
	/ɯɾ/	0.37	3.0	1.1	poor /ɯɾ/
/θ/	/s/	0.39	3.8	1.5	poor /s/
	/ɸ/	0.38	3.4	1.3	poor /ɸ/

In following this analysis, I will be able to examine the collected data from my experiment in Chapter 3 and classify the perceptual category mappings of Arabic plain and emphatic consonants in the minds of native speakers of American English, who have no experience with Arabic. In turn, I will be able to answer my first research question, listed above. The categorizations will also be used later in order to answer questions pertaining to perceptual models.

With regard to the second research question, my methodology and analysis seeks to examine the natural development of the perception of plain and emphatic sounds in Arabic by L2 learners and identify whether their perception results are affected by their L2 Arabic exposure. Chapter 4 contains a detailed explanation of the experiment and procedures. I used an identification experiment (ID) and a goodness-of-fit rating task in order to accomplish this. Fifty native speakers of American English learning Arabic as an L2 participated in this experiment. The participants came with varying L2 exposure levels and were divided into three distinct groups, as follows: thirty participants with only 5 weeks of Arabic L2 exposure were placed in the Novice Group, ten participants with 8 months of Arabic L2 exposure were placed in the Intermediate Group, and ten participants with about 12 months of Arabic L2 exposure were placed in the Advanced Group. The ID experiment was designed to target the perception of L2 Arabic plain and emphatic sounds for the aforementioned groups. In the ID task, participants were asked to listen to Arabic stimuli, containing Arabic plain or emphatic sounds and identify these consonant sounds in terms of Arabic categories that were presented to them. The participants were also asked to indicate how confident they were of their chosen selections by choosing one of seven numbers on a Likert scale. The selected numbers reflect the level of confidence of the chosen Arabic category; selecting the number 1 means that the listener is least confident, and selecting the number 7 means that the listener is most confident of his or her selection. Mean percent identification and goodness rating figures were generated in order to examine the collected data. Figure 2.1 below is a sample of how the data was presented for each group. In the figure, the grey bars represent the mean identification results for each one of the tested consonants. The grey line, on the other hand, represents the mean confidence goodness rating results for each one of the tested consonants. The lines on top of each bar represent the error bars.

Figure 2.1. Sample Presentation of Mean Percent L2 Identification and Goodness Rating Results.



After collecting the data, I attempted to make comparisons, in order to see whether the mappings between the Arabic sounds and Arabic categories show any differences between the three learner groups. In addition, I examined the confidence goodness ratings of the selected Arabic categories to see whether they demonstrated any differences across the three learner groups. The goodness ratings will be useful in the event that participants from different groups show similar Arabic identification patterns. In that case, differences between the confidence goodness ratings could be considered a secondary indication to the development pattern(s). For example, if participants in the Intermediate Group and Advanced Group identified Arabic plain /t/ as Arabic plain /t/ 90% and 92% of the time, respectively, one may examine the confidence goodness ratings and see whether they yield a different pattern. For instance, if participants in the Intermediate Group and Advanced Group had a 4 and 6.5 mean goodness ratings, respectively, for the above-selected Arabic plain /t/, then one can conclude that participants in the Advanced Group showed a better development pattern than the participants in the Intermediate Group. There is always a possibility that the scores from the ID task and the goodness-of-fit rating task will show parallel patterns within each group. An example of this would be if participants in the Novice Group identified Arabic plain /t/ as Arabic plain /t/ 50% of the time with a mean goodness rating of 3 out of 7. This is an indication that the confidence rating matches the mean identification score. I shall address the second research question by examining the results from the ID and goodness-of-fit rating tasks.

As for the third research question, my main goal is to examine the relationship between L1 and L2 categories for L2 learners over time. My methodology and analysis seek to examine and identify the mapping patterns between L1 English categories and L2 Arabic plain and emphatic consonants and investigate whether the mapping patterns are affected by L2 Arabic exposure. Chapter 4 contains a detailed explanation of the experiment and procedures that were used in order to collect and analyze the data. In order to address this question, I also used an identification experiment and a goodness of fit rating task. The same fifty native speakers of American English who participated in the previously mentioned experiment also participated in this experiment, entitled *L1 Labeling Task*. Similarly, L2 learners of Arabic, with varying L2 exposure levels, were divided into three distinct groups as follows: thirty participants with only 5 weeks of Arabic L2 exposure were placed in the Novice Group, ten participants with 8 months of Arabic L2 exposure were placed in the Intermediate Group, and ten participants with about 12 months of Arabic L2 exposure were placed in the Advanced Group. The ID experiment was designed in order to target the perceptual relationship of L2 Arabic plain and emphatic sounds with L1 English categories and establish mapping patterns for the three different learner groups.

In the ID task, participants were asked to listen to Arabic stimuli, containing Arabic plain or emphatic sounds, and identify these consonant sounds in terms of the English categories that were presented to them. The participants were also asked to indicate how similar they think their labeling was by choosing one of seven numbers on a Likert scale. The selected numbers reflect the level of similarity of the chosen English category; selecting number 1 means that the listener thinks that his or her English labeling is not similar to the Arabic stimulus, and selecting the number 7 means that the listener thinks that his or her English labeling is very similar to the Arabic stimulus. Mean percent identification and goodness rating tables similar to Table 2.1 were generated in order to examine the collected data for each group.

After gathering the data, I attempted to make comparisons in order to see whether the mapping patterns between the Arabic sounds and English categories show any differences between the three learner groups, whose L2 proficiency increases with more L2 exposure. In addition, I examined the similarity goodness ratings of the selected English categories to see whether they demonstrated any differences across the three learner groups. The goodness ratings will be useful in the event that participants from different groups show similar mapping patterns. In that case, differences between the confidence goodness ratings could be considered a secondary indication to the subtle change of mapping patterns. For example, if participants in the Intermediate Group and Advanced Group identified Arabic plain /d/ as English /d/ 80% and 81%, respectively, I can examine the similarity goodness ratings in this case and see whether they yield a different pattern. For instance, if participants in the Intermediate Group and Advanced Group had 3 and 6 mean goodness ratings, respectively, for the above selected English /d/, then I can conclude that participants in the Advanced Group showed a different mapping pattern from the Intermediate Group; even though their mean identification scores were similar, their higher goodness ratings indicate that they are less confused and more certain of the similarity of their English labeling. Thus, by examining the results from the ID and goodness-of-fit rating tasks, I will be able to address the third research question and see if the mapping patterns between L1 English and L2 Arabic change over time.

Finally, with regard to the fourth research question, I have used this data in order to attempt to test the SLM model directly and the PAM model indirectly. A detailed explanation of the methods and analysis are available in Chapter 5. As for testing the SLM, I adopt an approach that was used by Park and de Jong (2008). In their study, Park and de Jong used a quantitative approach to test the SLM; they quantified the degree to which L1 Korean categories are used in L2 English category identification. Forty native Koreans who had been studying English for seven years listened to English nonsense words that consist of the consonants /p b t d f v θ ð/, followed by the vowel /a/. They were asked to identify the consonants, once with Korean labels (Korean labeling task) and once with English labels (Roman/IPA labeling task). In addition, the listeners were asked to give gradient evaluations of the goodness of the selected labels to the presented stimuli; they were asked to mark how good they considered the label to be on a Likert sale from 1 to 7 (1 = not similar at all, and 7 = exact match). In the Korean labeling task, listeners identified and rated the presented English stimuli, in terms of a predetermined list that contained 13 Korean categories presented in Korean orthography. In the Roman/IPA labeling task, the Korean listeners were asked to choose the initial consonant they heard in the stimulus from a list of 15 presented alternatives. Example words were provided to help the listeners identify each presented English alternative. After each identification response, the Korean listeners were asked to indicate how confident they were of their selections, by choosing a number response from a 7-point Likert scale (1 = not confident at all, and 7 = very confident).

Based on the results from their experiment, Park and de Jong (2008) were able to present a detailed assessment of the degree to which the identification patterns of L2 English demonstrate a reliance on L1 Korean categories. This quantitative technique of assessment allows one to interpret the reliance explained earlier; if found present, it can be considered evidence for the entanglement of the L1 and L2 categories (Park and de Jong, 2016). Their analysis assumes that a given identification could be attributed to one of two possibilities, the first being reliance on an L1 category and the additional mapping of that L1 category onto the L2 response category (Park and de Jong, 2008, p. 708) and the second, the development of a (new) L2 category (p. 708). The analysis relies on the mapping data in order to distinguish these two possibilities. By doing so, Park and de Jong (2008) categorized the tested sounds into SLM's "new" and "similar" categories and were able to test SLM's predictions of L2 development. In Chapter 5 I will adopt Park and de Jong's analysis and test SLM's predictions.

The PAM, as explained in Chapter 1, is a cross-language perception model, rather than an L2 model interested in naïve monolinguals' perceptual pattern of nonnative sounds. Since the model relies on the assimilation patterns of nonnative segments in order to discriminate non-native sounds, PAM pairwise assimilation types between the non-native and native sounds were generated in order to describe the nature of the relationship between these sounds. PAM uses a degree of goodness fit that varies from poor to very good in order to define the discrimination of the non-native and native sounds. In order to test the PAM, I will use the results and data that were obtained from the cross-language perceptual task in Chapter 3 (explained above), since it involves naïve monolingual participants rather than L2 Arabic learners. I will generate predictions based on the findings of the experiment which was explained earlier in this chapter and test whether these PAM predictions are supported or not. In doing so, I will be able to address the fourth research question and see whether or not these findings support the models.

Chapter 3 Perceptual Category Mapping

In this chapter, I will attempt to establish perceptual category mapping between Arabic and English consonants in the minds of native speakers of American English, who have had no prior experience or exposure to Arabic. In doing so, I seek to describe and determine the perceptual classification of L2 Arabic plain and emphatic sounds in the minds of native speakers of English. As explained in Chapter 2, this experiment will follow the analysis of Guion *et al.* (2000), in order to classify the Arabic plain and emphatic sounds. The results of this experiment will allow us to answer my first research question.

3.1 Participants

3.1.1 Talkers

One 30-year-old female Arabic speaker and one 20-year-old female English speaker produced the stimuli. The Arabic speaker was born and raised in an Arab country (Jordan) and was exposed to Modern Standard Arabic at school and in college. The participant came to the U.S. three years prior to the recording date to attend a graduate school program and does not speak any languages besides Arabic and English. Her exposure to English began in the third grade, in a typical English as a Foreign Language (EFL) classroom setting. She was raised in a monolingual household and has no hearing or speaking impairments.

The English speaker was born and raised in the American Midwest. She speaks only English, though she had minimal exposure to Spanish in high school. She had no exposure to

Arabic, prior to the recording date, nor had she ever traveled to an Arab country. She grew up in a monolingual family and has no hearing or speaking impairments.

3.1.2 Listeners

Sixteen native speakers of American English (10 females and 6 males, between 19 and 28 years of age) participated in the experiment as listeners. All participants were monolingual English speakers, with minimal exposure to Spanish in high school. All participants were born and raised in the American Midwest and had no travel experience to an Arab country and no contact with Arabic, prior to recruiting them to participate in this experiment. They were raised in a monolingual household and had no hearing or speaking impairments at the time of their participation.

3.3 Stimuli

Two sets of stimuli were used in this experiment: Arabic stimuli and English stimuli. The criteria used for each set is described below, followed by an explanation of how the stimuli were recorded and processed.

3.3.1 Arabic Stimuli

The Arabic stimuli consisted of the following 14 Arabic consonants, /t, d, ð, s, $t^ʕ$, $d^ʕ$, $ð^ʕ$, $s^ʕ$, q, x, ɣ, ħ, ʕ, ʔ/ in CV and VCV syllable structures. All six short and long vowels /a, a:, u, u:, i, i:/ were used in the syllable templates filling the V positions to generate 12 possibilities (i.e., 6 vowels × 2 syllable structures) for each one of the 14 consonants for the syllable structures. I used the same vowel to fill the V positions within the same syllable structure (e.g., /tu/, /tu:/, /utu/, /u:tu:/). For example, there were 12 stimuli related to the target consonant /t/, as shown in Table 3.1. A total of 168 stimuli (= 14 Arabic consonants × 6 vowels × 2 syllable structures) were generated for this study.

Table 3.1. A Corpus of 12 /t/ Stimuli: /t/ with Six Short and Long Vowels in CV and VCV

	CV	**VCV**
Short Vowels	/**t**a/ - /**t**u/ - /**t**i/	/a**t**a/ - /u**t**u/ - /i**t**i/
Long Vowels	/**t**a:/ - /**t**u:/ - /**t**i:/	/a:**t**a:/ - /u:**t**u:/ - /i:**t**i:/

The Arabic female speaker produced the target stimuli in the carrier phrase "*ya-ktub* ***stimulus*** *al-yawm.*" ("*He writes* ***stimulus*** *today.*"). I included the short Arabic diacritics,[1] denoting the short vowels throughout the carrier phrases and the target syllables to avoid any orthographic mismatch representations while reading and recording the phrases. I also created foil stimuli with seven non-target consonants /b, ʒ, r, z, ʃ, f, m/. The same six short and long vowels in the CV and the VCV structures were used with these consonants, resulting in 84 foil stimuli (= 7 consonants × 6 vowels × 2 syllable structures) in total.

1 Arabic short vowels are represented by diacritics ◌َ ◌ُ ◌ِ which are placed on the consonants (the circles are a representation of a consonant slot).

3.3.2 English Stimuli

The English stimuli consisted of 7 consonants /t, d, ð, s, k, h, ʔ/ with the vowels /a, u, i/ in CV and VCV syllable structures. Since English does not have the short and long vowel distinction like Arabic, the English female speaker produced these stimuli in the carrier phrase "*I met* ***stimulus*** *today*." twice at different times. This repetition was intended to balance the number of stimuli between English and Arabic. For example, there were 12 tokens related to English /t/ (= 3 vowels × 2 repetitions × 2 syllable structures). As in the Arabic stimuli, I used the same vowel to fill the V positions within the same syllable structure (e.g., /tu/ and /utu/). In total, there were 84 stimuli (= 7 English consonants × 3 vowels × 2 syllable structures × 2 repetitions). I also created 6 foil stimuli in CV and VCV syllable structures in which C is not one of the tested consonants for the practice round that preceded the actual experiment.

The two talkers were recorded individually in a quiet room using Audacity (2.0.3) installed on a laptop and a headset microphone. Each stimulus in the carrier phrase was printed on a flashcard, and I handed over flashcards one at a time to the talker to read. If the talker misread the intended target stimuli, I placed the flashcard aside and passed it later to the talker to be read again. The target stimuli were extracted from the carrier phrases and the amplitude was normalized for later use in the experiment. Then, the stimuli were divided into two sets. One set consisted of the Arabic and English stimuli in CV syllable structure, and the other set consisted of the Arabic and English stimuli in VCV syllable structure. Using Praat, an identification task and a goodness-of-fit rating task were created for each of the two sets.

3.4 Procedure

The sixteen listeners participated in an identification and goodness-of-fit rating tasks. A detailed description of the tasks is presented in the section below.

3.4.1. Identification and Goodness-of-Fit Rating Tasks

I met with each participant individually in a quiet room to run the experiment. Participants were asked to perform the following tasks:

1) Label each Arabic and English consonant they hear in the stimulus by identifying an English form from the choices they see on a computer screen. The labeling choices included the following English consonants, /b, d, f, g, h, ʒ, k, l, m, n, p, r, s, t, z, θ, ð, ʔ/ and *other*. The "other" choice was included, in case participants thought no given English consonants were similar to the sound they heard. The English labeling consonants were placed in parenthesis in actual English words in an initial position for ease of recognition, e.g. (B)est, (D)oor, (F)ar, (G)ame, and they all appeared on the computer screen for each played stimulus. For instance, if the participant heard the English /ba/, he or she would have to identify the consonant in the syllable (/b/ in this case). If the participant chose *(B)est*, it was considered a correct match.

2) Determine how good participants consider the label they chose in (1) above. They were asked to click on one of 7 provided number choices on a Likert scale, with 1 being a poor match and 7 being an exact match. For instance, after selecting *(B)est* as indicated above, the listener had to click on one of seven numbers. Each number reflects the level of confidence to the already selected category. Both the parenthesized categories and the Likert scale appear on the same screen.

The English and Arabic CV set was presented first and followed by the English and Arabic VCV set. The stimuli were randomized within each set, and the listeners were given a 10-minute break after finishing the first set. The tasks were not timed, and it took each listener about 13 minutes to complete the tasks. Instructions were printed and handed out to listeners before the task, and they had a chance to go through a practice round to make sure they understood what they were supposed to do. The collected data were saved to a laptop for later analysis.

3.5 Data Analysis

The main goal of this study is to provide data that shows the degree of the perceived phonetic distance between L1 English and L2 Arabic sounds, based on a cross-language mapping experiment. This will allow me to answer the first research question: *What are the perceptual category mappings between Arabic plain and emphatic consonants with English consonants in the mind of native speakers of American English?*

To achieve this goal, I follow Guion *et al.*'s (2000) analysis. First, I generated matrices showing mean identification and goodness-of-fit ratings when listeners attempted to identify the English and Arabic consonants in terms of the English categories. I did this for each of the six Arabic vowels within each syllable structure. For example, for the short vowel /a/, mean values were generated when /a/ occurred in a CV structure and another mean values for when /a/ occurred in a VCV structure. I followed the same steps for the other vowels. Second, I generated fit index tables using the obtained proportion of classification (mean values of the selected category) and the mean goodness rating, classifying the Arabic sounds as “good”, “fair”, and “poor” fit in terms of English categories. To calculate the fit indices, I multiplied the proportion of identification by the mean goodness rating. Any proportion bigger than 30%, a percentage that was found to be a reliable predictor in Guion *et al.* (2000), was used in order to generate the fit indexes. For example, if the Arabic consonant /t/ was identified as English /t/ 80% of the time, with a mean goodness of 6.00, the fit index will be generated through multiplying the proportion of classification by the mean goodness (0.8 × 6 = 4.8). To determine the classification of the varied fit indexes in terms of English categories, I used the mean fit index for the *English* consonants with their standard deviation (SD) criterion. If the fit index for an Arabic consonant fell within 1.0 SD of the mean fit index for the English consonants, I considered it as a “good” fit or a good instance of the English category. If the fit index fell between 1 and 2 SDs, I classified it as “fair” fit to the English category. If the fit index was lower than 3 SD, I marked it as “poor” fit. The mean fit index for the English consonants was 4.5 with the SD 1.4. Based on that, Arabic consonants with a fit index of 3.1 (i.e., 4.5 – 1.4 = 3.1) and over were considered to have “good” fits, Arabic consonants with a fit index between 1.7 and 3.1 (i.e., 4.5 – (2 × 1.4) = 1.7) a “fair” fit, and Arabic consonants with a fit index lower than 1.7 a “poor” fit. I present 12 fit index tables in this paper; one for each of the six Arabic vowels in the CV and VCV syllable structures.

3.6 Results

In this section I present the results in light of the three factors that I decided to consider: prosodic location effects, vowel quality effects, and vowel duration effects. Thus, I present the results for each vowel in a different prosodic location. Also, it should be noted that even though I describe the results for all the tested sounds, the main focus of this experiment remains the four plain Arabic sounds /t, d, ð, **s/** and four emphatic sounds /tˤ, dˤ, ðˤ, **s**ˤ/. Table 3.2 presents the identification and goodness-of-fit rating results from 16 listeners for the Arabic consonants in terms of the English categories in the CV syllable structure with the short vowel /a/. The first column and the first row show the tested Arabic consonants and English labels used for the Arabic consonants, respectively. Two types of data are found in the table. One is the percentage of the frequency in which an Arabic consonant was labeled to an English category. The other is the average rating given to the selected English category and is shown in parentheses. Boldfaced data indicate the most frequently chosen English category for that Arabic consonant. For example, in Table 3.2, Arabic /t/ was labeled as English /d/ 75% of the time with an average rating of 5.4. Arabic /t/ was also classified as English /g/ 6.25% of the time, as English /t/ 12.5% of the time, and as English /θ/ 6.25% of the time. The most frequently chosen English /d/ label was boldfaced. Similar tables were generated for CV when V was /u/, /i/, /a:/, /u:/, and /i:/, for VCV when V were /a/, /u/, /i/, /a:/, /u:/ and /i:/. These tables are provided in Appendix A.

Table 3.2. Mean Percent Identifications and Goodness Ratings (in parentheses) of Arabic Consonant Stimuli in Terms of English Categories in CV Utterances when V was /a/.

Arabic Stimuli	**English Consonants**										
	/d/	**/f/**	**/g/**	**/h/**	**/r/**	**/s/**	**/t/**	**/θ/**	**/ð/**	**/ʔ/**	**Other**
/t/	**75 (5.4)**		6.25 (3)				12.5 (3)	6.25 (5)			
/d/	**87.5 (4.8)**								12.5 (7)		
/ð/								12.5 (3)	**87.5 (5.5)**		
/s/						**100 (5.6)**					
/tˤ/	12.5 (6)	**50 (5)**					37.5 (3.3)				
/dˤ/	**68.75 (4.8)**							6.25 (3)	12.5 (5)		12.5 (4)
/ðˤ/	12.5 (1)	12.5 (3)						12.5 (3)	12.5 (6)		**50(4.5)**
/sˤ/						**100 (6)**					
/q/				12.5 (2)						37.5 (5.3)	**50 (3.7)**
/x/				**62.5 (3)**							37.5 (5.3)
/ɣ/			12.5 (2)		37.5 (3.6)						**50 (4.7)**
/ħ/				**100 (5.7)**							

Table 3.3 below presents the fit indexes for the Arabic consonants in terms of English categories in the CV syllable structure when V was the short Arabic /a/. The table lists the Arabic consonants in the far left column, followed by the most commonly selected English categories. One will find that the proportions of identification and goodness ratings are the same as in Table 3.2. I multiplied these two for each Arabic consonant in order to obtain the fit index value (far right column). For example, Arabic /t/ was mostly identified as English /d/ with the highest proportion of 0.75 and the mean rating of 5.4. Thus, the fit index for Arabic /t/ was 4 (= 0.75 × 5.4). The degree of fit for each consonant was determined to be "good", "fair", or "poor" based on the pre-set SD criterion discussed earlier. Thus, the degree of fit for Arabic /t/ with English /d/ is "good" based on the SD criterion (cf., a value of 3.1 and above receives a "good" fit.).

Table 3.3. Fit Indexes for Arabic Consonants in Terms of English Categories in CV Syllable Structure When V was /a/.

Arabic Consonants	**Most Common Identification**	**Proportion of Identification**	**Goodness Rating**	**Fit Index**
/t/	/d/	0.75	5.4	4 Good
/d/	/d/	0.87	4.8	4.17 Good
/ð/	/ð/	0.87	5.5	4.78 Good
/s/	/s/	1	5.6	5.6 Good
/tˤ/	/f/ /t/	0.5 0.37	5 3.3	2.5 Fair 1.2 Poor
/dˤ/	/d/	0.68	4.8	3.26 Good
/ðˤ/	Other	0.5	4.5	2.25 Fair
/sˤ/	/s/	1	6	6 Good
/q/	/ʔ/	0.37	5.3	1.96 Fair
/x/	/h/	0.62	3	1.86 Fair
/ɣ/	/r/	0.37	3.6	1.33 Poor
/ħ/	/h/	1	5.7	5.7 Good
/ʕ/	Other	0.62	5.4	3.34 Good
/ʔ/	Other	0.75	3.5	2.6 Fair

Some Arabic consonants were mapped into two English categories. Thus, two fit indexes were generated for those Arabic consonants. For example, the Arabic consonant /tˤ/ was mapped into

English /f/ and /t/ with ‘fair’ and ‘poor’ fit indexes. Similar fit index tables were generated for CV when V were /u/, /i/, /a/, /u:/, and /i:/, and for VCV when V were /a/, /u/, /i/, /a:/, /u:/, and /i:/. These tables are found in Appendix B.

Similar mean percent identification and goodness rating tables were generated for when the participants attempted to identify the English consonants in terms of the English categories in CV and VCV with English /a, u, i/. These tables are found in Appendix C. The results show that there were no differences between the identifications and goodness ratings when different prosodic locations and vowels were used. As a result, a fit index (Table 3.4) was generated, based on the combined results and the SD criteria was then calculated based on it, as explained earlier.

Table 3.4. Fit Index for English Consonants in Terms of English Categories in CV and VCV Syllable Structures.

English Consonant	**Most Common Identification**	**Proportion of Identifications**	**Goodness Rating**	**Fit Index**
/t/	/t/	0.94	5.9	5.6
/d/	/d/	1	5.7	5.7
/ð/	/ð/	0.55	4.9	2.7
/s/	/s/	0.94	5.9	5.6
/k/	/k/	0.94	5.7	5.4
/h/	/h/	0.83	5.8	4.8
/ʔ/	Other	0.48	3.9	1.9

Table 3.5 below highlights the degree of fit results in terms of their prosodic location (CV and VCV) and different vowels (/a, u, i, a:, u:, i:/) for Arabic coronal plain obstruents /t/, /d/, /ð/, and /s/. The results show similar patterns for the coronal plain obstruents in general; the results are similar in CV and VCV in different vowel quality conditions, but long and short vowel differences are shown for some consonants. For example, Arabic /d/ is mapped onto English /d/ with a “good” degree of fit throughout all tested conditions such as prosodic location, vowel

quality difference, and vowel length difference. On the other hand, Arabic /t/ was mapped onto English /d/ with a "good" degree of fit when /t/ occurred in CV and VCV with short /a/. However, Arabic /t/ was mapped onto English /t/ and /d/ both with a "fair" degree of fit in CV and VCV with long /a:/.

Table 3.5. Major Responses (30% or more) and Fit Index for Arabic Coronal Plain Obstruents /t, d, ð, s/ in Terms of English Categories in Different Prosodic Locations (CV and VCV) with Arabic Short and Long Vowels /a, u, i, a:, u:, i:/.

Arabic Consonants	**Vowels**	**CV**		**VCV**	
		Short	**Long**	**Short**	**Long**
/t/	**/a/**	/d/ Good	/t/ Fair /d/ Fair	/d/ Good	/t/ Fair /d/ Fair
	/u/	/t/ Good	/t/ Good	/t/ Good	/t/ Good
	/i/	/t/ Good	/t/ Good	/t/ Good	/t/ Good
/d/	**/a/**	/d/ Good	/d/ Good	/d/ Good	/d/ Good
	/u/	/d/ Good	/d/ Good	/d/ Good	/d/ Good
	/i/	/d/ Good	/d/ Good	/d/ Good	/d/ Good
/ð/	**/a/**	/ð/ Good	/ð/ Good /θ/ Fair	/ð/ Good	/ð/ Good
	/u/	/ð/ Good	/ð/ Fair /θ/ Fair	/ð/ Good	/ð/ Fair /θ/ Fair
	/i/	/ð/ Good	/ð/ Fair	/ð/ Good	/ð/ Fair
/s/	**/a/**	/s/ Good	/s/ Good	/s/ Good	/s/ Good
	/u/	/s/ Good	/s/ Good	/s/ Good	/s/ Good
	/i/	/s/ Good	/s/ Good	/s/ Good	/s/ Good

Table 3.6 presents similar data to Table 3.5 (above) for the Arabic emphatics /tˤ/, /dˤ/, /ðˤ/, and /sˤ/. I observed similar patterns for these consonants as I did for their non-emphatic

counterparts. No prosodic location effect was observed, but I observed a vowel quality effect, especially for emphatic /tˤ/. Arabic /sˤ/ did not show any vowel quality effect. In addition, the results show a vowel length effect for some emphatic consonants. The Arabic consonant /sˤ/, for example, was mapped onto English /s/ with a good degree of fit in all tested conditions. However, the Arabic consonant /tˤ/ was mapped onto different English categories with varying degrees of fit between the short and long vowels. It was mapped to English /t/ with a "fair" fit when short /u/ was used, and to English /t/ but with a "good" fit when long /u:/ was used.

Table 3.6. Major Responses (30% or More) and Fit Index for Arabic /tˤ, dˤ, ðˤ, sˤ/ in terms of English Categories in Different Prosodic Locations (CV and VCV) with Arabic Short and Long Vowels /a, u, i, a:, u:, i:/.

Arabic Consonants	**Vowels**	**CV**		**VCV**	
		Short	**Long**	**Short**	**Long**
/tˤ/	**/a/**	/f/ Fair /t/ Poor	/b/ Fair	/f/ Fair /t/ Poor	/b/ Fair
	/u/	/t/ Fair	/t/ Good	/t/ Fair	/t/ Good
	/i/	/θ/ Poor	/ð/ Fair /θ/ Fair	/θ/ Poor	/ð/ Fair /θ/ Fair
/dˤ/	**/a/**	/d/ Good	/b/ Poor /ð/ Poor	/d/ Good	/b/ Poor /ð/ Poor
	/u/	/d/ Good	/d/ Poor	/d/ Good	/d/ Fair
	/i/	/b/ Fair	/θ/ Fair	/b/ Fair	/θ/ Fair
/ðˤ/	**/a/**	/Other/ Fair	/Other/ Fair	/Other/ Fair	/Other/ Fair
	/u/	/ð/ Poor /θ/ Poor	/θ/ Fair	/ð/ Poor /θ/ Poor	/θ/ Fair
	/i/	/ð/ Fair /θ/ Poor	/ð/ Poor /θ/ Poor	/ð/ Fair /θ/ Poor	/ð/ Poor /θ/ Poor

I observe similar patterns for the remaining Arabic consonants /q/, /x/, /ɣ/, /ħ/, /ʕ/, and /ʔ/. Thus, the results show similar English category selections and degrees of fit for the Arabic consonants between CV and VCV when V were short /a/, /u/, and /i/ and similar English category selection and degrees of fit for the Arabic consonants between CV and VCV when V were long /a:/, /u:/, and /i:/.

Next, I present results in order to examine the effects of vowels on the perception of the tested Arabic consonants. Table 3.7 below shows the obtained results and degrees of fit for the Arabic consonants /t, d, ð, s, sˤ/. The listed data in the table is a combination of the English category selections and their degrees of fit in CV and VCV. The results presented earlier in Tables 3.5 and 3.6 were combined here and presented as two categories — one category for both CV and VCV, when V were short Arabic vowels /a, u, i/, and a second category for both CV and VCV, when V were long Arabic vowels /a:, u:, i:/. I present the data in this way because I did not notice any differences between English selections and their degrees of fit between CV and VCV with all six Arabic vowels.

Generally speaking, the results in Table 3.7 below show no differences between the English mappings and their degrees of fit, when vowels were either short or long. For example, Arabic /d/ was identified as English /d/ with a good degree of fit in short /a/ and long /a:/ conditions. The same consonant /d/ was also identified as English /d/ with a good degree of fit in short /u/ and long /u:/ conditions. Another example from Table 3.7 is the Arabic consonant /sˤ/, which was identified as English /s/ with a good degree of fit when V were either short /i/ or long /i:/. The one exception is Arabic /t/. This consonant was mapped onto English /d/ in the short /a/ vowel condition and onto English /t/ and /f/ in the long /a:/ vowel condition.

Table 3.7. Major Responses and Fit Index for Arabic /t, d, ð, s, $s^ʕ$/ in Terms of English Categories with Arabic Short and Long Vowels /a, u, i, a:, u:, i:/.

Arabic Consonants	**Vowels**	**CV & VCV**	
		Short	**Long**
/t/	**/a/**	/d/ Good	/t/ Fair /f/ Fair
	/u/	/t/ Good	/t/ Good
	/i/	/t/ Good	/t/ Good
/d/	**/a/**	/d/ Good	/d/ Good
	/u/	/d/ Good	/d/ Good
	/i/	/d/ Good	/d/ Good
/ð/	**/a/**	/ð/ Good	/ð/ Good
	/u/	/ð/ Good	/ð/ Fair /θ/ Fair
	/i/	/ð/ Good	/ð/ Fair
/s/	**/a/**	/s/ Good	/s/ Good
	/u/	/s/ Good	/s/ Good
	/i/	/s/ Good	/s/ Good
/$s^ʕ$/	**/a/**	/s/ Good	/s/ Good
	/u/	/s/ Good	/s/ Good
	/i/	/s/ Good	/s/ Good

Table 3.8 below presents the same data as Table 3.7 above, but for the Arabic consonants /$t^ʕ$, $d^ʕ$, $ð^ʕ$/. The data in the table show differences in relation to the vowel length conditions. For instance, when short /a/ was used, Arabic /$t^ʕ$/ was identified as English /f/ and English /t/ with fair and poor degrees of fit, respectively, and as English /b/ with a fair degree of fit when long /a:/ was used. Differences are also present when short and long /u/ were used; Arabic /$t^ʕ$/ was mapped onto English /t/ with a fair fit when short /u/ was used, and as English /t/ but with a good fit when long /u:/ was used. Generally speaking, differences are present across all the tested

Arabic consonants in Table 8 /tˤ, dˤ, ðˤ/ between the short and long Arabic vowels. This is an indication that short and long vowels affected the mappings of these consonants differently.

Table 3.8. Major Responses and Fit Index for Arabic /tˤ, dˤ, ðˤ/ in Terms of English Categories with Arabic Short and Long Vowels /a, u, i, a:, u:, i:/.

Arabic Consonants	**Vowels**	**CV & VCV**	
		Short	**Long**
/tˤ/	**/a/**	/f/ Fair /t/ Poor	/b/ Fair
	/u/	/t/ Fair	/t/ Good
	/i/	/θ/ Poor	/θ/ Fair /ð/ Fair
/dˤ/	**/a/**	/d/ Good	/b/ Poor /ð/ Poor
	/u/	/d/ Good	/d/ Fair
	/i/	/b/ Fair	/θ/ Fair
/ðˤ/	**/a/**	/Other/ Fair	/Other/ Fair
	/u/	/θ/ Poor /ð/ Poor	/θ/ Fair
	/i/	/θ/ Poor /ð/ Fair	/θ/ Poor /ð/ Poor

The apparent patterns that arise from the results for the identification and goodness ratings suggest that perceptual mappings between L2 Arabic and L1 English consonants yield different outcomes when different factors are being considered. The observed findings suggest that the tested prosodic locations, CV and VCV, did not play a major role in the outcomes of the perceptual mappings between Arabic and English. Tables 3.5 and 3.6 show that the selections of the English categories by English listeners for the tested Arabic consonants did not differ between CV and VCV. The degrees of fit were also identical for the majority of the part between the established categories in the different prosodic locations.

Looking at the results in light of the vowel durations as a possible factor indicates that Arabic short and long vowels had an effect on the mapping results of *some* Arabic consonants. As presented in Table 3.7, the vowels had no effects on the English category selections and their degree of fit for the Arabic consonants /t, d, ð, s, $s^ʕ$/; the mappings were not affected by whether the short /a, u, i/ or the long Arabic vowels /a:, u:, i:/ were used except for Arabic /t/ which demonstrated differences for when short /a/ and long /a:/ were used. However, as presented in Table 3.8, the vowels had an effect on the mappings between Arabic and English consonants and their degrees of fit for the Arabic consonants /$t^ʕ$, $d^ʕ$, $ð^ʕ$/. This suggests that short-long vowel distinction should be taken into account when dealing with the consonants.

With regard to vowel qualities, the results in Table 3.7 show no vowel quality effect for the tested Arabic consonants, except Arabic /t/; Arabic /t/ shows a vowel quality effect when short /a/ and /a:/ are used. The results in Table 3.8, on the other hand, indicate that the vowel qualities had a more profound effect on the selected English categories and their degrees of fit.

Additionally, based on the mapping results, I notice that Arabic coronal plain obstruents /t, d, ð, s/ were mapped onto English categories with corresponding or adjacent places of articulation. Arabic emphatics /$t^ʕ$, $d^ʕ$, $ð^ʕ$/ were mapped onto more than one English category with varying degrees of fit; Arabic /$t^ʕ$/ with short Arabic /a/ was mapped onto English /f/ and /t/ both with a "poor" fit, Arabic /$t^ʕ$/ with long Arabic /i:/ was mapped onto English /θ/ and /ð/ both with a "fair" fit, Arabic / $d^ʕ$/ with long Arabic /a:/ was mapped onto English /b/ and /ð/ both with a "poor" fit, Arabic /$ð^ʕ$/ with short Arabic /u/ was mapped onto English /θ/ and /ð/ both with a "poor" fit, with short Arabic /i/ onto English /θ/ and /ð/ with "poor" and "fair" fits, respectively, and finally with long Arabic /i:/ onto English /θ/ and /ð/ with a "poor" fit for both (Table 3.8).

So, to answer the research question: *What are the perceptual category mappings between Arabic plain and emphatic consonants with English consonants in the mind of native speakers of American English?*, the perceptual mapping results allow us to identify which Arabic consonants may be considered to be “new” and which ones may be considered to be “similar” based on SLM. I set criteria for determining whether the Arabic consonants are “new” or “similar” based on the obtained degree of fit. If an Arabic consonant was mapped onto an English category with a “good” degree of fit, it would be fair to consider it as “similar” to the selected English category. If the Arabic consonant was mapped onto an English category with a “fair” or “poor” fit, I consider it to be “new” to the selected English category.

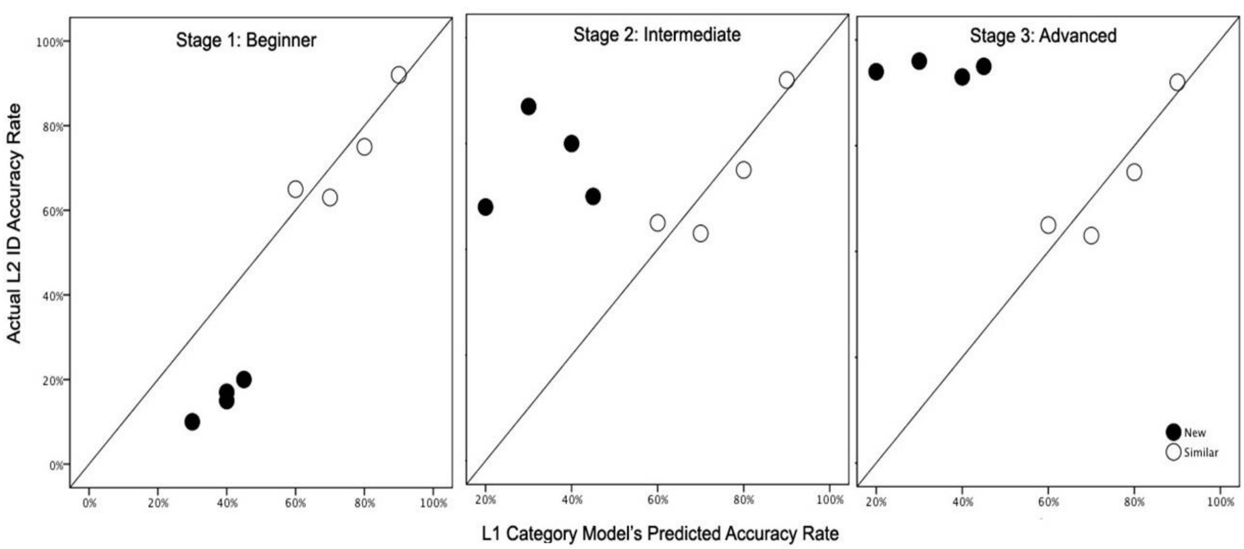
Based on this established criteria, the Arabic consonants /t, d, ð, s, sˤ/ are considered “similar” sounds to the English categories, based on their “good” degrees of fit when both /a/ and /a:/ were used. The Arabic consonants /tˤ, ðˤ, dˤ/ are considered “new” sounds to the English categories based on their degrees of fit (see Tables 3.7 and 3.8). It is important to keep in mind that this classification of sounds is contingent upon the used Arabic vowels. As I have explained earlier, some Arabic consonants demonstrate slight mapping onto L1 differences while others do not.

Chapter 4 Examining L2 Arabic Development

As explained previously, the way in which learners' L1 phonology and L2 phonology interact with each other is very intricate. Moreover, accurately measuring and understanding, how L2 segments develop in the minds of L2 learners has been the center of attention in the field of L2 acquisition. Thus, a number of studies have attempted to investigate the best possible ways of determining the overall L2 perceptual accuracy.

SLM presents an ideal outcome for the gradual learning of L2 sounds. The model generates predictions of accuracy, as illustrated in Figure 4.1 below. Figure 4.1 demonstrates the hypothetical learning phases of L2 sounds during three different L2 learning stages, beginner stage, intermediate stage, and advanced stage. The figure shows the gradual development in light of the two L2 sound types adopted from SLM, "new" and "similar".

Figure 4.1. Learning Stages According to SLM's Predictions.



The hollow shapes in the figure correspond to the "similar" sound types, and the dark shapes correspond to the "new" sounds. In the figure, accuracy performance of the "similar" sounds does not change across the three different learning stages. This is simply due to the fact that learners have associated these sounds with existing L1 categories, and further learning does not happen. As explained earlier, an L1 interference and/or a use of existing L1 categories is at work here for these categories. On the other hand, the accuracy performance of the “new” sounds undergoes a gradual increase from a low performance in the beginner stage to a higher performance in the intermediate stage and final better performance in the advanced stage. Notice that the performance of the “new” sounds surpasses the performance of the “similar” sounds in the advanced stage. This is an inevitable outcome to more effort being placed into learning the “new” sounds.

In light of that, this experiment is primarily interested in accounting for the perceptual developmental stages of L1-English learners of L2-Arabic. Following Park and de Jong (2008), the study will extend SLM into perception by examining groups of L2-Arabic learners, with varying levels of exposure, through an identification task and goodness-of-fit rating task. The current tasks will seek to generate quantitative predictions concerning L2 identification and attempt to answer the second and third research questions:

- *Does more L2 exposure over time lead to similar learning development with Arabic L2 plain and emphatic sounds?*
- *Do L2 learners’ L1-to-L2 mappings change over time?*

One benefit of using an identification task over a discrimination task is that the identification task enables researchers to conduct their studies in a naturalistic setting; this task is more ecological and reflective of real world situations among L2 learners than the discrimination tasks. Moreover, identification tasks enable us to rely on the use of “phonetic memory codes,” which are more reflective of the overall developmental stages unlike other tasks like the discrimination task which primarily tends to focus listeners' attention on the differences between stimuli at the low-level (Logan and Pruitt, 1995). Consequently, interpreting results obtained from an identification task will allow me to closely examine the learning process of the tested L2 segments, in particular, and to understand the perceptual process, in general.

4.1 Participants

4.1.1 Talkers

Four native speakers of Arabic (M = 2, F = 2) participated as talkers and produced the stimuli to test L2 perception of Arabic consonants by American learners of Arabic. The two male talkers were 25 and 28 years old, respectively, and the two female native talkers were 26 and 30 years old, respectively. The four participating talkers were born and raised in Jordan, an Arab country of the Levant region, and speak the same regional dialect, Ammani.[2] They were taught and exposed to MSA at school and in college, and they came to the United States 3–5 years prior to their participation in the experiment, in order to attend and complete their graduate studies at the University of Wisconsin-Milwaukee (UWM). They do not speak any other languages besides Arabic and English, and their exposure to English began in the third grade, in a typical EFL classroom setting. They were raised in monolingual households, and they do not have any hearing or speaking impairments. Four talkers were chosen to participate in recording the stimuli for the experiment because the sound signal differs from one talker to another.

[2] The Ammani dialect refers to the Arabic regional dialect spoken in the capital of Jordan, Amman.

4.1.2 Listeners

A total of fifty-five subjects participated as the listeners in this experiment (L2 Arabic learners = 50, Native Arabic speakers = 5). Fifty native speakers of American English who are L2 learners of Arabic at UWM participated in an identification and goodness-of-fit rating tasks. The L2 Arabic learners were divided into three different listener groups reflecting their Arabic L2 proficiency levels[3]. Thirty L2 Arabic learners (F = 18, M = 12) were placed in the Novice Group (NG). Their ages ranged from 18 to 20 years old (Mean = 18.8 years old), and they had been studying Arabic for five weeks at the time of their participation in the experiment, which is the time needed to cover the Arabic sound system in class at UWM. Ten L2 Arabic learners (F = 6, M = 4) were placed in the Intermediate Group (IG). Their ages ranged from 20 to 39 years old (Mean = 24 years old), and they had been studying Arabic for about 8 months at the time of their participation in the experiment. Ten L2 Arabic learners (F = 7, M = 3) were placed in the Advanced Group (AG). Their ages ranged from 22 to 25 years old (Mean = 23.3 years old), and they had been studying Arabic for about 12 months at the time of their participation in the experiment. All fifty L2 Arabic learners were born and raised in the Midwest, are undergraduate students at UWM, and started studying Arabic at different stages of their academic lives. They do not speak any other languages besides English, but they had minimal exposure to Spanish in high school. To ensure that all participants had the same amount of language exposure, it was made a criterion of the study that participants have no travel experience to any Arab country, were raised in monolingual families, and do not have any hearing or speaking impairments at the time of their participation in the experiment.

[3] The categorization of the listeners in the three groups reflects only the length of L2 exposure in the classroom. The given labelings (NG, IG, AG) do not reflect the listeners' actual L2 linguistic abilities according to L2 proficiency guidelines which entail their use of grammatical structures, L2 vocabulary…etc.

The remaining five listeners (F = 3, M = 2) out of the total fifty-five participants are native speakers of Arabic who participated as listeners in the Control Group (CG). Their ages ranged from 24 to 30 years old (Mean = 27.2 years old), and they also participated as listeners in the identification and goodness-of-fit rating tasks. All five listeners were born and raised in an Arab country of the Levant region (Jordan) and were taught and exposed to MSA at their schools and colleges. The participants had been living in the United States for a time varying from 4 to 6 years at the time of their participation in the experiment. Their exposure to English also began in the third grade, in a typical EFL classroom setting. All participants were born and raised in monolingual families, and they did not have any hearing or speaking impairments at the time of their participation in the experiment. Table 4.1 below presents a summary of the participating listeners and their basic information.

Table 4.1. Demographics of Participating Listeners.

Listener Groups	**Listeners' Information**				
	Number of Participants	**Gender**	**Age**	**Amount of L2 Exposure**	**Region**
Novice Group	30	F = 18 M = 12	18-20 Mean = 18.8	5 Weeks	Midwest
Intermediate Group	10	F = 6 M = 4	20-39 Mean = 24	8 Months	Midwest
Advanced Group	10	F = 7 M = 3	22-25 Mean = 23.3	12 Months	Midwest
Control Group	5	F = 3 M = 2	24-30 Mean = 27.2	Native Arabic Speakers	Jordan

4.3 Stimuli

The stimuli for this experiment consisted of the Arabic consonants /t, d, ð, s, $t^ʕ$, $d^ʕ$, $ð^ʕ$, $s^ʕ$/ being placed in CV and CV: syllable structures with the short Arabic vowel /a/ and long Arabic vowel /a:/. So, a total of 64 stimuli were generated (8 Arabic consonants x 2 syllable structures x 4 talkers = 64 stimuli). The target stimuli were produced in the carrier phrase:

Yaktub ***stimulus*** *belqalam.* (يَكتُبْ تَ بِالقَلَمْ)

He writes ***stimulus*** *with a pen.*

The Arabic diacritics[4] that denote the short vowels were included throughout the carrier phrases and the target syllables (see sample phrase in Arabic above). The reason for doing so is to avoid any orthographic mismatch representations while reading and recording the phrases. In addition, the 25-year-old male native Arabic talker produced six Arabic consonants /b, f, m, n, r, l/, which are not part of the tested consonants, to be used in the experiment as the practice set. These tokens were generated in CV and CV: syllable structures with the Arabic short vowel /a/ and the Arabic long vowel /a:/ (6 practice consonants x 2 syllable structures = 12 practice tokens). These practice consonants were placed in the same carrier phrase described above. As a result, this has brought the total number of recorded stimuli to 76 (= 64 target stimuli + 12 practice stimuli).

The four native Arabic talkers were met with individually and recorded in a quiet room at UWM using Audacity (2.0.3), installed on a laptop, and a head-mounted microphone (Logitech USB Headset H390). The carrier phrases were typed and printed on flashcards and handed over one at a time to each talker to be read. If any of the talkers misread any of the stimuli, the flashcard was placed aside and handed over later to the talker. Each talker was asked to record the 64 stimuli twice — one recording was done in a Normal Speech Rate (NSR), and the other recording was done in a Fast Speech Rate (FSR). In the NSR condition, talkers were instructed to

[4] Arabic short vowels are represented by diacritics ◌َ ◌ُ ◌ِ which are placed on the consonants (the circles are a representation of a consonant slot).

read the presented stimuli in a normal manner. Before recording the stimuli, the four talkers heard a sample recording of three phrases by the author in a NSR condition. Similarly, in the FSR condition, talkers were instructed to read the stimuli at a fast speech rate. Before recording the stimuli, the four talkers heard three phrases that were recorded by the author in a FSR. Praat was used to measure phrase durations and word intervals to make sure that the speech rates are comparable among the talkers. Having two different speech rates increases the difficulty of the identification and goodness-of-fit rating tasks to be performed later and prevents any difficulty ceiling effect(s). The talkers recorded the stimuli at a normal speech rate first, took a ten-minute break, and then recorded the stimuli at a fast speech rate. Instructions were printed and presented to the talkers and verbally repeated to ensure full understanding of what and how they were supposed to record the phrases. Before recording the stimuli, the talkers were asked to talk briefly about themselves, while the volume and software were tested.

The recorded phrases were saved on a laptop, coded accordingly, and the amplitudes were normalized using Praat. The recorded and normalized phrases were divided and placed into two sets. One contained the normal speech rate Arabic stimuli in the CV and CV: syllable structures with short and long /a, a:/, and the other contained the fast speech rate Arabic stimuli in the CV and CV: syllable structures with short and long /a, a:/. Using Praat (6.0), an identification task and a goodness-of-fit rating task for each one of the two sets were created. A practice set was also created which contained the 12 test stimuli.

4.4. Procedure

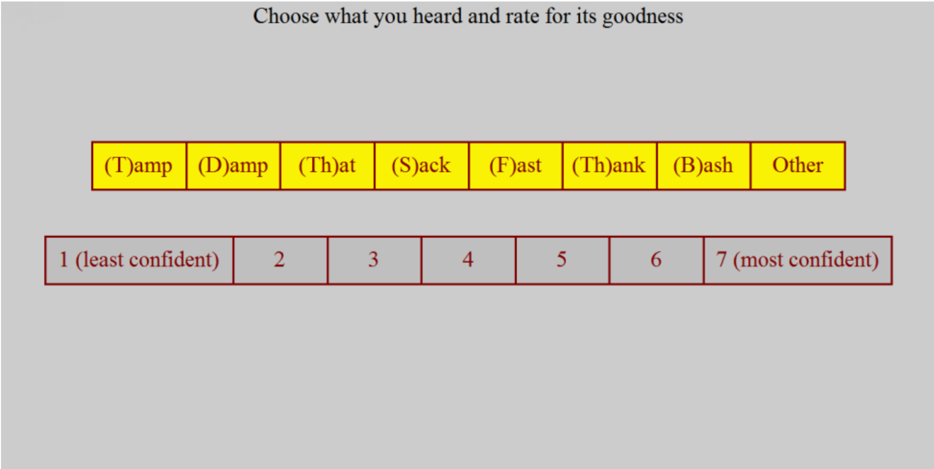
The fifty-five listeners participated in an identification task followed by a goodness-of-fit rating task. Each one of the participating listeners was met with individually in a quiet room at UWM to complete the tasks. Below is a detailed description of the two tasks.

The identification task consisted of two sub-tasks, an L1 labeling task and an L2 labeling task. Below is a detailed description of the two tasks. Even though the section below starts by describing the L1 labeling task, the L2 labeling task was performed first, in order to prevent listeners' L1 categories from affecting their L2 selection (such as if the L1 labeling task was performed first).

4.4.1 L1 Labeling and Goodness-of-Fit Rating Tasks

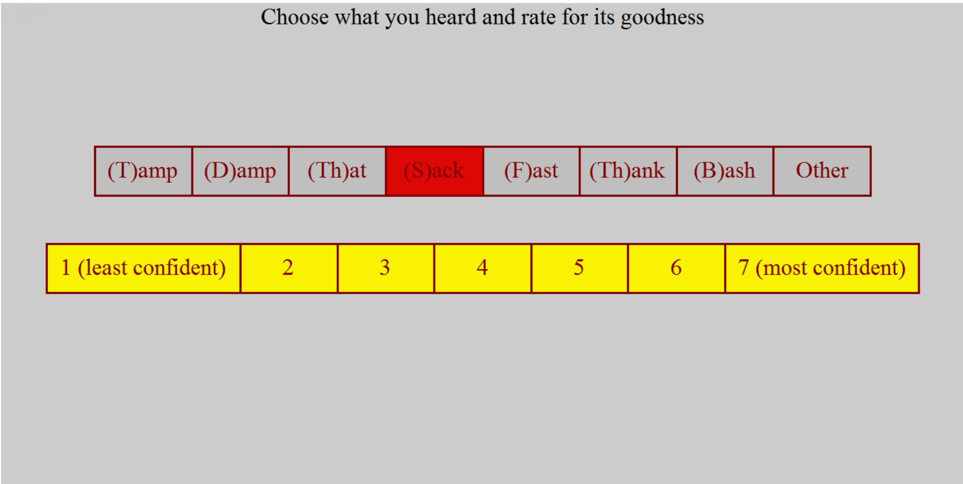
Three learner group listeners (NG, IG, and AG) participated in this task. All listeners were asked to label each one of the Arabic consonants they heard in the stimuli by identifying the consonants according to a provided list of English consonants. The labeling English categories included the following English consonants, /t, d, ð, s, f, θ, b/ and *other*. These categories were chosen based on the findings of the cross-language perceptual mapping between Arabic and English that was conducted earlier in Chapter 3. The English categories were placed in parentheses, in actual English words in an initial position in yellow blocks for the listeners' ease of recognition (see Figure 4.2 below).

Figure 4.2. Screenshot of the L1 Labeling and Goodness-of-Fit Rating Tasks Interface.



So for example, if a listener thought that he or she heard the Arabic consonant /d/ in either /da/ or /da:/, he or she will have to click on one of the given English categories on the computer screen. Once they choose and click on one of the blocks that contain the desired English category, the block will turn red indicating that it was selected (see figure 4.3 below).

Figure 4.3. Sample Screenshot of When an English Category is Selected in the L1 Labeling Task.



After selecting an English category from the provided list in the NSR and FSR, the listeners were asked to determine how similar they think their English selections are by clicking on one of seven provided number choices on a Likert scale. If a listener selected number (1) on the scale, that means that he or she thinks that his or her selection is not similar to what they heard. On the other hand, if a listener selected number (7) on the scale, that means that he or she thinks that his or her selection is very similar to what they heard. This goodness-of-fit rating task followed the identification of each presented stimuli in the L1 labeling task. Listeners are not able to click on a rating number on the scale without choosing one of the given categories first. The blocks in which the numbers are presented start as grey (see Figure 4.2 above). This means that the listener cannot click on the rating numbers, and he or she must choose a category for the stimuli that was heard. After a category has been selected, the rating blocks will turn yellow (see Figure 4.3 above), which means that the listeners can now rate the selected category.

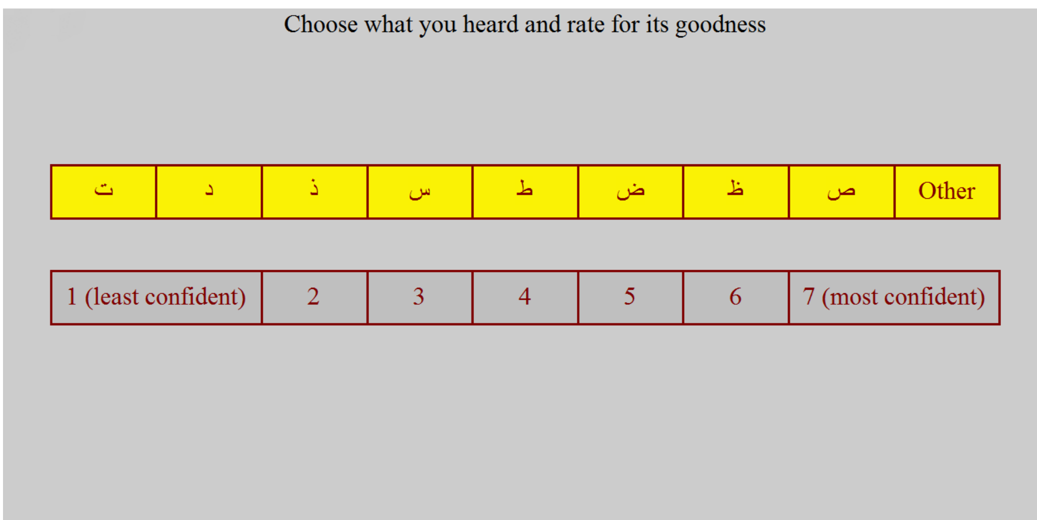
The instructions for this task were presented to the listeners on an instruction sheet before the task and were also verbally reiterated to ensure full understanding of the task. A practice round was conducted before the task to familiarize participants with the task and mouse commands. The participants listened and labeled the 64 stimuli in the FSR condition first. Upon completion of the FSR condition, listeners were given a ten-minute break and then proceeded to listen to and label the 64 stimuli in the NSR condition. Besides the increased level of difficulty, presenting the FSR condition first allowed for a reduction in the familiarity effect by the time this task was complete and listeners attempted the NSR condition. The procedures that were followed here are the same procedures that were applied in the next section for the L2 labeling and goodness-of-fit rating tasks.

All collected data from the identification and goodness-of-fit rating tasks were saved to a laptop and coded accordingly for analysis.

4.4.2 L2 Labeling and Goodness-of-Fit Rating Tasks

The three learner groups (NG, IG, and AG) and the control group (CG) participated in this task. The listeners were asked to label each one of the Arabic consonants that they heard in the stimuli, by identifying the consonants among the Arabic consonant choices provided. The labeling Arabic categories included the following Arabic consonants, /t, d, ð, s, $t^{ʕ}$, $d^{ʕ}$, $ð^{ʕ}$, $s^{ʕ}$/ and *other*. The Arabic categories were presented separately in Arabic orthography in yellow blocks (see Figure 4.4 below).

Figure 4.4. Screenshot of the L2 Labeling and Goodness-of-Fit Rating Tasks Interface.

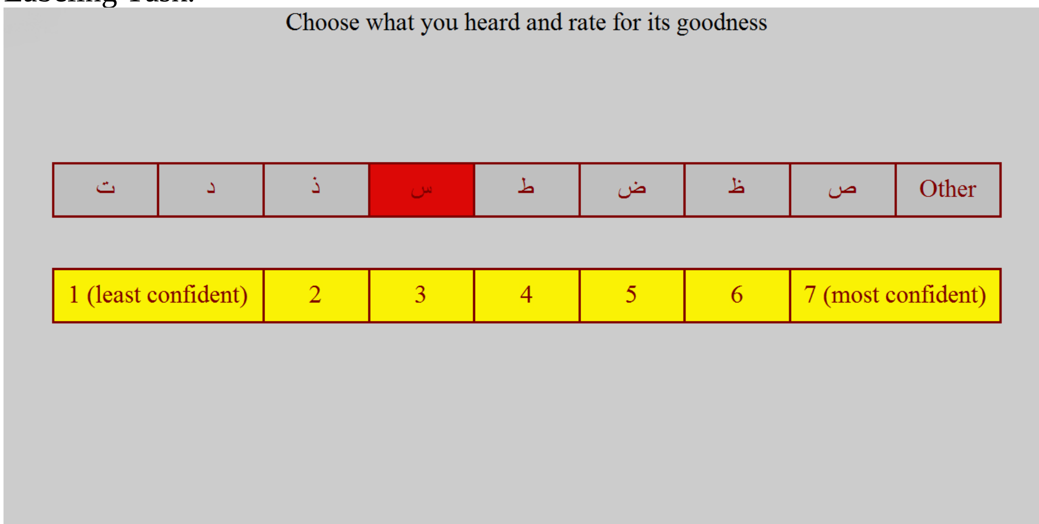


For instance, if a listener thought that he or she heard the Arabic consonant /s/ in either /sa/ or /sa:/, he or she would have to click on one of the given Arabic categories on the computer screen. Once he or she choose and selected of the blocks that contains an Arabic category, the block would turn red, indicating that it was selected (see Figure 4.5 below).

Similar to the procedure of the previous section, after selecting an Arabic category from the provided list in the NSR and FSR, the listeners were asked to determine how confident they were of their Arabic selections, by clicking on one of seven provided number choices on a Likert scale. If a listener selected number (1) on the scale, that means that he or she is least confident of the selection. On the other hand, if a listener selected number (7) on the scale, that means that he or she is most confident of the selection. This goodness-of-fit rating task also followed the identification of each presented stimuli in the L2 labeling task. Listeners were not be able to click on a rating number on the scale without choosing one of the given categories first. The blocks in which the numbers were presented start as grey (see Figure 4.4 above). This means that the listener cannot click on the rating numbers and he or she has to choose a category for the stimuli that they heard. After a category has been selected, the rating blocks will turn yellow (see Figure 4.5 below), and this means that the listeners can now rate the selected category.

The instructions for this task were presented to the listeners on an instruction sheet and were also verbally reiterated to ensure full understanding of the task. A practice round was conducted before the task to familiarize the participants with the task and mouse commands. Participants listened and labeled the 64 stimuli in the FSR condition first. At the completion of the FSR condition, the listeners were given a ten-minute break and then proceeded to listen to and label the 64 stimuli in the NSR condition.

Figure 4.5. Sample Screenshot of When an Arabic Category is Selected in the L2 Labeling Task.



4.5 Data Analysis

The main point of the current analysis is to present a detailed reading of the obtained results from the L1 and L2 labeling tasks and the goodness-of-fit rating tasks, which will allow me to answer my stated research questions. For the second research question: *Does more L2 exposure over time lead to similar learning development with Arabic L2 plain and emphatic sounds?*, I utilize comparisons to investigate whether the mappings between the Arabic sounds and Arabic categories in the L2 labeling task show any differences between the three learner groups. In addition, I will examine the goodness ratings of the selected Arabic categories and explore whether they demonstrate any differences across the three learner groups. In order to address the third research question: *Do L2 learners' L1-to-L2 mappings change over time?*, I examine the mappings between the Arabic sounds and the English categories in the L1 labeling task to see whether the mappings show any differences between the different learner groups. In addition, the similarity ratings that were obtained through the goodness ratings are also examined, in case I did not see any L1-to-L2 mapping pattern differences between the different learner groups. If I noticed any, changes in the overall similarity ratings would allow me to pinpoint any L1-to-L2 differences between the plain and emphatic sounds and discern whether the amount of L2 exposure affects this overall level of similarity.

In order to accomplish that, the analyses follow Park and de Jong's (2008) matrix analyses and examine the results obtained in the L1 labeling and goodness-of-fit rating tasks by presenting the data in matrix tables, showing the proportions of the English category selections and their overall goodness ratings in parentheses.

Table 4.2. Sample Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and Their Mean Ratings (Between Parentheses) in the NG Under NSR Condition with Long Vowel /a:/.

	English Categories							
Arabic Stimuli	**/b/**	**/d/**	**/ð/**	**/f/**	**/s/**	**/t/**	**/θ/**	**Other**
/dˤ/	2.5 (2.67)	**93.33 (6.13)**				4.17 (6)		
/d/		**95 (6.38)**				5 (6)		
/sˤ/					**100 (6.67)**			
/s/			0.83 (3)	0.83 (5)	**98.34 (6.64)**			
/tˤ/	1.67 (3.5)	9.16 (6.27)				**89.17 (6.19)**		
/t/						**100 (6.38)**		
/ðˤ/			**61.67 (5.61)**	4.17 (5.6)		5 (5.5)	29.16 (6.11)	
/ð/			**80 (6.01)**				20 (6)	

Modal responses are bolded.

Table 4.2 above is a sample matrix showing mean percentage labeling of Arabic consonants with English categories in the L1 labeling task for the NG under NSR condition with long Arabic vowel /a:/. The table also presents the mean goodness ratings for each selected English category in parentheses. Modal responses are bolded in the table for ease of recognition and comparison. This sample table also does not omit responses below chance level, which are determined to be any value less than 12.5% in the L1 labeling task (= 100 / 8 choices). For example, based on Table 4.2 above, one can see that when Arabic /dˤ/ was presented, NG listeners selected English /b/ 2.5% of the time, with a mean goodness rating of 2.67, English /d/ 93.33% with a mean goodness rating of 6.13, and English /t/ 4.17% with a mean goodness rating of 6. English selections /b/ and /t/, however, are ignored since they were less than the chance level.

In addition to considering the proportions and mean goodness ratings, the L1 labeling analyses also focus on the number of L1 categories chosen for an L2 category. As noted above, some Arabic L2 sounds will demonstrate one-to-one category mapping (after considering the chance level), like Arabic /dˤ/, while others will demonstrate a one-to-two category mapping, like Arabic /ðˤ/ which was selected as the English /ð/ 61.67% with a mean goodness rating of 5.61, and as English /θ/ 29.16% with a mean goodness rating of 6.11.

In addition to examining the proportions and number of category selections, the analyses examine the mean goodness ratings that were assigned for each selected L1 English category. Examining these patterns allows me to understand how learners' association with L1 and L2 categories change as L2 experience changes — this will help inform my understanding of the third research question above.

I present the findings of the L2 labeling task and its goodness-of-fit rating task in bar graphs, showing the accuracy rates and goodness ratings for each Arabic L2 consonant in the three participating groups under the specific speech rate and vowel conditions.

Figure 4.6. Sample Graph Showing L2 Labeling Results for the NG in NSR Condition with Long Arabic Vowel /a:/.

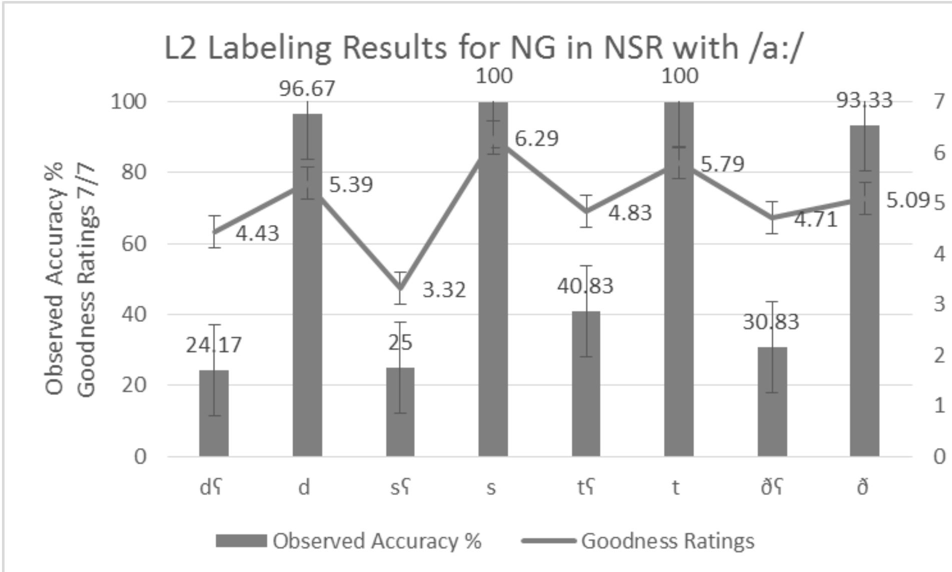


Figure 4.6 above is a sample graph of the L2 labeling task and goodness-of-fit rating task for the NG under the NSR and long Arabic vowel /a:/ conditions. The grey bars and the grey line represents the overall observed accuracy and the overall goodness rating percentage for each Arabic L2 stimulus, respectively. The lines on the bars in the figures represent one calculated standard error. For instance, in Figure 4.6, listeners in the NG labeled Arabic stimulus /d/ as Arabic category /d/ 96.67% with a mean goodness rating of 5.39 in the NSR and when Arabic vowel /a:/ was used.

Similar to the analyses of the L1 labeling task, I examine the mean goodness ratings that were assigned for each selected L2 Arabic category. Examining these patterns will inform an understanding of how learners' performance of specific L2 sounds changes and develops across time as their L2 experience changes.

4.6 Results

4.6.1 L1 Labeling and Goodness-of-Fit Rating Tasks

4.6.1.1 Normal Speech Rate

In this section, I describe and present results obtained from the L1 labeling task and goodness-of-fit rating task under the NSR condition with /a/ and /a:/ for the three listener groups. Based on the results, I notice that there are no substantial differences in the obtained proportions and goodness ratings for each listener group when short /a/ and long /a:/ were used; the average goodness ratings were 6.10 and 6.21 for /a/ and /a:/, respectively, for the NG, 6.08 and 6.25 for /a/ and /a:/, respectively, for the IG, and 6.52 and 6.54 for /a/ and /a:/, respectively for the AG. The overall goodness ratings of the AG seem to be a little higher than the NG and IG. Also, for all three groups, I noticed that all labelings demonstrated a one-to-one Arabic to English mapping pattern except for Arabic /ðˤ/ and /ð/, which demonstrated a one-to-two mapping pattern.

Tables 4.3 and 4.4 present the proportion of L1 English labeling chosen for each L2 Arabic stimulus, with its mean goodness rating in the NG with /a:/ and /a/, respectively. The number of labels selected for each Arabic stimulus show how many English categories are related to a specific Arabic L2 sound. The proportion of each selected category indicates the level of connection strength between the Arabic and English categories. The numbers in parentheses under the proportions represent the overall goodness rating that was assigned to the selected English categories, and they, too, indicate the strength of connection between the Arabic and English categories.

Table 4.3. Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and their Mean Ratings (Between Parentheses) in the NG Under NSR Condition with Long Vowel /a:/.

	English Categories							
Arabic Stimuli	**/b/**	**/d/**	**/ð/**	**/f/**	**/s/**	**/t/**	**/θ/**	**Other**
/dˤ/	2.5 (2.67)	**93.33 (6.13)**				4.17 (6)		
/d/		**95 (6.38)**				5 (6)		
/sˤ/					**100 (6.67)**			
/s/			0.83 (3)	0.83 (5)	**98.34 (6.64)**			
/tˤ/	1.67 (3.5)	9.16 (6.27)				**89.17 (6.19)**		
/t/						**100 (6.38)**		
/ðˤ/			**61.67 (5.61)**	4.17 (5.6)		5 (5.5)	29.16 (6.11)	
/ð/			**80 (6.01)**				20 (6)	

Table 4.4. Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and their Mean Ratings (Between Parentheses) in the NG under NSR Condition with Short Vowel /a/.

	English Categories							
Arabic Stimuli	**/b/**	**/d/**	**/ð/**	**/f/**	**/s/**	**/t/**	**/θ/**	**Other**
/dˤ/	3.33 (4)	**94.17 (5.92)**	1.67 (4.5)			0.83 (5)		
/d/	2.5 (4.67)	**92.5 (6.21)**				5 (6.17)		
/sˤ/					**100 (6.64)**			
/s/					**100 (6.67)**			
/tˤ/		4.17 (5.6)				**95.83 (6.11)**		
/t/						**100 (6.16)**		
/ðˤ/			**85 (5.94)**				14.17 (5.76)	0.83 (2)
/ð/			**67.5 (5.98)**	1.67 (5)	0.83 (1)		30 (5.75)	

On the one hand, a high mean goodness means that the listeners felt that there was a high connection between the L2 Arabic stimulus and the English L1 category that they selected. On the other hand, a low mean goodness rating means that the listeners did not find a strong connection between the produced Arabic stimulus and what they selected, but they did not have any better choice to select.

In terms of the vowel condition, the tables above show no substantial differences, with respect to the selected English categories for the Arabic L2 stimuli. The highest proportions and mean goodness ratings are almost identical for the selected categories with both short Arabic /a/ and long Arabic /a:/. For example, when the short /a/ was used, English category /d/ was selected 94.17% with a mean goodness of 5.92 for the Arabic stimulus /dˤ/. This is almost identical in outcome to when the long /a:/ was used; English category /d/ was selected 93.33% with a mean goodness of 6.13 for the same Arabic stimulus /dˤ/. The same applies to all other Arabic stimuli; the proportions and mean goodness ratings were almost identical for the highest selected English categories.

Regarding the number of selected English categories per Arabic Stimulus, the tables show no difference under the short vowel condition and long vowel condition. For example, after considering the chance level (any value less than 12.5% will not be accounted for), all Arabic sounds in both tables, except Arabic /ðˤ/ and /ð/, demonstrate a one-to-one mapping pattern with no substantial difference in the overall proportion. For instance, in Table 4.3 with long /a:/, Arabic /dˤ/ was labeled as English /d/ 93.33%, and in Table 4.4 with short /a/, Arabic /dˤ/ was labeled as English /d/ 94.17%. The same pattern applies to all other Arabic categories. However, Arabic /ðˤ/ and /ð/ in both tables were mapped to two English categories, rather than one. For example, in Table 4.3, Arabic /ðˤ/ was labeled as English /ð/ 61.67% and as English /θ/ 29.16%.

Similarly, in Table 4.4, Arabic /ð^{ʕ}/ was labeled onto the same English categories 85% and 14.17%, respectively. Arabic /ð/ was also labeled onto two English categories in both tables; 80% as English /ð/ and 20% as English /θ/ in Table 4.3, and 67.5% as English /ð/ and 30% as English /θ/ in Table 4.4. More than one English category selection per Arabic stimulus means that the listeners were confused about the identity of the heard Arabic sound and chose other English categories at some point.

With regard to the mean goodness ratings, the results in both tables indicate no substantial differences in the overall given goodness ratings. For example, from the same examples given above, in Table 4.3 with long /a:/, Arabic /d^{ʕ}/ was labeled as English /d/ 93.33% with a goodness rating of 6.13, and in Table 4.4 with short /a/, Arabic /d^{ʕ}/ was labeled as English /d/ 94.17% with a goodness rating of 5.92. The same pattern was observed for Arabic sounds that were mapped onto two English categories; no substantial differences are present in the overall goodness ratings.

Tables 4.5 and 4.6 above present the proportion of English labeling that was chosen for each L2 Arabic stimulus, with its mean goodness rating in the IG with /a:/ and /a/, respectively. In terms of vowel condition, the tables above show no substantial differences, with respect to the selected English categories for the Arabic L2 stimuli. The highest proportions and mean goodness ratings are almost identical for the selected categories with both short Arabic /a/ and long Arabic /a:/. For example, when the short /a/ was used in Table 4.6, English category /d/ was selected 97.5% with a mean goodness of 6.08 for the Arabic stimulus /d/. This is almost identical to when the long /a:/ was used in Table 4.5; English category /d/ was selected 100%, with a mean goodness of 6.25 for the same Arabic stimulus /d/. The same applies to all other Arabic stimuli;

the proportions, and mean goodness ratings were almost identical for the highest selected English categories.

Table 4.5. Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and Their Mean Ratings (in Parentheses) in the IG Under NSR Condition with Long Vowel /a:/.

English Categories								
Arabic Stimuli	**/b/**	**/d/**	**/ð/**	**/f/**	**/s/**	**/t/**	**/θ/**	**Other**
/dˤ/		**100 (6.23)**						
/d/		**100 (6.25)**						
/sˤ/					**100 (6.53)**			
/s/					**100 (6.48)**			
/tˤ/			2.5 (6)			**97.5 (6)**		
/t/						**100 (6.2)**		
/ðˤ/			**72.5 (6.03)**				27.5 (6.36)	
/ð/		2.5 (6)	**80 (6.22)**				17.5 (6.29)	

Table 4.6. Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and their Mean Ratings (Between Parentheses) in the IG under NSR Condition with Short Vowel /a/.

	English Categories							
Arabic Stimuli	**/b/**	**/d/**	**/ð/**	**/f/**	**/s/**	**/t/**	**/θ/**	**Other**
/dˤ/		**95 (6.11)**	2.5 (5)			2.5 (7)		
/d/		**97.5 (6.08)**				2.5 (5)		
/sˤ/					**100 (6.3)**			
/s/					**100 (6.55)**			
/tˤ/			2.5 (6)			**97.5 (6.03)**		
/t/						**100 (6.1)**		
/ðˤ/			**77.5 (6.13)**				22.5 (5.67)	
/ð/			**70 (6.39)**				30 (5.5)	

Similar to the findings in the NG, in terms of the number of selected English categories per Arabic stimulus, the data in the tables show no difference under the short and long vowel conditions. After considering the chance level, for example, all Arabic sounds in both tables except Arabic /ðˤ/ and /ð/ demonstrate a one-to-one mapping pattern with no substantial difference.

With regard to the mean goodness ratings, the results in both tables also indicate no substantial difference in the overall given goodness ratings. For example, Arabic /s/ was mapped onto English /s/ 100%, with a goodness rating of 6.48 in Table 4.5, and it was also mapped onto English /s/ 100% with a goodness rating of 6.55 in Table 4.6. The same pattern was observed for

Arabic sounds that were mapped onto two English categories, with no substantial differences in the overall goodness ratings.

Comparing the results of the NG with the results of the IG in the tables above reveals no difference in terms of the highly selected English labels, even though the proportions changed at times for a number of selected English categories.

Table 4.7 below presents the proportion of L1 English labeling that was chosen for each L2 Arabic stimulus, with its mean goodness rating in the AG with long vowel /a:/. And Table 4.8 below presents the proportion of English labeling that was chosen for each L2 Arabic stimulus, with its mean goodness rating in the AG with short vowel /a/. Similar to the results in the NG and IG, the results in the tables below indicate that there are no differences among the highly selected English categories, when short vowel /a/ and long vowel /a:/ were used.

Table 4.7. Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and Their Mean Ratings (Between Parentheses) in the AG Under NSR Condition with Long Vowel /a:/.

	English Categories							
Arabic Stimuli	**/b/**	**/d/**	**/ð/**	**/f/**	**/s/**	**/t/**	**/θ/**	**Other**
/dˤ/		**100 (6.55)**						
/d/		**100 (6.65)**						
/sˤ/					**100 (6.7)**			
/s/					**100 (6.65)**			
/tˤ/			2.5 (6)			**97.5 (6.41)**		
/t/						**100 (6.63)**		
/ðˤ/			**70 (6.39)**				30 (6.5)	
/ð/			**82.5 (6.58)**				17.5 (6.43)	

Table 4.8. Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and their Mean Ratings (Between Parentheses) in the AG under NSR Condition with Short Vowel /a/.

	English Categories							
Arabic Stimuli	**/b/**	**/d/**	**/ð/**	**/f/**	**/s/**	**/t/**	**/θ/**	**Other**
/dˤ/		**100 (6.6)**						
/d/		**97.5 (6.56)**				2.5 (7)		
/sˤ/					**100 (6.63)**			
/s/					**100 (6.73)**			
/tˤ/						**100 (6.4)**		
/t/						**100 (6.55)**		
/ðˤ/			**75 (6.47)**				25 (6.4)	
/ð/			**70 (6.54)**				30 (6.33)	

Similar to the findings in the NG and the IG, in terms of the number of selected English categories per Arabic stimulus, the tables above show no difference under the short and long vowel conditions. For example, after considering the chance level, all Arabic sounds in both tables except Arabic /ðˤ/ and /ð/ demonstrate a one-to-one mapping pattern with no substantial difference in the overall proportion.

The mean goodness rating results in both tables also indicate no substantial differences in the overall given goodness ratings. For example, Arabic /t/ was mapped onto English /t/ 100% with a goodness rating of 6.63 in Table 4.7 and was also mapped onto English /t/ 100% with a goodness rating of 6.55 in Table 4.8. The same pattern was observed for Arabic sounds that were mapped onto two English categories, no substantial differences are present in the overall goodness ratings.

By comparing the results of the AG with the results of NG and IG in the tables above I can see that there are no differences in terms of the commonly selected English categories; there was no substantial change in the proportions of the selected English categories among the three groups, and the same English categories were selected for the Arabic stimuli. All labelings demonstrated a one-to-one mapping pattern, except Arabic /ðˤ/ and /ð/, which both demonstrated a one-to-two mapping pattern onto English /ð/ and /θ/. With regards to the goodness ratings, the results also show no substantial differences in the given goodness for the three listener groups. The goodness ratings of the AG, however, seem to be a little higher than the NG and IG.

4.6.1.2 Fast Speech Rate

In this section, I will describe and present results obtained from the L1 labeling task and goodness-of-fit rating task, under the FSR condition with /a/ and /a:/ for the three listener groups. The results indicate that there are no substantial differences in the obtained proportions and goodness ratings between the NG and IG when short /a/ and long /a:/ were used. Results from the AG, however, show some differences in the selected categories and proportions between /a/ and /a:/. The goodness ratings of the IG and AG seem to be higher than the goodness ratings of the NG; the average goodness ratings were 5.69 and 5.72 for /a/ and /a:/, respectively, for the NG, 6.24 and 6.31 for /a/ and 6.31, respectively, for the IG, and 6.23 and 6.20 for /a/ and /a:/, respectively, for the AG. Also, for the NG and IG groups, all labelings demonstrated a one-to-one Arabic to English mapping pattern, except for Arabic /ðˤ/ and /ð/ which demonstrated a one-to-two mapping pattern, with some slight differences between /a/ and /a:/. On the other hand, the AG labelings mainly demonstrate a one-to-two mapping pattern when /a:/ was used and mainly a one-to-one mapping pattern when /a/ was used.

Tables 4.9 and 4.10 below present the proportions of English labelings that were chosen for each L2 Arabic stimulus with their mean goodness ratings in the NG with /a:/ and /a/, respectively. Similar to what was explained in the NSR section, the number of labels selected for each Arabic stimulus show how many English categories are related to a specific Arabic L2 sound. The proportion of each selected category indicates the level of connection strength between the Arabic and English categories. The numbers in parentheses under the proportions represent the overall goodness rating that was assigned to the selected English categories and they, too, indicate the strength of connection between the Arabic and English categories.

In terms of the vowel condition, the tables below show no substantial differences with respect to the selected English categories for the Arabic L2 stimuli. The highest proportions and mean goodness ratings are almost identical for the selected categories with both the short Arabic /a/ and long Arabic /a:/. For instance, when the short /a/ was used, the English category /d/ was selected 100% with a mean goodness of 5.8 for the Arabic stimulus /dˤ/. This is almost identical to when the long /a:/ was used; English category /d/ was selected 100% with a mean goodness of 5.89 for the same Arabic stimulus /dˤ/. The same applies to all other Arabic stimuli; the proportions and mean goodness ratings were almost identical for the highest selected English categories.

Table 4.9. Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and their Mean Ratings (Between Parentheses) in the NG Under FSR Condition with Long Vowel /a:/.

	English Categories							
Arabic Stimuli	**/b/**	**/d/**	**/ð/**	**/f/**	**/s/**	**/t/**	**/θ/**	**Other**
/dˤ/		**100 (5.89)**						
/d/		**100 (5.8)**						
/sˤ/					**100 (6.09)**			
/s/					**100 (6.25)**			
/tˤ/		5 (6)				**95 (5.55)**		
/t/						**100 (5.46)**		
/ðˤ/			**60 (5.32)**				40 (5.63)	
/ð/			**95 (5.54)**				5 (5.67)	

Table 4.10. Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and their Mean Ratings (in Parentheses) in the NG under FSR Condition with Short /a/.

	English Categories							
Arabic Stimuli	**/b/**	**/d/**	**/ð/**	**/f/**	**/s/**	**/t/**	**/θ/**	**Other**
/dˤ/		**100 (5.8)**						
/d/	0.83 (5)	**99.17 (5.44)**						
/sˤ/					**100 (6.08)**			
/s/					**100 (6.07)**			
/tˤ/		11.67 (5.64)				**88.33 (5.95)**		
/t/						**100 (5.72)**		
/ðˤ/			**95 (5.44)**				5 (5.17)	
/ð/			**65 (5.49)**				35 (5.24)	

In terms of the number of selected English categories per Arabic Stimulus, the tables show almost similar findings under the short vowel condition and long vowel condition. After considering the chance level, all Arabic sounds demonstrated a one-to-one Arabic to English mapping pattern with no substantial difference in the in the overall proportion, except for Arabic /ðˤ/ in the long vowel condition and Arabic /ð/ in the short vowel condition — both of which demonstrated a one-to-two mapping pattern. For instance, demonstrating a one-to-one mapping pattern, in Table 4.9, Arabic /dˤ/ was mapped onto English /d/ 100%, and in Table 4.10 Arabic /dˤ/ was also mapped onto English /d/ 100%. On the other hand, demonstrating a one-to-two mapping pattern, in Table 4.9, Arabic /ðˤ/ was mapped onto English /ð/ 60% and onto English /θ/ 40%, and in Table 4.10, Arabic /ð/ was mapped onto English /ð/ 65% and onto English /θ/ 35%.

With regard to the mean goodness ratings, the results in both tables indicate no substantial differences in the overall given goodness ratings. From the same examples given in Table 4.9 above, with long /a:/, Arabic /dˤ/ was labeled as English /d/ 100% with a goodness rating of 5.89, and in Table 4.10 with short /a/, Arabic /dˤ/ was labeled as English /d/ 100% with a goodness rating of 5.8. The same pattern was observed for the Arabic sounds that were mapped onto two English categories, with no substantial differences in the overall goodness ratings.

Tables 4.11 and 4.12 below present the proportion of English labeling that was chosen for each L2 Arabic stimulus, with its mean goodness rating in the IG with /a:/ and /a/, respectively. In terms of the vowel condition, the tables below show no substantial differences with respect to the selected English categories for the Arabic L2 stimuli. The highest proportions and mean goodness ratings are almost identical for the selected categories with /a/ and /a:/.

Table 4.11. Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and their Mean Ratings (Between Parentheses) in the IG Under FSR Condition with /a:/.

English Categories								
Arabic Stimuli	**/b/**	**/d/**	**/ð/**	**/f/**	**/s/**	**/t/**	**/θ/**	**Other**
/dˤ/		**95 (6.29)**	5 (6)					
/d/		**100 (6.3)**						
/sˤ/					**100 (6.15)**			
/s/					**100 (6.5)**			
/tˤ/		2.5 (7)	2.5 (3)			**95 (6.29)**		
/t/		2.5 (6)				**97.5 (6.41)**		
/ðˤ/			**80 (5.97)**				20 (6.5)	
/ð/			**82.5 (6)**				17.5 (6.71)	

Table 4.12. Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and Their Mean Ratings (Between Parentheses) in the IG under FSR Condition with /a/.

English Categories								
Arabic Stimuli	**/b/**	**/d/**	**/ð/**	**/f/**	**/s/**	**/t/**	**/θ/**	**Other**
/dˤ/		**100 (6.25)**						
/d/		**100 (6.3)**						
/sˤ/					**100 (6.18)**			
/s/					**100 (6.45)**			
/tˤ/			5 (4.5)			**92.5 (5.97)**		2.5 (6)
/t/		2.5 (5)				**97.5 (6.23)**		
/ðˤ/			**90 (6.17)**				10 (6.75)	
/ð/			**67.5 (6.19)**				32.5 (6.46)	

When the short /a/ was used in Table 4.12, English category /d/ was selected 100% with a mean goodness of 6.3 for the Arabic stimulus /d/. This is identical to when the long /a:/ was used in Table 4.11; English category /d/ was selected 100% with a mean goodness of 6.3 for the same Arabic stimulus /d/. The same applies to all other Arabic stimuli; the proportions and mean goodness ratings were almost identical for the highest selected English categories.

In terms of the number of selected English categories per Arabic Stimulus, the tables show no difference under the short and long vowel conditions, except for Arabic /ðˤ/ in Table 4.12. After considering the chance level, all Arabic sounds in both tables demonstrate a one-to-one mapping pattern, with no substantial difference in the overall proportion. Arabic /ðˤ/ and /ð/ in Table 4.11, however, demonstrate a one-to-two mapping pattern while Arabic /ð/ in Table 4.12 demonstrates the same one-to-two mapping pattern, Arabic /ðˤ/ in Table 4.12 demonstrates a one-to-one mapping pattern.

With regard to the mean goodness ratings, the results in both tables also indicate no substantial differences in the overall given goodness ratings.

Table 4.13 below presents the proportion of English labeling that was chosen for each L2 Arabic stimulus with its mean goodness rating in the AG with long vowel /a:/. Whereas Table 4.14 below presents the proportion of English labeling that was chosen for each L2 Arabic stimulus with its mean goodness rating in the AG with short vowel /a/. The results in the two tables indicate no differences among the commonly selected English categories for when short vowel /a/ and long vowel /a:/ were used.

In terms of the number of selected English categories per Arabic Stimulus, the tables below show a different pattern from what had been observed above. In Table 4.13, Arabic /d/,

/sˤ/, /s/, and /t/ demonstrated a one-to-one mapping pattern. On the other hand, in the same table, Arabic /dˤ/, /tˤ/, /ðˤ/, and /ð/ demonstrated a one-to-two mapping pattern. In Table 4.14, only Arabic /ð/ demonstrated a one-to-two mapping pattern; all other Arabic sounds demonstrated a one-to-one mapping pattern.

With regard to the mean goodness ratings, the results in both tables also indicate no substantial differences in the overall given goodness ratings.

Table 4.13. Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and Their Mean Ratings (in Parentheses) in the AG Under FSR Condition with Long Vowel /a:/.

	English Categories							
Arabic Stimuli	**/b/**	**/d/**	**/ð/**	**/f/**	**/s/**	**/t/**	**/θ/**	**Other**
/dˤ/	5 (6)	**72.5 (6.14)**	15 (6)			7.5 (6.67)		
/d/		**97.5 (6.23)**	2.5 (7)					
/sˤ/					**100 (6.15)**			
/s/					**100 (6.5)**			
/tˤ/		2.5 (7)	17.5 (5.71)			**75 (6.2)**	5 (7)	
/t/		2.5 (6)	2.5 (3)			**95 (6.39)**		
/ðˤ/			**80 (5.97)**				20 (6.5)	
/ð/		2.5 (5)	**80 (5.94)**				17.5 (6.71)	

Table 4.14. Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and Their Mean Ratings (in parentheses) in the AG under FSR Condition with Short Vowel /a/.

English Categories								
Arabic Stimuli	**/b/**	**/d/**	**/ð/**	**/f/**	**/s/**	**/t/**	**/θ/**	**Other**
/dˤ/	2.5 (6)	**90 (6.33)**	2.5 (7)			2.5 (5)	2.5 (4)	
/d/		**95 (6.26)**	5 (7)					
/sˤ/					**97.5 (6.18)**			2.5 (5)
/s/					**97.5 (6.44)**			2.5 (6)
/tˤ/			10 (5.75)			**87.5 (5.91)**	2.5 (7)	
/t/		2.5	2.5 (7)			**92.5 (6.24)**	2.5 (7)	
/ðˤ/			**90 (6.17)**				10 (6.75)	
/ð/			**67.5 (6.11)**				32.5 (6.46)	

According to a comparison of the results of the AG and the results of the NG and IG in the tables above, there are no differences in terms of the commonly selected English categories when /a/ and /a:/ were used in the NG and IG. The AG, however, showed lower proportions for some of the selected English categories when long /a:/ was used. Clearly, there appears to be more confusion among the listeners in the AG group, which is manifested by the differences in the number of selected categories for the Arabic consonants; this, in turn, affected the overall proportions. This could be linked to the increased level of difficulty of the FSR task. For instance, in the AG, when long /a:/ was used, Arabic /dˤ/, /tˤ/, /ðˤ/, and /ð/ demonstrated a one-to-two Arabic to English mapping pattern. Arabic /d/, /sˤ/, /s/, and /t/, on the other hand, demonstrated a one-to-one mapping pattern. Contrary to these findings, when short /a/ was used, all Arabic consonants demonstrated a one-to-one mapping pattern, except for Arabic /ð/, which demonstrated a one-to-two mapping pattern. The results of the NG and IG showed no differences except for Arabic /ðˤ/ and /ð/, as explained earlier.

With regard to the goodness ratings, the three groups did not show any substantial differences among the given goodness ratings to the modal proportions.

Finally, in comparing the results in the FSR with the NSR for the same three groups, the NG demonstrates similar mapping patterns between Arabic and English categories for both vowels, with one exception. All Arabic consonants show a one-to-one Arabic to English mapping pattern, except Arabic /ðˤ/ and /ð/, which show a one-to-two mapping pattern in the NSR and FSR. In the FSR, however, Arabic /ð/ shows a one-to-one mapping pattern with /a:/ and Arabic /ðˤ/ shows a one-to-one mapping pattern with /a/. Results from the IG and AG also revealed similar mapping patterns between Arabic and English categories for both vowels, with similar NG exceptions.

With regards to the goodness ratings, the results show that the overall mean goodness ratings were slightly higher in in both the NSR and FSR for the AG (M = 6.4, SD = .40) than in the NG (M = 5.9, SD = .233) and IG (M = 6.2, SD = .554). To confirm my suspicions, I conducted a one-way ANOVA to compare the effect of L2 exposure on goodness ratings of labeled L1 English categories. The results indicate that there was a significant L2 exposure effect on the goodness results of labeled L1 English categories at the p <.01 level for the three groups $F(2, 47) = 5.9$, $p = .005$.

Figure 4.7. Mean L1 Goodness for Three Learner Groups in NSR and FSR.

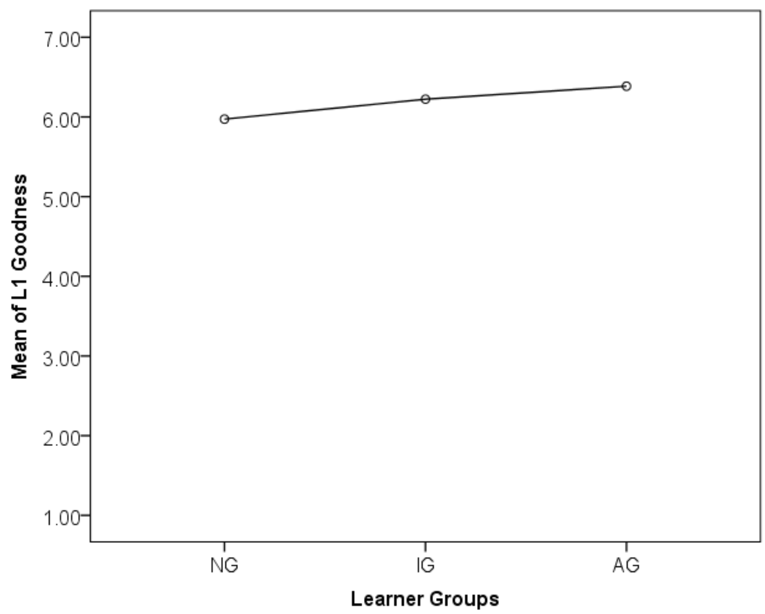


Figure 4.7 above shows the significant difference that was found with regards to the L1 goodness ratings among the three learner groups. It is clear that the goodness ratings of the IG were higher than those of the NG and the goodness ratings of the AG were higher than those of both the NG and IG.

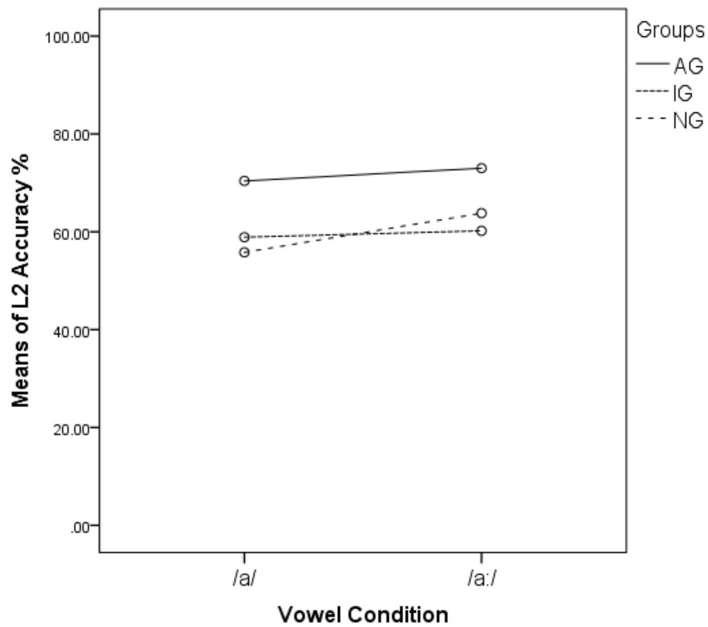
To address my third research question: *Do L2 learners' L1-to-L2 mappings change over time?*, it appears that the L1-to-L2 mappings are not affected by more L2 exposure, based on the mapping pattern results; the proportions did not demonstrate differences between the three listener groups. Any minor differences in the FSR could be attributed to listener confusion, created by the increased difficulty of the FSR task. On the other hand, the overall goodness ratings were higher in the AG than the NG and IG, indicating that the relationship between L1 and L2 sounds becomes stronger with more L2 experience.

4.6.2. L2 Labeling and Goodness-of-Fit Rating Tasks

4.6.2.1 Normal Speech Rate

This section will describe and present results obtained from the L2 labeling task and goodness-of-fit rating task, under the NSR condition with /a/ and /a:/ for the three listener groups. Based on the results, the three listener groups show better overall accuracies in the long /a:/ condition than short /a/ condition. I analyzed the data by running a mixed-design ANOVA with a within-subjects factor of *vowel duration* (short vowel /a/ vs. long vowel /a:/) and a between-subject factor of *groups* (NG vs. IG vs. AG). The results indicate that there was a main effect of *vowel duration*, $F(1,47) = 8.36$, $p = .006$, without any significant interaction among the *groups*, $F(2, 47) = 3.08$, $p = .055$. This means that the accuracies were higher for the long vowel condition /a:/ than for the short vowel /a/ condition, and this trend was generally true for all three groups. Since I observe a trend of interaction ($p = .055$), I examined the accuracy patterns among the groups. The accuracy difference between the long and short vowels was larger among the NG than other groups (see Figure 4.8 below), suggesting that the short vowel /a/ condition was more difficult than the long vowel /a:/ condition for the NG. However, as time goes on and with more L2 exposure, L2 accuracies for the short vowel /a/ condition improve and start to catch up with the long vowel /a:/ condition, suggesting that L2 exposure affects overall L2 accuracy with the short vowel /a/, as seen in the figure.

Figure 4.8. L2 Mean Accuracy Comparisons between NG, IG, and AG in NSR with Short and Long Vowel Conditions.



Accuracies of plain sounds were higher than accuracies of emphatic sounds in the three groups. I ran a mixed-design ANOVA with the *consonants* (plain versus emphatic) and *vowels* (short /a/ versus long /a:/) as the within-subjects factor and *learner groups* (NG vs. IG vs. AG) as the between-subject factor. The results indicate that there was a main effect of consonant type, $F(1,47) = 243$, $p = .000$, with a significant interaction among the groups, $F(2, 47) = 85$, $p = .000$. This means that the accuracies were higher for the plain sounds than the emphatic sounds, and this trend was not true for all three groups in general. Unlike the accuracy results of plain sounds in the IG and AG, accuracy results of plain sounds in the short vowel condition in the NG were lower than the long vowel condition. The results also showed a significant interaction between the vowels and consonants at the $p < .01$ level: $F(1, 47) = 7.7$, $p = .008$.

Figures 4.9, 4.10, and 4.10 show the mean accuracy results for the NG, IG, and AG, respectively. The x-axis indicates the two vowel conditions that were used: short and long. The y-axis indicates the mean L2 accuracy. The solid line represents the L2 emphatic sounds, while the dotted line represents the L2 plain sounds. A comparison of figures reveals that, within each group, the performance of the plain consonants was higher than that of the emphatic consonants. One exception is in the AG, where performance of the plain consonants in the long vowel condition is similar to the performance of the emphatic consonants in the long vowel condition. For example, the mean performance of the NG was 99% and 30% for the plain and the emphatic consonants, respectively, in the short vowel condition (Figure 4.9). The same observation applies to the other groups. However, the magnitude of difference between the plain and emphatic sounds seems to be decreasing with more L2 exposure. In addition, the performance was generally higher when the long vowel condition was applied for both plain and emphatic consonants. In Figure 4.11, the AG's performance of the emphatic sounds with long /a:/ was higher than when short /a/ was used: 73% versus 64%.

Figure 4.9. L2 Mean Accuracy Comparison between Plain and Emphatic Sounds in NSR with Short and Long Vowel Conditions for the NG.

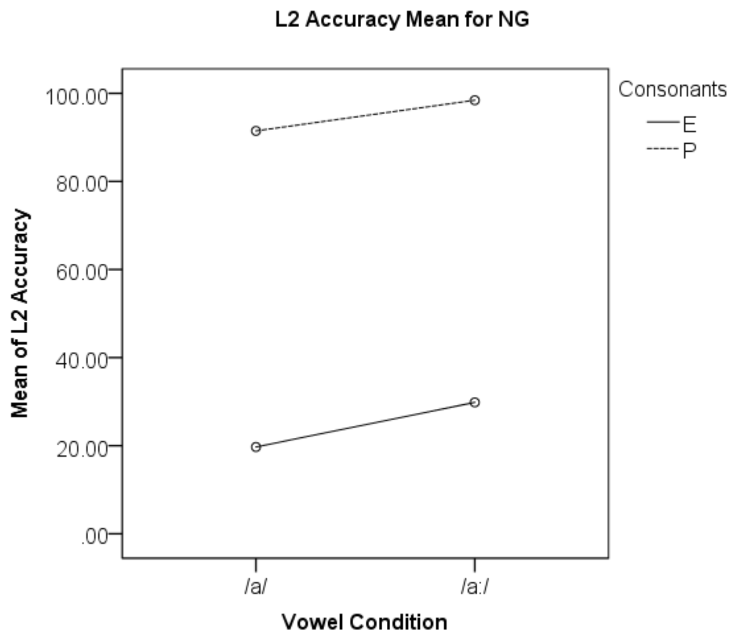


Figure 4.10. L2 Mean Accuracy Comparison between Plain and Emphatic Sounds in NSR with Short and Long Vowel Conditions for the IG.

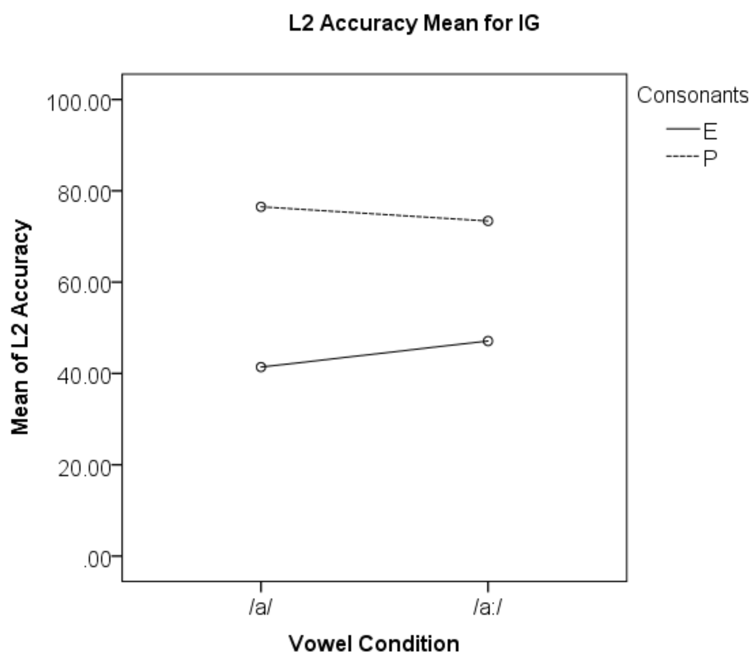
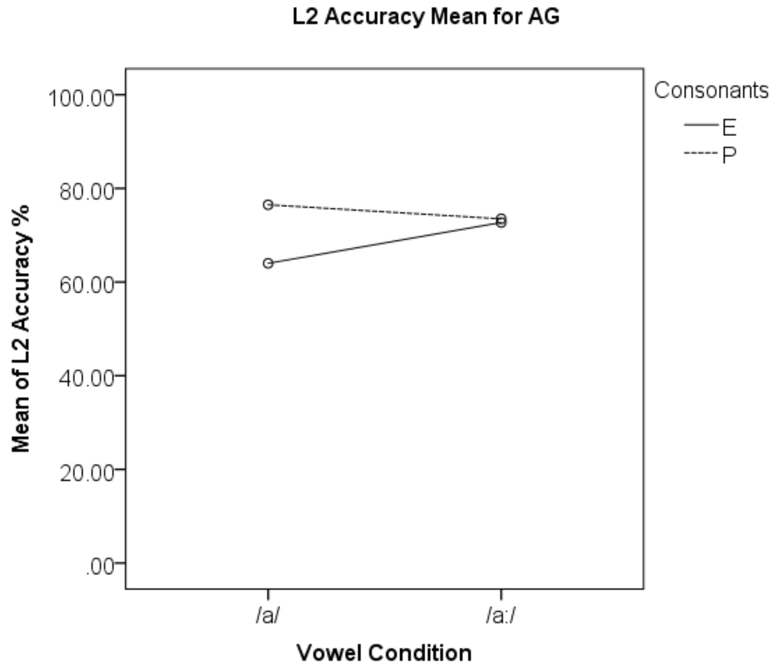


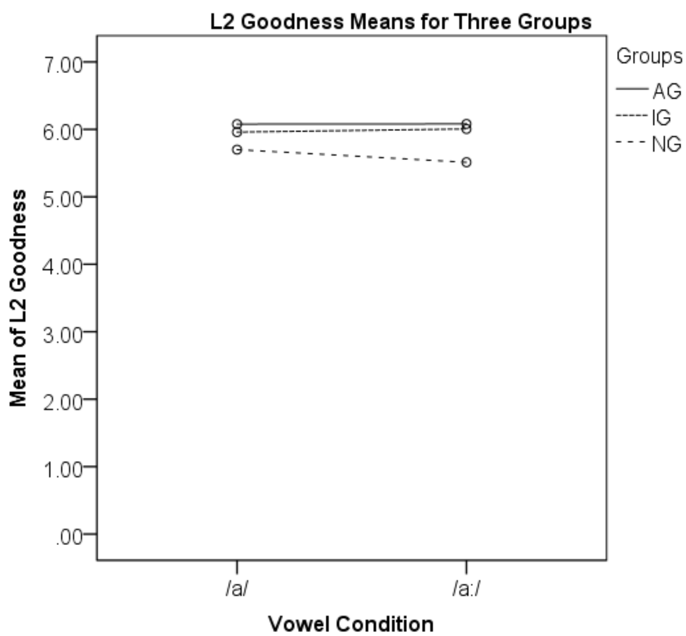
Figure 4.11. L2 Mean Accuracy Comparison between Plain and Emphatic Sounds in NSR with Short and Long Vowel Conditions for the AG.



The abovementioned results indicate that the performance of the emphatic sounds improves with more L2 exposure. As for the plain sounds, it seems that the performance is not affected much with L2 exposure.

With regard to goodness ratings, the results show no difference between goodness ratings when /a/ and /a:/ were used within each group. I analyzed the data by running a mixed-design ANOVA with *vowel duration* (short vowel /a/ vs. long vowel /a:/) as the within-subjects factor and *listener groups* (NG vs. IG vs. AG) as the between-subjects factor. The results obtained indicate that there was no main effect of *vowels*, $F(1, 47) = 1.68$, $p = .201$, with a significant interaction among the groups: $F(2, 47) = 5.73$, $p = .006$. This trend, however, was not true for the NG. Figure 4.12 below demonstrates the mean goodness scores for the three learner groups in NSR with short and long vowel conditions.

Figure 4.12. L2 Mean Goodness Comparisons between NG, IG, and AG in Short and Long Vowel Conditions in NSR.



The x-axis in the figure represents the two short and long vowel conditions, and the y-axis represents the mean goodness for the L2 consonants. It is not very clear from the figure whether the obtained mean in the short vowel /a/ condition is higher than the long /a:/ condition. In the same figure, the performance is highest between the NG and the AG. There is a slight increase between the IG and AG. This also suggests that L2 exposure positively affects the overall confidence ratings given for L2 sounds.

With regard to the plain and emphatic consonants, the goodness of plain sounds was better than the goodness of emphatic sounds in the three groups. I ran a mixed-design ANOVA with the *consonants* and *vowel duration* (plain vs. emphatic; short /a/ vs. long /a:/) as the within-subjects factor, and *groups* (NG vs. IG vs. AG) as the between-subjects factor. The results indicate that there was a main effect of consonant type, $F(1, 47) = 5.12$, $p = .028$, without a significant interaction among the groups, $F(2, 47) = 2.9$, $p = .065$. This means that the goodness was higher for the plain sounds than for the emphatic sounds, and this trend was also true for all the three groups in general. The results also showed no significant interaction between the *vowels* and *consonants*: $F(1, 47) = 1.68$, $p = .201$.

Figures 4.13, 4.14, and 4.15 show the mean goodness results for the NG, IG, and AG, respectively. The x-axis indicates the two vowel conditions that were used: short and long. The y-axis indicates the mean of L2 goodness. The blue line represents the L2 emphatic sounds, while the green line represents the L2 plain sounds. By comparing the figures within the NG group, one can see that the performance of the plain consonants was higher than the emphatic consonants. For example, the mean performance of the NG was 5.8 for the plain consonants in the short vowel condition versus 5.5 for the emphatic consonants in the same short vowel condition (Figure 4.13). In addition, Figures 4.14 and 4.15 demonstrate that mean goodness for both plain and emphatic, under the two vowel conditions, are almost identical, suggesting no change of performance with more L2 exposure.

Figure 4.13. L2 Mean Goodness Comparison between Plain and Emphatic Sounds in NSR with Short and Long Vowel Conditions for the NG.

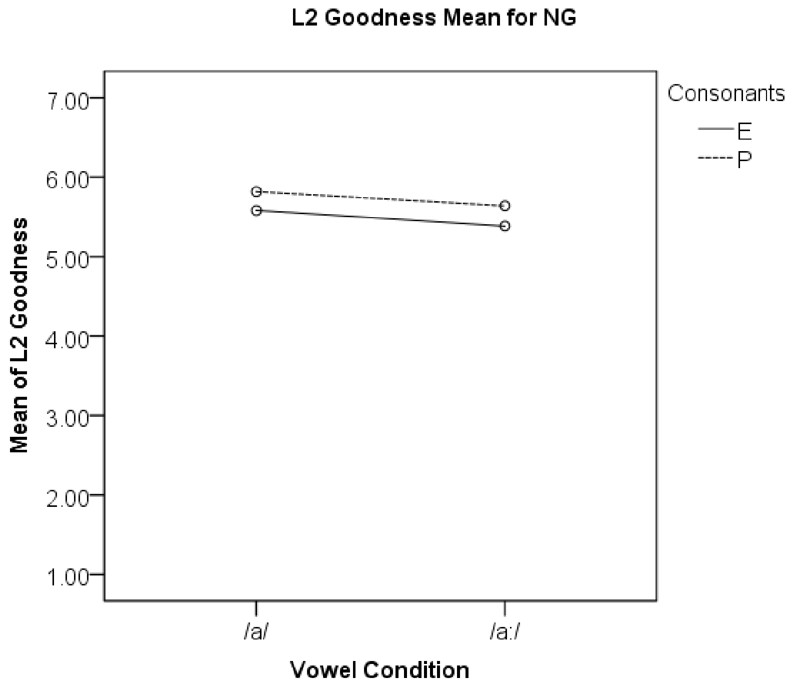


Figure 4.14. L2 Mean Goodness Comparison between Plain and Emphatic Sounds in NSR with Short and Long Vowel Conditions for the IG.

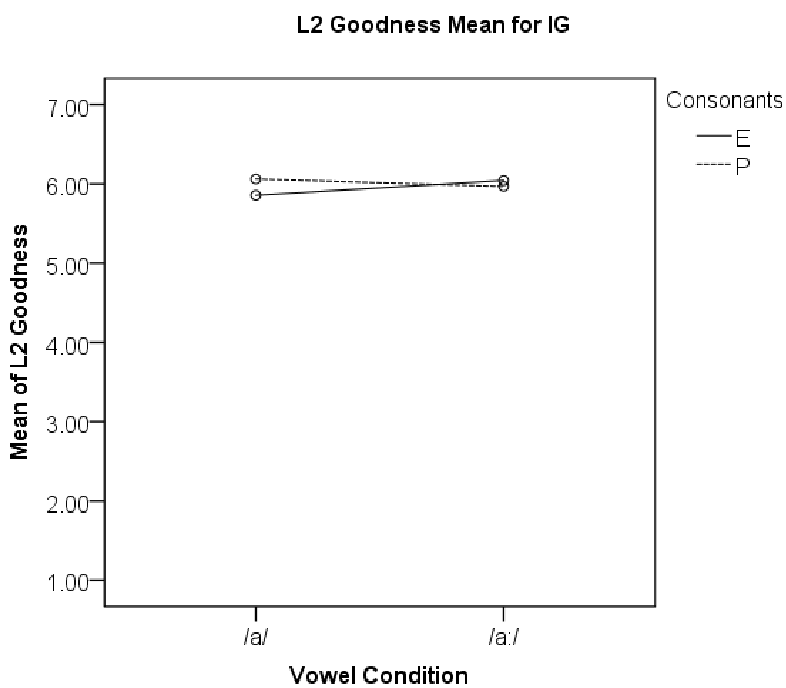
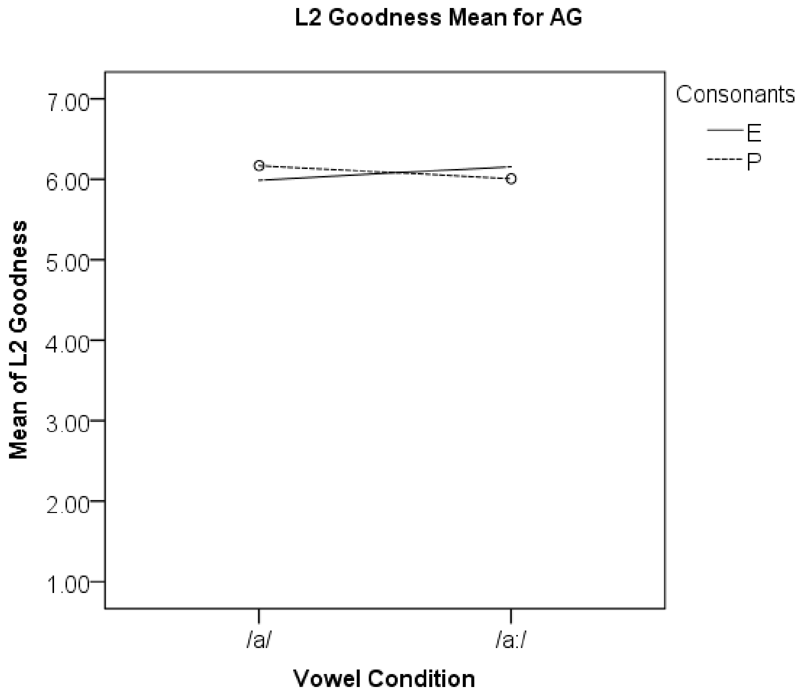


Figure 4.15. L2 Mean Goodness Comparison between Plain and Emphatic Sounds in NSR with Short and Long Vowel Conditions for the AG.



As mentioned above, Figures 4.16 and 4.17 present the proportion of L2 Arabic labeling chosen for each Arabic stimulus, with its mean goodness rating in the NG with /a:/ and /a/, respectively. The grey bars in the graph represent the overall accuracy percentage for each Arabic L2 stimulus. (In other words, the grey bars represent in numbers how successful listeners were at accurately identifying the Arabic stimuli.) In addition, the grey line represents the overall goodness rating for each selected Arabic L2 category, or in other words, listeners' level of confidence of their Arabic labeling. The error bars in the figures represent one calculated standard error. Based on the results in the two figures, the accuracies in Figure 4.16 seem to be slightly better than the accuracies in Figure 4.17, especially for the emphatic consonants, indicating better performance when long /a:/ was used. For example, when /a:/ was used in Figure 4.16, Arabic /d/ was accurately labeled onto Arabic /d/ 96.67% versus 74.17% when /a/

was used in Figure 4.17. To confirm this, I conducted a paired-samples t-test, and the results show that there was a significant difference in the scores for the long vowel (M = 63.8, SD = 5.8) and short vowel (M = 55.8, SD = 4.2) conditions: $t(29)=-5.16$, p = .000.

Figure 4.16. L2 Labeling Results for the NG in NSR Condition with Long Arabic Vowel /a:/.

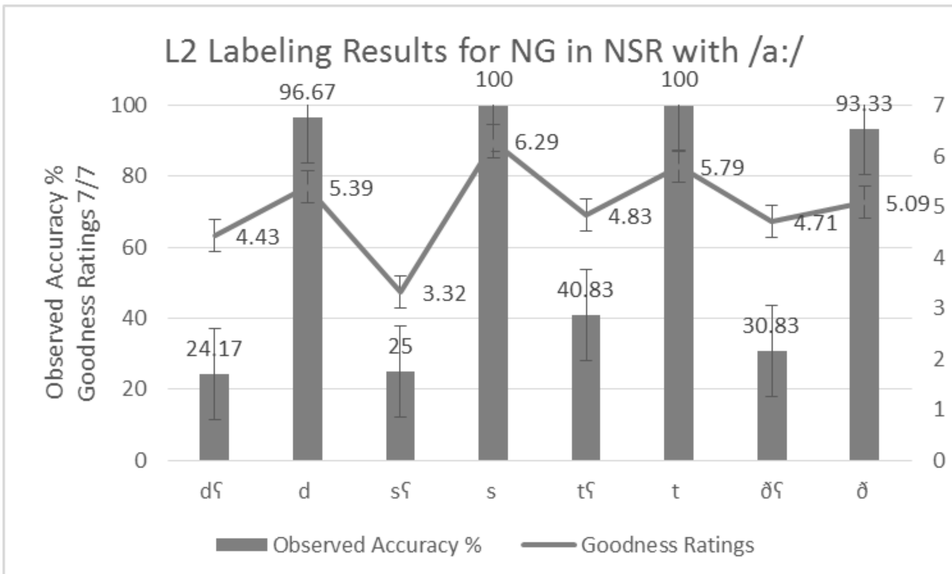
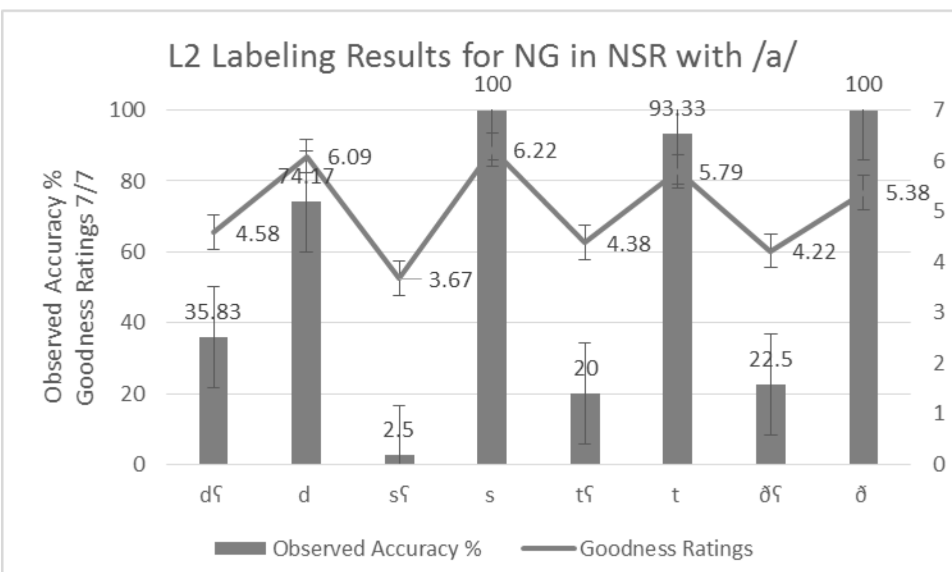


Figure 4.17. L2 Labeling Results for the NG in NSR Condition with Short Arabic Vowel /a/.



In addition, the results show that the accuracies for the plain consonants were better than the accuracies for the emphatic consonants in both figures. For instance, in Figure 4.16, Arabic plain /d/ was accurately labeled onto Arabic plain /d/ 96.67% of the time. Arabic emphatic /d^ʕ/, however, was accurately labeled onto Arabic emphatic /d^ʕ/ only 24.17% of the time. Results from a paired-samples t-test show that there was a significant difference in the scores for the plain sounds (*M* = 0.95, *SD* = 0.26) and emphatic sounds (*M* = 0.25, *SD* = 0.55) conditions; *t*(29)=-57.73, *p* = .000.

With regard to the goodness ratings, the results in both figures show that there were no differences between the given ratings, when /a/ and /a:/ were used. The results of a paired-samples t-test show that there was no significant difference in the scores for the long vowel (*M* = 5.51, *SD* = .28) and short vowel (*M* = 5.7, *SD* = .171) conditions; *t*(29)=-3.4, *p* = .161. The results also show that the goodness ratings of the plain sounds (*M* = 5.63, *SD* = .23) were higher than the goodness ratings of the emphatic sounds (*M* = 5.38, *SD* = .37) in the long vowel condition; *t*(29)=-5.14, *p* = .000. Similarly, the results show that the goodness ratings of the plain sounds (*M* = 5.81, *SD* = .20) were higher than the goodness ratings of the emphatic sounds (*M* = 5.58, *SD* = .22) in the short vowel condition *t*(29)=-5, *p* = .000. In Figure 4.16, for example, the Arabic /s/ received a high goodness rating of 6.22 corresponding to the perfect 100% accuracy. The Arabic /s^ʕ/, on the other hand, received a low goodness rating of 3.67, corresponding to the low 2.5% accuracy. Thus, the higher the accuracies, the higher the goodness ratings and vice versa. Figure 4.18 below is an indication to this positive correlation between the accuracies and overall goodness ratings with a relatively high r squared value *(R^2 = 0.85).*

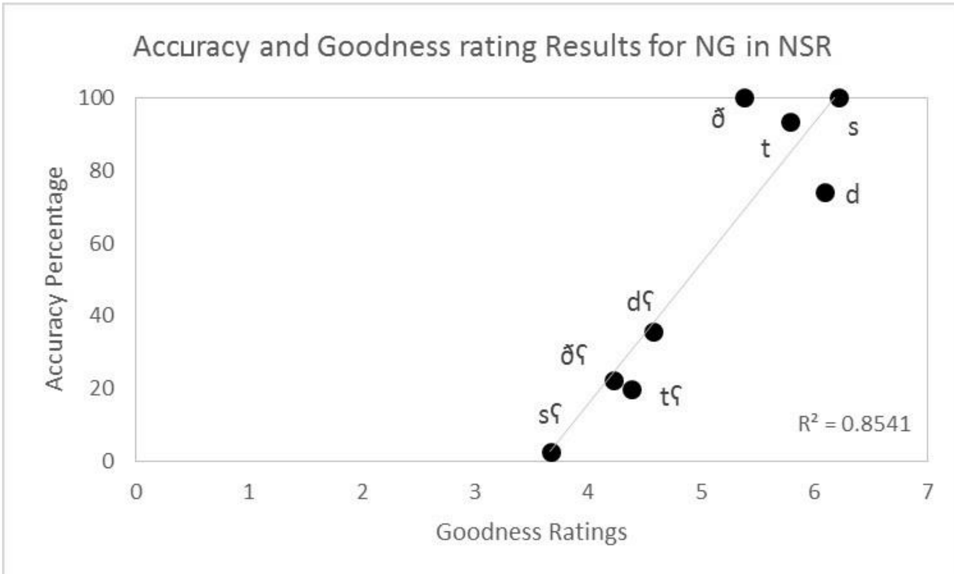


Figure 4.18. Sample of Positive Correlation between Accuracies and Goodness Ratings for the NG in the NSR with Short /a/.

Figures 4.19 and 4.20 diagram the proportion of L2 Arabic labeling chosen for each Arabic stimulus, with its mean goodness rating in the IG with /a:/ and /a/, respectively. Based on the results in the two figures, the accuracies in Figure 4.19 seem to be slightly better than the accuracies in Figure 4.20, indicating better performance when long /a:/ was used. (When /a:/ was used in Figure 4.19, Arabic /d/ was accurately labeled onto Arabic /d/ 70% versus 60% when /a/ was used in Figure 4.20.)

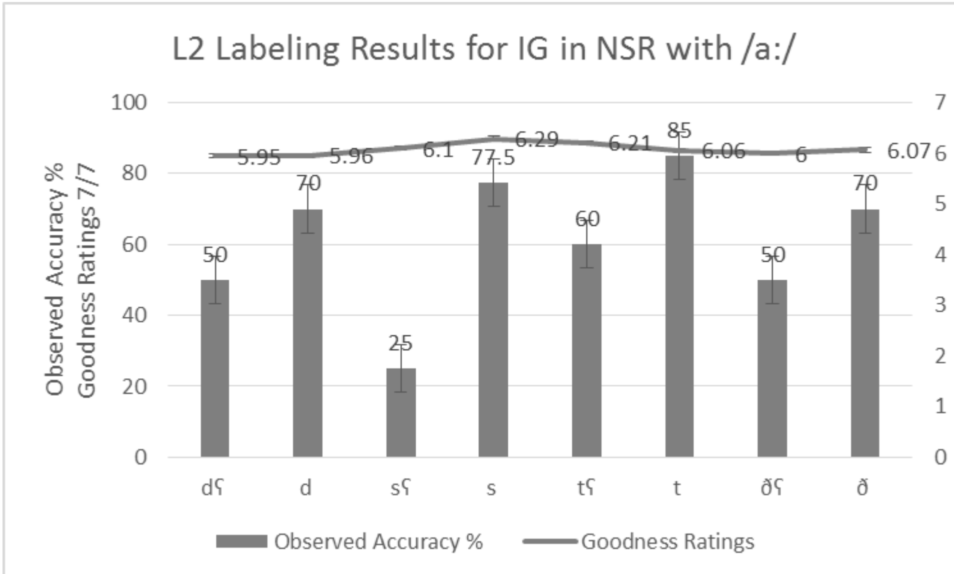


Figure 4.19. L2 Labeling Results for the IG in NSR Condition with Long Arabic Vowel /a:/.

Results from a paired-samples t-test, however, show that there was no significant difference in the scores for the short vowel (*M* = 59, *SD* = 8.6) and long vowel (*M* = 60, *SD* = 8.9) conditions; $t(9)=-.48$, $p = .638$.

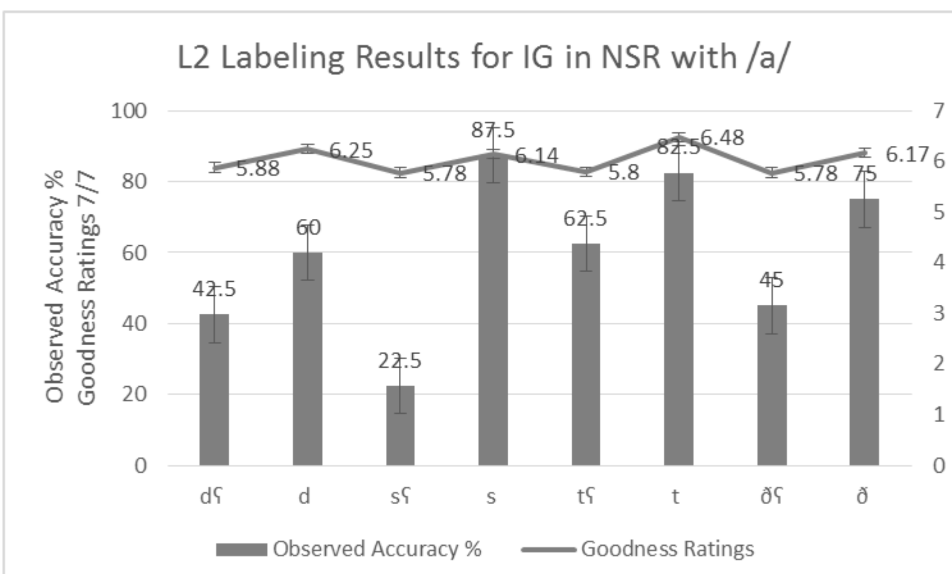


Figure 4.20. L2 Labeling Results for the IG in NSR Condition with Short Arabic Vowel /a/.

In addition, the results show that the accuracies for the plain consonants were higher than the accuracies for the emphatic consonants in both figures. For instance, in Figure 4.20, Arabic plain /d/ was accurately labeled onto Arabic plain /d/ 60% of the time. Arabic emphatic /$d^ʕ$/, however, was accurately labeled onto Arabic emphatic /$d^ʕ$/ only 42.5% of the time. The same finding was observed for all plain and emphatic consonants in both figures. The results of a paired-samples t-test show that there was a significant difference in the scores for the emphatic sounds (M = 44, SD = 12) and plain sounds (M = 75, SD = 16) conditions; $t(9)=-4$, p = .003.

With regard to the goodness ratings, the results in both figures show that there were no differences between the given ratings for when /a/ (M = 5.95, SD = .747) and /a:/ (M = 6, SD = .684) were used $t(9)=1.14$, p = .282. The results show that the goodness ratings of the plain sounds (M = 6, SD = .80) were close to the goodness ratings of the emphatic sounds (M = 5.8, SD = .76), $t(9)=-.471$, p = .649. In Figure 4.18, for example, the Arabic plain /d/ received a goodness rating of 5.96, while the Arabic emphatic /$d^ʕ$/ received 5.95. This could possibly be due to the fact that listeners' confidence grew with more L2 exposure and that was reflected directly by the decreased rating difference between the plain and emphatic sounds.

Figures 4.21 and 4.22 below present the proportion of L2 Arabic labeling chosen for each Arabic stimulus, with its mean goodness rating in the AG with /a:/ and /a/, respectively. Based on the results in the two figures, the accuracies in Figure 4.21 seem to be higher, in general, than the accuracies in Figure 4.22, indicating better performance when long /a:/ was used. For example, when /a:/ was used in Figure 4.21, Arabic /ð/ was accurately labeled onto Arabic /ð/ 70% versus 55% when /a/ was used in Figure 4.22. The results of a paired-samples t-test, however, shows that there was no significant difference in the scores for the short vowel (M = 70, SD = 16.5) and long vowel (M = 73, SD = 16.2) conditions; $t(9)=-.943$, p = .370.

Figure 4.21. L2 Labeling Results for the AG in NSR Condition with Long Arabic Vowel /a:/.

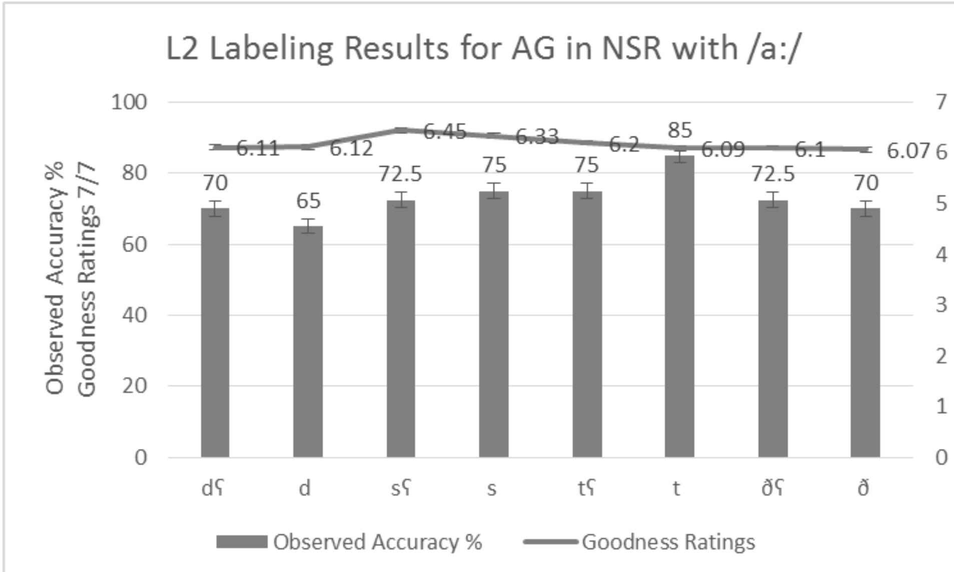
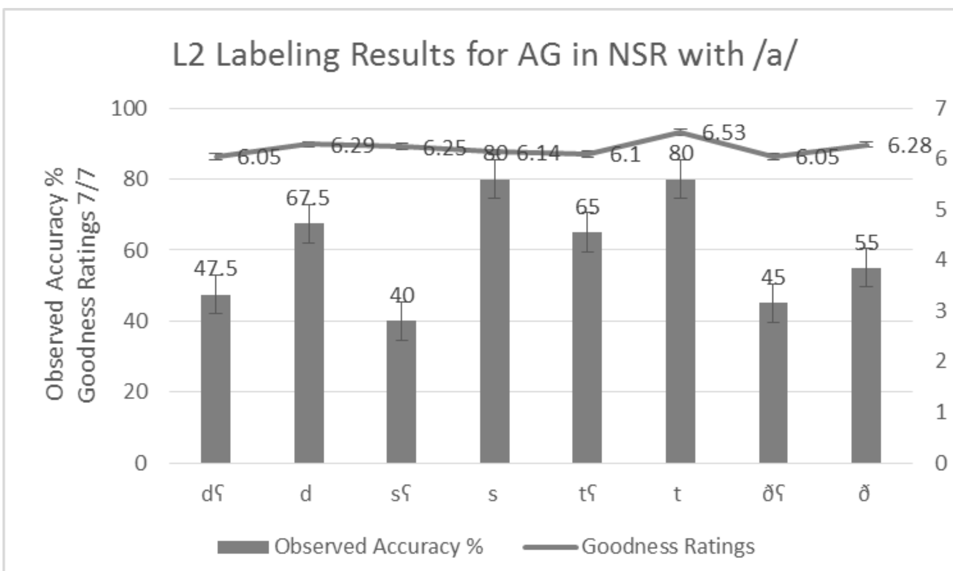


Figure 4.22. L2 Labeling Results for the AG in NSR Condition with Short Arabic Vowel /a/.



In addition, the results show that the accuracies for the plain consonants were in general better than the accuracies for the emphatic consonants. The results in 4.21 show that the accuracies for some emphatics in fact slightly surpassed the accuracies of their plain counterparts. For instance, in Figure 4.21, Arabic emphatic /dˤ/ was accurately labeled onto Arabic plain /dˤ/ 70% of the times. Arabic plain /d/, however, was accurately labeled onto Arabic plain /d/ 60% of the times. When I conducted a paired-samples t-test, the results show that there was no significant difference in the scores for the emphatic sounds (M = 68.75, SD = 19.8) and plain sounds (M = 75.31, SD = 16.4) conditions; $t(9)=-1.12$, p = .289.

With regard to the goodness ratings, the results in both figures show that there were no differences between the given ratings for when /a/ and /a:/ were used. This time, however, the results show that the goodness ratings of the plain and emphatic sounds were high and close to each other. I conducted a paired-samples t-test and the results show that there was no significant difference in the scores for the emphatic sounds (M = 6.06, SD = .60) and plain sounds (M = 6.08, SD = .67) conditions; $t(9)=-.132$, p = .898.

Finally, by comparing the results of the three learner groups above I notice the following with regard to the three main points, accuracy differences between short /a/ and long /a:/, accuracy differences between plain and emphatic consonants, and differences between goodness ratings. First, results from the NG show that the accuracies were better in the long vowel /a:/ condition than in the short vowel /a/ condition. Results from the IG and AG were generally better when long /a:/ was used but the difference between the two vowel conditions did not reach a statistical significance.

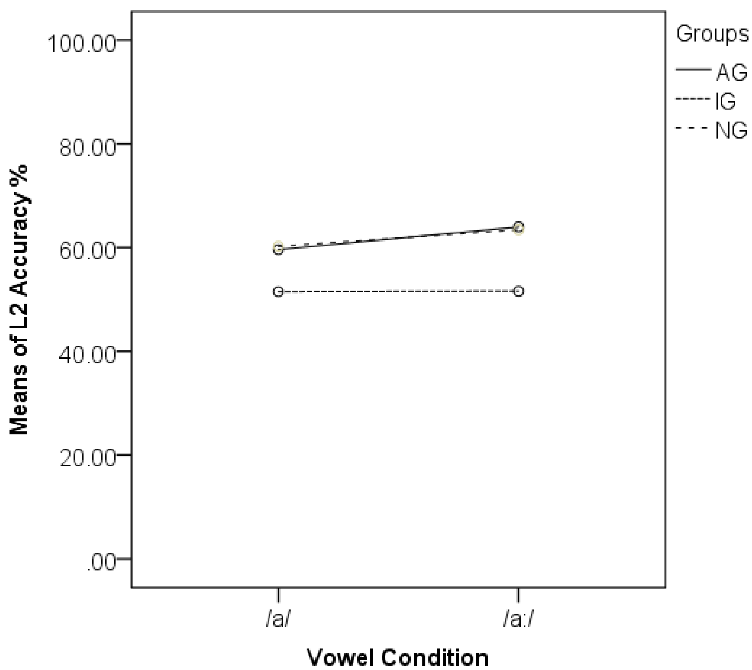
Second, results from the NG show that the accuracies of the plain sounds were better than the accuracies of the emphatic sounds in both /a/ and /a:/ conditions (Figures 4.19 and 4.20). Similarly, results for the IG show that accuracies of the plain sounds were better than the emphatic sounds (Figures 4.21 and 4.22). As for the AG, the accuracies of the plain sounds were in general better than the accuracies of the emphatic sounds but I did not see any statistical significance. In the /a:/ condition, accuracies of emphatic sounds surpassed or were close to the accuracies of the plain sounds (Figure 4.22). It is also worth mentioning that accuracies of the emphatics in the AG were better than the accuracies of emphatics in the NG and IG. This increased level of accuracy performance could be directly attributed to more L2 experience; learners in the AG have more exposure and experience with the emphatic sounds and could identify them better than learners in the NG and IG.

Third, with regard to the goodness ratings, the three learner groups show no difference in the goodness ratings in the /a/ and /a:/ conditions within each group. In addition, the results show that goodness ratings of plain sounds were better than goodness ratings of emphatic sounds in the NG. In the IG, goodness ratings of plain sounds were close to goodness ratings of emphatic sounds (Figures 4.19 and 4.20). Lastly, results from the AG also show that goodness ratings of plain and emphatic sounds were both high and close to each other (Figures 4.21 and 4.22). I believe that this is an indication that the level of confidence of the selected L2 categories grows with more L2 exposure. In other words, L2 listeners become more confident of their selected categories as their L2 exposure increases.

4.6.2.2. Fast Speech Rate

In this section, I will describe and present results obtained from the L2 labeling task and goodness-of-fit rating task under the FSR condition with /a/ and /a:/ for the three listener groups. Based on the results, the three learner groups do not show much difference in terms of the accuracies between /a/ and /a:/. I analyzed the data by running a mixed-design ANOVA with the *vowel duration* (short /a/ vs. long /a:/) as the within-subjects factor, and *groups* (NG vs. IG vs. AG) as the between-subjects factor. The obtained results indicate that there was no main effect of vowels, $F(1, 47) = 3.31$, $p = .075$, with no significant interaction among the groups, $F(2, 47) = .674$, $p = .515$. This means that the accuracies were not different for the long vowel condition /a:/ and the short vowel /a/ condition, and this trend was true for all the three groups in general. The accuracy difference between the long and the short vowels was larger between the NG and AG (see Figure 4.23 below), suggesting that the short vowel /a/ condition was more difficult than the long vowel /a:/ condition for the NG and AG. This may be due to the increased difficulty of the FSR task.

Figure 4.23. L2 Mean Accuracy Comparisons between NG, IG, and AG in FSR with Short and Long Vowel Conditions.



Accuracies of plain sounds were better than accuracies of emphatic sounds in the NG, IG, and AG. I ran a mixed-design ANOVA with the *consonants* and *vowel duration* (plain vs. emphatic vs. short /a/ vs. long /a:/) as the within-subjects factor, and *groups* (NG vs. IG vs. AG) as the between-subjects factor. The obtained results indicate that there was a main effect of consonant type, $F(1, 47) = 734$, $p = .000$, with significant interaction among the groups, $F(2, 47) = 112$, $p = .000$. This means that the accuracies were higher for the plain sounds than the emphatic sounds. The results also show significant interaction between the vowels and consonants at the $p < .01$ level, $F(1, 47) = 9$, $p = .004$. The accuracies were higher in the long vowel condition than the short vowel condition.

Figures 4.24, 4.25, and 4.26 show the mean accuracy results for the NG, IG, and AG, respectively. The x-axis indicates the two vowel conditions that were used, short and long. The y-axis indicates the mean of L2 accuracy. The blue line represents the L2 emphatic sounds while the green line represents the L2 plain sounds. By comparing the figures to each other I can see that, within each group, the performance of the plain consonants was somewhat higher than the emphatic consonants when short /a/ was used. In the long vowel condition, however, performance of the plain consonants is similar to the performance of the emphatic consonants in the NG. For example, the mean performance of the NG was 29% for the plain consonants in the short vowel condition versus 23% for the emphatic consonants in the same short vowel condition (Figure 4.24). The same observation applies to the other groups. This suggests that the long vowel /a:/ condition was more difficult than the short vowel /a/ condition. This was supported when I noticed that the performance was generally higher when the long vowel condition was applied for both plain and emphatic consonants. For example, in Figure 4.25, performance of the emphatic sounds with long /a:/ was higher than when short /a/ was used ($M = 79$ versus $M = 29$), and performance of the plain sounds with long /a:/ was higher than when short /a/ was used ($M = 74$ versus $M = 29$).

Figure 4.24. L2 Mean Accuracy Comparison between Plain and Emphatic Sounds in FSR with Short and Long Vowel Conditions for the NG.

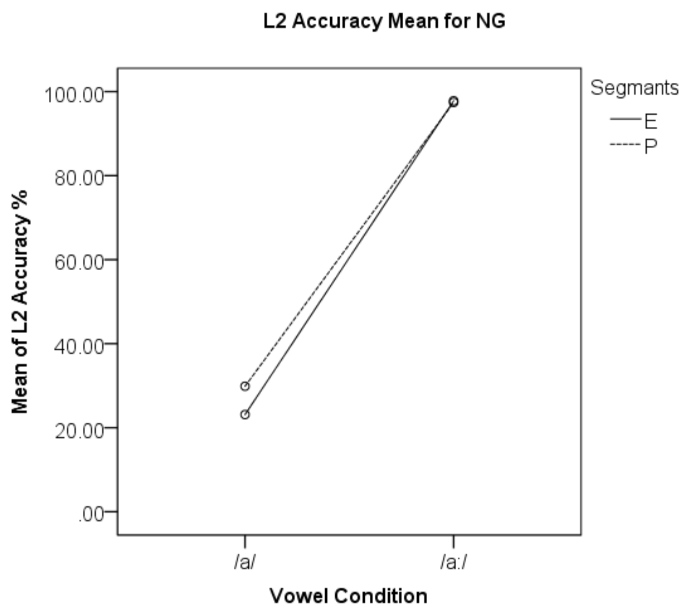


Figure 4.25. L2 Mean Accuracy Comparison between Plain and Emphatic Sounds in FSR with Short and Long Vowel Conditions for the IG.

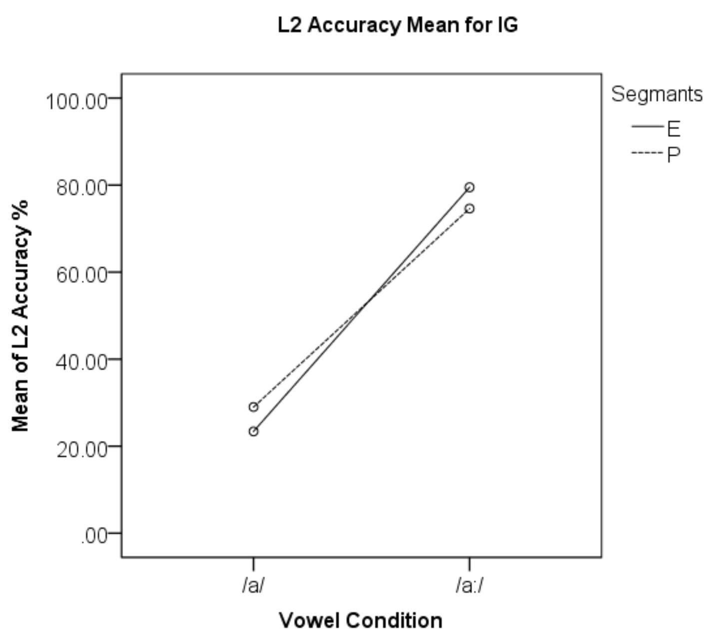
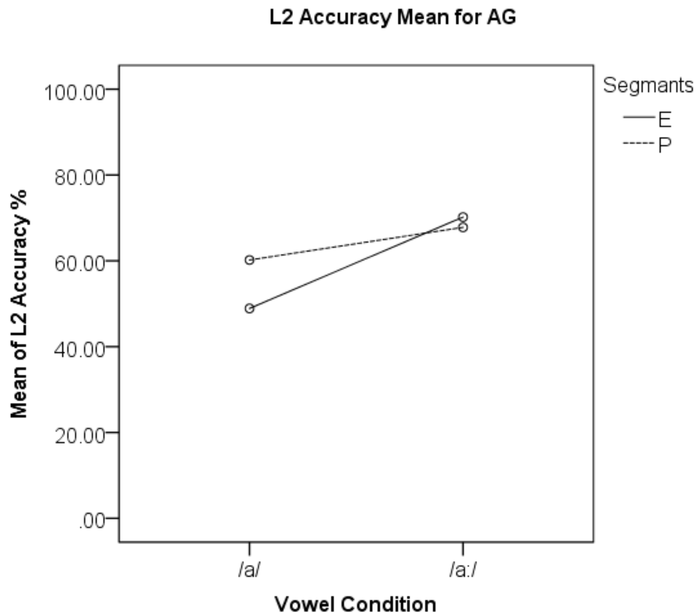
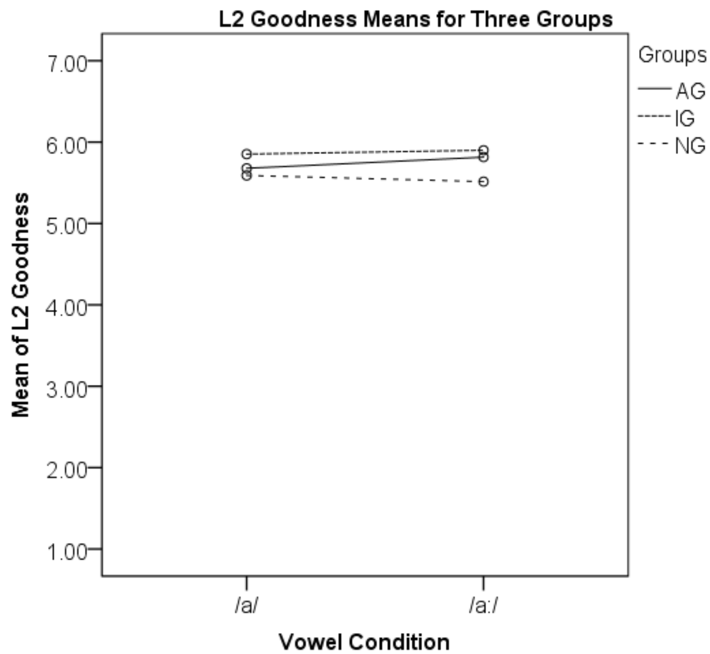


Figure 4.26. L2 Mean Accuracy Comparison between Plain and Emphatic Sounds in FSR with Short and Long Vowel Conditions for the AG.



With regard to goodness ratings, the results show no difference between goodness ratings when /a/ and /a:/ were used within each group. I analyzed the data by running a mixed-design ANOVA with *vowel duration* (short vowel /a/ vs. long vowel /a:/) as the within-subjects factor, and *groups* (NG vs. IG vs. AG) as the between-subjects factor. The obtained results indicate that there was no main effect of vowels, $F(1, 47) = 1.17$, $p = .284$, but a significant interaction among the groups was found, $F(2, 47) = 4.43$, $p = .017$. This trend, however, was not true for the NG. Figure 4.27 below demonstrates the mean goodness scores for the three learner groups in FSR with short and long vowel conditions. The x-axis in the figure represents the two short and long vowel conditions. On the other hand, the y-axis represents the mean goodness for the L2 consonants. I can notice from the figure that the obtained mean in the short vowel /a/ condition is slightly higher than the long /a:/ condition. In the same figure, I can notice that the performance improvement is most between the NG and the AG. There is a slight performance improvement between the IG and AG. This also suggests that L2 exposure positively affects the overall given goodness of L2 sounds.

Figure 4.27. L2 Mean Goodness Comparisons between NG, IG, and AG in Short and Long Vowel Conditions in FSR.



With regard to the plain and emphatic consonants, goodness of plain sounds were better than goodness of emphatic sounds in the three groups. I ran a mixed-design ANOVA with the *consonants* and *vowel duration* (plain vs. emphatic; short /a/ vs. long /a:/) as the within-subjects factor, and *groups* (NG vs. IG vs. AG) as the between-subjects factor. The obtained results indicate that there was a main effect of *consonant* type, $F(1, 47) = 9.29$, $p = .004$, with significant interaction among the *groups*, $F(2, 47) = 27.29$, $p = .000$. This means that the goodness were higher for the plain sounds than the emphatic sounds, but this trend was not true for all the three groups. The results also showed no significant interaction between the vowels and consonants, $F(1, 47) = .067$, $p = .798$.

Figures 4.28, 4.29, and 4.30 show the mean goodness results for the NG, IG, and AG, respectively. The x-axis indicates the two vowel conditions that were used, short and long. The y-axis indicates the mean of L2 goodness. The blue line represents the L2 emphatic sounds while the green line represents the L2 plain sounds. By comparing the figures to each other I can see that, within the NG group, the performance of the emphatic consonants was higher than the plain consonants when short /a/ was used; ($M = 5.3$) for emphatic consonants in the short vowel condition versus ($M = 5.1$) for the plain consonants in the same short vowel condition (Figure 4.28). In addition, by looking at Figures 4.28 and 4.29 I can see that mean goodness for both plain and emphatic under the two vowel conditions are almost identical suggesting no change of performance with more L2 exposure. However, performance of plain consonants was slightly better than the emphatic consonants this time.

Figure 4.28. L2 Mean Goodness Comparison between Plain and Emphatic Sounds in FSR with Short and Long Vowel Conditions for the NG.

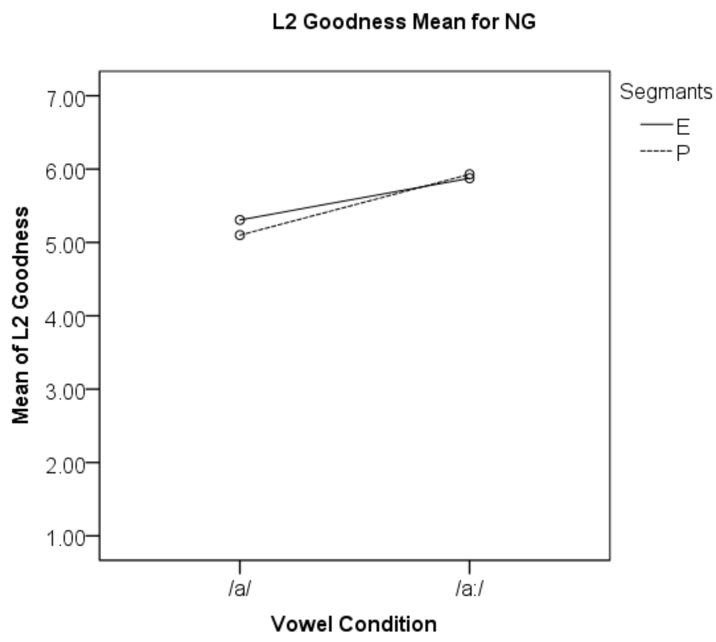


Figure 4.29. L2 Mean Goodness Comparison between Plain and Emphatic Sounds in FSR with Short and Long Vowel Conditions for the IG.

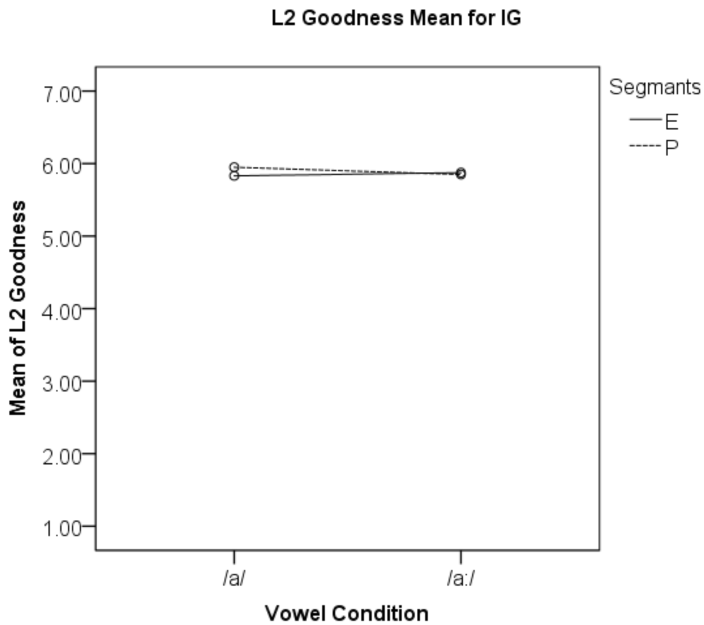
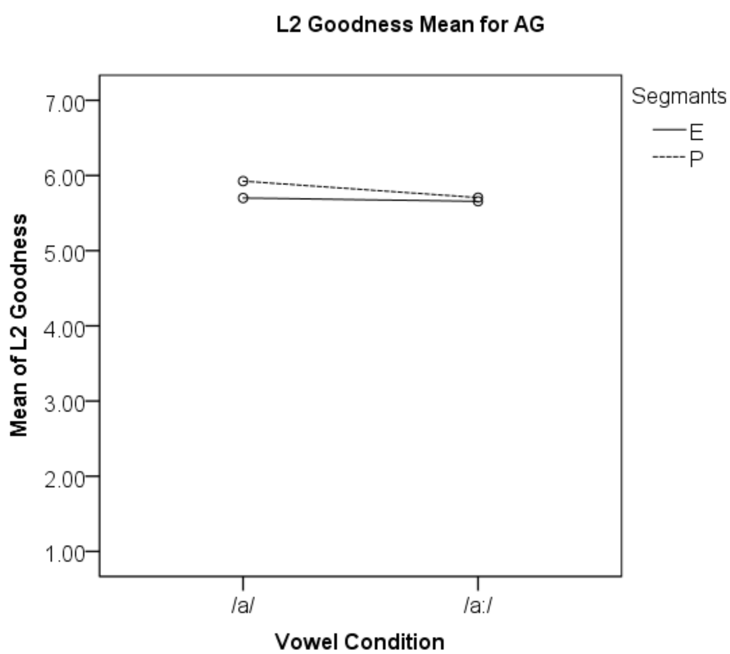


Figure 4.30. L2 Mean Goodness Comparison between Plain and Emphatic Sounds in FSR with Short and Long Vowel Conditions for the AG.



Based on what has been mentioned above, Figures 4.31 and 4.32 present the proportion of L2 Arabic labeling chosen for each Arabic stimulus with its mean goodness rating in the NG with /a:/ and /a/, respectively. Similar to what was explained in the NSR section, the grey bars in the graph represent the overall accuracy percentage for each Arabic L2 stimulus. In other words, the grey bars represent in numbers how successful listeners were at accurately identifying the Arabic stimuli. In addition, the grey line represents the overall goodness rating for each selected Arabic L2 category. In other words, the grey line represents the listeners' level of confidence of their Arabic labeling. Based on the results in the two figures, the accuracies in Figures 4.31 and 4.32 do not show much difference in terms of the accuracies which indicates that the /a:/ and /a/ conditions did not have an effect. I conducted a paired-samples t-test and the results show that there was a significant effect in the scores for the long vowel (M = 63.4, SD = 7.7) and short vowel (M = 60.2, SD = 6.4) conditions; $t(29)=-2.76$, p = .010.

Figure 4.31. L2 Labeling Results for the NG in FSR Condition with Long Arabic Vowel /a:/.

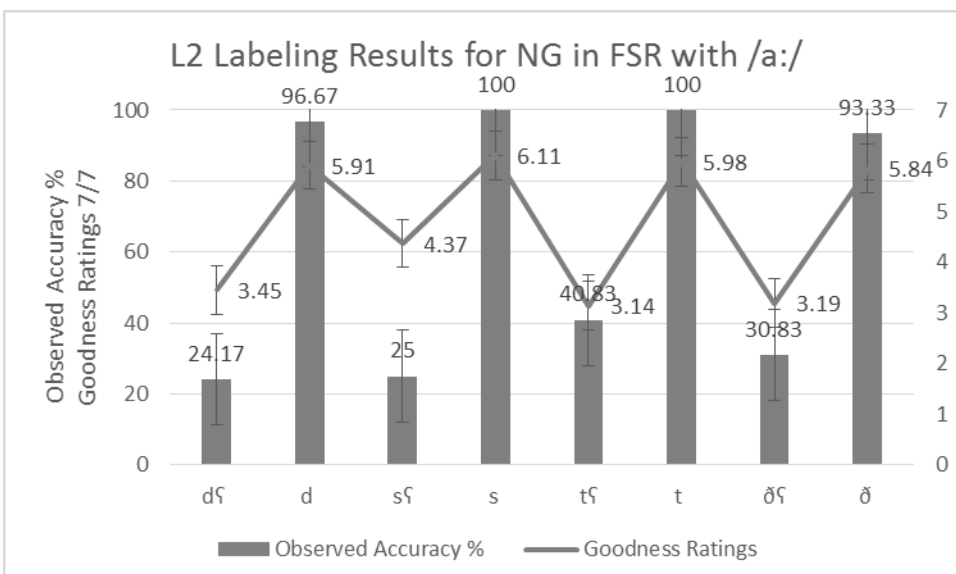
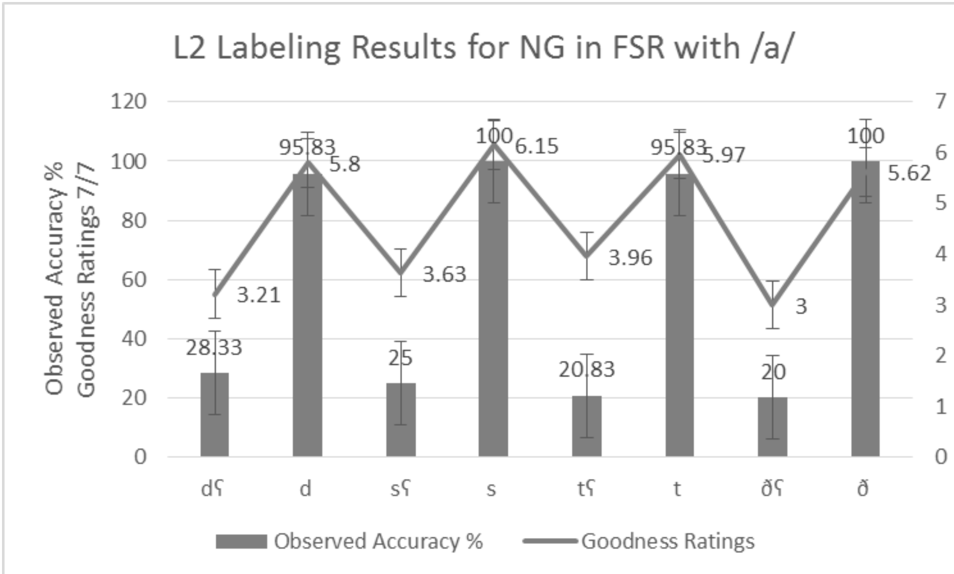


Figure 4.32. L2 Labeling Results for the NG in FSR Condition with Short Arabic Vowel /a/.



In addition, the results show that the observed accuracies of the plain consonants were better than the accuracies of the emphatic consonants in both figures. For instance, in Figure 4.31, Arabic plain /d/ was accurately labeled onto Arabic plain /d/ 96.67% of the times. Arabic emphatic /$d^ʕ$/, however, was accurately labeled onto Arabic emphatic /$d^ʕ$/ only 24.17% of the times. Results from a paired-samples t-test show that there was a significant difference in the scores for the plain sounds (M = 97, SD = 4.7) and emphatic sounds (M = 26.8, SD = 10.33) conditions; $t(29)=-41.57$, p = .000.

With regard to the goodness ratings, the results in both figures show that there were no differences between the given ratings for when /a/ and /a:/ were used. I conducted a paired-samples t-test and the results show that there was a significant difference in the scores for the long vowel (M = 5.51, SD = .23) and short vowel (M = 5.59, SD = .30) conditions at the $p < .05$ level; $t(29)=-2.28$, p = .030. The results also show that the goodness ratings of the plain sounds were higher than the goodness ratings of the emphatic sounds. I conducted a paired-samples t-test and the results show that there was a significant difference in the scores for the emphatic sounds (*M* = 5.2, *SD* = .30) and plain sounds (*M* = 5.9, *SD* = .34) conditions; *t*(29)=-9.49, *p* = .000. As explained earlier, this could be an indication to the positive correlation between the accuracies and overall goodness ratings; the higher the accuracies the higher the goodness ratings and vice versa. In Figure 4.32, for example, Arabic /s/ received a high goodness rating of 6.15 corresponding to the perfect 100% accuracy. Arabic /sˤ/, on the other hand, received a low goodness rating of 3.63 corresponding to the low 25% accuracy.

Figures 4.33 and 4.34 present the proportion of L2 Arabic labeling chosen for each Arabic stimulus with its mean goodness rating in the IG with /a:/ and /a/, respectively. Based on the results in the two figures, the accuracies in Figures 4.33 and 4.34 do not show much difference in terms of the accuracies which indicates that the /a:/ and /a/ conditions did not have an effect on the accuracies. Results from a paired-samples t-test show that there was no significant difference in the scores for the short vowel (*M* = 51.5, *SD* = 14.5) and long vowel (*M* = 51.6, *SD* = 10.5) conditions; *t*(9)=-.022, *p* = .983.

In addition, the results show that the accuracies of the plain consonants were better than the accuracies of the emphatic consonants in both figures. For instance, in Figure 4.34, Arabic plain /d/ was accurately labeled onto Arabic plain /d/ 82.5% of the times. Arabic emphatic /dˤ/, however, was accurately labeled onto Arabic emphatic /dˤ/ only 30% of the times. Results from a paired-samples t-test show that there was a significant difference in the scores for the emphatic sounds (*M* = 26.56, *SD* = 8.49) and plain sounds (*M* = 77.5, *SD* = 14.64) conditions; *t*(9)=-14.22, *p* = .000.

Figure 4.33. L2 Labeling Results for the IG in FSR Condition with Long Arabic Vowel /a:/.

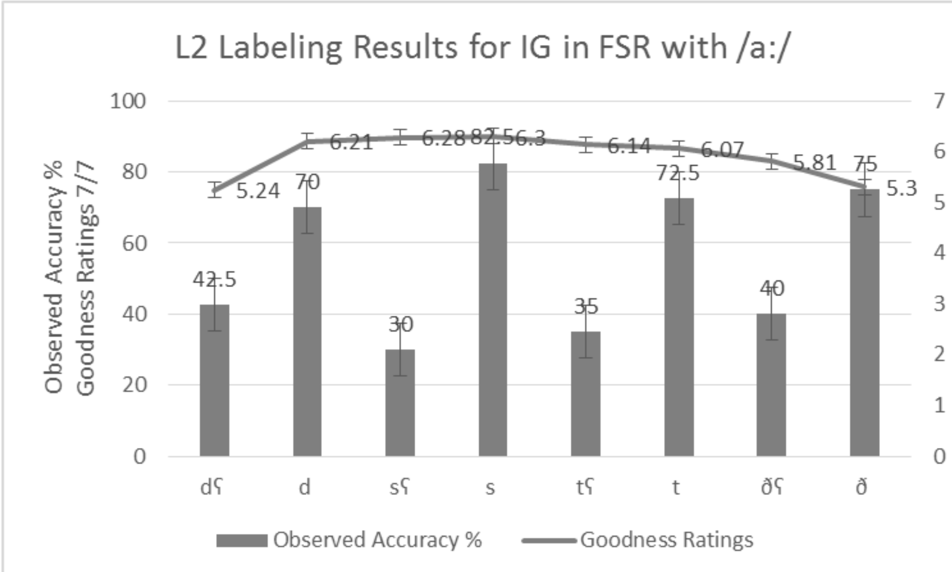
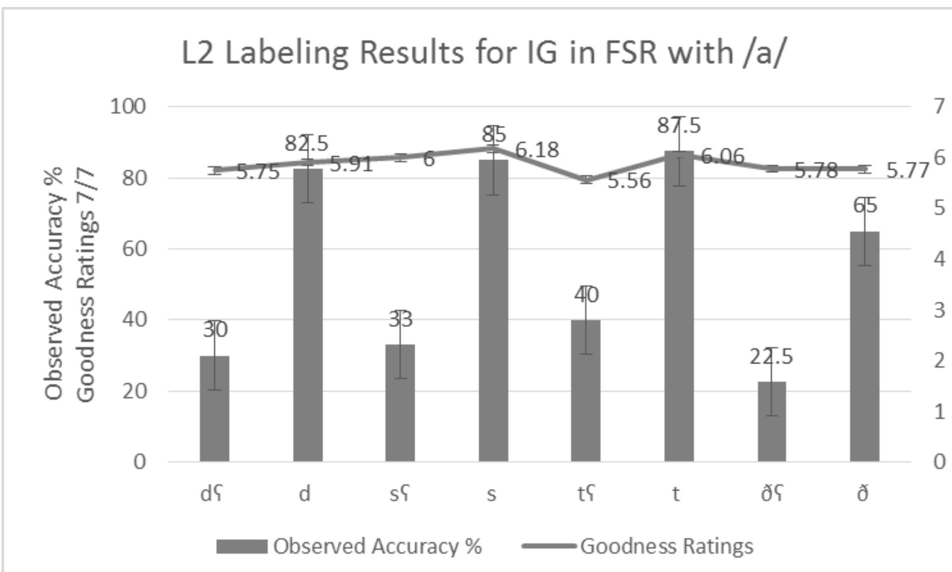


Figure 4.34. L2 Labeling Results for the IG in FSR Condition with Short Arabic Vowel /a/.



With regard to the goodness ratings, the results in both figures show that there were no differences between the given ratings for when /a/ (M = 5.85, SD = .383) and /a:/ (M = 5.9, SD = .437) were used, $t(9)=-.778$, p = .456.

The results show that the goodness ratings of the plain sounds (*M* = 5.86, *SD* = .378) were not different from the goodness ratings of the emphatic sounds (*M* = 5.89, *SD* = .461), *t*(9)=.333, *p* = .747. In Figure 4.34, for example, Arabic plain /s/ received a goodness rating of 6.18 while Arabic emphatic /sˤ/ received 6.

Figures 4.35 and 4.36 below present the proportion of L2 Arabic labeling chosen for each Arabic stimulus with its mean goodness rating in the AG with /a:/ and /a/, respectively. Based on the results in the two figures, the accuracies in Figures 4.35 and 4.36 do not show much difference in terms of the accuracies which indicates that the /a:/ and /a/ conditions did not have an effect on the accuracies. I conducted a paired-samples t-test and the results show that there was no significant difference in the scores for the short vowel (*M* = 59.6, *SD* = 15.72) and long vowel (*M* = 64, *SD* = 13.35) conditions; *t*(9)=-1.73, *p* = .116.

In addition, the results show that the accuracies of the plain consonants were in general better than the accuracies of the emphatic consonants. For example, in Figure 4.35, Arabic plain /s/ was accurately observed as Arabic /a/ 80% of the time while Arabic emphatic /sˤ/ was accurately observed as 55% of the time. When I conducted a paired-samples t-test, the results show that there was a significant difference in the scores for the emphatic sounds (*M* = 55, *SD* = 11.89) and plain sounds (*M* = 69.38, *SD* = 18.32) conditions; *t*(9)=-3.60, *p* = .006.

Figure 4.35. L2 Labeling Results for the AG in FSR Condition with Long Arabic Vowel /a:/.

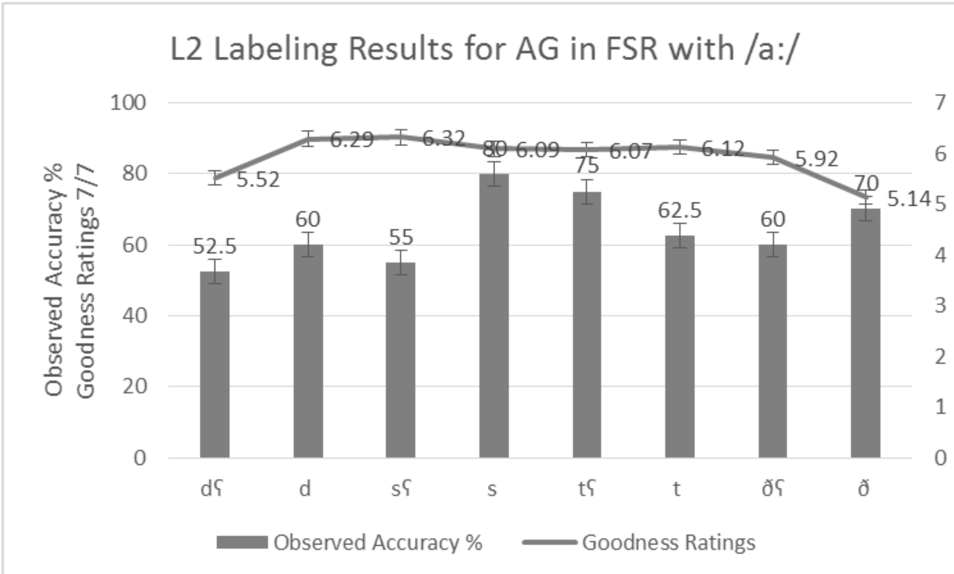
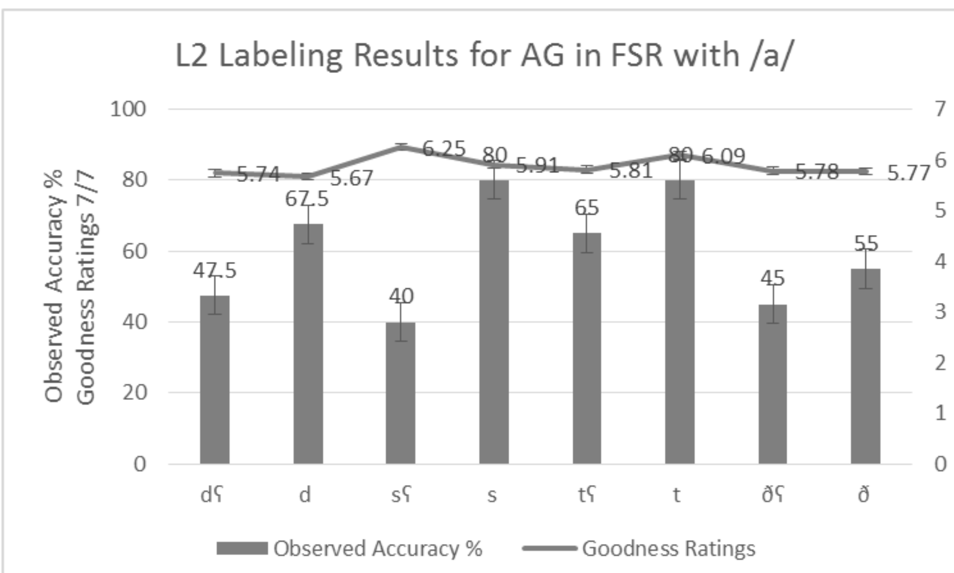


Figure 4.36. L2 Labeling Results for the AG in FSR Condition with Short Arabic Vowel /a/.



With regard to the goodness ratings, the results in both figures show that there were no differences between the given ratings for when /a/ and /a:/ were used. I conducted a paired-samples t-test and the results show that there was no significant difference in the scores for the short vowel (*M* = 5.67, *SD* = .381) and long vowel (*M* = 5.81, *SD* = .460) conditions; $t(9)=-1.52$, $p = .161$.

Similarly, the results show that the goodness ratings of the plain and emphatic sounds were close to each other with no much differences. I conducted a paired-samples t-test and the results show that there was no significant difference in the scores for the emphatic sounds (*M* = 5.81, *SD* = .447) and plain sounds (*M* = 5.68, *SD* = .411) conditions; $t(9)=1.27$, $p = .233$.

By comparing the results of the three learner groups above I notice the following with regard to the three main points, accuracy differences between short /a/ and long /a:/, accuracy differences between plain and emphatic consonants, and differences between goodness ratings. First, all results from the NG, IG, and AG show that the accuracies did not show much difference between the /a/ and /a:/ conditions.

Second, results from the NG show that the accuracies of the plain sounds were better than the accuracies of the emphatic sounds in both /a/ and /a:/ conditions (Figures 4.31 and 4.32). Similarly, results for the IG show that accuracies of the plain sounds were better than the emphatic sounds in both /a/ and /a:/ conditions (Figures 4.33 and 4.34). As for the AG, the results show that the accuracies of the plain sounds were in general better than the accuracies of the emphatic sounds but the difference was not statistically significant at the $p < .05$ level. It is also worth mentioning that accuracies of the emphatics in the AG were better than the accuracies of emphatics in the NG and IG. This increased level of accuracy performance could be directly attributed to more L2 experience; learners in the AG have more exposure and experience with the emphatic sounds and could identify them better than learners in the NG and IG.

Third, with regard to the goodness ratings, results from the three learner groups show no difference in the goodness ratings between the /a/ and /a:/ conditions within each group. In addition, results from the NG show that goodness ratings of plain sounds were higher than goodness ratings of emphatic sounds (Figures 4.31 and 4.32). In the IG, results show that goodness ratings of plain sounds were not different from goodness ratings of emphatic sounds (Figures 4.33 and 4.34). Similarly, results from the AG show that goodness ratings of plain and emphatic sounds were similar to each other and did not show differences between them (Figures 4.35 and 4.36). As I explained earlier, this is an indication that the confidence level of the chosen sounds is growing with more L2 exposure. As a matter of fact, L2 listeners' levels of confidence is getting better for the emphatic sounds suggesting a subtle learnability trend.

Now, by comparing the results in the NSR with the FSR for the same three groups I notice the following. First, in the NSR, the three groups show better accuracies in the /a:/ condition than in the /a/ condition. However, in the FSR, the three groups show no difference in the accuracies for when /a:/ and /a/ were used. Showing no differences between the two vowel conditions in the FSR could be attributed to the increased difficulty of the task.

Second, in the NSR, results from the NG and IG show that the accuracies of the plain sounds were better than the accuracies of the emphatic sounds in both /a/ and /a:/ conditions. As for the AG, the results show that the accuracies of the plain sounds were in general better than the accuracies of the emphatic sounds. In the /a:/ condition, accuracies of emphatic sounds surpassed or were close to the accuracies of the plain sounds. In addition, accuracies of the emphatics in the AG were better than the accuracies of emphatics in the NG and IG. Similarly, in the FSR, results from the NG and IG show that the accuracies of the plain sounds were better than the accuracies of the emphatic sounds in both /a/ and /a:/ conditions. As for the AG, similar to the NSR, the results show that the accuracies of the plain sounds were in general better than the accuracies of the emphatic sounds but it was not very significant. In addition, the accuracies of the emphatics in the AG were better than the accuracies of emphatics in both NG and IG. The increased level of performance with the emphatic sounds could be linked to the level of L2 exposure, more L2 experience helped learners better identify L2 emphatics. In addition, the results indicate that the accuracy levels for plain and emphatic sounds were higher in the NSR than the FSR for all three groups. This difference could also be attributed to the difficulty of the task, NSR is more natural and learners were able to label L2 sounds more accurately than in the FSR which is more difficult and could have cause uncertainty and confusion.

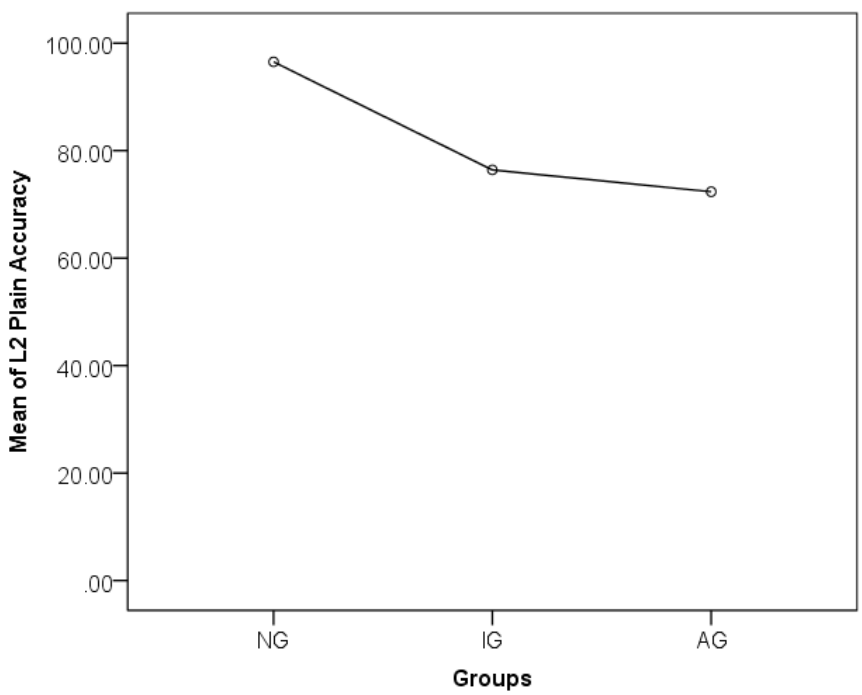
Third, with regard to the goodness ratings in the NSR, results from the three learner groups show no difference in the goodness ratings in the /a/ and /a:/ conditions within each group. Results from the NG show that goodness ratings of plain sounds were higher than goodness ratings of emphatic sounds. In the IG, results show that goodness ratings of plain sounds were close to goodness ratings of emphatic sounds. Lastly, results from the AG show that goodness ratings of plain and emphatic sounds were also both high and close to each other. As for the FSR, results also show no difference in the goodness ratings between the /a/ and /a:/ conditions within each group. Results from the NG also show that goodness ratings of plain sounds were higher than goodness ratings of emphatic sounds. In the IG, results show that goodness ratings of plain sounds were not different from goodness ratings of emphatic sounds. Lastly, results from the AG show that goodness ratings of plain and emphatic sounds were similar to each other and did not show differences between them.

4.7. Development of Plain and Emphatic Sounds

Based on the results presented above, now I can turn to the second research question that I posted earlier in this chapter: *Does more L2 exposure over time lead to similar learning development with Arabic L2 plain and emphatic sounds?* In order to answer this question, I need to look at the overall development of the plain and emphatic sounds in the three different learning groups (NG, IG, and AG). This could be achieved by examining the accuracy results from the L2 labeling task and the goodness ratings.

Figure 4.37 below presents the mean percentage accuracies obtained from the L2 identification task for all three learner groups for the plain sounds in the NSR and FSR. In this figure, I can see that the mean results for the NG are higher the mean results of the IG and AG. In addition, mean results of the IG are slightly higher than mean results of the AG.

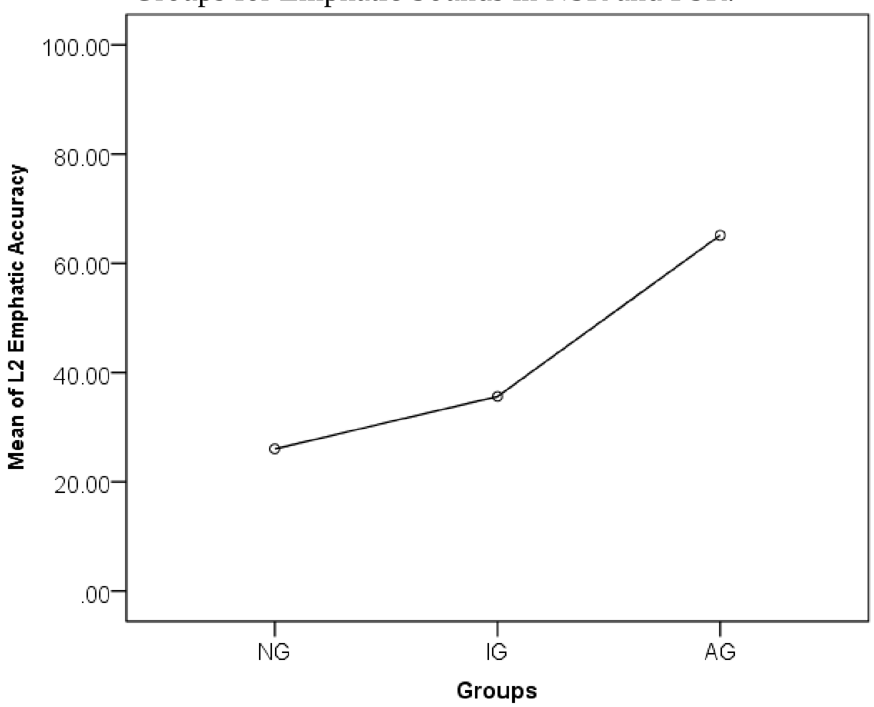
Figure 4.37. Mean L2 Accuracies for Three Learner Groups for Plain Sounds in NSR and FSR.



The interpretation that comes out of observing this pattern is that the three learner groups did not show any L2 learning development for the L2 plain sounds. As a matter of fact, the pattern indicates a decrease in performance among the three groups starting from the NG. To confirm my speculations, I conducted a one-way ANOVA to compare the effect of L2 exposure on accuracy of labeling L2 plain sounds. The results show that there was a significant effect of L2 exposure on the accuracy of labeling the plain sounds at the $p < 0.005$ level for the three groups, $F(2, 47) = 30.71$, $p = .000$. Post hoc comparisons indicate that the mean score for the NG ($M = 96.51$, $SD = 2.34$) was significantly different from the IG ($M = 76.4$, $SD = 14.66$) and the AG ($M = 72.34$, $SD = 16.55$) at $p = < .01$ level. Even though the results were significant, but that does not interpret into a development of learning plain sounds. On the contrary, the results do not support a learning pattern and I can say that the plain sounds do not show any improvement with more L2 exposure. This decrease in accuracy among the learner groups could be attributed to a perceptual confusion that is due to the learners' attempt to distinguish the differences between the plain and emphatic consonants. Another potential explanation sheds light on the vowel condition /a/ and /a:/. As explained in an earlier chapter, vowel quality of /a/ and /a:/ is different than the English vowels and even though the main task asks listeners to focus only on the consonants of the stimuli, listeners' focus may be after all deviating from the main task and listeners were actually paying more or equal attention to the vowels used in the syllable structures. As a consequence, accuracy rates fluctuated a little for the plain Arabic sounds. As for the better performance of plain sounds in the NG, it could be possible that they had a bias towards the sounds they were hearing. In other words, it may be possible that listeners in the NG were selecting plain categories for when they are confused since they are similar to their L1 sounds.

On the other hand, Figure 4.38 below presents the mean percentage accuracies obtained from the L2 identification task for all three learner groups for the emphatic sounds in the NSR and FSR. In this figure, I can see that the mean results for the AG are higher the mean results of the IG, and the mean results of the IG are higher than the mean results of the NG. The interpretation that comes out of observing this pattern is that the three learner groups did this time show L2 learning development for the L2 emphatic sounds. I notice a gradual development in the performance of the learners starting from the NG.

Figure 4.38. Mean L2 Accuracies for Three Learner Groups for Emphatic Sounds in NSR and FSR.

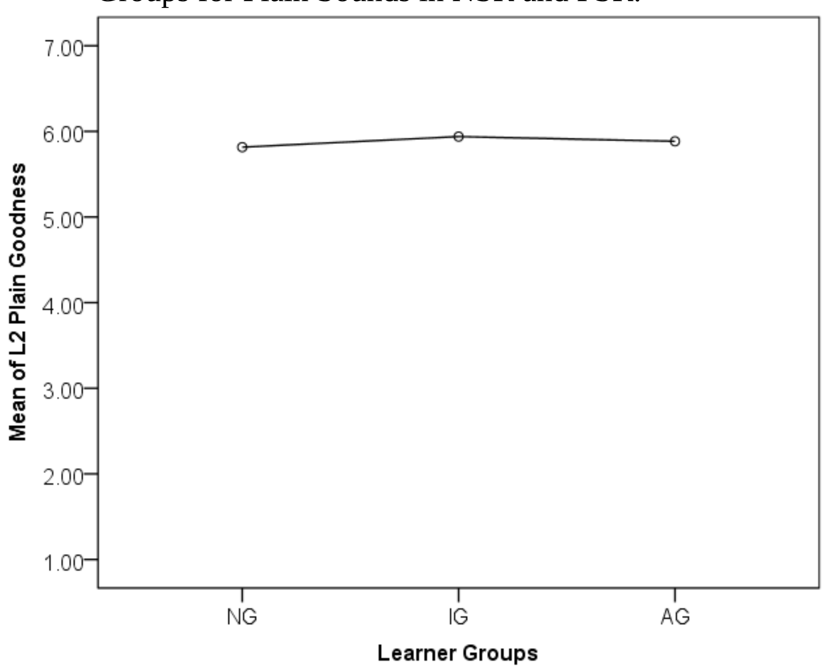


To confirm my speculations, I conducted a one-way ANOVA to compare the effect of L2 exposure on accuracy of labeling L2 empathic sounds. The results show that there was a significant effect of L2 exposure on the accuracy of labeling the emphatic sounds for the three groups $F(2, 47) = 90.05$, $p = .000$. Post hoc comparisons indicate that the mean score for the AG

(M = 65.15, SD = 12.60) was significantly different from the IG (M = 35.62, SD = 8.32) and the NG (M = 26.04, SD = 5.69) at p = < .01 value.

The goodness ratings of the plain and emphatic sounds should demonstrate a pattern similar to what I observed in the Figures 4.37 and 4.38 above in order to solidify my speculations with regard to the development of the plain and emphatic sounds explained above. In other words, for the plain sounds, the goodness results should not show any significant differences between the given goodness ratings among the three learner groups. On the other hand, for the emphatic sounds, the results should demonstrate a significant and a gradual increase in the goodness performances between the three learner groups. Based on that, Figure 4.39 below presents the mean goodness scores obtained from the L2 identification task for all three learner groups for the plain sounds in the NSR and FSR. In this figure, I can see that the mean results for the three learner groups are almost identical with not much difference between them.

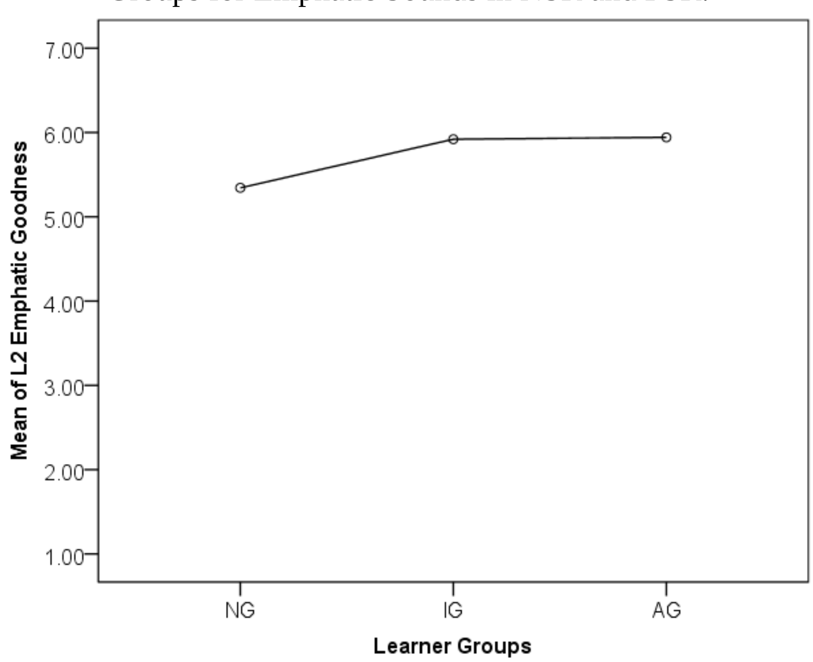
Figure 4.39. Mean L2 Goodness for Three Learner Groups for Plain Sounds in NSR and FSR.



The interpretation that comes out of observing this pattern is similar to what I said earlier about the results in Figure 4.37; the three learner groups were equally confident about their labelings of the L2 plain sounds and did not show any L2 learning development. To confirm this statement, I conducted a one-way ANOVA to compare the effect of L2 exposure on goodness ratings of labeled L2 plain sounds. The results show that there was no significant effect of L2 exposure on the goodness results of labeled L2 plain sounds for the three groups $F(2, 47) = .494$, $p = .613$. Post hoc comparisons indicate that the mean score for the NG ($M = 5.8$, $SD = .196$) was not significantly different from the IG ($M = 5.9$, $SD = .546$) and the AG ($M = 5.8$, $SD = .492$) at the $p = .613$ level.

On the other hand, Figure 4.40 below presents the mean goodness scores obtained from the L2 identification task for all three learner groups for the emphatic sounds in the NSR and FSR. In this figure, I can see that the mean results for the three learner groups demonstrate a gradual increased pattern.

Figure 4.40. Mean L2 Goodness for Three Learner Groups for Emphatic Sounds in NSR and FSR.



The interpretation that comes out of observing this pattern is similar to what I said earlier about the results in Figure 4.38; the three learner groups show a gradual increase in their confidence about their labelings of the L2 plain sounds reflecting a L2 learning development. To confirm this statement, I conducted a one-way ANOVA to compare the effect of L2 exposure on goodness ratings of labeled L2 emphatic sounds. The results show that there was a significant effect of L2 exposure on the goodness results of labeled L2 emphatic sounds level for the three groups $F(2, 47) = 14.797$, $p = .000$. Post hoc comparisons indicate that the mean score for the NG ($M = 5.34$, $SD = .211$) was significantly different from the IG ($M = 5.92$, $SD = .572$) and the AG ($M = 6$, $SD = .511$) at the $p = .000$ level.

So, to answer my research question, based on these results, I can now generalize and say that the four plain sounds did not show L2 development with more L2 exposure; accuracy results differed but did not show a gradual learning pattern between the three learner groups. On the other hand, I can say that the four emphatic sounds demonstrated gradual development with more L2 exposure; accuracy results reflected better performance as L2 experience increased. The results could also help us address the first research question and look at it from a different angle in light of SLM. In chapter 3, I attempted to answer the first research question: *What are the perceptual category mappings of Arabic plain and emphatic consonants?* As I discussed earlier in this chapter, SLM categorizes L2 sounds as either “new” or “similar” sounds. Based on these two categories, L2 learners are expected to show specific learning patterns; “similar” sounds will not show much learnability improvement as language proficiency increases while “new” sounds will improve gradually with more exposure and surpass the performance of the “similar” sounds at some point. Based on that, I can now interpret my results in the figures above. Figures 4.38 and 4.40 demonstrate the latter pattern. This gradual improvement with more exposure leads us to believe that these four emphatic sounds fall under the “new” sounds category. On the other hand, Figures 4.37 and 4.39 demonstrate the first pattern. This lack of performance improvement or somewhat stable accuracy regardless of language exposure leads us to believe that these four plain sounds fall under the “similar” sounds category.

To summarize, this chapter attempted to examine the learning development of Arabic plain and emphatic consonants and whether L1-to-L2 mappings are affected by L2 exposure. An L1 labeling task and L2 labeling task followed by a goodness-of-fit rating task were conducted in order to see how three leaner groups with three varying L2 exposure levels perform. The results were analyzed according to the tested conditions, NSR versus FSR, and short vowel /a/ versus long vowel /a:/ conditions. After examining the results, it was found that the L1-to-L2 mappings are not affected by more L2 exposure; the proportions did not demonstrate differences between the three learner groups. However, the overall goodness ratings were higher in the AG than the NG and IG indicating that the similarities between the L1 and L2 categories are rated differently with more L2 exposure. I believe that the L2 listeners are becoming more sensitive to the L2 categories and that is affecting their goodness ratings of the selected categories. This is particularly clear from the observed pattern of the emphatic sounds. L2 listeners’ goodness increased for these sounds which indicates that they are becoming more sensitive to these sounds. . Also, it was found that the plain consonants did not show much improvement as L2 exposure increased. On the other hand, the results showed that the emphatic consonants showed gradual improvement as L2 exposure increased. The results were also interpreted in light of SLM and it was found that the plain sounds show a pattern that corresponds to the “similar” sounds, and the emphatic consonants show a pattern that corresponds to the “new” sounds.

Chapter 5 General Discussion

In this chapter I will discuss the findings of the current study and present answers to my proposed research questions. Each section below explains the answers to each one of the four research questions in light of the results that were presented and discussed in the earlier chapters of the study. The chapter will also discuss the results in light of the current perceptual models SLM and PAM. In addition, the results will be applied to Park and de Jong's (2008) analysis for comparison purposes.

5.1. Arabic Plain and Emphatic Perceptual Category Mappings with English (Answer to Research Question 1)

This section presents an answer to the first research question of the study: *What are the perceptual category mappings between Arabic plain and emphatic consonants with English consonants in the mind of native speakers of American English?* The results from the perceptual category mapping experiment in chapter 3, which used an identification task and a goodness-of-fit rating task in order to establish perceptual category mappings between Arabic and English categories in the mind of native speakers of American English with no prior experience with Arabic, were analyzed following Guion *et al.* (2000) and the tested consonants were categorized into categories based on the SLM framework. The results indicate that with both short /a/ and long /a:/, the Arabic consonants /t, d, ð, s, $s^ʕ$/ are considered "similar" sounds to the English categories based on their "good" degrees of fit, that the Arabic consonants /$t^ʕ$, $ð^ʕ$/ are "less similar" sounds to the English categories based on their "fair" degrees of fit, and that the Arabic consonant /$d^ʕ$/ is a "new" sound to the English categories based on its "poor" degree of fit. No prosodic location effect was observed between the two syllable structures CV and VCV. However, vowel qualities and durations affected the English category selection and the degree of fit for some, but not all, of the Arabic consonants. It could be possible that listeners were paying attention to the acoustic details of the vowels and that had affected their category selection(s). I conclude that the Arabic vowels must be taken into account and are a crucial factor to consider with regards to the perception of the Arabic consonants.

5.2. Arabic Plain and Emphatic L2 Development (Answer to Research Question 2)

This section presents an answer to the second research question of the study: *Does more L2 exposure over time lead to similar learning development with Arabic L2 plain and emphatic sounds?* As I discussed the results from the L2 labeling task and goodness-of-fit rating task in chapter 4, the answer is "No". Arabic plain and Arabic emphatic sounds did not show a similar learning development with more L2 exposure. I examined the overall development of the plain and emphatic sounds in the three different learner groups (NG, IG, and AG) by examining the accuracy results from the L2 labeling task and the goodness ratings. The results of the plain sounds showed that the NG's performance was higher than the performance of the IG, and the performance of the IG was higher than the performance of the AG (Figure 4.37). This was an indication that the three learner groups did not show any L2 learning development for L2 Arabic plain sounds with more exposure. This finding was supported by a one-way ANOVA to compare the effect of L2 exposure on accuracy of labeling L2 plain sounds. A significant effect of L2 exposure on the accuracy of labeling the plain sounds was found for the three groups. This significance, however, was not an indication of L2 development. On the contrary, as I explained above, this was due to the decrease of accuracy performance rather than a positive increase that you would expect with more L2 exposure. As I explained earlier, this decrease in the accuracy of performance for the plain sounds among the learner groups could be either attributed to a perceptual confusion that is due to the learners' attempt to distinguish the differences between the plain and emphatic consonants, or it could be attributed to the vowel conditions /a/ and /a:/. As I explained in chapter 1, vowel quality of /a/ and /a:/ is different than the English vowels and even though the main task asks listeners to focus only on the consonants of the stimuli, listeners' focus may have deviated from the main task and listeners were actually paying more or equal attention to the vowels used in the syllable structures. Consequently, accuracy rates fluctuated for the plain Arabic sounds among the learner groups. So, Arabic L2 plain sounds did not show any learning improvement with more L2 exposure.

On the other hand, I also examined the overall development of the emphatic sounds in the three different learner groups (NG, IG, and AG) by examining the accuracy results from the L2 labeling task and the goodness ratings. The results showed that the accuracy performance of the AG was higher than the accuracy performance of the IG, and the accuracy performance of the IG was higher than the accuracy performance of the NG (Figure 4.38). This was an indication that the three learner groups did this time show a gradual L2 learning development for L2 Arabic emphatic sounds with more L2 exposure. This was supported by a one-way ANOVA to compare the effect of L2 exposure on accuracy of labeling L2 emphatic sounds.

My answer to the research question was also supported by the findings of the goodness ratings of the L2 labeling task. The goodness ratings of the plain and emphatic sounds demonstrated similar result patterns to the accuracy result patterns. In other words, the high goodness ratings of the plain sounds indicate that the L2 listeners were very confident about their chosen L2 categories. As for the emphatic sounds, the gradual increase in the goodness ratings of these sounds indicate that L2 listeners are becoming more confident about their chosen L2 categories with more L2 exposure. Results from a one-way ANOVA to compare the effect of L2 exposure on goodness ratings of labeled L2 plain sounds showed that there was no significant effect of L2 exposure on the goodness results of labeled L2 plain sounds for the three groups (Figure 4.39). On the other hand, there was a significant effect of L2 exposure on the goodness results of labeled L2 emphatic sounds for the three groups (Figure 4.40).

With regards to the differences between the NSR condition and the FSR condition for the three learner groups, my findings show that in the NSR, the three learner groups show better accuracies in the /a:/ condition than in the /a/ condition. A mixed-design ANOVA with a within-subjects factor of *vowel duration* (short vowel /a/ vs. long vowel /a:/) and a between-subject factor of *groups* (NG vs. IG vs. AG) was performed. The results showed that there was a main effect of *vowel duration* without any significant interaction among the *groups*. This means that the accuracies were higher for the long vowel condition /a:/ than for the short vowel /a/ condition. This leads us to believe that the perception of sounds is easier with a long vowel than a short vowel. I also noticed that the accuracy difference between the long and the short vowels was larger among the NG than other groups (Figure 4.8). However, as time goes on, and with more L2 exposure, L2 accuracies for the short vowel /a/ condition improved and started to catch up with the long vowel /a:/ condition suggesting that L2 exposure affects the overall L2 accuracy with short vowel /a/. In the FSR, the three learner groups showed no difference in the accuracies for when /a:/ and /a/ were used. According to a mixed-design ANOVA with the *vowel duration* (short /a/ vs. long /a:/) as the within-subjects factor, and *groups* (NG vs. IG vs. AG) as the between-subjects factor, there was neither main effect of *vowel duration* nor significant interaction among the *groups*. I also noticed that the accuracy difference between the long and the short vowels was larger among the NG and AG (Figure 4.23), suggesting that the short vowel /a/ condition was more difficult than the long vowel /a:/ condition for the NG and AG. I believe that this may be due to the increased difficulty of the FSR task.

As for the accuracy performance of plain and emphatic sounds in the NSR and FSR for the three learner groups, my findings show that in the NSR, the accuracies of the plain sounds were better than the accuracies of the emphatic sounds in both /a/ and /a:/ conditions in the three learner groups. A mixed-design ANOVA with the *consonants* (plain vs. emphatic) and *vowels* (short /a/ vs. long /a:/) as the within-subjects factor and *groups* (NG vs. IG vs. AG) as the between-subject factor was performed. The results showed that there was a main effect of consonant type. This means that the accuracies were higher for the plain sounds than the emphatic sounds (Figures 4.9, 4.10, and 4.11). The results showed no significant interaction between the vowels and consonants. In the FSR, accuracies of plain sounds were also better than accuracies of emphatic sounds in the three learner groups. A mixed-design ANOVA with the *consonants* and *vowel duration* (plain vs. emphatic vs. short /a/ vs. long /a:/) as the within-subjects factor, and *groups* (NG vs. IG vs. AG) as the between-subjects factor was performed. There was a main effect of consonant type (Figures 4.24, 4.25, and 4.26). In both NSR and FSR conditions, the results from the AG showed that the accuracies of the plain sounds were in general better than the accuracies of the emphatic sounds but I did not see any significance.

As for the goodness ratings of plain and emphatic sounds in the NSR and FSR for the three learner groups, in the NSR condition average goodness ratings of plain sounds were better than those of emphatic sounds in the three groups. I conducted a mixed-design ANOVA with the *consonants* and *vowel duration* (plain vs. emphatic vs. short /a/ vs. long /a:/) as the within-subjects factor, and groups (NG vs. IG vs. AG) as the between-subjects factor. The results indicate that there was a main effect of consonant type. This means that the goodness were higher for the plain sounds than the emphatic sounds (Figures 4.13, 4.14, and 4.15). In the FSR condition, my results showed that goodness of plain sounds were better than goodness of emphatic sounds in the three groups. A mixed-design ANOVA with the consonants and vowel duration (plain vs. emphatic vs. short /a/ vs. long /a:/) as the within-subjects factor, and groups (NG vs. IG vs. AG) as the between-subjects factor was performed. The results showed that there was a main effect of consonant type (Figures 4.28, 4.29, and 4.30). This means that the goodness ratings were higher for the plain sounds than the emphatic sounds.

So, the answer to my second research question is "No". L2 Arabic plain and emphatic sounds did not show similar learning development across time with more L2 exposure. The four plain sounds did not show L2 development with more L2 exposure; accuracy results differed but did not show a gradual learning pattern between the three learner groups. On the other hand, the four emphatic sounds demonstrated a gradual development with more L2 exposure; accuracy results reflected better performance as L2 experience increased.

5.3. English L1 to Arabic L2 Mappings (Answer to Research Question 3)

In this section, I present an answer to the third research question of the study: *Do L2 learners' L1-to-L2 mappings change over time?* As I discussed the results from the L1 labeling task and goodness-of-fit rating task in chapter 4, the answer is "No". The results showed that L1-to-L2 mappings are not affected by more L2 exposure based on the mapping pattern results; the obtained proportions did not demonstrate differences between the three tested learner groups. I believe that some minor differences that were found in the FSR condition could be attributed to a listener confusion created by the increased difficulty of the FSR task. The results, however, unlike the mapping pattern results, showed that the overall goodness ratings were higher in the AG than the NG and IG reflecting a better performance across time. This increase in the goodness ratings leads us to believe that there is a subtle change of L1-to-L2 mappings across time manifested by how learners determined the similarity levels between the L1 and L2 sounds.

By comparing the results in the NSR and FSR conditions for the three groups, I noticed that the NG demonstrated similar mapping patterns between Arabic and English categories for both vowels /a/ and /a:/ with one exception. All Arabic consonants showed a one-to-one Arabic to English mapping pattern except Arabic /ðˤ/ and /ð/ which showed a one-to-two Arabic to English mapping pattern in the NSR and FSR. The main confusion for these two sounds was created by the Arabic voiceless fricative /θ/, which shares the same place and manner of articulation. In the FSR, however, Arabic /ð/ showed a one-to-one mapping pattern with /a:/ and Arabic /ðˤ/ showed a one-to-one mapping pattern with /a/. I believe this was due to the increased difficulty of the FSR condition. Results from the IG and AG also showed similar mapping patterns between Arabic and English categories for both vowels with similar NG exceptions.

As for the goodness ratings in the NSR and FSR for the three groups, my results showed that the overall mean goodness ratings were higher in the AG than in the IG and NG in both the NSR and FSR. A one-way ANOVA to compare the effect of L2 exposure on goodness ratings of labeled L1 English categories was performed. The results showed that there was a significant effect of L2 exposure on the goodness results of labeled L1 English categories for the three groups (Figure 4.7). This is an indication to the gradual change in the L1-to-L2 mappings between English and Arabic with more L2 exposure. I believe that the more exposure to L2

Arabic affected, not the mapping patterns, but the levels of similarity between L1 and L2 sounds. In other words, L2 listeners became more sensitive to the differences between the L1 and L2 sounds as their exposure increased.

I also found that by comparing the results of the AG with the results of the NG and IG in the FSR, listeners of the AG showed lower proportions for some of the selected English categories when long /a:/ was used (Tables 4.10, 4.12, and 4.14). I interpret this as a confusion among the AG listeners due to the increased level of difficulty of the FSR.

So, the answer to the third research question is “No”. L1-to-L2 mappings are not affected and do not change by more L2 exposure based on the mapping pattern results. However, the answer is “Yes” based on the similarity ratings. Overall goodness ratings were higher in the AG than the IG, and goodness ratings were higher in the IG than the NG. This, I believe, is an indication to a subtle change across time with more L2 exposure.

5.4. Predicting Identification Patterns from Category Mappings

As I explained earlier in chapter 2, Park and de Jong (2008) proposed an approach to quantify the degree to which L1 categories are used in L2 category identification. This approach was used in order to examine the identification of English obstruents by L2 Korean learners of English. By following their analysis, I present a detailed assessment to the degree in which the identification patterns of L2 Arabic demonstrate a reliance on L1 English categories. This quantitative technique of assessment allows to interpret the reliance explained earlier, if found present, as evidence for the entanglement of the L1 and L2 categories (Park and de Jong, submitted. The analysis relies on the mapping data in order to distinguish these two possibilities. More specifically, the analysis generates predictions based on the obtained mapping data as to what a null hypothesis would predict, that is if listeners use their L1 categories for the L2 identification of segments in terms of identification accuracy. Based on these predictions, the analysis assumes that identification accuracies greater than what is expected based on the obtained mapping data would be attributable to a development beyond relying on L1 categories. In this section I combine my results from the L1 labeling task and L2 labeling task and apply them to Park and de Jong's analysis to examine the relationship between them.

Park and de Jong's (2008) analysis relies on two approaches for generating L2 accuracy predictions from the mapping data. The first approach models the probability of having a direct link between the L2 response and L2 production through summing up the probabilities of getting the same L2 response by the means of each L2 to L1 mapping. In other words, the probable paths that are involved here are the probable path(s) between the produced L2 and its mapped category L1, and the probable path(s) between that L1 category and a selected L2 response category. So, the analysis relies on summing up these paths that start from an L2 production onto L2 categories but go through one or, more often, multiple L1 categories. To clarify, below I present examples of how the analysis works assuming that native English listeners participated in an L1 identification task where they listened to Arabic stimuli and labeled them according to English categories. So, for instance, let's say that L2 Arabic /$d^ʕ$/ was labeled 93.33% of the time as L1 English /d/, and that the L1 English category /d/ was also selected for other L2 Arabic consonants at varying percentages; 95% /d/ and 9.16% /$t^ʕ$/, as shown in Table 5.1 below. This means that English category /d/ has connections to three Arabic categories. As a result, the connection between Arabic /$d^ʕ$/ and English /d/ is not 93.33%. Rather, it is only 47% (=

0.93/(0.93 + 0.95+ 0.09)). Further, I may predict that if listeners are asked to identify Arabic /dˤ/, the probability to get this identification correctly will be 43% (= 0.93 x 0.47) under the assumption that the listeners identify Arabic /dˤ/ using only English /d/ category.

Table 5.1. Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and Their Mean Ratings (in parentheses). Modal Responses Are Bolded.

	English Categories						
Arabic Stimuli	**/b/**	**/d/**	**/ð/**	**/f/**	**/s/**	**/t/**	**/θ/**
/dˤ/	2.5 (2.67)	**93.33 (6.13)**				4.17 (6)	
/d/		**95 (6.38)**				5 (6)	
/sˤ/					**100 (6.67)**		
/s/			0.83 (3)	0.83 (5)	**98.34 (6.64)**		
/tˤ/	1.67 (3.5)	9.16 (6.27)				**89.17 (6.19)**	
/t/						**100 (6.38)**	
/ðˤ/			**61.67 (5.61)**	4.17 (5.6)		5 (5.5)	29.16 (6.11)
/ð/			**80 (6.01)**				20 (6)

The second approach of Park and de Jong (2008) is a duplicate of the first approach with one exception. It considers the mean goodness rating to weight the L1-to-L2 category mapping. Park and de Jong argue that using the mean goodness ratings will increase the estimation of the match between the L1 and L2 categories (Cebrian, 2006). The second approach uses an analysis similar to Guion *et al.*'s (2000) fit index in which each one of the categorical mappings will be weighted by the mean goodness rating of every one of the cases in which a specific L1 category is applied to a stimulus of a specific L2 category. The model attempts to increase the "dynamic range of the goodness ratings" by removing the lower end of the response scale through subtracting a lower threshold determined to be the lowest mean value for any of the L1 category mappings which were used more than 12.5%; the chance level that was determined for this experiment (Park and de Jong, 2016).

So, following Park and de Jong's (2008) analysis, the protocols for accuracy and specific error predictions based on L1 labeling that are adopted are as follows:

1) Predictions of accuracy based on the confusion between L1 and L2
Probability of accuracy where category A is perceived as category A = ∑ (probability of category A being perceived as L1 category X x probability of L1 category X being associated with category A). (Park and de Jong, 2008).

2) Predictions of specific errors based on the confusion between L1 and L2
Probability of accuracy where category A is perceived as category B = ∑ (probability of category A being perceived as L1 category X x probability of L1 category X being associated with category B). (Park and de Jong, 2008).

Additionally, predictions that incorporate the goodness ratings are generated following a third protocol. As explained earlier, "each categorical mapping is weighted by its mean goodness rating of all the cases in which that L1 category is applied to a stimulus of a specific L2 category" (Park and de Jong, 2008, p. 710). The protocol for weighting the connection between L1 and L2 according to their mean similarity rating is:

3) Weighted proportion of L2 category X in L1 category Y
Proportion of L2 category X in L1 category Y = {probability of L2 category X is perceived as L1 category Y x (its mean similarity rating score – 3.5)}/∑ {probability of all L2 categories associated with L1 category Y x (its mean similarity rating score – minimum mean goodness ratings for the chosen categories)}. (Park and de Jong, 2008).

So, based on the algorithms mentioned above, the analysis yields two possible outcomes. The first outcome, called the *unweighted model*, demonstrates a bi-directional mapping relationship of two folds. First, it shows the mapped relationship between the produced Arabic

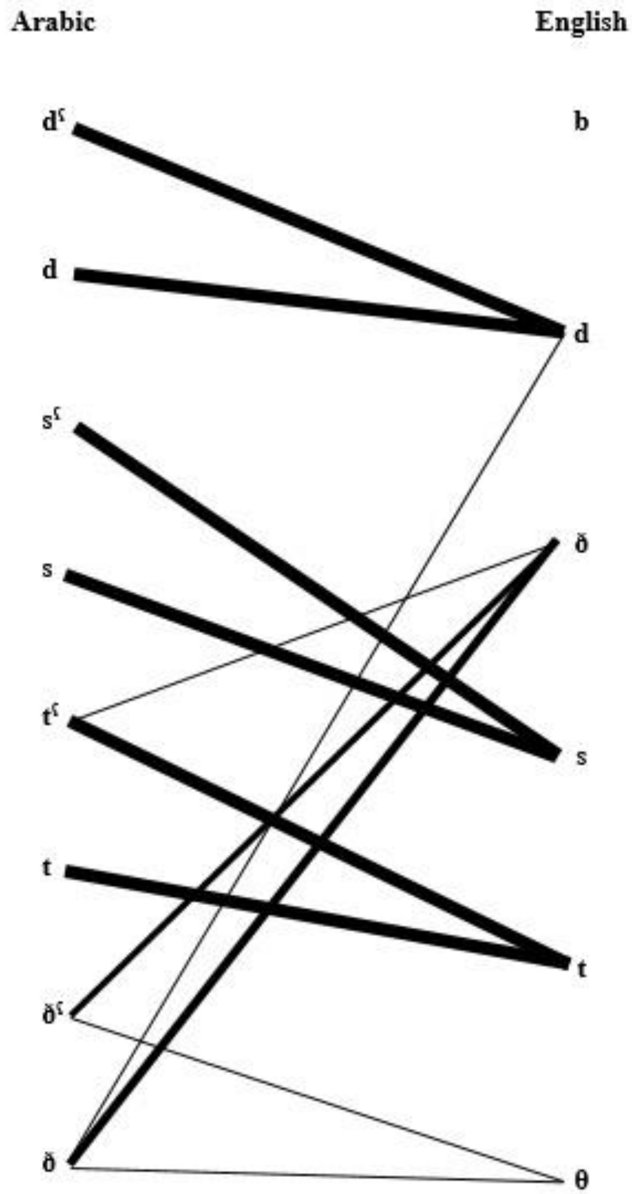
L2 consonants and their selected English L1 categories; this is based on the L1 Labeling Task. Second, the model maps the probability of an Arabic L2 category in an English L1 category; this is based on summing up the path from an Arabic L2 consonant to an English L1 category in addition to all the paths that link Arabic L2 consonants to the same English L1 category, as explained earlier. For instance, consider Table 5.2 below.

Table 5.2. Matrix Showing Mean Percentage Labeling of Arabic Consonants with English Categories and Their Mean Ratings (in parentheses).

	English Categories				
Arabic Stimuli	**/d/**	**/ð/**	**/s/**	**/t/**	**/θ/**
/dˤ/	**100 (6.23)**				
/d/	**100 (6.25)**				
/sˤ/			**100 (6.53)**		
/s/			**100 (6.48)**		
/tˤ/		2.5 (6)		**97.5 (6)**	
/t/				**100 (6.2)**	
/ðˤ/		**72.5 (6.03)**			27.5 (6.36)
/ð/	2.5 (6)	**80 (6.22)**			17.5 (6.29)

This table presents the proportion of English labeling chosen by the listeners for each Arabic stimulus with its mean goodness rating presented in parentheses. Bolded proportions and goodness ratings indicate that this English L1 category was selected the most by the listeners. The obtained data in this table serve as the L2 to L1 perceptual mappings of the Arabic consonants to English categories (Figure 5.1).

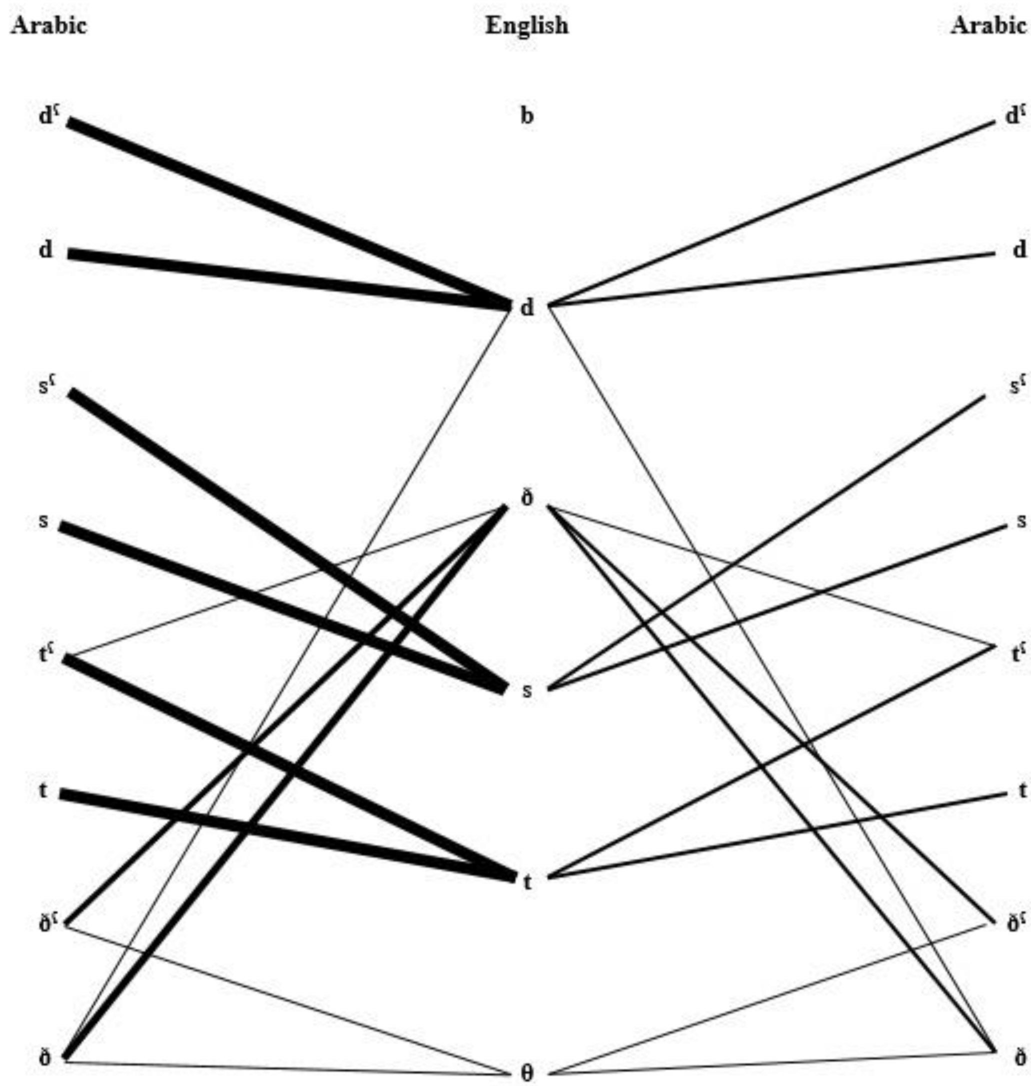
Figure 5.1. Perceptual Mapping of Arabic Consonants to English Categories. Lines' Widths Correspond to the Proportion of Times the English Category was Selected for the Arabic Production.



In Figure 5.1, the lines connecting the Arabic consonants on the left with the English categories on the right are synonymous to the shown proportions in Table 5.1. The thickness of each line indicates the overall obtained proportion; the thicker the line the higher the obtained proportion, and the thinner the line the lower the obtained proportion. So, for example, the thickness of the line that connects Arabic /dˤ/ with English category /d/ reflects the overall 100% proportion that was obtained and demonstrated in Table 5.1. It is important to note that it is possible to see multiple or more paths leading to the same L1 English category. In Figure 5.1, for instance, there are two lines leading to the English category /d/, one from Arabic /dˤ/, and another one from Arabic /d/. This means that English /d/ was selected as an L1 category in the labeling task for two Arabic consonants.

Second, the unweighted model shows the mapped relationship between the English L1 categories and the Arabic L2 consonants. In other words, Figure 5.2 plots the proportion of times an Arabic L2 consonant is selected and labeled with an English L1 category, while now, a reverse mapping is applied in order to include the proportion of times an Arabic L2 label corresponded with the same English L1 category. This is done by summing up the proportion of an Arabic L2 consonant with all other instances in which that English L1 category was labeled with other Arabic L2 consonants. The result will yield a second mapping probability between English and Arabic categories added to the right of the mappings in Figure 5.1, as shown in Figure 5.2. So, for instance, in order to map the probability between English /d/ and Arabic /dˤ/ I sum up, from left to right in Figure 5.1, all the paths that lead to choosing English /d/ as a category. In this particular case, English /d/ was selected as a category for L2 Arabic /dˤ/, /d/, and /ð/ (see Figure 5.1). I sum the paths up and divide them by the proportion of English label for Arabic category. Of these, only 49% (=100/(100 + 100 + 2.5)) are cases of Arabic /dˤ/ mapped, left to right, with English /d/, as shown below in Figure 5.2. As a result, a complete and symmetrical bi-directional perceptual mapping will be generated.

Figure 5.2. Bi-Directional Perceptual Mapping between English and Arabic Categories. Lines' Widths Correspond to the Proportion of Times the English Category was Selected for the Arabic Production.

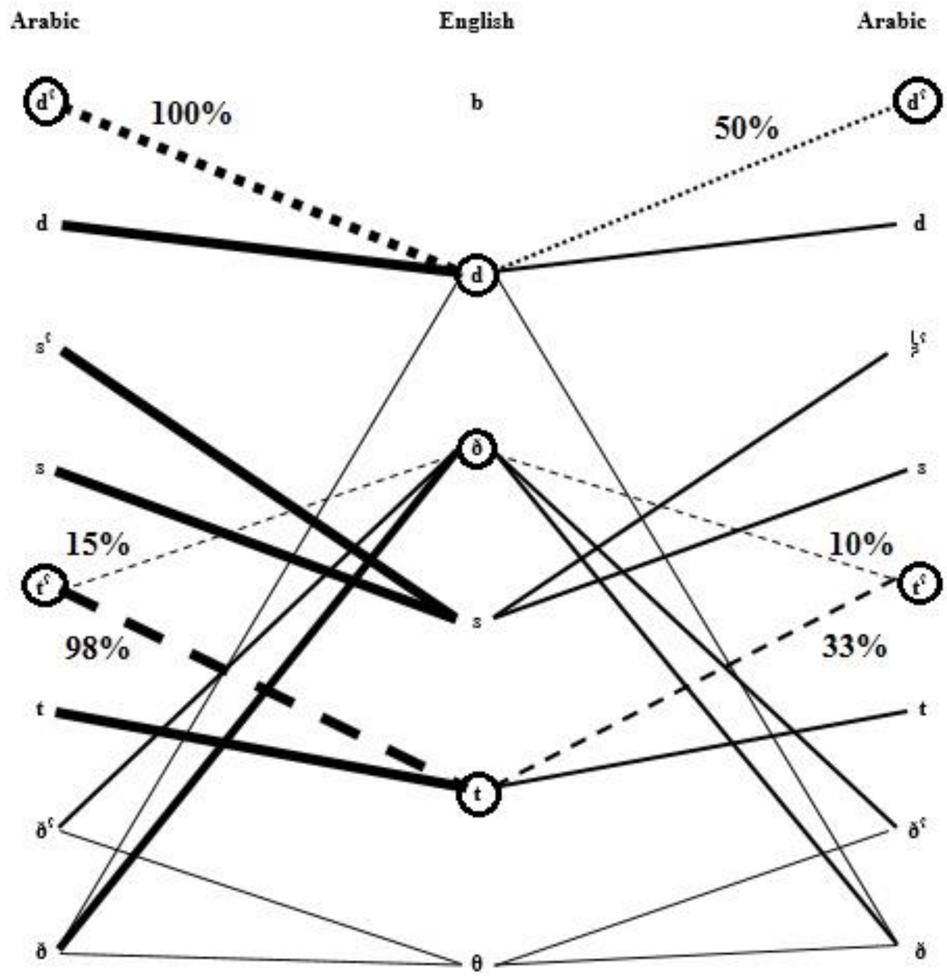


It is worth mentioning that the lines in the second perceptual mapping become less thick. This is a natural outcome due to the fact that the highly probable mappings from Arabic L2 consonants to English L1 categories become less probable from English L1 categories to Arabic L2 categories. For example, in Figure 5.2, the path connecting Arabic /dˤ/ with English /d/ means that Arabic /dˤ/ was heard as English /d/. However, there are paths connecting English /d/ with Arabic /d/ and /ð/. This means that Arabic /d/ and /ð/ were also heard as English /d/. Thus, the path connecting English /d/ with Arabic /dˤ/ will be less thick indicating less probability from English to Arabic following that specific path.

The next step in the analysis is generating specific predictions about English identification performance. In order to do that, the model relies on two factors, first, the model uses the proportions that connect the produced Arabic consonants with the selected English categories. Second, the model uses the proportions of Arabic categories in each English category. These are the probable data that were used in the bi-directional perceptual mapping that was explained earlier. So, by performing this analysis, two sets of predictions will be produced; predictions of accuracy, and predictions of errors. The predictions of accuracy consider the path connecting the produced Arabic stimulus with one or more English category which, in turn, connects to the correct Arabic category.

For example, in Figure 5.3 below, to calculate the predicted accuracy with which Arabic /dˤ/ is accurately perceived by English listeners, two paths are considered in order to do that. The first path is the one that connects Arabic stimuli /dˤ/ with English category /d/, and the second path is the one that connects English /d/ with the Arabic response /dˤ/. Both paths are represented by dotted lines in the figure. As a result, the probability that Arabic /dˤ/ is identified correctly as Arabic /dˤ/, based on the English categories, is calculated by multiplying the proportion of times that Arabic stimuli /dˤ/ were perceived as English category /d/ (100%) by the proportion of times the English category /d/ corresponds to Arabic /dˤ/ (50%), which is 50%.

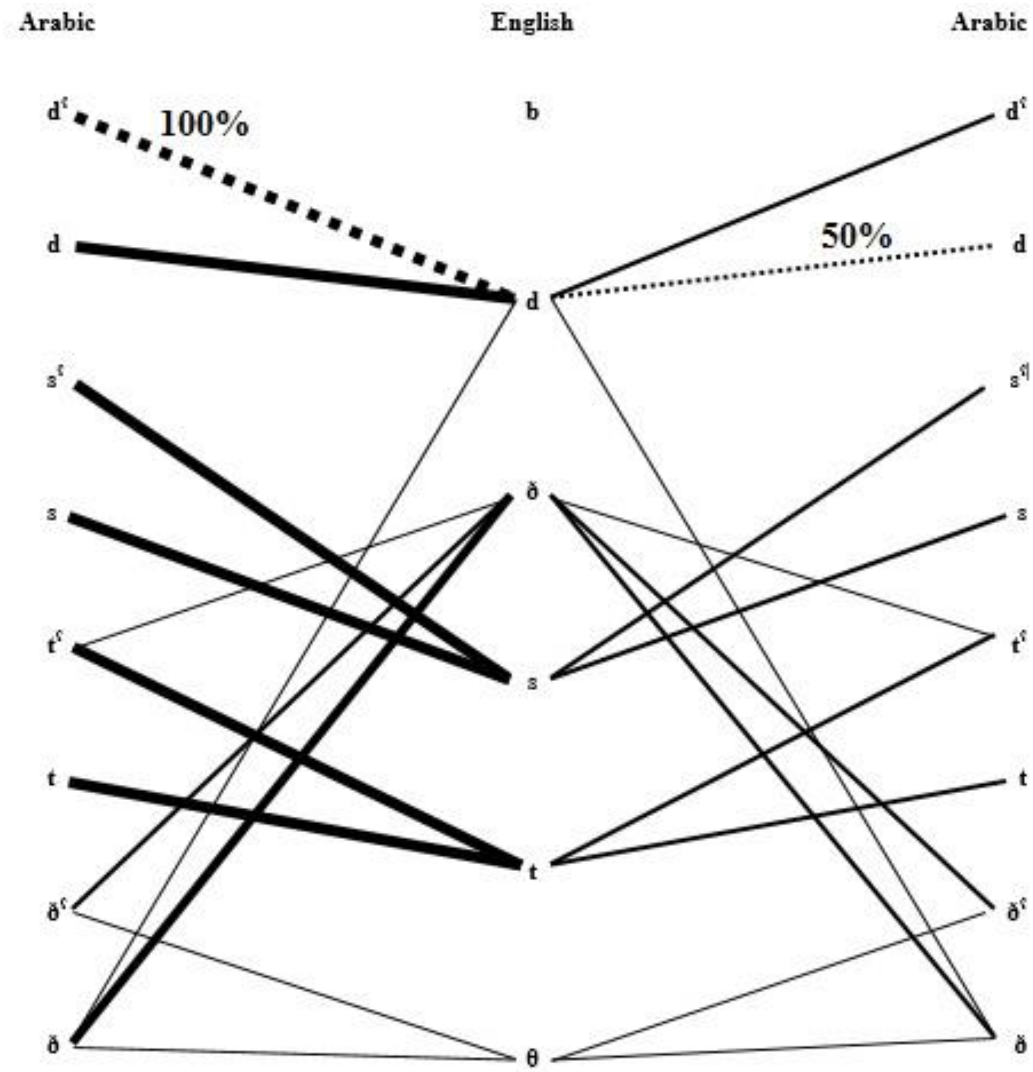
Figure 5.3. Sample Bi-Directional Perceptual Mapping Between English and Arabic Categories Showing Predicted Accuracy Calculations.



Park and de Jong's prediction of accuracy approach also considers multiple paths that connect an Arabic stimulus with more than one English category as long as the reverse mapping also connects the English categories to the same Arabic category. In Figure 5.3, for instance, to calculate the predicted accuracy with which Arabic /tˤ/ is perceived accurately by English listeners, four paths are considered in order to do accomplish that. The first path connects Arabic /tˤ/ with English category /ð/, and then English /ð/ to Arabic /tˤ/. The third and fourth paths connect Arabic /tˤ/ with English /t/, and then English /t/ to Arabic /tˤ/. All four paths are represented by dash lines in Figure 5.3 above. As a result, the probability that Arabic /$t^ʕ$/ is identified correctly as Arabic /$t^ʕ$/, based on the English categories, is calculated by multiplying the proportion of times that Arabic stimuli /$t^ʕ$/ were perceived as English category /ð/ (15%) by the proportion of times the English category /ð/ corresponds to Arabic /$t^ʕ$/ (10%), added to the multiplication of the proportion of times that Arabic stimuli /$t^ʕ$/ were perceived as English /t/ (98%) by the proportion of times the English category /t/ corresponds to Arabic /$t^ʕ$/ (33%), which is 33.84%.

The second set of predictions in which the model generates is the predictions of errors. These predictions are generated in the same way as the predictions of accuracy were generated. For example, in Figure 5.4 below, the predicted error rate at which English listeners misidentify Arabic /$d^ʕ$/ as Arabic /d/ is predicted by multiplying the proportions of times in which Arabic /$d^ʕ$/ connects to English /d/ (100%) by the proportion of times English /d/ connect to Arabic /d/ (50%), which yields 50% error rate. The paths associated with this calculation are represented by dotted lines in the figure below. Multiple paths connecting Arabic stimuli with multiple English categories are also calculated in the same way as explained above in the predictions of accuracy.

Figure 5.4. Sample Bi-Directional Perceptual Mapping Between English and Arabic Categories Showing Predicted Error Calculations.



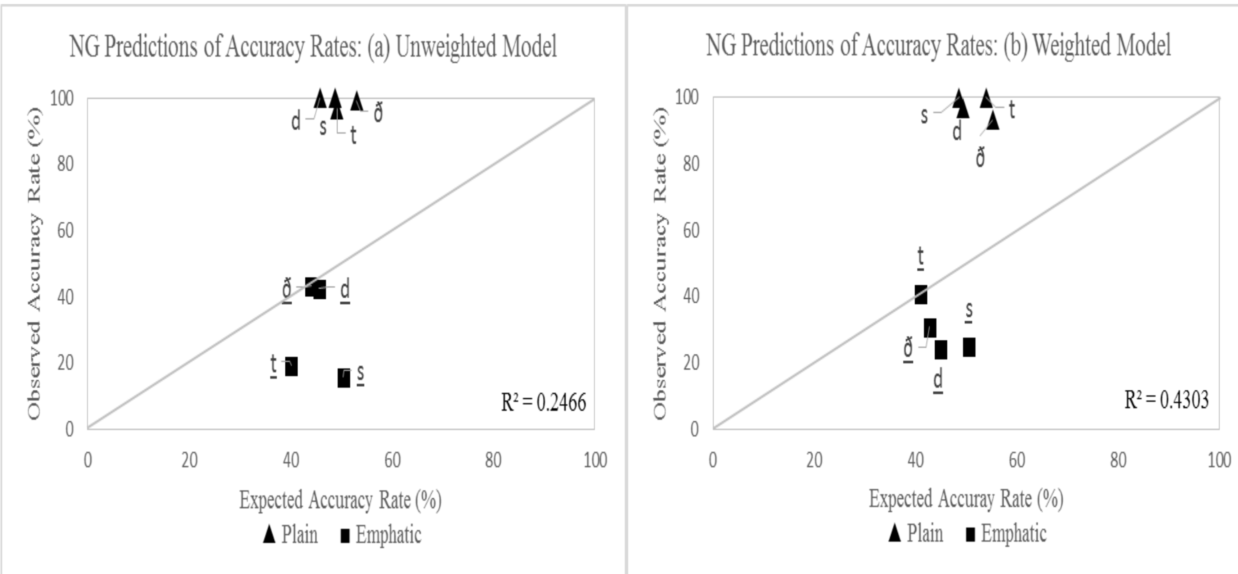
The second outcome in which the analysis generates is called the *weighted model*. In this model, the predicted accuracy and error rates incorporate the goodness rating data by weighting the probabilities of each connection between the English labels and the Arabic response labels by the adjusted goodness ratings (Park and de Jong, 2008). The accuracy and error rates are calculated using the same method that was explained earlier in the unweighted model. Using the goodness ratings in this calculation allow us to account for the gradual perceptual development in L2 in addition to augmenting the estimation of the connection between L1 and L2 categories as explained earlier.

Based on their analysis, the researchers were able to categorize the tested sounds into SLM's "new" and "similar" categories and extend SLM from production onto perception. I examine the relationship between L2 Arabic sounds and L1 English categories in light of Park and de Jong (2008) by presenting results after relating L2 identification performance in the L2 labeling task to the predictions that were calculated as explained earlier based on the data from the L1 labeling task. In other words, the figures below plot the accuracy and error predictions against the actual performance of the listeners in the *unweighted* and *weighted* models. As explained earlier, the predictions in the weighted model are calculated by weighting the probabilities of each connection between the English labels and the Arabic response labels by the adjusted goodness ratings.

5.4.1. Normal Speech Rate

Figure 5.5 below plots the accuracy in the L2 labeling task for each one of the four target plain and four target emphatic Arabic consonants against calculated predictions of both the unweighted and weighted models when the long vowel /a:/ was used for the NG. Left panel (a) plots the predictions without weighting them by the goodness ratings, while panel (b) on the right plots the predictions after weighting them by the goodness ratings. The x-axis corresponds to the predicted accuracy rates calculated form the L1 labeling task, and the y-axis to the observed accuracy rates that were obtained from the L2 labeling task. Plain sounds data are represented by triangle shapes and emphatic sounds data are represented by square shapes.

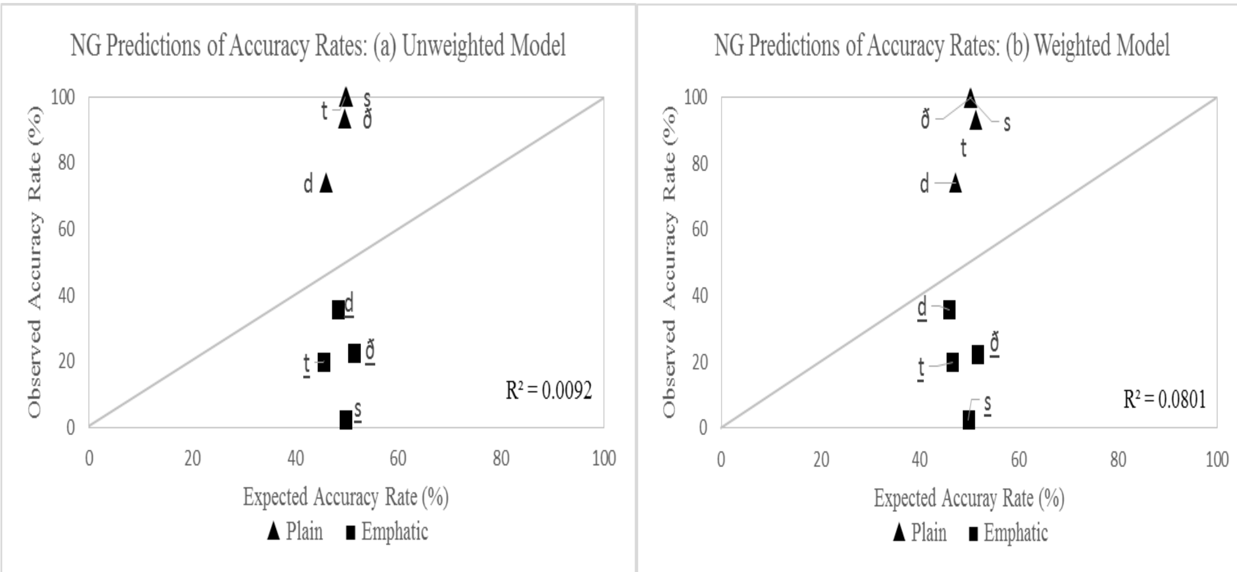
Figure 5.5. Accuracy Rate Predictions for NG Based On the L1 and L2 Labeling Tasks in NSR with Long Vowel /a:/ Condition. Left Panel (a) Plots Predictions without Weighting by Goodness Ratings, While Right Panel (b) Plots Weighted Predictions. The Line ($x = y$) Indicates An Exact Prediction By The Model.



According to Park and de Jong (2008), segments that happen to fall on the diagonal line in the figure indicate that listeners were relying on their L1 categories to identify those L2 sounds. Based on the figure above, I notice that only Arabic /ðˤ, dˤ, tˤ/ demonstrate a reliance on the L1 categories. The other sounds fall far from the line which means that listeners did not rely on their L1 categories in order to identify these sounds. In addition, as I explained earlier, plain sounds had higher observed accuracy rates than the emphatic sounds. Comparing panels (a) and (b) does not show much difference in terms of the reliance on the L1 categories.

Figure 5.6 below plots the accuracy in the L2 labeling task for each one of the four target plain and four target emphatic Arabic consonants against calculated predictions of both the unweighted and weighted models when the short vowel /a/ was used for the NG. Similarly, left panel (a) plots the predictions without weighting them by the goodness ratings, while panel (b) on the right plots the predictions after weighting them by the goodness ratings.

Figure 5.6. Accuracy Rate Predictions for NG Based on the L1 and L2 Labeling Tasks in NSR with Short Vowel /a/ Condition. Left Panel (a) Plots Predictions without Weighting by Goodness Ratings, While Right Panel (b) Plots Weighted Predictions. The Line ($x = y$) Indicates an Exact Prediction by the Model.



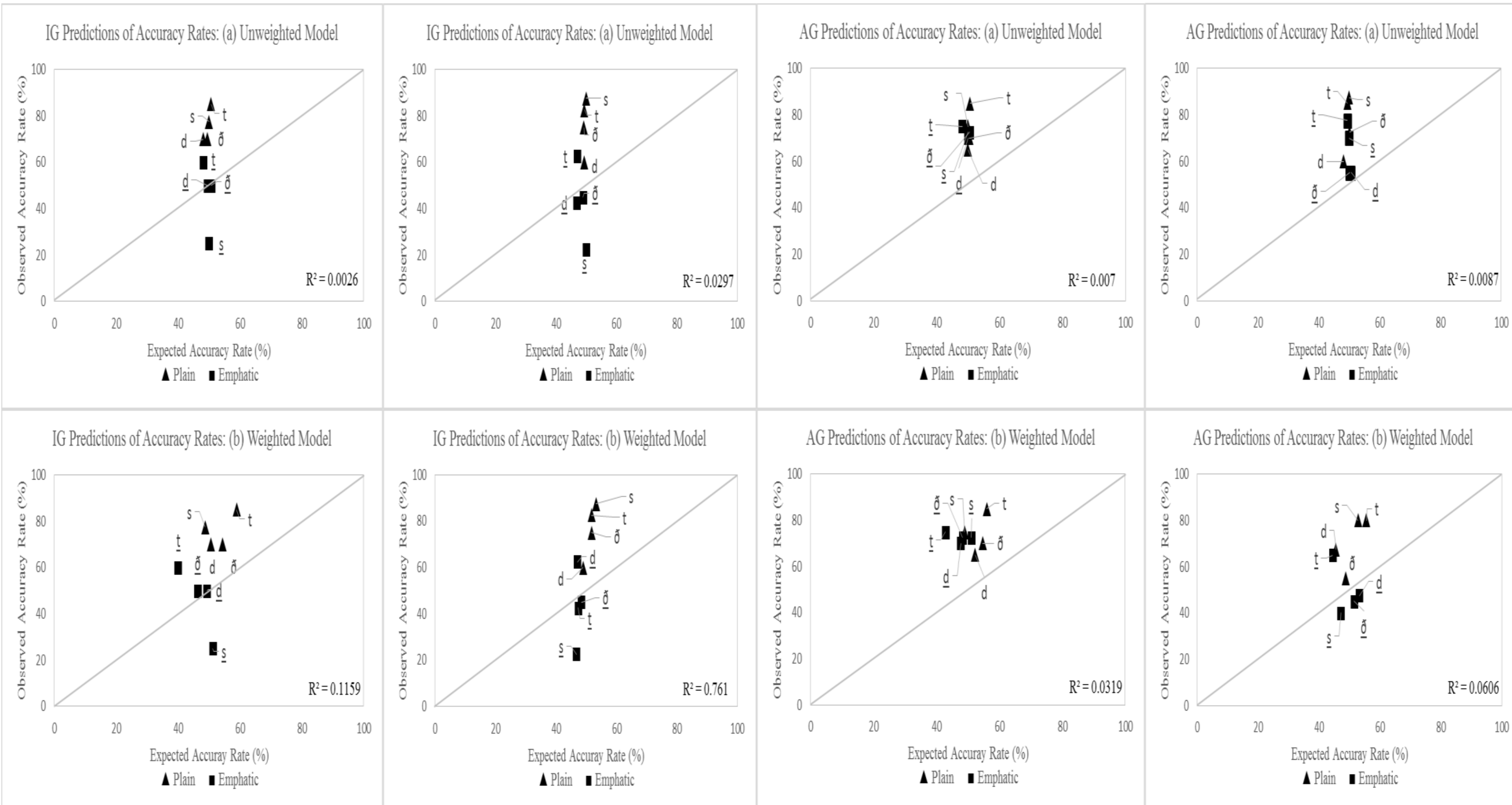
Based on the figure above, I notice that none of the sounds demonstrate a reliance on L1. Only Arabic /dˤ/ demonstrated some reliance on L1. Also, plain sounds had higher observed accuracy rates. Comparing panels (a) and (b) does not show any difference in terms of the reliance on the L1 categories. By comparing the long vowel condition in Figure 5.5 with the short vowel condition in Figure 5.6 I notice a slight difference between the two conditions in terms of reliance on L1; more reliance was observed on the long vowel condition. Also, overall observed performance of the listeners in the long vowel condition was better than the short vowel condition for both the plain sounds and emphatic sounds.

Figure 5.7 below plots the accuracy in the L2 labeling task for each one of the four target plain and four target emphatic Arabic consonants against calculated predictions of both the unweighted and weighted models when long /a:/ (top row) and shot /a/ (bottom row) were used for the IG and AG. Based on the figure, I also notice that there was also not much reliance on L1 by listeners. Listeners of the IG, however, showed some reliance on L1 for the Arabic sounds /ðˤ, dˤ/ and /tˤ/ in the short vowel condition. The results also did not show much difference between the two vowel conditions. Similar to what has been found earlier, the observed accuracy of the plain sounds was still better than the emphatic sounds. However, performance of the emphatic sounds in the AG was better than the IG which, in turn, was better than the NG.

So, by comparing the results from the three listener groups, I notice that there was no much reliance on the listeners' L1 in order to identify L2 Arabic sounds. Arabic /ðˤ, dˤ/ were the only two sounds that showed reliance on L1. To put this under SLM's scope, SLM states that the development patterns of "new" L2 sounds are more difficult to predict based on L1 models because they do not rely on L1 categories. As a result, I can state that Arabic /ðˤ, dˤ/ who showed reliance on L1 fall under SLM's "similar" category since they rely on L1. Also, based on the figures, I did not observe much difference between panels (a) and (b). That is an indication that using the goodness ratings with the weighted models did not impact the outcome of the results.

Finally, by comparing the observed accuracies for the three groups, the results in the figures indicate that listeners in all three groups performed better with the plain sounds than the emphatic sounds. The NG's performance actually surpassed the performance of the IG and AG for the plain sounds. As for the emphatic sounds, I can see a gradual increase in the performance among the three groups; the NG had the lowest performance followed by the IG and then the AG.

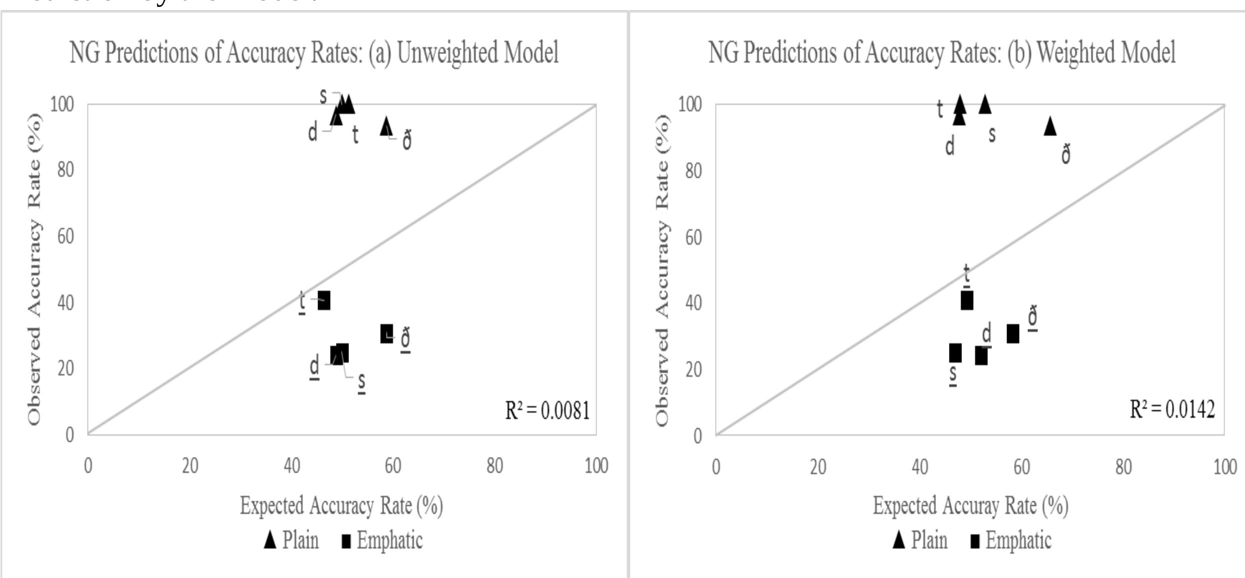
Figure 5.7. Accuracy Rate Predictions for IG and AG Based on the L1 and L2 Labeling Tasks in NSR with Long /a:/ (top row) and Short /a/ (bottom row) Condition. Panels (a) Plot Predictions Without Weighting by Goodness Ratings, While Panels (b) Plot Weighted Predictions.



5.4.2. Fast Speech Rate

Similar to what has been presented above, Figure 5.8 below plots the accuracy in the L2 labeling task for each one of the four target plain and four target emphatic Arabic consonants against calculated predictions of both the unweighted and weighted models when the long vowel /a:/ was used for the NG.

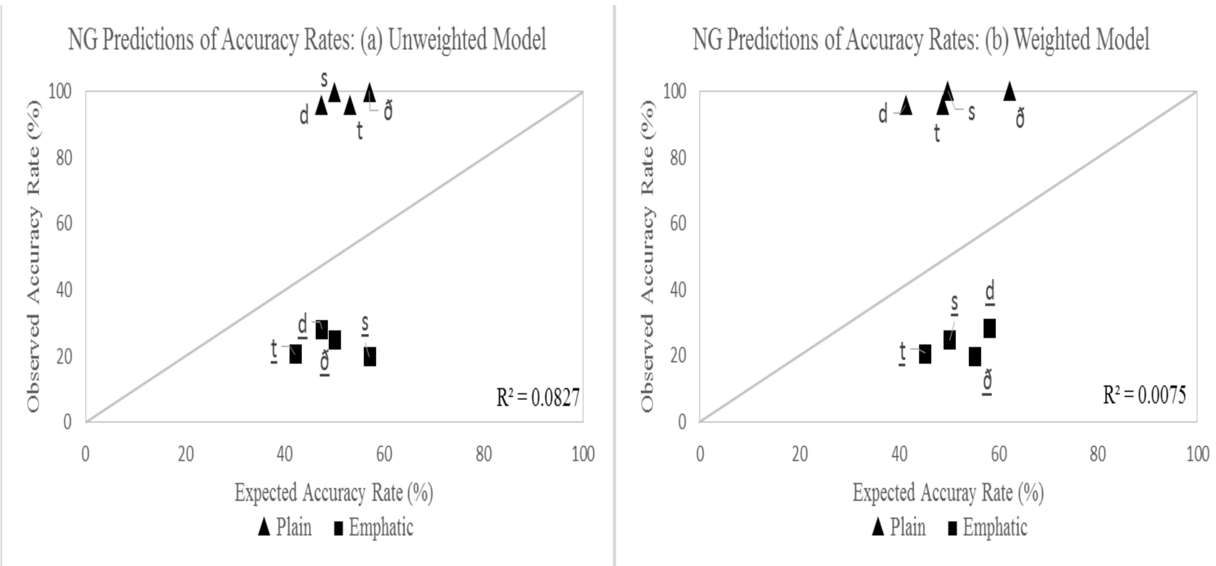
Figure 5.8. Accuracy Rate Predictions for NG Based on the L1 and L2 Labeling Tasks in FSR with Long Vowel /a:/ condition. Left panel (a) Plots Predictions without Weighting by Goodness Ratings, While Right Panel (b) Plots Weighted Predictions. The Line ($x = y$) Indicates an Exact Prediction by the Model.



Based on the figure above, I notice that only Arabic emphatic /tˤ/ demonstrates a reliance on L1. The other L2 sounds, especially the plain sounds, fall far from the diagonal line which means that listeners did not rely on their L1 categories in order to identify these sounds. In addition, as I have explained earlier, plain sounds had higher observed accuracy rates than the emphatic sounds. Also, comparing panels (a) and (b) does not show much difference in terms of the reliance on the L1 categories.

Figure 5.9 below plots the accuracy in the L2 labeling task for each one of the four target plain and four target emphatic Arabic consonants against calculated predictions of both the unweighted and weighted models when the short vowel /a/ was used for the NG.

Figure 5.9. Accuracy Rate Predictions for NG Based on the L1 and L2 Labeling Tasks in FSR with Short Vowel /a/ Condition. Left Panel (a) Plots Predictions without Weighting by Goodness Ratings, While Right Panel (b) Plots Weighted Predictions.



Based on the figure above, I notice that none of the sounds demonstrate a reliance on L1. Comparing panels (a) and (b) does not show any difference in terms of the reliance on the L1 categories. By comparing the long vowel condition in Figure 5.8 with the short vowel condition in Figure 5.9 I notice a slight difference between the two conditions in terms of reliance on L1; more reliance was observed on the long vowel condition. Also, overall observed performance of the listeners in the long vowel condition was better than the short vowel condition for both the plain sounds and emphatic sounds.

Figure 5.10. Accuracy Rate Predictions for IG and AG Based on the L1 and L2 Labeling Tasks in FSR with Long /a:/ (top row) and Short /a/ (bottom row) Condition. Panels (a) Plot Predictions Without Weighting by Goodness Ratings, While Panels (b) Plot Weighted Predictions.

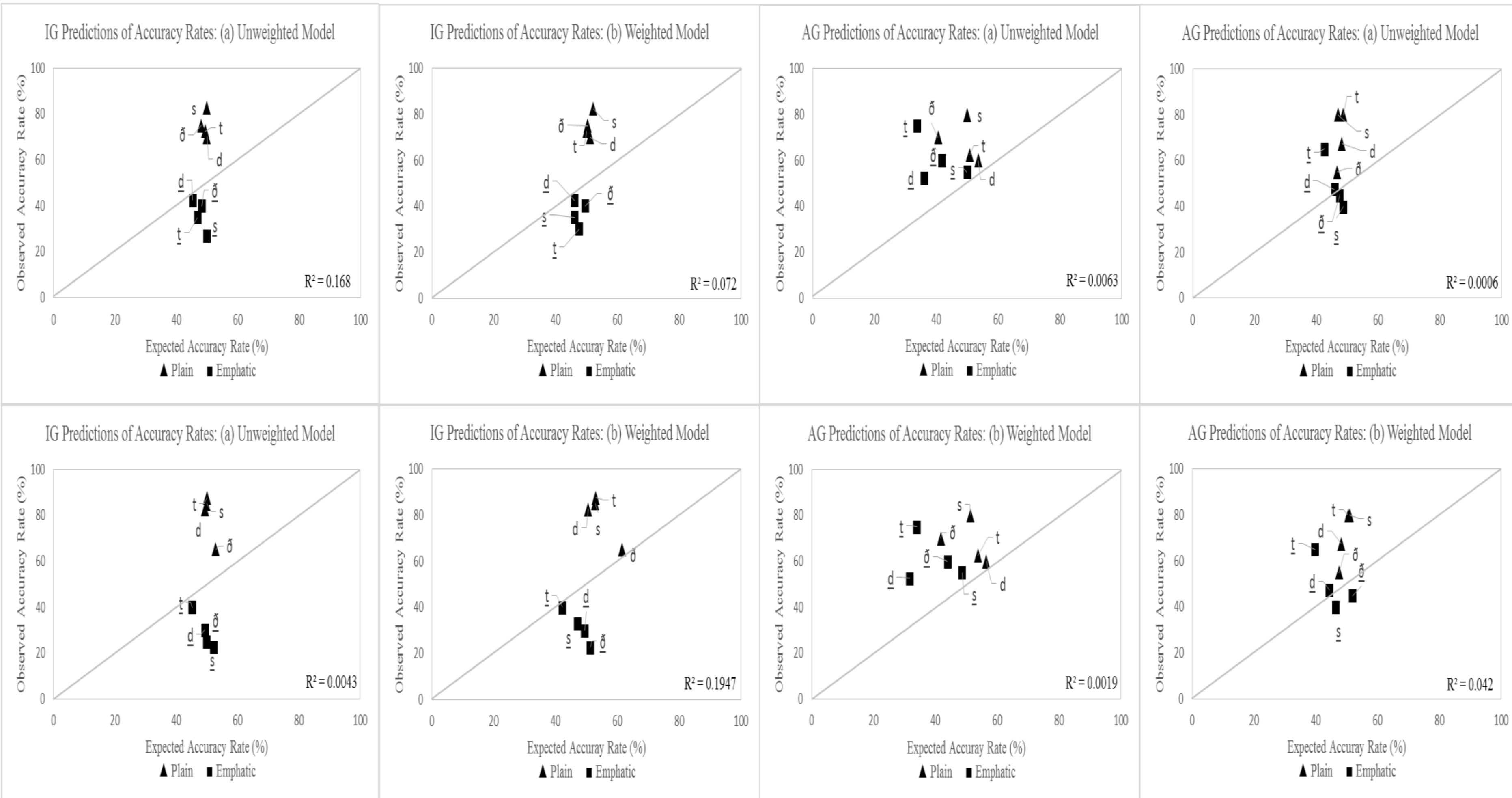


Figure 5.10. above plots the accuracy in the L2 labeling task for each one of the four target plain and four target emphatic Arabic consonants against calculated predictions of both the unweighted and weighted models when long /a:/ (top row) and shot /a/ (bottom row) were used for the IG and AG. Similar to what was found in the NSR condition, I also notice that there was also not much reliance on L1 by listeners. Listeners of the IG, however, showed some reliance on L1 for the Arabic sounds /dˤ/ in the long vowel condition, and /tˤ/ in the short vowel condition in both the unweighted and weighted models. As for the AG, the results show that Arabic emphatics /ðˤ, dˤ/ showed reliance on L1 in the weighted models, while Arabic /d/ and /sˤ/ showed some reliance on L1 in the unweighted models. The results also did not show much difference between the two vowel conditions. Similar to what has been found earlier, the observed accuracy of the plain sounds was still better than the emphatic sounds. However, performance of the emphatic sounds in the AG was better than the IG which, in turn, was better than the NG.

By comparing the observed accuracies for the three groups, the results in the figures indicate that listeners in all three groups performed better with the plain sounds than the emphatic sounds. The NG's performance actually surpassed the performance of the IG and AG for the plain sounds. As for the emphatic sounds, I can see a gradual increase in the performance among the three groups; the NG had the lowest performance followed by the IG and then the AG.

In general, my results did not come out as Park and de Jong's model expects them to be. It was not very clear to what extent L2 learners are using and facilitating their L1 categories in order to perceive the tested L2 sounds. The results showed that L2 learners performed better than what the model predicts; there was no relationship between the mappings and L2 perceptions.

However, based on the observed accuracies, the results were successful at showing how L2 exposure affects the overall learnability of L2 emphatic sounds; learners of the AG showed the highest performance rates followed by learners in the IG and then the NG. This is an indication to the overall L2 development that is linked to more L2 exposure. On the other hand, the observed accuracy results of L2 plain sounds did not demonstrate any performance development.

5.5. The Results in Light of SLM and PAM/PAM-L2 (Answer to Research Question 4)

In this section of the chapter, I discuss the fourth research question of the study: *How can the current perceptual models be evaluated based on the results?* Earlier in chapter 1, I presented a detailed explanation of the SLM and the tenet that was established by Flege (1988, 1992, and 1995). Interpreting the results in light of SLM could be established by successfully being able to categorize the tested L2 Arabic sounds into "similar" and "new" categories based on the obtained results from the identification task. As explained earlier, SLM predicts that for the "similar" sounds, learners will show no change in their L2 learnability of these sounds regardless whether their L2 proficiency increases or not. On the other hand, "new" sounds will demonstrate gradual L2 learnability that keeps improving with more L2 exposure. I noticed based on the observed accuracy results (L2 labeling task) earlier that the emphatic sounds demonstrated a gradual increase in their performance starting with the NG. Results showed that the AG had the highest performance rates among the three learner groups. This, in my opinion, is an indication that these emphatic sounds meet the "new" sounds criteria laid out by the SLM. On the other hand, my results showed that the plain sounds did not demonstrate any improvement in their performance across time with more L2 exposure. I conclude that these sounds must fall under the “similar” category according to the SLM because their performance did not change across time.

The results of the plain sounds in fact demonstrated a decrease in the performance pattern across time for all three learner groups. SLM attributes this change simply to the fact that “phonetic systems reorganize in response to sounds encountered in an L2 through the addition of new phonetic categories, or through the modification of old ones” (Flege, 1995: p. 233). In other words, this inconsistency in the performance of the plain sounds may be a direct result to the modification that is happening to their phonetic system.

One final issue that remains is to explain the differences between the categorization of the plain and emphatic sounds in the cross-language category mapping experiment in chapter 3 and the predictions of accuracy discussed above. I believe that the difference lies in the level of experience with L2. In the cross-language category mapping task, I was interested in naïve listeners who have no experience with Arabic. On the other hand, in the identification task, I was interested in L2 learners with varying L2 exposure (NG, IG, and AG). My explanation emanates from SLM’s claim that “as L2 learners gain experience in the L2, they may gradually discern the phonetic difference between certain L2 sounds and the closest L1 sound(s)” (Flege, 1995: p. 263).

With regard to the PAM-L2, I have explained earlier in chapter 1 that the model is interested in two distinct groups: Late L2 learners and Naïve monolinguals. Examining these two groups enables PAM-L2 to account for the amount of L2 exposure in learning nonnative (or L2) sounds. Moreover, I explained that PAM-L2 is primarily interested in the possibility of assimilation at the phonological level rather than the phonetic level which interests the SLM. In other words, PAM-L2 incorporates the influence of L2 learners’ developing phonetic and phonological knowledge of L2 which allows for perceptual assimilation at the gestural, phonetic, and phonological level (Chang, 2010). So, PAM-L2 relies on the gestural dimensions rather than the acoustic properties in order to form cross-language similarity which eventually allows perceptual assimilation to take place.

To examine the PAM-L2's predictions let us consider the mapping results from chapter 3 and the L1 labeling task in chapter 4. In chapter 3, I have established a cross-language perceptual mapping based on results obtained from naïve American English native speakers. The results showed, based on a fit index, the degree of similarity between an Arabic sound and a native L1 English category. Based on the results, /t-tˤ/ and /ð-ðˤ/ fell under the two category assimilation type (TC) and /d-dˤ/ and /s-sˤ/ fell under the category goodness assimilation type (CG). However, by examining the results from the AG in the in the L1 labeling task in Table 4.7, I can obtain their fit index by multiplying the proportion of identification with the goodness rating. Based on that fit index, the above assimilation types, /t-tˤ/ and /ð-ðˤ/, become among the category goodness assimilation type (CG). In addition, /d-dˤ/ and /s-sˤ/ fall under the same (CG) assimilation type. Even though the PAM-L2 requires a discrimination task in order to assess the success of its predictions, I can see to a certain extent the difference in assimilation types based on the L1 labeling task between the naïve English listeners and experience L2 Arabic learners. This is an indication to a preliminary perceptual development that needs to be addressed further with a suitable methodology. If I were to conduct a discrimination task, based on the assimilation types aforementioned, I would expect the naïve listeners to show a better performance for /t-tˤ/ and /ð-ðˤ/ than for /d-dˤ/ and /s-sˤ/. On the other hand, listeners from the AG would demonstrate a similar performance for all four assimilation type cases.

Chapter 6 Conclusion

In Chapter 1, I explained that perception and production theoretical models, such as the SLM and PAM/PAM-L2 have attempted to find answers to the difficulties faced by L2 learners, when they try to learn an L2. These models have proposed that such difficulties are related to the relationship between sounds in a learner's L1 and those in the target language. The chapter also compared and contrasted the SLM and the PAM, observing the SLM's interest in the phonetic/acoustic properties of the speech signal, versus the PAM's main interest in the actual articulatory gestures of the sounds; in addition, the SLM, as I observed, is interested in examining L2 learners, while the PAM is interested in naïve listeners. I have also presented a number of studies which attempted to modify the models and make them account for L2 perception development such as the PAM-L2 and Park and de Jong (2008). The chapter also showed that the aim of the study is to understand the degree to which Arabic is perceived by nonnative American English listeners, whether they were naïve monolinguals or L2 language learners. In the second chapter, I discussed the research questions in light of the theoretical models and I explained the general methods that will be used in order to answer the research questions.

Guion et al.'s (2000) analysis was used in Chapter 3 in order to answer the first research question: *What are the perceptual category mappings between Arabic plain and emphatic consonants with English consonants in the mind of native speakers of American English?* An identification task and a goodness-of-fit task were used in order to establish perceptual category mappings between the Arabic plain and emphatic sounds and the English categories in the mind of naive American English monolinguals. The results allowed us to classify the Arabic plain and emphatic sounds under the scope of the SLM. The results showed that the Arabic consonants /t, d, ð, s, $s^ʕ$/ are considered "similar" sounds to the English categories based on their "good" degrees of fit when both /a/ and /a:/ were used. The Arabic consonants /$t^ʕ$, $ð^ʕ$/ are considered "less similar" sounds to the English categories based on their "fair" degrees of fit. Finally, the Arabic consonant /$d^ʕ$/ is considered a "new" sound based on its "poor" degree of fit when /a/ and /a:/ were used. Also, the results demonstrated that there was no prosodic location effect for when CV and VCV were used.

An identification task (L2 labeling task) and a goodness-of-fit task were used in Chapter 4 with three different learner groups in order to answer the second research question: *Does more L2 exposure over time lead to similar learning development with Arabic L2 plain and emphatic sounds?* Based on the results, the answer was "No". It was shown that Arabic plain and Arabic emphatic sounds did not demonstrate a similar learning development with more L2 exposure. The overall development of the plain and emphatic sounds in the three different learner groups (NG, IG, and AG) was examined by studying the accuracy results from the L2 labeling task and the goodness ratings. The results of the plain sounds showed that the NG's performance was higher than the performance of the IG, and the performance of the IG was higher than the performance of the AG. I deduced that this was an indication to a no L2 learning development with more L2 exposure. On the other hand, the results of the emphatic sounds showed that the accuracy performance of the AG was higher than the accuracy performance of the IG, and the accuracy performance of the IG was higher than the accuracy performance of the NG. The results from the goodness-of-fit rating task also supported the results of the labeling task (accuracy performance). Goodness ratings of the plain sounds did not show any significant increase among the three learner groups while goodness ratings of the emphatic sounds demonstrated a significant and a gradual increase in the goodness performances between the three learner groups. Also, the results showed that the accuracies were higher for the long vowel condition /a:/ than for the short vowel /a/ condition, which leads us to believe that the perception of sounds is easier with long vowels than short vowels.

An identification task (L1 labeling task) and a goodness-of-fit task were used in order to answer the third research question: *Do L2 learners' L1-to-L2 mappings change over time?* Based on the results, the answer was "No." My results showed that L1-to-L2 mappings were not affected by more L2 exposure, based on the mapping pattern results; the obtained proportions did not demonstrate any differences between the three tested learner groups. However, the results from the overall goodness ratings were higher in the AG than the NG and IG, reflecting better performance over time. This increase in the goodness ratings lead me to believe that there was a subtle change of L1-to-L2 mappings over time, manifested by how learners determined the similarity levels between L1 and L2 sounds.

I have also attempted to analyze the results of the L1 labeling task and L2 labeling task in Chapter 4 by following Park and de Jong's (2008) quantitative technique in order to examine the degree to which L1 categories are used in L2 category identification. The results did not come out as Park and de Jong's model predicts. It was not very clear to what extent L2 learners are using and facilitating their L1 categories in order to perceive the tested L2 sounds. However, based on the observed accuracies (L2 labeling task), the results were successful at revealing how L2 exposure affects the overall learnability of L2 emphatic sounds; learners of the AG showed the highest performance rates and lowest error rates, followed by learners in the IG, and then the NG. Finally, I have concluded that only the observed accuracy results of L2 emphatic sounds demonstrated L2 performance development; plain sounds did not demonstrate any performance development.

In Chapter 5, I discussed the fourth research question: *How can the current perceptual models be evaluated based on the results?* I concluded that, for the SLM, the accuracy results (L2 labeling task) showed that the emphatic sounds demonstrated a gradual increase in their performance, starting with the NG, and that the AG had the highest performance rates among the three learner groups. That is an indication that the emphatic sounds met the “new” sounds criteria laid out by the SLM. In regard to plain sounds, however, the results showed no demonstrable improvement in their performance over time, with more L2 exposure. I concluded that these sounds must fall under the “similar” category, according to the SLM, because their performance did not change across time. I also explained that the difference of Arabic sound categorization between Chapters 3 and 4 lies in the level of experience with L2. In the cross-language category mapping task (Chapter 3), I was interested in naïve listeners who have no experience with Arabic. On the other hand, in the identification task (Chapter 4), I was interested in L2 learners with varying L2 exposure (NG, IG, and AG).

With regard to the PAM-L2, I indicated that the results in the L1 labeling tasks showed category assimilation type differences between the naïve English monolingual listeners (Chapter 3) and the AG L2 Arabic listeners (Chapter 4). I have concluded that this was indicative of preliminary perceptual development.

Lastly, another direction of future work could investigate the accuracy of the PAM-L2 predictions, in light of the findings in this study. In addition, it would be intriguing to see how the findings of the perception tasks in this study compare to future findings in production experiments which test the same Arabic sounds.

REFERENCES

Al-Ani, S. H. (1970). *Arabic phonology: An acoustical and physiological investigation.* Netherlands: Mouton and Co.

Al-Ani, S. and M. El-Dalee. (1984). Tafkhim in Arabic: the Acoustic and Psychological Parameters. In proceedings of the *Tenth International Congress of Phonetic Sciences,* 385-389. Dordrecht.

Ali, L. & Daniloff, R. (1972). A contrastive cinefluorographic investigation of the articulation of emphatic-non emphatic cognate consonants. *Studia Linguistica*, 26, 81-105.

Allatif, O. 2008. Contrôle des corrélats temporels et spectraux de la quantité vocalique: de l'arabe syrien de l'Euphrate au français de Savoie. Ph.D. dissertation, Préparée à l'ex-Institut de la Communication Parlée (GIPSA-LAB) dans le cadre de l'Ecole Doctorale: Langues, littératures et sciences humaines.

Al Mahmoud, M. (2013). Discrimination of Arabic contrasts by American learners. *Studies in Second Language Learning and Teaching,* 2, 261-292.

Al-Masri, M., & Jongman, A. (2004). Acoustic correlates of emphasis in Jordanian Arabic: Preliminary results. In *Proceedings of the 2003 Texas Linguistics Society Conference. Somerville, MA: Cascadilla Proceedings Project* (pp. 96-106).

Almbark, R. (2012). The perception and production of SSBE vowels by Syrian Arabic learners: The foreign language model. (Unpublished doctoral dissertation). University of York, York, UK.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 345-360.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of phonetics*, *20*(3), 305-330.

Best, C. T. (1995). A direct realist view on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 171- 204). Timonium, MD: York Press.

Best, C. T. (1999). Development of Language-Specific Influences on Speech Perception and Production in Pre-verbal Infancy. *International Congress of Phonetic Sciences,* 14, 1261–1263.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener’s native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775-794.

Best, C. T. & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementaries. In O. S. Bohn and M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Flege* (pp. 13-34). Philadelphia, PA: John Benjamins B. V.

Bin-Muqbil, M. S. (2006). *Phonetic and phonological aspects of Arabic emphatics and gutturals* (Doctoral dissertation, University of Wisconsin-Madison).

Bohn, O., Flege, J., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437-470.

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, *3*(01), 219-252.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*(3), 372-387.

Chang, C. B. (2010). *First Language Phonetic Drift During Second Language Acquisition* (Doctoral dissertation, University of California, Berkeley).

Eckman, F. (1987). Markedness and the contrastive analysis hypothesis. In I. Georgette & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 55-69). Cambridge, MA: Newbury House.

Flege, J. E. and J. Hillenbrand (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America 76*(3), 708‑721.

Flege, J. E. (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. E. (1992). Speech learning in a second language. In C. Ferguson, L. Menn, and C. Stoel-Gammon (Eds), *Phonological Development: Models, Research, and Application* (pp. 565-604). Timonium, MD: York Press.

Flege, J. E. (1995). Second-language Speech Learning: Theory, Findings, and Problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language research* (pp. 229-273). Timonium, MD: York Press.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics*, *25*(4), 437-470.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in second language acquisition*, *23*(04), 527-552.

Furman, N., Goldberg, D., & Lusin, N. (2010). Enrollments in Languages Other than English in United States Institutions of Higher Education, Fall 2009. In *Modern Language Association*. Modern Language Association. 26 Broadway 3rd Floor, New York, NY 10004-1789.

Gass, S. M., & Selinker, L. (2001). *Second language acquisition: An introductory course.* London: Lawrence Earlbaum.

Ghazeli, S. (1977). *Back consonants and backing coarticulation in Arabic*, Ph.D. Dissertation, University of Texas at Austin.

Giannini, A. & Pettorino, M. (1982). *The emphatic consonants in Arabic*: Speech laboratory report IV. Naples: Istituto Universitario Orientale.

Guion, S., Flege, J., Akahane-Yamada, R., & Pruitt, J. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, 107, 2711–2725.

Khattab, G., F. Al-Tamimi and B. Heselwood. (2006). Acoustic and Auditory differences in the /t/- /T/ Opposition in Male and Female Speakers of Jordanian Arabic. In *Perspectives on Arabic Linguistics XVI: Papers from the sixteenth annual symposium on Arabic linguistics,* ed. Boudelaa, S., 131-160. Cambridge, UK: John Benjamins.

Khattab, G., & Al-Tamimi, J. (2008). Durational cues for gemination in Lebanese Arabic. *Language and Linguistics*, *22*, 39-55.

Kotby, M. N., Saleh, M., Hegazi, M., Gamal, N., Abdel, S. M., Nabil, A., & Fahmi, S. (2011). The Arabic vowels: features and possible clinical application in communication disorders. *Folia phoniatrica et logopaedica: official organ of the International Association of Logopedics and Phoniatrics (IALP)*, *63*(4), 171.

Lababidi, Z., & Park, H. (2016). L1-English tense-lax vowel system influence on L2-Arabic short and long vowel learning. In *Perspectives on Arabic Linguistics XXVIII: Papers from the Annual Symposium on Arabic Linguistics, Gainesville, Florida, 2014* (Vol. 4, p. 63). John Benjamins Publishing Company.

Ladefoged, P., & Johnson, K. (2011). *A course in phonetics*. Boston, MA: Wadsforth.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor, MI: University of Michigan Press.

Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. *Speech perception and linguistic experience: Issues in cross-language research*, 351-377.

MacKain, K., Best, C., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Linguistics, 2*, 369-390.

Most, T., Levin, I., & Sarsour, M. (2008). The effect of modern standard Arabic orthography on speech production by Arab children with hearing loss. *Journal of deaf studies and deaf education*, *13*(3), 417-431.

Newman, D. (2002). The phonetic status of Arabic within the world's languages: the uniqueness of the lughat al-daad. *Antwerp papers in linguistics. 100*, 65-75.

Newman, D., & Verhoeven, J. (2002). Frequency analysis of Arabic vowels in connected speech. *Antwerp papers in linguistics. 100*, 77-86.

Park, H. & de Jong, K. (2008). Perceptual category mapping between English and Korean prevocalic obstruents: Evidence from mapping effects in second language identification skills. *Journal of Phonetics*, 36, 704-723.

Park, H. & de Jong, K. (submitted). Perceptual category mapping between English and Korean obstruents in non-CV positions: Prosodic location effects in second language identification skills. *Journal of Phonetics*, (under review).

Pilus, Z. (2005). Perception of voicing in English word-final consonants: examining the perceptual assimilation model. *Malaysian Journal of ELT Research*, 1-12.

Polivanov, E. (1931). La perception des sons d’une langue´etrang`ere (The perception of non-native language sounds.). *Travaux du Cercle Linguistique de Prague, 4*, 79-96.

Ryding, K. C. (2005). *A reference grammar of modern standard Arabic*. Cambridge university press.

Saadah, E. (2011). The production of Arabic vowels by English L2 learners and heritage speakers of Arabic. Unpublished Ph.D. Dissertation, University of Illinois at Urbana-Champaign.

Towell, R., & Hawkins, R. (1994). *Approaches to second language acquisition*. Clevedon, UK: Multilingual Matters.

Trubetzkoy, N. (1939). *Principles of phonology* (C. A. Baltaxe, Trans.). Berkeley, CA: University of California Press.

Watson, J. C. (2002). *The phonology and morphology of Arabic*. Oxford University Press on Demand.

APPENDICES

Appendix A: Identifications and Goodness Ratings for Arabic Stimuli.

Table A-1: Mean percent identifications and goodness ratings (in parentheses) of Arabic consonant stimuli in terms of English Categories in CV utterances when V was /u/.

Arabic Consonants	**English Consonants**														
	/b/	**/d/**	**/f/**	**/g/**	**/h/**	**/ʒ/**	**/k/**	**/r/**	**/s/**	**/t/**	**/z/**	**/θ/**	**/ð/**	**/ʔ/**	**Other**
/t/		12.5 (7)								**87.5 (5.8)**					
/d/		**100 (5.5)**													
/ð/												25 (2.5)	**75 (5)**		
/s/									**87.5 (6)**		12.5 (4)				
/tˤ/		25 (4.5)								**50 (5)**		12.5 (3)			12.5 (4)
/dˤ/	6.25 (7)	**81.25 (5)**								12.5 (2)					
/ðˤ/												**37.5 (3.6)**	**37.5 (4)**		25 (3)
/sˤ/									**100 (5.25)**						
/q/				**37.5 (4)**		12.5 (1)	25 (4)								25 (3.5)
/x/					**75 (3.8)**										25 (4.5)
/ɣ/			12.5 (2)					25 (3.5)							**62.5 (4.6)**
/ħ/					**62.5 (4)**				12.5 (4)						25 (6)
/ʕ/								**75 (3.6)**							25 (6.5)
/ʔ/					12.5 (2)										**87.5 (4.7)**

Table A-2: Mean percent identifications and goodness ratings (in parentheses) of Arabic consonant stimuli in terms of English Categories in CV utterances when V was /i/.

Arabic Consonants	**English Consonants**															
	/b/	**/d/**	**/f/**	**/g/**	**/h/**	**/k/**	**/l/**	**/p/**	**/r/**	**/s/**	**/t/**	**/z/**	**/θ/**	**/ð/**	**/ʔ/**	**Other**
/t/											**100 (6.1)**					
/d/		**100 (5.5)**														
/ð/	12.5 (4)												25 (4.5)	**62.5 (5.4)**		
/s/										**100 (5.8)**						
/tˤ/								6.25 (5)			18.75 (4)		**50 (3)**			25 (3.5)
/dˤ/	**43.75 (4.3)**	12.5 (3)											6.25 (2)	12.5 (2)		25 (5)
/ðˤ/			12.5 (4)										37.5 (4)	**50 (5.2)**		
/sˤ/										**100 (4.6)**						
/q/		6.25 (2)					6.25 (2)		**50 (3)**					12.5 (2)		25 (4)
/x/			25 (4)	12.5 (2)	**50 (3.2)**								12.5 (2)			
/ɣ/									**43.7 (3.3)**			6.25 (1)		25 (2)		25 (5)
/ħ/					**75 (5)**	12.5 (1)										12.5 (4)
/ʕ/							12.5 (4)		37.5 (4)							**50 (5.5)**
/ʔ/					12.5 (2)											**87.5 (4.5)**

Table A-3: Mean percent identifications and goodness ratings (in parentheses) of Arabic consonant stimuli in terms of English Categories in CV utterances when V was /a:/.

Arabic Consonants	English Consonants												
	/b/	**/d/**	**/f/**	**/g/**	**/h/**	**/r/**	**/s/**	**/t/**	**/z/**	**/θ/**	**/ð/**	**/ʔ/**	**Other**
/t/		**50 (4.2)**						**50 (4.6)**					
/d/		**87.5 (5.4)**										12.5 (5)	
/ð/										37.5 (4.3)	**62.5 (5.2)**		
/s/							**75 (5.8)**		25 (5)				
/tˤ/	**43.75 (4)**	31.25 (4.5)	12.5 (4)					12.5 (4)					
/dˤ/	**25 (6)**	12.5 (7)								12.5 (3)	**25 (5.5)**		25 (5.5)
/ðˤ/	12.5 (1)				12.5 (1)						12.5 (1)		**62.5 (4.6)**
/sˤ/							**87.5 (5.8)**		12.5 (5)				
/q/				25 (3.5)								12.5 (3)	**62.5 (4)**
/x/				6.25 (2)	**43.75 (4)**								**50 (5.2)**
/ɣ/				12.5 (4)		**43.75 (3.6)**							**43.75 (5)**
/ħ/					**100 (4.5)**								
/ʕ/				12.5 (3)	12.5 (2)								**75 (4.1)**
/ʔ/					25 (2.5)							12.5 (6)	**62.5 (4.6)**

Table A-4: Mean percent identifications and goodness ratings (in parentheses) of Arabic consonant stimuli in terms of English Categories in CV utterances when V was /u:/.

Arabic Consonants	**English Consonants**											
	/d/	**/f/**	**/g/**	**/h/**	**/r/**	**/s/**	**/t/**	**/z/**	**/θ/**	**/ð/**	**/ʔ/**	**Other**
/t/							**100 (5.6)**					
/d/	**100 (6)**											
/ð/									**50 (5.2)**	37.5 (5.6)		12.5 (6)
/s/						**100 (5.7)**						
/tˤ/	12.5 (7)						**87.5 (5.2)**					
/dˤ/	**50 (5)**						25 (4)					25 (3.5)
/ðˤ/		25 (4)							**37.5 (5.3)**	25 (5)		12.5 (6)
/sˤ/						**87.5 (5.8)**		12.5 (5)				
/q/			25 (3.5)								12.5 (3)	**62.5 (4)**
/x/			12.5 (2)	37.5 (3.3)								**50 (6.5)**
/ɣ/			12.5 (3)		25 (3)							**62.5 (4.4)**
/ħ/				**100 (4.8)**								
/ʕ/			12.5 (4)		25 (3.5)							**62.5 (4.2)**
/ʔ/				25 (2.5)							12.5 (3)	**62.5 (4.2)**

Table A-5: Mean percent identifications and goodness ratings (in parentheses) of Arabic consonant stimuli in terms of English Categories in CV utterances when V was /i:/.

Arabic Consonants	**English Consonants**										
	/b/	**/d/**	**/f/**	**/g/**	**/h/**	**/r/**	**/s/**	**/t/**	**/θ/**	**/ð/**	**Other**
/t/								**100 (6)**			
/d/		**87.5 (5.7)**								12.5 (5)	
/ð/	12.5 (4)									**50 (4)**	37.5 (5.6)
/s/							**100 (5.8)**				
/tˤ/								25 (4.5)	**37.5 (4.6)**	**37.5 (4.6)**	
/dˤ/		12.5 (5)				12.5 (5)			**37.5 (4.3)**	25 (3.5)	12.5 (4)
/ðˤ/	12.5 (3)		12.5 (4)						**25 (5.5)**	**25 (5)**	25 (4)
/sˤ/							**100 (3.7)**				
/q/	25 (5.5)			**50 (3.7)**		12.5 (7)					12.5 (4)
/x/					**56.25 (5)**				6.25 (3)		37.5 (5.6)
/ɣ/				25 (2)		**50 (4)**				12.5 (6)	12.5 (3)
/ħ/					**100 (4.8)**						
/ʕ/				12.5 (1)		25 (4)					**62.5 (4.6)**
/ʔ/					12.5 (2)						**87.5 (4.5)**

Table A-6: Mean percent identifications and goodness ratings (in parentheses) of Arabic consonant stimuli in terms of English Categories in VCV utterances when V was /a/.

Arabic Consonants	**English Consonants**													
	/b/	**/d/**	**/g/**	**/h/**	**/ʒ/**	**/k/**	**/r/**	**/s/**	**/t/**	**/z/**	**/θ/**	**/ð/**	**/ʔ/**	**Other**
/t/									**100 (5.7)**					
/d/		**100 (6.3)**												
/ð/	6.25 (4)									25 (6.5)	12.5 (6)	**43.75 (6)**		12.5 (5)
/s/								**100 (6.7)**						
/tˤ/		31.25 (6.5)						6.25 (1)	**50 (5.7)**					12.5 (7)
/dˤ/		**75 (4.8)**							25 (6.5)					
/ðˤ/	6.25 (4)									25 (6)		**56.25 (4.2)**		12.5 (7)
/sˤ/								**100 (6.2)**						
/q/						**100 (5.6)**								
/x/				**62.5 (5)**			12.5 (6)							25 (6.5)
/ɣ/			25 (4)				**31.25 (5.5)**							**43.75 (5.2)**
/ħ/				**100 (5.7)**										
/ʕ/				12.5 (2)	6.25 (4)		12.5 (7)							**68.75 (4.2)**
/ʔ/				12.5 (3)									25 (5)	**62.5 (4)**

Table A-7: Mean percent identifications and goodness ratings (in parentheses) of Arabic consonant stimuli in terms of English Categories in VCV utterances when V was /u/.

Arabic Consonants	English Consonants												
	/d/	**/g/**	**/h/**	**/k/**	**/l/**	**/r/**	**/s/**	**/t/**	**/z/**	**/θ/**	**/ð/**	**/ʔ/**	**Other**
/t/								**100 (6.1)**					
/d/	**87.5 (6)**							12.5 (5)					
/ð/										37.5 (6)	**62.5 (4.6)**		
/s/							**87.5 (5.7)**		12.5 (7)				
/tˤ/	12.5 (7)							**87.5 (6.4)**					
/dˤ/	**75 (6.3)**							25 (6.5)					
/ðˤ/										**62.5 (6.4)**	37.5 (5.6)		
/sˤ/							**100 (6.2)**						
/q/		25 (5)		**62.5 (5.8)**									12.5 (7)
/x/		12.5 (6)	**62.5 (3.8)**										25 (5)
/ɣ/		12.5 (6)	**31.25 (3)**	6.25 (5)		**31.25 (4)**							18.75 (7)
/ħ/			**50 (5)**		12.5 (2)					12.5 (1)			25 (6.5)
/ʕ/						12.5 (3)						12.5 (7)	**75 (5.1)**
/ʔ/				12.5 (6)								25 (4)	**62.5 (4)**

Table A-8: Mean percent identifications and goodness ratings (in parentheses) of Arabic consonant stimuli in terms of English Categories in VCV utterances when V was /i/.

Arabic Consonants	English Consonants													
	/d/	**/f/**	**/g/**	**/h/**	**/k/**	**/l/**	**/r/**	**/s/**	**/t/**	**/z/**	**/θ/**	**/ð/**	**/ʔ/**	**Other**
/t/	25 (6.5)								**75 (6.1)**					
/d/	**87.5 (6.4)**								12.5 (7)					
/ð/										12.5 (4)	12.5 (6)	**62.5 (6)**		12.5 (7)
/s/								**100 (6)**						
/tˤ/	25 (7)								**75 (3.6)**					
/dˤ/						6.25 (3)	25 (2.5)		12.5 (7)		6.25 (5)	**37.5 (5.5)**		12.5 (6)
/ðˤ/	6.25 (6)	6.25 (5)					12.5 (3)				12.5 (5)	**62.5 (6)**		
/sˤ/								**87.5 (5.8)**		12.5 (7)				
/q/			25 (6)		**37.5 (4)**									37.5 (5.6)
/x/			12.5 (1)	**37.5 (4)**	25 (6)									25 (7)
/ɣ/							**87.5 (5.1)**					12.5 (1)		
/ħ/				**87.5 (4.8)**										12.5 (7)
/ʕ/							25 (2)						25 (4)	**50 (5)**
/ʔ/				12.5 (6)		12.5 (5)	12.5 (2)						12.5 (6)	**50 (4.2)**

Table A-9: Mean percent identifications and goodness ratings (in parentheses) of Arabic consonant stimuli in terms of English Categories in VCV utterances when V was /a:/.

Arabic Consonants	**English Consonants**												
	/b/	**/d/**	**/g/**	**/h/**	**/k/**	**/r/**	**/s/**	**/t/**	**/z/**	**/θ/**	**/ð/**	**/ʔ/**	**Other**
/t/								**100 (6.1)**					
/d/		**100 (6)**											
/ð/		**37.5 (2.3)**							12.5 (7)	25 (6.5)	12.5 (5)		12.5 (6)
/s/							**87.5 (6.5)**		12.5 (7)				
/tˤ/		25 (6)			12.5 (7)			**62.5 (3.6)**					
/dˤ/		**87.5 (5.2)**									12.5 (6)		
/ðˤ/	6.25 (5)									12.5 (6)	**43.75 (4.3)**		37.5 (4.6)
/sˤ/						6.25 (6)	**75 (5.8)**		18.75 (7)				
/q/			25 (5.5)		**62.5 (5)**								12.5 (7)
/x/			12.5 (1)	**56.25 (4.5)**		6.25 (6)							25 (5.5)
/ɣ/			12.5 (6)			**43.75 (4)**					6.25 (6)		37.5 (2.6)
/ħ/				**100 (5.5)**									
/ʕ/				12.5 (5)		25 (6)							**62.5 (4.4)**
/ʔ/			12.5 (6)	12.5 (3)								25 (4)	**50 (5)**

Table A-10: Mean percent identifications and goodness ratings (in parentheses) of Arabic consonant stimuli in terms of English Categories in VCV utterances when V was /u:/.

Arabic Consonants	**English Consonants**										
	/d/	**/g/**	**/h/**	**/k/**	**/r/**	**/s/**	**/t/**	**/θ/**	**/ð/**	**/ʔ/**	**Other**
/t/							**100 (6.6)**				
/d/	**100 (6.5)**										
/ð/	**50 (6.5)**							12.5 (5)	37.5 (5.6)		
/s/						**75 (6.6)**		25 (5.5)			
/tˤ/	25 (6)						**75 (6)**				
/dˤ/	**87.5 (5.8)**						12.5 (6)				
/ðˤ/						12.5 (5)		37.5 (6.3)	**50 (5.5)**		
/sˤ/						**100 (5.8)**					
/q/		12.5 (7)		**75 (4.8)**							12.5 (4)
/x/		12.5 (1)	**62.5 (4.4)**	12.5 (6)							12.5 (6)
/ɣ/		**31.25 (4)**	12.5 (5.5)		**31.25 (5.5)**						25 (7)
/ħ/			**100 (4.7)**								
/ʕ/			12. 5 (5)		**50 (3.5)**						37.5 (5.6)
/ʔ/			12.5 (6)							37.5 (5)	**50 (4.5)**

Table A-11: Mean percent identifications and goodness ratings (in parentheses) of Arabic consonant stimuli in terms of English Categories in VCV utterances when V was /i:/.

Arabic Consonants	**English Consonants**												
	/d/	**/g/**	**/h/**	**/k/**	**/l/**	**/r/**	**/s/**	**/t/**	**/z/**	**/θ/**	**/ð/**	**/ʔ/**	**Other**
/t/								**100 (6.1)**					
/d/	**100 (6.2)**												
/ð/	12.5 (5)		6.25 (6)							**43.75 (6.3)**	31.25 (5.5)		6.25 (6)
/s/							**100 (6)**						
/tˤ/	**31.25 (6.5)**							**50 (4)**		12.5 (3)			6.25 (3)
/dˤ/	12.5 (4)				6.25 (3)	**37.5 (4.6)**		18.75 (7)		12.5 (5)	12.5 (5)		
/ðˤ/	**31.25 (5)**					18.75 (4.5)				12.5 (6)	**37.5 (6)**		
/sˤ/							**87.5 (6.2)**		12.5 (7)				
/q/				**43.75 (4.6)**		18.75 (4)				12.5 (4)			25 (7)
/x/			**50 (5.7)**							25 (2)			25 (5.5)
/ɣ/		12.5 (6)				**75 (5.5)**					12.5 (4)		
/ħ/			**87.5 (5.1)**										12.5 (7)
/ʕ/			12.5 (5)									12.5 (7)	**75 (4.5)**
/ʔ/			12.5 (4)									25 (5.5)	**62.5 (4)**

Appendix B: Fit Index Tables for Arabic Stimuli.

Table B-1: Fit indexes for Arabic consonants in terms of English categories in CV syllable structure when V was /u/.

Arabic Consonants	Most Common Identification	Proportion of Identifications	Goodness Rating	Fit Index
/t/	/t/	0.87	5.8	5 Good
/d/	/d/	1	5.5	5.5 Good
/ð/	/ð/	0.75	5	3.75 Good
/s/	/s/	0.87	6	5.22 Good
/tˤ/	/t/	0.5	5	2.5 Fair
/dˤ/	/d/	0.81	5	4 Good
/ðˤ/	/ð/	0.37	4	1.48 Poor
	/θ/	0.37	3.6	1.33 poor
/sˤ/	/s/	1	5.25	5.25 Good
/q/	/g/	0.37	4	1.48 Poor
/x/	/h/	0.75	3.8	2.85 Fair
/ɣ/	Other	0.62	4.6	2.85 Fair
/ħ/	/h/	0.62	4	2.48 Fair
/ʕ/	/r/	0.75	3.6	2.7 Fair
/ʔ/	Other	0.87	4.7	4 Good

Table B-2: Fit indexes for Arabic consonants in terms of English categories in CV syllable structure when V was /i/.

Arabic Consonants	Most Common Identification	Proportion of Identifications	Goodness Rating	Fit Index
/t/	/t/	1	6.1	6.1 Good
/d/	/d/	1	5.5	5.5 Good
/ð/	/ð/	0.62	5.4	3.34 Good
/s/	/s/	1	5.8	5.8 Good
/tˤ/	/θ/	0.5	3	1.5 Poor
/dˤ/	/b/	0.43	4.3	1.84 Fair
/ðˤ/	/ð/	0.5	5.2	2.6 Fair
	/θ/	0.37	4	1.48 poor
/sˤ/	/s/	1	4.6	4.6 Good
/q/	/r/	0.5	3	1.5 poor
/x/	/h/	0.5	3.2	1.6 Fair
/ɣ/	/r/	0.43	3.3	1.4 Poor
/ħ/	/h/	0.75	5	3.75 Good
/ʕ/	Other	0.5	5.5	2.75 Fair
	/r/	0.37	4	1.48 poor
/ʔ/	Other	0.87	4.5	3.9 Good

Table B-3: Fit indexes for Arabic consonants in terms of English categories in CV syllable structure when V was /a:/.

Arabic Consonants	Most Common Identification	Proportion of Identifications	Goodness Rating	Fit Index
/t/	/t/ /d/	0.5 0.5	4.6 4.2	2.3 Fair 2.1 Fair
/d/	/d/	0.87	5.4	4.69 Good
/ð/	/ð/ /θ/	0.62 0.37	5.2 4.3	3.22 Good 1.59 Fair
/s/	/s/	0.75	5.8	4.35 Good
/tˤ/	/b/	0.43	4	1.72 Fair
/dˤ/	/b/ /ð/	0.25 0.25	6 5.5	1.5 Poor 1.37 Poor
/ðˤ/	Other	0.62	4.6	2.85 Fair
/sˤ/	/s/	0.87	5.8	5 Good
/q/	Other	0.62	4	2.48 Fair
/x/	/h/	0.43	4	1.72 Fair
/ɣ/	/r/	0.43	3.6	1.54 Fair
/ħ/	/h/	1	4.5	4.5 Good
/ʕ/	Other	0.75	4.1	3 Good
/ʔ/	Other	0.62	4.6	2.85 Fair

Table B-4: Fit indexes for Arabic consonants in terms of English categories in CV syllable structure when V was /u:/.

Arabic Consonants	Most Common Identification	Proportion of Identifications	Goodness Rating	Fit Index
/t/	/t/	1	5.6	5.6 Good
/d/	/d/	1	6	6 Good
/ð/	/θ/	0.5	5.2	2.6 Fair
	/ð/	0.37	5.6	2 Fair
/s/	/s/	1	5.7	5.7 Good
/tˤ/	/t/	0.87	5.2	4.52 Good
/dˤ/	/d/	0.5	5	1.5 Poor
/ðˤ/	/θ/	0.37	5.3	1.96 Fair
/sˤ/	/s/	0.87	5.8	5 Good
/q/	Other	0.62	4	2.48 Fair
/x/	/h/	0.37	3.3	1.22 Poor
/ɣ/	Other	0.62	4.4	2.72 Fair
/ħ/	/h/	1	4.8	4.8 Good
/ʕ/	Other	0.62	4.2	2.6 Fair
/ʔ/	Other	0.62	4.2	2.6 Fair

Table B-5: Fit indexes for Arabic consonants in terms of English categories in CV syllable structure when V was /i:/.

Arabic Consonants	Most Common Identification	Proportion of Identifications	Goodness Rating	Fit Index
/t/	/t/	1	6	6 Good
/d/	/d/	0.87	5.7	4.95 Good
/ð/	/ð/	0.5	4	2 Fair
/s/	/s/	1	5.8	5.8 Good
/tˤ/	/ð/	0.37	4.6	1.7 Fair
	/θ/	0.37	4.6	1.7 Fair
/dˤ/	/θ/	0.37	4.3	1.59 Fair
/ðˤ/	/ð/	0.25	5	1.25 Poor
	/θ/	0.25	5.5	1.37 Poor
/sˤ/	/s/	1	3.7	3.7 Good
/q/	/g/	0.5	3.7	1.85 Fair
/x/	/h/	0.56	5	2.8 Fair
/ɣ/	/r/	0.5	4	2 Fair
/ħ/	/h/	1	4.8	4.8 Good
/ʕ/	Other	0.62	4.6	2.85 Fair
/ʔ/	Other	0.87	4.5	3.9 Good

Table B-6: Fit indexes for Arabic consonants in terms of English categories in VCV syllable structure when V was /a/.

Arabic Consonants	Most Common Identification	Proportion of Identifications	Goodness Rating	Fit Index
/t/	/d/	0.75	5.4	4 Good
/d/	/d/	0.87	4.8	4.17 Good
/ð/	/ð/	0.87	5.5	4.78 Good
/s/	/s/	1	5.6	5.6 Good
/tˤ/	/f/	0.5	5	2.5 Fair
	/t/	0.37	3.3	1.22 Poor
/dˤ/	/d/	0.68	4.8	3.26 Good
/ðˤ/	Other	0.5	4.5	2.25 Fair
/sˤ/	/s/	1	6	6 Good
/q/	/ʔ/	0.37	5.3	1.96 Fair
/x/	/h/	0.62	3	1.86 Fair
/ɣ/	Other	0.5	4.7	2.35 Fair
	/r/	0.37	3.6	1.33 Poor
/ħ/	/h/	1	5.7	5.7 Good
/ʕ/	Other	0.62	5.4	3.34 Good
/ʔ/	Other	0.75	3.5	2.62 Fair

Table B-7: Fit indexes for Arabic consonants in terms of English categories in VCV syllable structure when V was /u/.

Arabic Consonants	Most Common Identification	Proportion of Identifications	Goodness Rating	Fit Index
/t/	/t/	0.87	5.8	5 Good
/d/	/d/	1	5.5	5.5 Good
/ð/	/ð/	0.75	5	3.75 Good
/s/	/s/	0.87	6	5.22 Good
/tˤ/	/t/	0.5	5	2.5 Fair
/dˤ/	/d/	0.81	5	4 Good
/ðˤ/	/ð/	0.37	4	1.48 Poor
	/θ/	0.37	3.6	1.33 Poor
/sˤ/	/s/	1	5.25	5.25 Good
/q/	/g/	0.37	4	1.48 Poor
/x/	/h/	0.75	3.8	2.85 Fair
/ɣ/	Other	0.62	4.6	2.85 Fair
/ħ/	/h/	0.62	4	2.48 Fair
/ʕ/	/r/	0.75	3.6	2.7 Fair
/ʔ/	Other	0.87	4.7	4 Good

Table B-8: Fit indexes for Arabic consonants in terms of English categories in VCV syllable structure when V was /i/.

Arabic Consonants	Most Common Identification	Proportion of Identifications	Goodness Rating	Fit Index
/t/	/t/	1	6.1	6.1 Good
/d/	/d/	1	5.5	5.5 Good
/ð/	/ð/	0.62	5.4	3.34 Good
/s/	/s/	1	5.8	5.8 Good
/tˤ/	/θ/	0.5	3	1.5 Poor
/dˤ/	/b/	0.43	4.3	1.84 Fair
/ðˤ/	/ð/	0.5	5.2	2.6 Fair
	/θ/	0.37	4	1.48 Poor
/sˤ/	/s/	1	4.6	4.6 Good
/q/	/r/	0.5	3	1.5 Poor
/x/	/h/	0.5	3.2	1.6 Fair
/ɣ/	/r/	0.43	3.3	1.41 Poor
/ħ/	/h/	0.75	5	3.75 Good
/ʕ/	Other	0.5	5.5	2.75 Fair
/ʔ/	Other	0.87	4.5	3.9 Good

Table B-9: Fit indexes for Arabic consonants in terms of English categories in VCV syllable structure when V was /a:/.

Arabic Consonants	Most Common Identification	Proportion of Identifications	Goodness Rating	Fit Index
/t/	/t/ /d/	0.5 0.5	4.6 4.2	2.3 Fair 2.1 Fair
/d/	/d/	0.87	5.4	4.69 Good
/ð/	/ð/	0.62	5.2	3.22 Good
/s/	/s/	0.75	5.8	4.35 Good
/tˤ/	/b/	0.43	4	1.72 Fair
/dˤ/	/b/ /ð/	0.25 0.25	6 5.5	1.5 Poor 1.37 Poor
/ðˤ/	Other	0.62	4.6	2.85 Fair
/sˤ/	/s/	0.87	5.8	5 Good
/q/	Other	0.62	4	2.48 Fair
/x/	/h/	0.43	4	1.72 Fair
/ɣ/	/r/	0.43	3.6	1.54 Fair
/ħ/	/h/	1	4.5	4.5 Good
/ʕ/	Other	0.75	4.1	3 Good
/ʔ/	Other	0.62	4.6	2.85 Fair

Table B-10: Fit indexes for Arabic consonants in terms of English categories in VCV syllable structure when V was /u:/.

Arabic Consonants	**Most Common Identification**	**Proportion of Identifications**	**Goodness Rating**	**Fit Index**
/t/	/t/	1	5.6	5.6 Good
/d/	/d/	1	6	6 Good
/ð/	/θ/ /ð/	0.5 0.37	5.2 5.6	2.6 Fair 2 Fair
/s/	/s/	1	5.7	5.7 Good
/tˤ/	/t/	0.87	5.2	4.52 Good
/dˤ/	/d/	0.5	5	2.5 Fair
/ðˤ/	/θ/	0.37	5.3	1.96 Fair
/sˤ/	/s/	0.87	5.8	5 Good
/q/	Other	0.62	4	2.48 Fair
/x/	/h/	0.37	3.3	1.22 Poor
/ɣ/	Other	0.62	4.4	2.72 Fair
/ħ/	/h/	1	4.8	4.8 Good
/ʕ/	Other	0.62	4.2	2.6 Fair
/ʔ/	Other	0.62	4.2	2.6 Fair

Table B-11: Fit indexes for Arabic consonants in terms of English categories in VCV syllable structure when V was /i:/.

Arabic Consonants	**Most Common Identification**	**Proportion of Identifications**	**Goodness Rating**	**Fit Index**
/t/	/t/	1	6	6 Good
/d/	/d/	0.87	5.7	4.95 Good
/ð/	/ð/	0.5	4	2 Fair
/s/	/s/	1	5.8	5.8 Good
/tˤ/	/ð/	0.37	4.6	1.7 Fair
	/θ/	0.37	4.6	1.7 Fair
/dˤ/	/θ/	0.37	4.3	1.59 Fair
/ðˤ/	/ð/	0.25	5	1.25 Poor
	/θ/	0.25	5.5	1.37 Poor
/sˤ/	/s/	1	3.7	3.7 Good
/q/	/g/	0.5	3.7	1.85 Fair
/x/	/h/	0.56	5	2.8 Fair
/ɣ/	/r/	0.5	4	2 Fair
/ħ/	/h/	1	4.8	4.8 Good
/ʕ/	Other	0.62	4.6	2.85 Fair
/ʔ/	Other	0.87	4.5	3.91 Good

Appendix C: Identifications and Goodness Ratings of English Stimuli.

Table C-1: Mean percent identifications and goodness ratings (in parentheses) of English consonant stimuli in terms of English Categories in CV.

	English Categories																		
	/b/	**/d/**	**/f/**	**/g/**	**/h/**	**/ʒ/**	**/k/**	**/l/**	**/m/**	**/n/**	**/p/**	**/r/**	**/s/**	**/t/**	**/z/**	**/θ/**	**/ð/**	**/ʔ/**	**Other**
/t/														**100 (5.2)**					
/d/		**100 (5.2)**																	
/ð/																44.5 (5.1)	**55.5 (4.3)**		
/s/													**94.45 (5.2)**		5.5 (5)				
/k/				5.5 (2)			**94.45 (4.9)**												
/h/					**83.33 (5)**														16.67 (4)
/ʕ/					27.8 (2.4)	5.5 (1)												**61.2 (4.3)**	5.5 (4)

Table C-2: Mean percent identifications and goodness ratings (in parentheses) of English consonant stimuli in terms of English Categories in VCV.

English Stimuli	English Categories																		
	/b/	**/d/**	**/f/**	**/g/**	**/h/**	**/ʒ/**	**/k/**	**/l/**	**/m/**	**/n/**	**/p/**	**/r/**	**/s/**	**/t/**	**/z/**	**/θ/**	**/ð/**	**/ʔ/**	**Other**
/t/													5.5 (1)	**89 (6.5)**	5.5 (2)				
/d/		**100 (6.2)**																	
/ð/		5.5 (6)	5.5 (7)												5.6 (6)	27.8 (5.8)	**55.6 (5.4)**		
/s/													**94.5 (6.5)**		5.5 (6)				
/k/				5.5 (5)			**94.5 (6.5)**												
/h/					**83.4 (6.3)**							5.5 (7)				5.6 (6)			5.5 (5)
/ʕ/		5.5 (6)			5.5 (7)						5.5 (7)	22.23 (6.5)						**33.47 (3.4)**	27.8 (4.8)

Table C-3: Mean percent identifications and goodness ratings (in parentheses) of English consonant stimuli in terms of English Categories in CV and VCV.

English Stimuli	**English Categories**																		
	/b/	**/d/**	**/f/**	**/g/**	**/h/**	**/ʒ/**	**/k/**	**/l/**	**/m/**	**/n/**	**/p/**	**/r/**	**/s/**	**/t/**	**/z/**	**/θ/**	**/ð/**	**/ʔ/**	**Other**
/t/													5.5 (1)	**94.5 (5.9)**	5.5 (2)				
/d/	**100 (5.7)**																		
/ð/		2.7 (3)	2.7 (3.5)												2.8 (3)	36.2 (5.5)	**55.6 (4.9)**		
/s/													**94.5 (5.9)**		5.5 (5.5)				
/k/				5.5 (3.5)			**94.5 (5.7)**												
/h/					**83.4 (5.8)**							2.7 (3.5)				2.8 (3)			11.1 (4.5)
/ʕ/		2.7 (3)			16.8 (4.7)	2.7 (0.5)					2.8 (3.5)	11.1 (3.3)						**48 (3.9)**	16 (4.4)

CURRICULUM VITAE

Zafer Lababidi
zafer.lababidi@fulbrightmail.org

RESEARCH INTERESTS

Arabic Phonetics, Arabic as a Second Language, Arabic as a Foreign Language, Arabic sociophonetics, Speech perception and production, dialectal variation in Arabic, second language acquisition, and technology applications for Arabic.

EDUCATION

- **University of Wisconsin-Milwaukee**

  Ph.D. in Linguistics (2016).

  Dissertation: *The L2 Perceptual Mapping of Arabic and English Consonants by American English Learners.*

  Committee Chair: Professor Hanyong Park.

- **University of Wisconsin-Milwaukee**

  Certificate in the teaching of English as a second language (TESOL), (2012).

- **University of Wisconsin-Milwaukee**

  M.A in Foreign Languages, Literature, and Translation. Linguistics concentration (2011).

- **Homs University**

  Bachelor of Arts in English Literature and Translation (2007).

PROFESSIONAL EXPERIENCES

Sept. 2009-Present **University of Wisconsin-Milwaukee. Wisconsin, USA.**

Arabic Instructor

Teach Arabic as a foreign language, Arabic 101, 102, 201, and 202.

Design the course description, syllabus, and the supporting material.

Organize the cultural events throughout the academic year.

Aug. 2009-Present **University of Wisconsin-Milwaukee. Wisconsin, USA.**

Language Resource Center.

Graduate Research Assistant

Implemented the use of technology in language classrooms.

Utilized 2.0 web tools for language educators to support in-class language and culture teaching. Designed activities and supporting materials for the Arabic language learners.

Aug. 2008- June 2009 **Radford University. Virginia, USA.**

Fulbright Arabic Instructor

Started the Arabic language program on campus with the support of a Fulbright scholarship and taught cultures of the Middle East.

July 2007- July 2008 **Kempinski Hotels and Resorts. Dead Sea, Jordan.**

Human Resources and Training Departments.

Language Teacher and Trainer

Taught English as a foreign language and organized training sessions for the hotel's staff. Responsibilities also included assisting HR officers with office work.

May 2006- June 2007 **Al-Noor Language Center. Homs, Syria.**

Language Instructor & Interpreter

Taught English as a foreign language at the intermediate and advanced levels.

Worked as an interpreter for visiting delegations from European universities.

PUBLICATIONS

• Lababidi, Z. & Park, H. (in review). The Perceptual Mapping between Arabic and English Consonants.

• Lababidi, Z. & Park, H. (2016). L1-English tense-lax Vowel System Influence on L2-Arabic Short and Long Vowel Learning. *Perspectives on Arabic Linguistics 28,* Edited by Youssef Haddad & Eric Potsdam. John Benjamins Publishing Company.

CONFERENCE PRESENTATIONS

• Lababidi, Z. & Park, H. (2015). Perceptual Mapping between Arabic and English Consonants. 29th Annual Symposium on Arabic Linguistics (ASAL29) University of Wisconsin-Milwaukee. April.

• Lababidi, Z & Park, H. (2014). Cross-language Perceptual Mapping between Arabic and English Consonants. 19th Mid-Continental Phonetics & Phonology Conference (MidPhon) University of Wisconsin-Madison. September.

• Lababidi, Z. & Park, H. (2014). L1-English tense-lax Vowel System Influence on L2-Arabic Short and Long Vowel Learning. 28th Annual Symposium on Arabic Linguistics (ASAL28) meeting. University of Florida, Gainesville. March.

• Lababidi, Z. & Sun, M. (2012). Achieving Cultural Competency While Playing. National Council of Less Commonly Taught Languages (NCOLCTL) at the university of Wisconsin-Madison. April.

• Lababidi, Z. & Sun, M. (2011). Integrating Culture into a Language Course through Technology. International Association for Language Learning Technology (IALLT) at the University of California Irvine. June.

• Lababidi, Z. & Sun, M. (2010). Teaching Culture through Technology. Wisconsin Association for Language Teachers (WAFLT). Appleton, Wisconsin. November.

• 2009. Toward a Sustainable Education. Appalachian Studies Association conference (ASA). Shawnee State University. Portsmouth, OH. March.

• 2008. Benefits of Placed-based Education. Appalachian Regional Committee (ARC) annual meeting. Washington D.C. April.

WORKSHOPS

• 2014. Teach For America - New Jersey 20th Anniversary Summit. Newark, New Jersey. March.

• 2012. STARTALK summer Teachers' Program. New York University. New York, New York. June.

• 2010. National Middle East Language Resource Center workshop (NMELRC). Seminar for Arabic as a second language instructors by Prof. Mahmoud Al Batal. University of Texas, Austin. June.

• 2010. Intensive five-day workshop on how to administer the Oral Proficiency Interview OPI. ACTFL Sponsored. University of Wisconsin-Milwaukee. June.

• 2009. Chicago Language Symposium. University of Chicago. Chicago, Illinois. November.

HONORS/SCHOLARSHIPS/AWARDS

• 2015. Recipient of the Chancellor's Graduate Student Award, University of Wisconsin-Milwaukee.

• 2013-2014. Student Success Award, University of Wisconsin-Milwaukee.

• 2008-2015. Recipient of the Graduate Teaching Assistant Scholarship. University of Wisconsin-Milwaukee.

• 2008-2009. Recipient of the USA Department of State's Fulbright Foreign Language Teaching Assistant (FLTA). Radford University. Radford, Virginia.

• 2009. Selected as Exceptional FLTA of the year by the Fulbright committee and was invited as an honorary guest to the annual workshop in Cairo, Egypt. July.

MEMBERSHIPS

• Arabic Linguistics Society. (ALS).

• National Council of Less Commonly Taught Languages (NCOLCTL).

• Wisconsin Association of Foreign Language Teachers (WAFLT).

• Linguistics Society of America (LSA).

• American Council on the Teaching of Foreign Languages (ACTFL).

DEVELOPED MATERIAL

Interactive Arabic resource page with materials for acquiring the language: http://uwm.edu/language-resource-center/resources/arabic/

OTHER SERVICES

• 2015. Hosting and organizing the 29th Annual Symposium on Arabic Linguistics (ASAL) at the University of Wisconsin-Milwaukee.

• 2012. Summer common reading group instructor at the University of Wisconsin-Milwaukee.

• 2011. Assisted in the preparation of *Perspectives on Arabic Linguistics XXII-XXIII*.

• 2011~ 2013. Organize a one-day workshop for Plymouth high school students on how to study and learn Arabic as a second language. University of Wisconsin-Milwaukee.

LANGUAGES

Arabic: Native language skills. English: Fluent.

COMPUTER SKILLS

Microsoft Office: Word, Excel, and PowerPoint. Web 2.0 applications for language instruction and audio/visual software, Adobe Premier Pro, Praat, and Audacity.

REFERENCES

References will be provided upon request.