The Acquisition of English /ɪ/ by Spanish Speakers

via Text-to-Speech Synthesizers: A Quasi-Experimental Study

Fernanda Soler Urzúa

A Thesis

in

the Department

of

Education

Presented in Partial Fulfillment of the Requirements

For the Degree of Master of Arts (Applied Linguistics) at

Concordia University

Montreal, Quebec, Canada

August 2011

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Fernanda Soler Urzúa

Entitled: The Acquisition of English /ɪ/ by Spanish Speakers via Text-to-Speech Synthesizers: A Quasi-Experimental Study

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Applied Linguistics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Joanna White	Chair
Marlise Horst	Examiner
Pavel Trofimovich	Examiner
Walcir Cardoso	Supervisor

Approved by Pavel Trofimovich
Chair of Department or Graduate Program Director

Brian Lewis
Dean of Faculty

Date August 25, 2011

ABSTRACT

The Acquisition of English /ɪ/ by Spanish Speakers

via Text-to-Speech Synthesizers: A Quasi-Experimental Study

Fernanda Soler Urzúa

A plausible explanation for learners having difficulties with the acquisition of L2 phonology is the idea that L2 speech is processed through the L1 and perceived in relation to it. L2 learners sometimes fail to perceive the differences between L1 and L2 segments; consequently, they are unable to acquire new sounds. In this context, the concept of perceptual salience takes on added importance because learners might be able to establish differences between L1 and L2 sounds if they are perceptually prominent in the L2 input. Some researchers suggest that multimedia environments are beneficial because the language input can be highlighted in many ways and thus render opaque forms more salient to the learner. This study investigates the extent to which pedagogical instruction using text-to-speech (TTS) technology as a means to enhance the aural input assists learners in the acquisition of the English /ɪ/. Three groups of learners of the same L1 (Spanish) and similar English proficiency were pre-tested on their ability to perceive and produce the target vowel by means of different tasks (two for each ability). Each group was subjected to a different instructional condition: TTS-based instruction, non-TTS based instruction and regular classroom instruction. The TTS group performed tasks intended to develop their perception of the target forms via TTS; the non-TTS group performed the same tasks, but receiving input from the researcher; and the third group worked on listening comprehension tasks. It was hypothesized that the TTS group would outperform the other two groups in terms of perception and production. After completing the treatments, the three groups were tested on their productive and perceptual abilities in relation to the target sound. Two weeks later, the participants received the same tests. The results obtained showed that the TTS group significantly outperform the non-TTS group in one of the pronunciation tasks. However, their performance in the other tasks in the post-tests was not significantly different from the other groups. These results are discussed with respect to the hypotheses proposed and in relation to the relevant theory and previous studies. The limitations of the study together with suggestions for future research and its implications for ESL teaching are also addressed.

ACKNOWLEDGEMENTS

There are many people whom I would like to thank for their encouragement and tremendous support during the process of giving shape to this research and writing this thesis. First of all, I would like to express my deepest gratitude to my supervisor, Dr. Walcir Cardoso, without whom this thesis would not have been completed. When I first approached him, without my being a student of his yet, he enthusiastically received my 'yet-to-be-refined' research idea. His unconditional and solid support and his ever-encouraging mood, as well as his time, guidance and expertise have been there along the entire process of completing this research. Dr. Cardoso has set an impressive example of what is to be an excellent professor to me, one that inspires, encourages and enlightens students.

I would also like to sincerely thank the members of my committee, Dr. Pavel Trofimovich and Dr. Marlise Horst, for their insightful comments and help when this investigation was still a project. In addition, I would like to thank Randall Halter for providing me with assistance with the statistical procedures to be carried out with my 'messy' data.

My special thanks go to some of the faculty members and students of the Foreign Languages Department in Universidad de Concepción: Thanks to the Vice Dean of the Faculty of Arts and Humanities, María Edith Larenas, for allowing me go back to my former university and conduct my research there; thanks to the English instructors Pamela Arévalo and Ricardo Vera for letting me intervene in their groups to carry out the experiments that were part of this research. I would also like to extend my gratitude to the participants of this study, without whom this thesis would not have been even possible.

Thanks to those former colleagues at Universidad de Concepción who warmly welcomed me back home and showed a deep interest in this research.

I would also like to thank Neospeech Inc. who donated the licences for the TTS software and voices used in this study. Thanks to Mr. Alexander Konyari, manager of budget and facilities of Concordia University Libraries, for donating the recyclable bags that the participants of this study received as compensation for their participation.

My friends in Montreal have also shown an unconditional support during this process. They have taken very good care of distracting and relaxing me when I needed to have my mind 'cleared'. Thanks to those friends, especially Dr. Bruno Giordano, for listening and entered into technical discussions with me when I 'needed' to let people know what my research was about. I would also like to give special thanks to my friend Chelsea O'Meara, whose professional help with the rating of the speech samples resulted of immense value to this work. Thanks, Chels, for being next to me since this project started.

I would also like to express my heartfelt and eternal gratitude to my parents, Mauricio and Pilar, and my sisters, Alejandra, Paloma and Masiel, who have always encouraged me to 'take the plunge' and start up new projects in my life. Many thanks for sticking with me and for always showing me all your love and support. Thanks mom and dad for instilling confidence and a sense of entitlement into me; thanks for teaching me to contest my own ideas and those of others.

Finally I would like to deeply thank my beloved husband, partner and friend, Miguel Torres. Thanks for introducing me to the world of computer programming, which I never thought I would be into. Without your invaluable help, it would have been extremely difficult to complete the design of the computer tasks of the experiment by myself. Thanks for listening to me for hours talking about my research project and my concerns about it. Thanks for sticking with me during this fantastic journey with your unconditional love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Phonology is a crucial part of linguistic competence, since the phonological inventory of a given language influences speakers' speech patterns and governs systematic phonetic variation (Czaykowska-Higgins & Dobrovolsky, 2009). Phonology also underlies speakers' productive (speaking) ability, and it has been found to be intimately linked to speakers' identity (e.g., Gatbonton & Trofimovich, 2008). However, learning second language (L2) phonology (i.e., being able to perceive and produce L2 speech in an intelligible or native-like manner) poses persistent difficulties for many L2 learners.

Many theories and hypotheses have been elaborated over the years in an attempt to explain L2 learners' difficulties with the acquisition of L2 phonology. However, the idea that production errors are caused because L2 speech is processed through the native language (L1) and perceived in relation to it is widely accepted in the field of second language acquisition (SLA) (e.g., Best, 1993, 1995; Brown, 1998, 2000; Eckman, 2004; Flege, 1995, 2002, 2003; Hallé & Best, 2007; Hancin-Bhatt, 1994). In this context, the most influential theoretical proposals are Best's Perceptual Assimilation Model or PAM (1995) and Flege's Speech Learning Model or SLM (1995, 2002, 2003); see also Escudero (2006) for a similar view. Briefly, the PAM argues that an L2 listener perceptually compares the articulatory gestures of L2 sounds with those of the L1 (e.g., places of articulation, points of articulation, tongue height, tongue backness or frontness). This model also predicts different patterns of assimilation (i.e., perception) according to the features of the segments in the L1. These patterns may vary in their degree of accuracy depending on the way in which the learner perceives the L2 input. Similarly, the

SLM hypothesizes that L2 speech is processed through the existing L1 phonetic categories, and it predicts that if L2 learners are able to perceive some of the phonetic properties that make an L2 segment distinct from an L1 one, they will be likely to produce it more accurately and form a new phonetic category in their phonological repertoire.

But what makes learners perceive the properties of L2 sounds and make a distinction between non-native and native segments? A plausible explanation relates to perceptual salience, which refers to the overall perceptibility of a language form (Collins, Trofimovich, White, Cardoso & Horst, 2009; Ellis, 2008; Goldschneider & DeKeyser, 2001). L2 learners might have difficulties in acquiring certain L2 sounds because they are not easily perceivable, even though these sounds may be frequent forms in the aural input. It has previously been shown that highly frequent L2 forms still pose considerable challenges for L2 learners, possibly because they are not perceptually prominent in speech (e.g., the allomorphs of the regular simple past tense morpheme *–ed*). In fact, Collins et al. (2009) suggest that perceptual salience is apparently one of the most determining factors that influence how well language forms, such as the morpheme *–ed* and its associated allomorphs, are acquired.

From the predictions of the PAM and the SLM in relation to the acquisition of L2 segments, and taking perceptual salience into account, one could hypothesize that enhancing the aural input in the L2 might lead to an improved accuracy in the production of L2 sounds, since L2 learners could have more opportunities to perceive some of the phonetic features that make L2 sounds different from those found in the inventory of their L1s.

The aural input that is not always easily perceivable to the L2 learner could certainly be enhanced in several ways. However, the idea of exploring the effectiveness of technological tools that have not been studied to a great extent bears particular relevance to SLA, since these might prove useful to assist L2 learning. This study investigates the effectiveness of a text-to-speech synthesizer (i.e., technology that reproduces written text orally) as a means to highlight the aural input and explore its influence on the acquisition of a problematic English vowel by Spanish learners of L2 English: /ɪ/.

The Current Study

Over the years, it has been observed that Spanish learners of L2 English encounter difficulties when learning some English sounds. Vowels seem to be especially problematic, but one that appears to represent a constant challenge is /ɪ/, as the one found in the word *bit* [bɪt]. Possibly due to L1 transfer and perceptual difficulties, Spanish speakers tend to pronounce this phoneme as the Spanish /i/, such as the one found in *sí* [si] (i.e., *yes*), which in turn is very similar to English /i/, the one found in *beat* [bi:t].

Considering the acquisition of /ɪ/ as the linguistic target for this study, the purpose of this investigation is to explore the extent to which pedagogical instruction using a text-to-speech synthesizer (TTS) helps in the phonological development of this vowel by enhancing the aural input to which language learners are exposed.

The rationale behind the choice of this vowel lies in the following factors: As mentioned earlier, Spanish speakers encounter major difficulties in producing it (e.g., Cenoz & García Lecumberri, 1999; García Pérez, 2003). A second factor is the concept of functional load (FL), which refers to the "measure of the work that two phonemes (or distinctive features) do in keeping utterances apart" (King, 1967, p. 831). According to Munro and Derwing (2006), the concept of FL has been extended in the field of applied linguistics to contrasting segments with the aim of establishing a hierarchy in terms of their relevance to pronunciation. The authors state that the "frequency of minimal pairs, the neutralizations of phonemic distinctions in regional varieties, segmental position within a word and the probability of occurrence of individual members of a minimal pair" are factors that have been taken into account when establishing FL (Munro & Derwing, 2006, p. 522). The /ɪ/-/i/ contrast has a high FL; specifically, it corresponds to an 8 in a 10-point scale, where 1 means weak FL, and 10 high FL (Levis & Cortes, 2008). That is, words with this contrast that constitute minimal pairs are rather common. Because of being a high FL contrast, the alternation between /ɪ/ and its counterpart, /i/, might cause communication breakdowns. Celce-Murcia, Brinton and Goodwin (1996) illustrate this phenomenon by presenting the following anecdotal situation: "The learner is discussing an incident in which her child had choked on something and could not breathe. 'He swallowed a pill,' she says. 'What kind of peel?' asks the native speaker. 'An aspirin,' says the learner. 'Oh, a *pill*! I thought you said *peel*,' responds the native speaker" (p. 131). Situations like this one would clearly confuse a listener in a context where there are not sufficient cues to understand what the speaker intended to say. In addition, Jenkins (2002) states that the alternation between short and long vowels, and explicitly /i/-/ɪ/, hinders mutual intelligibility. She suggests that this vowel contrast should be included in the *Lingua Franca Core*, a phonological syllabus for learners of English as an international language.

Another reason to study /ɪ/ is that its acquisition does not seem to have been studied using TTS technology, either in lab or instructional settings. Finally, the fact of choosing a single vowel allows for a more comprehensive investigation of the phenomenon and a more focused investigation of the effects of TTS.

In the following section, I will provide a review of both Spanish and English vowel systems, so that we can have a better understanding of the phonetic and phonological properties of the vocalic sounds in both languages, and thus establish the differences that might constrain the acquisition of /ɪ/ by Spanish learners of L2 English.

Spanish vs. English Vocalic System

The English vowel system is fairly large in comparison to the one found in Spanish. Although the repertoire of vowels depends on the variety of English, there are at least 14 phonemic vowels in North American English, namely /æ/, /a/, /ɑ/, /ʌ/, /ɛ/, /e/, /ə/, /ɜ/, /i/, /ɪ/, /o/, /ɔ/, /ʊ/ and /u/ (Ladefoged, 2005), whereas the Spanish system has only five: /a/, /e/, /i/, /o/ and /u/ (Finch & Ortiz, 1982; Hualde, 2005; Odisho, 1992; Stockwell & Bowen, 1965). The two vocalic systems are illustrated in Figures 1 and 2.

.

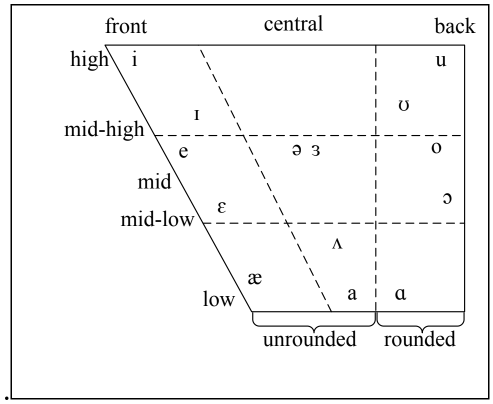

Figure 1: North American English vowel system.

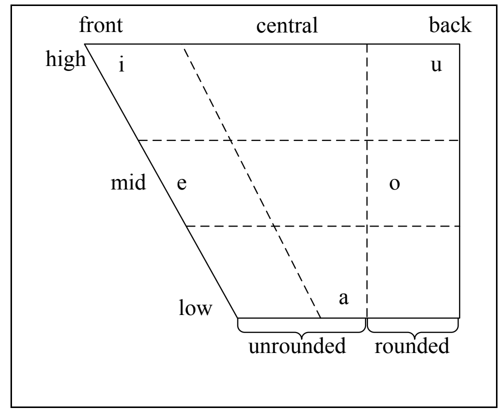

Figure 2: Spanish vowel system.

Other differences arise when we take a closer look at quality (i.e., height, frontness, roundedness) and quantity (i.e., length) of the vowels in both systems. In terms of quality, some sounds that appear to be similar still differ in some features. For example, Spanish /u/ is non-diphthongal and retracted, whereas English /u/ is slightly diphthongal (phonetically represented by a following ‘ː’ [uː] or by ‘w’ [uw] to indicate increased vowel length) and less retracted than its Spanish counterpart (Hualde, 2005).

Suprasegmental factors, such as stress and rhythm, account for further differences between English and Spanish vowels, since quantity and quality of vocalic sounds are influenced differently in Spanish and English by these two factors. In English, vowels in syllables with primary stress reach their maximum quantity and optimal quality, whereas in syllables with secondary or no stress, vowels are dramatically reduced in both quantity and quality (Avery & Ehrlich, 1992; Odisho, 1992). The fact that stress influences vowel quality and quantity in English contributes to the movement of unstressed vowels to the center of the vowel spectrum (Odisho, 1992). This is what Odisho (1992) calls a *centripetal* vowel system, in which all unstressed vowels have the tendency to undergo some kind of *schwaization* or reduction. In contrast, he defines the Spanish vowel system as *centrifugal*, since vowels resist moving to the center.

In terms of rhythm, which relates to the notion of alternation between stressed and unstressed syllables in a sentence (Odisho, 1992), English and Spanish are diametrically opposed. English has a stress-timed rhythm, whereas Spanish has a syllable-timed rhythm. In a stress-timed language such as English, stressed syllables tend to occur at regular time intervals (Richards & Smith, 2002). In contrast, in syllable-timed languages such as Spanish, "all syllables, whether stressed or unstressed, tend to occur at regular time intervals and the time between stressed syllables will be shorter or longer in proportion to the number of unstressed syllables" (Roach, 2000, p. 135). The tendency of English to stress vowels at regular time intervals also contributes to the reduction of unstressed vowels, which in turn affects vowel quality and quantity. In the case of Spanish, vowels are resistant to change in quality and quantity because the language emphasizes syllables; as a result, vowels are unaffected by reduction. For example, in the English word 'calendar' [ˈkæ.lən.dɚ], the vowel that carries the stress in this case is /æ/ (in the first syllable), and the other two are reduced. If we were to pronounce this word stressing the second syllable, we would probably say [kə'læn.dɚ] and the first vowel would undergo reduction. Conversely, the vowels in the Spanish word 'calendario' (i.e., *calendar*) [ka.lenˈda.rjo] would be pronounced distinctly, no matter where the word stress be placed.

In summary, the size of the phonological repertoires, as well as the segmental and suprasegmental characteristics of each vowel system account for substantial differences between English and Spanish. These differences might be some of the reasons why Spanish speakers have difficulties with the acquisition of the English /ɪ/.

The English /ɪ/ and the Spanish /i/

As mentioned earlier, there are vowels in Spanish that resemble some of the vowels that comprise the English system; however, they do not have exactly the same quality and quantity. The English /ɪ/ and the Spanish /i/ share the quantity feature, i.e., they are both short vowels, but they still differ in qualitative features. Spanish /i/ is a high-front vowel, but English /ɪ/ is lower and more centralized. Figure 3 shows a close-up of the distribution of these segments in the articulatory space.

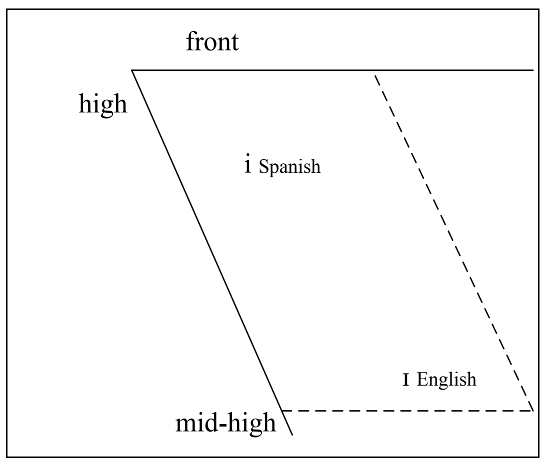

Figure 3: Distribution of English /ɪ/ and Spanish /i/ in the articulatory space.

The qualitative differences between these vowels, as illustrated above, might constrain the acquisition of the English /ɪ/ by Spanish learners of L2 English. However, there are some theoretically-oriented approaches that could also explain the reasons why Spanish speakers encounter difficulties with this sound. In Chapter 2, I will review some of the approaches that have been proposed by SLA researchers to explain some of the constraints in acquiring L2 phonology.

CHAPTER 2

LITERATURE REVIEW

L2 Phonology Acquisition: An Overview of Models and Hypotheses

Researchers working in the field of L2 phonology have proposed different explanations for L2 learners' difficulties with acquiring segmental (e.g., individual phones and phonemes) and suprasegmental (e.g., stress, rhythm, intonation) aspects of an L2. In this attempt, they have formulated several proposals in the form of hypotheses and models of acquisition. For example, Penfield and Roberts (1959) and later Lenneberg (1967) proposed that there was a critical age period for learning an L2 (Critical Period Hypothesis). These researchers suggested that because of the loss of brain plasticity, a native-like pronunciation in an L2 was difficult to achieve after puberty.

Other researchers have relied on markedness to explain the difficulty in achieving a native-like pronunciation. Eckman (1977) defines markedness as the following: "a phenomenon A in some language is more marked than B if the presence of A in a language implies the presence of B, but the presence of B does *not* imply the presence of A" (p. 320). He elaborated the Markedness Differential Hypothesis (MDH), which mainly consisted on establishing degrees of difficulties in the acquisition of L2 sounds based on the comparison of the L1 and L2 systems in terms of markedness. Later, Eckman (1991) re-elaborated the MDH and proposed the Structural Conformity Hypothesis, which was also based on principles of markedness, but with less emphasis on the L1 and L2. This new hypothesis was based on the assumption that interlanguages

(ILs) are also languages, so the principles of universal markedness would also apply to these developing ILs.

Other linguists have tried to explain L2 phonology acquisition in terms of its development over time (e.g., Cardoso, 2007; Escudero, 2000; Major, 1986, 2001; Morrison, 2008, 2009). However, the most relevant models for the present study are those that are concerned with the influence of perception of L2 sounds on L2 phonological acquisition, since this investigation will also address perception and the role of the aural input (via TTS) in production. As mentioned earlier, the most influential models that deal with non-native perception and L2 phonology acquisition are the Perceptual Assimilation Model or PAM by Best (1995) and the Speech Learning Model or SLM by Flege (1995, 2002, 2003). The two models broadly share the notion of L2 perception (i.e. they both assume that L2 speech is perceived in relation to the L1), but only the SLM deals with production. Another approach that has been advocated as a factor constraining L2 acquisition from the point of view of perception and deserves attention because of its relevance to this study is perceptual salience. In the next section, a review of these acquisition perspectives and their implications for the current investigation will be provided.

The Perceptual Assimilation Model

The PAM (Best, 1995) assumes that L2 learners base their perception on the articulatory gestures of the L2 and then they process L2 speech through an L1 filter. In this context, L2 learners are expected to assimilate (i.e., perceive) a new L2 sound (a) as an L1 category, (b) as an uncategorizable sound which can fall somewhere in the articulatory space in which a given L1 sound is produced, (c) or as a non-speech (i.e., non-phonemic) sound. In her model, Best also elaborated perceptual patterns for L2 contrasts, which follow from the three possible assimilations just outlined above. In the first pattern, called *two-category type* (TC type), both members of an L2 contrast are assimilated as two different L1 categories and, in this case, the discrimination between the L2 sounds and those of the L1 is expected to be excellent. In the second one, called *category-goodness type* (CG type), L2 learners assimilate both members of the L2 contrast to only one L1 sound, but this same sound will be acceptable in comparison to one of the members of the non-native contrast, and, at the same time, it will deviate from the other L2 sound. Here, discrimination between L1 and L2 sounds is expected to be moderate to very good. In the third case, referred to as *single-category type* (SC type), the two members of the L2 contrast are also assimilated to one L1 sound, but this L1 sound can be either equally acceptable or equally deviant from the L2 ideals. In this case, discrimination is expected to be poor. The three remaining assimilation patterns for L2 contrasts relate to the uncategorizable or non-assimilable L2 sounds, but they will not be discussed here as they do not relate to the feature under investigation.

The PAM and the Acquisition of /ɪ/

Following the PAM assimilation patterns for contrasts, there are three possibilities for the assimilation of the English /ɪ/ by Spanish speakers. The most frequently observed case seems to be the SC type. That is when learners assimilate both members of the contrast /ɪ/-/i/ as the Spanish /i/ (see (a) in Table 1), which is different from /ɪ/ and slightly different to the English /i/. The production of /i/ might be also considered acceptable for the target sound /i/, in which case there would be a CG type of assimilation (see (b) in Table 1). The third possibility is the TC type, since Spanish speakers might assimilate

English /i/ as the Spanish /i/, and /ɪ/ as the Spanish /e/. It is important to mention that this type of assimilation might be regarded as the “ideal”, since /e/ shares more similarities in terms of quality and quantity with English /ɪ/ (Escudero, 2000). Table 1 displays the types of assimilation postulated by Best (1995) contrasting both languages.

	(a) SC type		(b) CG type		(c) TC type	
English	/i/	/ɪ/	/i/	/ɪ/	/i/	/ɪ/
Spanish	/i/		/i/		/i/	/e/

Table 1: PAM and the /i/-/ɪ/ contrast: Possible assimilation patterns.

The Speech Learning Model

The SLM (Flege, 1995, 2002, 2003) is also concerned with perception, but it also includes production of L2 speech. Similar to the PAM, this model assumes that the sounds of a new language will be processed through an L1 filter. That is, the already existing phonetic categories of the L1, which are stored in long-term memory and remain unchanged over the life span, will serve as a point of reference to perceive L2 sounds. Another assumption that is made explicit in this model is that bilinguals have difficulties in discerning the differences between L1 and L2 sounds that co-exist in a common phonological space, as in the case with English /ɪ/ and Spanish /i/. However, the model predicts that if an L2 learner is able to perceive at least some phonetic properties that make an L2 sound different from the closest L1 sound, she will be able to establish a new L2 phonetic category in her phonological repertoire. As a consequence, the more phonetic differences noticed between closely related L1 and L2 sounds, the more likely it is for the learner to acquire the new L2 sounds and produce them more accurately.

The SLM and the /i/-/ɪ/ Contrast Acquisition

Recall that English /ɪ/ and Spanish /i/ are different in terms of quality. In addition, the Spanish vowel system lacks /ɪ/, which co-exists in almost the same phonological space as the Spanish /i/. According to the SLM (Flege, 1995, 2002, 2003), this is one of the cases in which perception and production of the L2 sounds becomes difficult. However, if Spanish learners of English are able to perceive some of the phonetic features that make /ɪ/ different from Spanish /i/, they will be more likely to produce this L2 sounds more accurately.

Perceptual Salience

In broad terms, salience refers to the general perceived prominence of a stimulus (Ellis, 2006). Ellis states that despite the fact that this prominence may be somewhat related to a physically measurable property, salience is associated with a more subjective experience with a stimulus, so what might be salient for an individual might not be salient for others.

As mentioned earlier, the concept of salience in SLA relates to the overall perceptibility of a language form, i.e., it refers to the "ease with which a structure is heard or seen" (Dulay, Burt & Krashen, 1982, p.32). According to Deumert (2003), our ability to perceive salient structures plays a crucial role in our learning in general, not only when we learn languages. Collins et al. (2009) support this claim stating that perceptual salience is apparently one of the most determining factors in language learning because it influences the way in which L2 learners interact with the input in the L2. If certain phonetic features in the L2 input are perceptually prominent, it might be possible that L2 learners will be able to establish new phonetic categories for a given non-native sound and, therefore, produce it with more accuracy. In fact, there is previous empirical evidence that supports the notion that prominent L2 input can influence perception. McCandliss, Fiez, Protopapas, Conway and McClelland (2002) manipulated the language input by repeatedly exposing Japanese learners to English /ɹ/ and /l/ and found positive results in terms of L2 perception.

Perceptual Salience and the Acquisition of /ɪ/

Perceptual salience may well explain the difficulty of acquisition of the English /ɪ/. Phonetically speaking, /ɪ/ is a short, non diphthongal phoneme that often occurs in unstressed syllables; consequently, one may conclude that this vowel is weak (i.e., reduced) in speech. In addition, /ɪ/ is less sonorous than its counterpart /i/, according to the sonority hierarchy of vowels, which states that high peripheral vowels are more sonorous than high central vowels (De Lacy, 2006). The lack of prominence of /ɪ/ might contribute to its difficulty of acquisition; therefore, it might prevent the formation of a new L2 phonetic category for this foreign phoneme.

The Relationship between the SLM, the PAM and Perceptual Salience

Despite the fact that the PAM makes detailed predictions based on the articulatory gestures of L2 segments, and that the SLM establishes predictions based on the phonetic properties of L2 sounds, both models support the idea that L2 speech is processed through an L1 filter and still propose that target-like acquisition is possible. This means that L2 learners perceive L2 segments in relation to the phonetic categories established for L1 sounds, and then produce L2 speech according to their perceptions of these categories. As indicated in the previous section, the concept of perceptual salience is related to some extent to the SLM and PAM because it addresses the influence of language prominence on the perception and the acquisition of L2 sounds. Drawing from these ideas, it seems logical that enhancing the aural input might lead to a better perception of L2 segments, as input might allow learners to notice the difference between L1 and L2 segments and, therefore, produce them more accurately.

In light of the findings of their study, in which a set of language forms that was more difficult to acquire by L2 learners were found to be less perceptually salient, Collins et al. (2009) suggest that researchers should explore new forms of enhancing the input to increase the availability and accessibility of language forms that pose constant challenges to L2 learners. Certainly, the language input can be enhanced in several ways, for example, by exposing students to multiple instances of the forms we want them to learn via repetition or by highlighting these forms using stress and intonation. However, a multimedia environment in which learners are exposed to the L2 input might have tremendous advantages (Chapelle, 2003). Technology can help make the L2 input prominent by automatically taking on repetitive tasks –something that might not be very appealing to L2 teachers– or it can help us to visually highlight the L2 input, as is the case of text-to-speech synthesizers (TTS).

It is important to mention that the current study was not set up to test the validity of the PAM, SLM or the concept of perceptual salience. Instead, these approaches were adopted in order to establish a theoretically-sound link between the use of technology and L2 instruction, as suggested by Chapelle (2009).

Before presenting the TTS application that was used in the current study, I will review below some of the research carried out to investigate the acquisition of /ɪ/ in the context of Hispanophones learning English as an L2.

Previous Research on the Acquisition of English /ɪ/

Several studies have been devoted to the investigation of the acquisition of /ɪ/ in terms of production and perception. As will be discussed below, the outcomes of these studies are not uniform: while some studies found that speakers whose mother tongue lacks /ɪ/ can produce it and perceive it in a native-like manner, others revealed that, even though learners cannot produce it, they can still perceive it.

Cenoz and García Lecumberri (1999) set up a study to investigate whether pronunciation training had any effect on the perception of several English vowels and diphthongs (N=19), including both /i/ and /ɪ/. The participants of this study included 109 university students in Spain. Around 70% of them had Spanish as their L1, and the remaining 30% were Basque-Spanish bilinguals. Participants, whose self-reported level of English was intermediate, were pre-tested on their ability to discriminate English vowels and then subjected to a treatment of 28 hours during 14 weeks, which mainly consisted of aural discrimination practice and transcription of sounds. After being post-tested, participants exhibited an improvement on their ability to discriminate English vowels. However, one of the most difficult vowels to identify was /i/, and no major difficulties were found with /ɪ/. Despite the fact that the authors reported a significant influence of training on the overall perception of English vowels, the improvement of isolated vowel sounds was not statistically tested; therefore, it is not possible to conclude whether the improvement in perception of some vowels, namely English /ɪ/, was significant.

Also with the aim of exploring the effects of training on the acquisition of English vowels, García Pérez (2003) elaborated a study to investigate the ability to perceive and produce English vowel contrasts, including /i/ and /ɪ/ among a larger inventory of vowels. Before and after a three-week training period that consisted of teaching participants strategies to identify the members of each contrast and to produce them, learners' perception and production were assessed. Thirty-two participants, whose mother tongue was Spanish and whose level of English proficiency was intermediate, were randomly assigned to an experimental and a control group. The results of the post-test showed that the participants in the experimental group significantly outperformed the controls in the perception of all contrasts, including /i/ and /ɪ/. However, in terms of production, none of the groups showed a significant improvement, and no significant differences were found between the controls and the experimental group, which suggested that pronunciation training had no relevant effects on their ability to produce /ɪ/.

Similarly, Bion, Escudero, Rauber and Baptista (2006) set up a study to investigate the relationship between vowel production and perception. It is important to mention that the participants of this study (N = 17) were L1 speakers of Brazilian Portuguese (BP), a language that also lacks /ɪ/, so the phenomena observed with L1 BP speakers might be similar to those observed in L1 Spanish speakers learning English. The participants, who had a high proficiency in English, were tested on their ability to perceive and produce two vowel contrasts, one of which included the /i/-/ɪ/ set. The results of this study revealed that participants had a very accurate perception of both /i/

and /ɪ/ and could differentiate them in 96% of the instances. However, participants had great difficulties in their production, thus suggesting that their production of these vowels was not directly related to their perception. In other words, the participants' perception of the contrast was better developed than their ability to produce the phonemes.

Similar results were obtained by the same researchers (Rauber, Escudero, Bion & Baptista, 2005) in an experiment carried out to investigate whether poor discrimination (i.e., perception) of vowels was related to poor production. As in the previous study, the participants (*N*= 16) were L1 speakers of BP who had a high proficiency in English and had taught English for at least five years. The results of the production task revealed that half of the participants produced /i/ and /ɪ/ distinctly, but the authors note that this distinction was "small". No tests were carried out to show whether this difference was statistically significant. In terms of perception, the results revealed that participants had almost no difficulties in discriminating this vowel contrast, since they were able to perceive it correctly in 93.8% of the instances. Again, the relationship between perception and production in this study was somewhat weak. However, the authors suggested that the ability to perceive vowels with accuracy might precede its correct production.

Different results were found by Flege, Bohn and Jang (1997), whose main purpose was to explore the effects of language experience in the production and perception of a large set of English vowels. The researchers recruited 90 participants with different L1 backgrounds; however, because of the scope of this study, only the results obtained by Spanish speakers (*N*=20) involving /ɪ/ will be reported. Participants were divided into two groups: experienced and inexperienced subjects. After testing participants' accuracy in producing and perceiving vowels, the researchers concluded that there were some differences associated with their experience with the target language. Experienced subjects perceived and produced English vowels with more accuracy, whereas inexperienced subjects encountered more difficulties in both perceiving and producing vowels, thus suggesting that their production was related to their perception: The better the participants perceived /ɪ/, the better their performance in production.

As mentioned at the outset of this section, the results obtained in these studies exploring the acquisition of /ɪ/ are not uniform. While in some studies proficient subjects were able to perceive the difference between /ɪ/ and /i/, but not to produce /ɪ/ accurately (e.g., Bion et al., 2006; García Pérez, 2003; Rauber et al., 2005), others have found that learners could perceive and produce these sounds with relative accuracy at the same time (e.g., Flege et al., 1997).

From the conflicting results of the research carried out to explore the acquisition of /ɪ/, one may conclude that this English vowel poses persistent difficulties to Spanish learners of L2 English, both in terms of perception and production. As a result, adopting a pedagogical approach that emphasizes increased exposure to this vowel by manipulating the aural input and consequently increase its overall perceptibility might be helpful to overcome these difficulties. One of the approaches that can be adopted to accomplish this purpose is the use of a text-to-speech synthesizer (TTS). In the next section, I will present this tool followed by a review of some studies that have explored the effectiveness of TTS in different areas of L2 learning.

What is a Text-to-Speech Synthesizer?

A text-to-speech synthesizer (TTS) is a computer program designed to generate speech from written text automatically. TTSs work by modules that serve the function of decoding the text and transforming it into speech (Dutoit & Cerňak, 2005; Handley, 2009; Lee, 1969; Šef & Gams, 2003). Briefly, they are designed with a set of two modules. While the first module transforms the text into phonemes, the second processes the phonemes, and transforms them into speech (Handley, 2009). According to Sisson (2007), the speech output is usually generated by three methods, namely formant, articulatory, and concatenative synthesis. Formant synthesis is based on the acoustic properties of speech sounds, whereas articulatory synthesis simulates the movements of the vocal tract to generate speech (Sisson, 2007). Additionally, concatenative methods generate speech based on pre-recorded chunks of human voice that are then linked together in order to reproduce it (Carlson, 1995).

It is important to mention that the quality of the output of TTS has improved substantially over the years. In the past, the speech output was of low quality (Carlson, 1995), whereas nowadays it sounds more natural (e.g., voices produced by AT&T, Neospeech and Acapella). The fact that some of the current TTS programs have a more natural-sounding output represents an advantage because they could be used as a 'more natural' means to enhance the L2 aural input and, therefore, help learners perceive some of the phonetic properties of /ɪ/ and, consequently, the acoustic differences between this vowel and the Spanish equivalent /i/.

TTS and Computer-Assisted Language Learning

Chapelle (1998, 2003, 2007, 2009) recommends there be a stronger link between computer-assisted language learning (CALL) and SLA theory, i.e., CALL-based pedagogy should be implemented taking into account what it is hypothesized to facilitate L2 learning. Chapelle (2001, 2009) has outlined a series of criteria that CALL tasks should satisfy to be used in pedagogically-sound L2 instruction; these include the potential for learning to take place, learner fit, meaning focus, authenticity, positive impact and practicality. Interestingly, the study of TTS systems in CALL-based pedagogy has received very little attention in the field of SLA. Despite the scarcity of investigations assessing TTS suitability for L2 learning (e.g., Handley, 2009; Stratil, Burkhardt, Jarrat & Yandle, 1987; Stratil, Weston & Burkhardt, 1987), the results suggest an overall positive effect on L2 learning. However, these studies have assessed the speech of languages other than English (i.e., French and Spanish), and as Handley (2009) states, the challenges for TTS systems are different depending on the language being used; therefore, the results of the previous research may not necessarily apply to English TTS synthesis.

Previous Research on TTS and its Influence on Second Language Learning

Very few studies have been conducted to investigate the effectiveness of TTS in different areas of second language learning, particularly in L2 phonology. After an exhaustive search in different academic databases and search engines, it was possible to find four studies that dealt with the use of this tool in assisting L2 learning. These studies addressed the development of L2 writing skills, L2 vocabulary in conjunction with reading comprehension, and L2 phonology. For example, Kirstein (2006) found that TTS was an effective tool to assist L2 writing because it allowed learners to pay more attention to the semantic aspects of their written work. Proctor, Dalton & Grisham (2007) also found positive –although not significant– evidence for a tutoring system that had a TTS feature on learners' vocabulary gains and reading comprehension.

In the area of L2 phonology only two (non peer-reviewed) studies have been documented. The first is that of Hincks (2002), who set up an investigation to find out whether the use of some of the features of a TTS could help on the acquisition of the stress patterns of two English words that are often mispronounced by Swedish speakers: '*component*' and '*parameter*'. She recruited 13 participants from a technical English class and, while being recorded, had them read aloud a text in which there were three instances for each of the words. After this task, she instructed the students to type the target words into TTS software in their L1 (Swedish) and do a series of changes to the pitch and duration of the vowels until their pronunciation in the application sounded English-like. Students were then instructed to listen to the words repeated times. Four weeks after this task, the researcher had the students read the texts used in the pre-test again and recorded them. The first target word, '*component'*, was accurately pronounced in more than 80% of the instances, while in the pre-test the same words were correctly pronounced only around 55% of the time. The word *'parameter'*, which had an accuracy percentage of 20% in the pre-test, was then pronounced correctly 60% of the time. Although these results show some improvement on pronunciation, the researcher did not run statistical tests to check whether this gain was significant.

The second study that investigated the development of L2 pronunciation via TTS is Kiliçkaya's (2008), whose purpose was to find out whether the use of accent reduction software and a TTS application improved learners' pronunciation skills in L2 English. Participants (N = 35), whose L1 was Turkish, were pre- and post-tested on their pronunciation of single words and declarative sentences in English. Participants were divided into a control group, which followed traditional instruction; an experimental group under traditional instruction plus the accent reduction software; and a second experimental group using accent reduction software and a TTS synthesizer, following traditional instruction as well. The author did not present a detailed description of the treatments to which participants were subjected; however, she stated that traditional instruction involved the use of a pronunciation textbook together with a CD for the listening activities. Kiliçkaya found that the group that used both software, i.e., accent reduction and TTS, significantly outperformed the other groups in the pronunciation of declarative sentences in the post-test. Although not many details are presented in Kiliçkaya's (2008) article, it is possible to infer that the use of TTS synthesizers might be an effective way of improving L2 learners' pronunciation skills.

From the review above, it is evident that there is a substantial research gap regarding the use of TTS and its effects on SLA. Most of the studies reviewed in this section explored the effects of several tools or functions of computer programs in conjunction with TTS, making it difficult to be precise about the effectiveness of TTS in SLA. Therefore, more formal studies, in which the use of the TTS can be better controlled, are needed in order to confirm its potential effects.

An inspection of the few available studies suggests that this technology has the potential to be used as a tool to assist L2 learning. Besides, and as mentioned earlier, TTS can both take on repetitive tasks and visually highlight forms in the L2 input (e.g., those that are being spoken). The efficiency of TTS might also be an important factor when considering its use, since learners can paste any text in the application and listen to it, without teachers having the need to obtain custom recordings for their lessons. In addition, there are several TTS applications freely available on the internet, so learners could easily have them installed in a computer or use them in a web-based, on-line environment. The use of TTS in L2 instruction might also help teachers save time on finding material for listening activities, since they could use a written text and have learners listen to it. In this context, the current research explores the usefulness of this tool on the acquisition of the English /ɪ/ by Spanish speakers in an instructional setting.

Research Questions

In order to explore whether TTS-based instruction facilitates the acquisition of /ɪ/ by Spanish learners of L2 English, I propose the following research questions:

1. Will a group receiving TTS-based instruction show a greater improvement in the perception of English /ɪ/ than a group receiving instruction without a TTS and a control group?
2. Will a group receiving TTS-based instruction show a greater improvement in the production of English /ɪ/ than a group receiving instruction without a TTS and a control group?

Hypotheses

Based on the theoretical perspectives adopted in this study, as well as on general CALL literature and previous –albeit scarce– research on TTS, I hypothesize that the use of a TTS will be an effective way to assist in the development of L2 phonology. Via TTS and its related features, learners may perceive the phonetic properties that distinguish the Spanish /i/ from the English /ɪ/ and, consequently, produce the target sound with higher accuracy. The proposed main hypotheses are:

1. The TTS group will significantly outperform the non-TTS group and the control group in terms of perception of English /ɪ/.
2. The TTS group will significantly outperform the non-TTS group and the control group in terms of production of English /ɪ/.

The following chapter describes the methodology adopted in this study to answer the research questions and test the proposed hypotheses.

CHAPTER 3

METHODOLOGY

Participants

Forty-seven native speakers of Spanish studying English as a foreign language voluntarily agreed to take part in this study. All participants were first-year undergraduate students in a translation and interpreting program at a university in the south of Chile. The sample consisted of 40 females and 7 males, whose ages ranged from 17 to 23 years (*M* = 18.5, *SD* = 1.2). Participants were part of three intact classes of the same English language course which was part of their program and was intended to be taken by first-year students with low and low-intermediate proficiency in English. Participants attended this English course during eight hours per week. Each class was randomly assigned to a different condition: a group receiving TTS-based instruction (17 participants; 16 females and 1 male); a group receiving instruction without TTS (16 participants; 14 females and 2 males); and a control group (14 participants; 10 females and 4 males).

Participants' Background

After signing a letter of informed consent to participate in the current study (see Appendix A), in which no reference to perception or pronunciation of sounds was made, participants filled out a language background questionnaire (LBQ) (see Appendix B). The questionnaire provided information on their history of language learning (i.e., the number of years they studied English in primary and secondary education, whether they had attended a bilingual school during primary or secondary education, whether they had enrolled in a language school to study English, whether they were learning a third language (L3) at the moment of the study, and the number of hours of exposure to English outside the classroom). In this questionnaire, participants were also asked to self-rate their levels of English proficiency in speaking, listening, reading and writing skills in a 7-point scale, where 1 corresponded to very low proficiency, and 7 corresponded to advanced proficiency.

In primary school, participants had studied English for a mean of 5.7 years (*SD*= 1.9), whereas in secondary school, they had studied English for a mean of 4 years (*SD*= *.2*). Seven participants had attended bilingual schools during their primary education, whereas five participants had done so during their secondary education. Fifteen participants also stated they had taken communicative English lessons in language schools. All research participants were simultaneously studying an L3 at the moment of the study: 37 were taking a German language course, whereas the remaining 16 were taking French at the same university. When asked about English exposure through television, music and reading, they reported they were exposed to English outside the classroom from two to 56 hours per week (*M*= 15.9; *SD*= 13.1). In summary, the group can be considered to be skilled and experienced language learners.

Regarding participants' self-rating of English proficiency, the mean for speaking skills was 3.8 (*SD*= 1.5); for listening, 4.4 (*SD*= 1.4); for reading 4.8 (*SD*= 1.4); and for writing, 4.4 (*SD*= 1.4).

Data Collection Materials

The materials used in this study included (a) a forced-identification task, (b) a discrimination task, (c) a word-list read-aloud task, and (d) a passage read-aloud task.

The first two tasks were intended to elicit information on participants' perceptual skills, whereas the read-aloud tasks were intended to elicit pronunciation of the target vowel. Each of these is described in detail below.

Forced-identification Task

The forced-identification task (FIT), adapted from Escudero (2000), was intended to find out whether participants were able to identify the English /ɪ/ when they heard either /ɪ/ or /i/. During the experiment, participants heard 20 monosyllabic English pseudowords via a COBY CV18523 headphone connected to a laptop computer, and were asked to decide whether the vowel sound of every "word" they heard was more similar to the vowel sound found in the word *sheep*, i.e., /i/, or *ship*, i.e., /ɪ/. To carry out this task, participants were shown two pictures on a computer screen, one depicting a sheep and another one depicting a ship; they had to click on either image after each aural stimulus. The use of pictures instead of words in this task was intended to reduce orthographic bias (Escudero, 2000; Morrison, 2008). It is worth mentioning that this task was pilot-tested by a native speaker of English (NS). The NS carried out this task in two instances, separated by a three-week interval. In both pilot testing sessions, the NS solved the task with 100% accuracy.

The researcher decided to use pseudowords in order to avoid the possibility of participants' experience with specific words influencing results. Only pseudowords in a CVC or CCVC template were used, of which ten tokens were used to elicit information on their perception of /ɪ/. Table 2 shows the pseudowords used in the first perception task.

Pseudwords with /ɪ/	Pseudwords with /i/
dis	dreef
kib	fleef
bik	deeb
rit	jeet
zif	feek
glik	geef
flid	bleeb
prif	prees
tib	cleed
clib	keet

Table 2: Pseudowords included in the FIT.

The researcher designed this task using PowerPoint 2007 and Visual Basic 6.3 in order to create an auto-executable presentation that could automatically play the sounds and store the participants' names and responses in a log file. All pseudowords were recorded in a laptop computer from a speech synthesizer, VoiceText 3.11.1.0, using an audio program, Audacity 1.3-12 beta. Figure 4 shows a screenshot of the interface of the experiment.

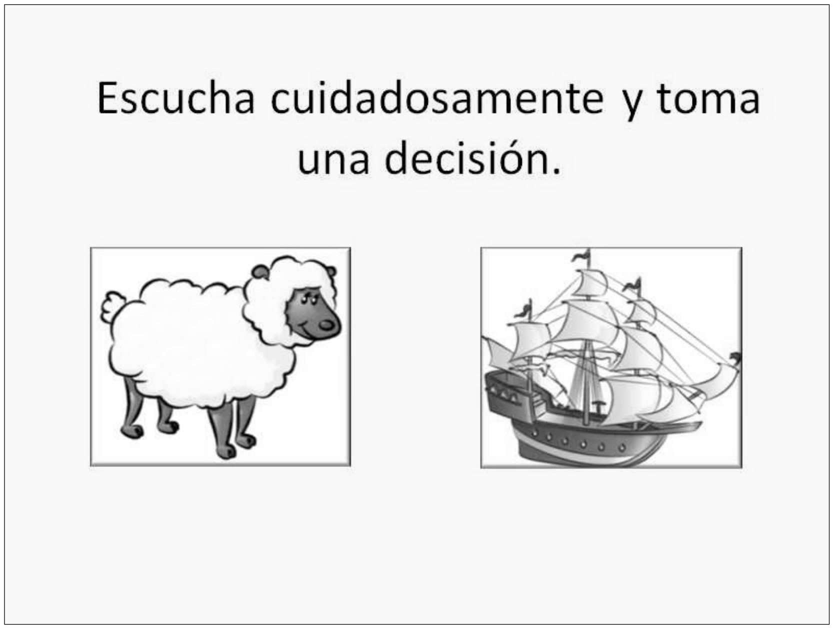

Figure 4: Screenshot of the FIT.

Discrimination Task

The discrimination task (DT) was also adapted from Escudero (2000) and designed using PowerPoint 2007 and Visual Basic 6.3 for the same reasons described in the description of the FIT experiment. The DT was intended to elicit information about the participants' ability to differentiate between the target vowel, /ɪ/, and /i/. In this experiment, participants heard ten pairs of English monosyllabic pseudowords through the same headphone as in the FIT, and they were asked to decide whether the pairs of "words" they heard had the same vowel sound or a different one. Five pairs contained pseudoword with the same vowel sound, i.e., /ɪ/, and five pairs contained both /ɪ / and /i/ (see Table 3). On the screen, there were two large buttons with the words *iguales* (i.e., same) and *diferentes* (i.e., different), and participants were asked to click on either button according to what they heard. Figure 5 shows a screenshot of this task.

Same-vowel pairs	Different-vowel pairs
sid-gis	preeb-drit
mif-fis	crit-creef
nis-vik	pleek-flib
hif-pib	plid-dif
lis-jid	bif-bleef

Table 3: Pseudowords included in the DT.

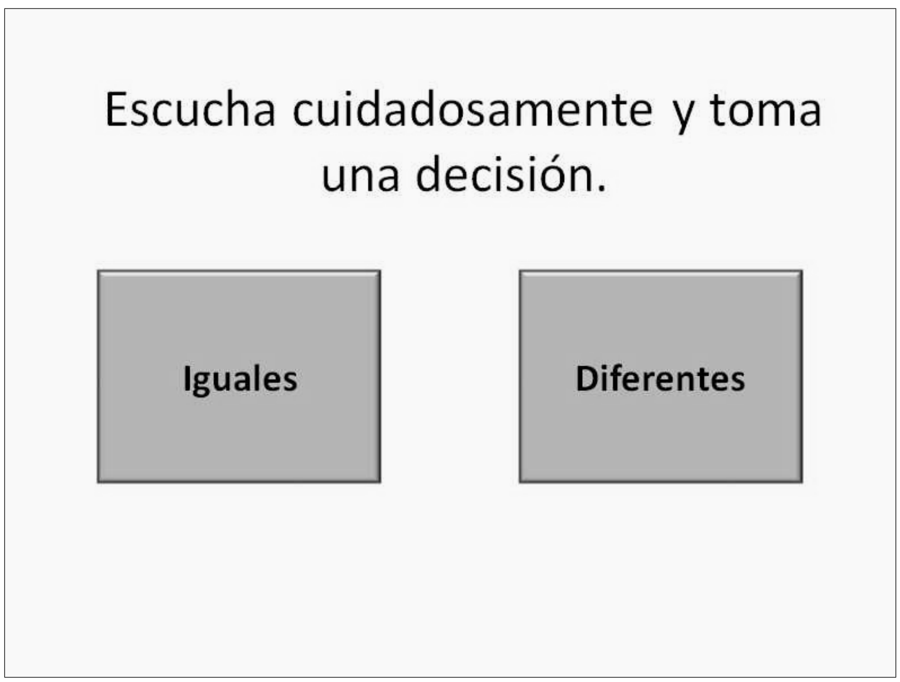

Figure 5: Screenshot of the DT.

The DT was also pilot-tested by the same NS as in the FIT. She solved the task twice, each time separated by a three-week interval. The NS also solved this task with 100% accuracy in both instances.

Word-list Read-aloud Task

The word-list read-aloud task (WLT) was one of the tasks intended to elicit production of the target vowel. Participants were given a list of 20 English monosyllabic words, which included ten tokens for the target vowel, /ɪ/, and ten distractors. They were asked to read the words aloud at a normal pace, leaving a short pause before reading the next one. Participants were audio-recorded using an H4 Zoom digital recorder at a sampling rate of 44.1 kHz and a lavaliere microphone Audio Technica ATQ8531.

Only monosyllabic words following a CVC or CCVC template were selected for this task. This way, the pronunciation of /ɪ/ was less likely to be influenced by phonological factors such as stress and vowel reduction. /s/ + consonant onset clusters were also avoided, since it has previously been observed that Spanish speakers might insert an epenthetic /e/ before such clusters, transforming monosyllabic words into disyllabic ones (e.g., Carlisle, 1998). In order to compensate for participants' potential unfamiliarity with the words in the list, rather commonly-used words were used. Seven out of ten words belonged to the 1000 most common words of English (K1 band), while three belonged to the 2000 most common words of English (K2 band) (Nation & Heatley, 1994). The vocabulary profile of these words was assessed using Web Vocabprofile (Cobb, n.d.). Table 4 shows the words included in the WLT, separated by K bands (see Appendix C for the WLT with all distractors included).

K1 words	K2 words
bill	hit
hill	pick
kill	thick
miss	
sit	
wish	
with	

Table 4: Words included in the WLT separated by K bands.

Passage Read-aloud Task

The passage read-aloud task (PT) was also intended to elicit the production of the target vowel. Nine monosyllabic words containing /ɪ/ were included in the PT. In this task, participants were also asked to read the passage aloud at a normal pace. They were also audio-recorded using the same digital recorder, microphone and sampling rate described above in the context of the WLT task.

The words that contained the target sound also followed a CVC or CCVC template. Again, commonly used words were used in this task. Six tokens belonged to the K1 band, while the remaining three belonged to the K2 band. Similarly, no tokens with /s/ + consonant onset clusters were included in this task. Table 5 shows the words included separated by K bands (see Appendix D for the complete passage).

K1 words	K2 words
big	bit
fill	list
fit	trip
fix	
give	
live	

Table 5: Words included in the PT separated by K bands.

Procedures for Data Collection

The study followed a pre-test/post-test/delayed post-test design and the data collection was carried out during seven consecutive weeks. During week one, participants signed the consent forms, responded to the language background questionnaire (LBQ), and were pre-tested on their perceptual and productive skills of the target vowel, /ɪ/, by means of the above-mentioned tasks. From week two to five, participants underwent the treatment phase. During week five, participants were post-tested with the same tasks as in the pre-test, and finally, in week seven the delayed post-test was carried out. Table 6 shows the timeline of the data collection procedure of the current study.

Week	Procedure
Week 1	Consent form
	LBQ
	Pre-test
Week 2 to 5	Experimental treatments
Week 5	Immediate post-test
Week 7	Delayed post-test

Table 6: Timeline of the data collection procedure.

Pre- and Post-testing Sessions

All participants were pre- and post-tested individually in a quiet room at the Foreign Languages Department at the university where the study took place; each testing lasted approximately 15 minutes. During the pre- and post-tests, participants solved all four tasks described earlier. In order to avoid task ordering effects, the researcher used a counterbalanced design within each group in each of the testing instances, i.e., all participants carried out the tasks in different orders.

Before starting with the forced-identification task (FIT), participants were asked to name the objects in the pictures, i.e., “sheep” and “ship”, to know whether participants made any difference between the vowels in both words. For this purpose, the researcher showed each participant a picture of a sheep and a picture of a ship and asked in Spanish: “What is this?” or “What do you call this in English?”, never mentioning the objects displayed in the pictures. During the pre-test, several students responded “boat” when referring to ship. In that case, the researcher would prompt the student to say the intended target word by saying: "No, that's not a boat. A boat looks different. So what do you call this?" until they produced *ship*. Thirty-seven participants appeared to produce a quantity difference when pronouncing both words, i.e., they produced [ʃip] for ship and [ʃi:p] for sheep, whereas ten participants did not produce the vowels distinctly. Despite the fact that the use of pictures in this task was intended to reduce orthographic bias, the researcher decided to show these participants the written version of the objects in the pictures for them to produce the words and check whether they knew there was a difference between the vowels. After reading the words for the objects, the remaining ten participants made a quantity difference. Once this process was finished, the researcher gave the participants the instructions in Spanish to start the task. It is important to mention that the procedure carried out before participants started this task was the only available option to ensure that participants solved the task properly.

Before starting with the discrimination task, participants were given the instructions by the researcher in Spanish. These were also displayed on the computer screen, together with some examples of how the task was to be completed, also in Spanish.

For both production tasks, i.e., the word-list read-aloud task and the passage read-aloud task, participants were asked to wear a lavaliere microphone in order to be recorded. To avoid physical contact with the participants, the microphone was pinned to a tie, and they were asked to wear and adjust the tie so that the microphone was close enough to their mouths. The instructions for both tasks were written in Spanish on the sheets where the word-list and the passage were printed. The researcher also gave the participants oral instructions to carry out the tasks in Spanish.

When participants completed the last testing session during week seven, they were given a small compensation for their participation in the study. This was a recyclable bag with the Concordia University logo, donated by the Concordia University Libraries to the researcher.

Experimental Treatments

The experimental treatments consisted of four 45-minute sessions once a week, during four consecutive weeks. As mentioned above, the experiment took place during weeks two to five. All lessons took place in two of the computer labs at the Foreign Languages Department of the university and were carried out by the researcher. Each lab was equipped with 20 computers, each one with a headset. In the following three sections, I will provide a detailed explanation of the materials used and the activities conducted by each group.

Materials

TTS application

The group receiving TTS-based instruction (henceforth TTS group), worked with VoiceText 3.11.1.0, (see Figure 6 for a screenshot of the application). This TTS application highlights the text as the machine reads it, so that the listeners see what is being said. It also provides different voices (male and female) that have to be previously installed in the computer where the application will be used. In the current study, three voices were used: Julie, Paul and Kate (sampling rate = 16 kHz). In the setup of the program, the user can adjust the speed, the volume, the pitch of the speech output and also the pauses between sentences and after commas. The licenses of the TTS and the voices for this study were donated by NeoSpeech to the researcher.

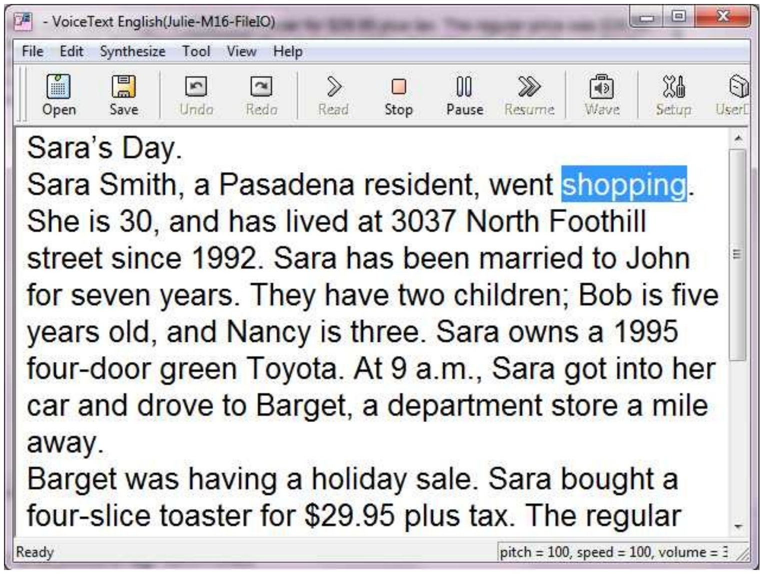

Figure 6: Screenshot of VoiceText 3.11.1.0.

Short stories and follow-up activities

The TTS-based instruction group (TTS group) and the group receiving instruction without the TTS (non-TTS group) worked with four short stories (Appendix E), one per experimental lesson. Three of them were taken and adapted from Chang (2011). A fourth story was created by the researcher. Each story contained monosyllabic words with the target vowel, /ɪ/. The first story contained five target words, whereas the remaining three contained six target words each. See Table 7 for a list of the target words included in the four passages.

Story 1	Story 2	Story 3	Story 4
Smith	with	sing	six
lived	kick	drink	milk
milk	kids	kill	still
since	rip	still	fish
kids	think	miss	bit
	spill	bit	bin

Table 7: Monosyllabic words with /ɪ/ included in the experimental treatments.

The stories were used in the first activity of each lesson. Furthermore, there were seven additional follow-up activities. The first one consisted of five comprehension questions related to the story that the participants worked with. This activity was included as a distractor, so that the participants did not notice that the study was about the /ɪ/ sound. For the second follow-up activity, the researcher selected several words from the passages, all monosyllabic words containing /ɪ/, plus an additional set of monosyllabic words containing /i/ and other vowels, and had the participants listen to them. The third activity was also a listening task, in which the same words selected for the previous task were paired up to have students listen to them. The next activity consisted of having the participants listen to the same pairs as in the previous task and decide whether they had the same or a different vowel sound. For the fifth activity, the researcher used the words used in the two previous activities and grouped them in sets of three words to have participants decide which one had a different vowel sound in each of these sets. In activity six, participants had to write three sentences with any of the words used in activity five. Finally, in activity seven, participants had the original story with gaps and had to complete the missing phrases by listening to the story again. All missing phrases contained a monosyllabic word with the target vowel. A detailed explanation of the way in which the activities were used in each group will be provided in the next two sections (see Appendix F for a sample lesson of the TTS group with all activities included).

Experimental Treatment TTS Group

Each student belonging to the TTS group worked on a computer wearing a headset via which they listened to TTS-based oral output produced by VoiceText. During the first part of the first experimental lesson, the researcher taught the participants how to use the application, i.e., they learned to paste text into the application, change its font size, change the speed of the speech output and play the text. Students were instructed not to change any other features of the application, such as pitch and the length of sentence pauses.

During the four experimental lessons, the participants worked on the stories and the additional follow-up activities described in the previous section. All activities were uploaded to a website where participants could download them in .doc format. Each lesson consisted of having the participants paste the story selected for each lesson in VoiceText and then listen to it three times at a 90% speed rate, using the three voices available, i.e., Julie, Kate and Paul (a different voice each time). They then answered the five comprehension questions related to each story without looking at the written document. After the activity was completed, they were asked to check the answers by rereading the written version of the story. The following activity consisted of having the participants paste and listen three times to the words selected from the passage at a 70% speed rate, each time with a different voice. The participants were then instructed to paste the pairs of words (those previously selected by the researcher) in the application and listen to them three times, using the speed rate they wanted, but using the three voices available each time. In the next activity, participants were asked to leave the previous pairs of words pasted in VoiceText, chose any of the three voices available, set the speed at 70% and then decide whether the vowel of the pairs of words sounded the same or different. For this purpose, the researcher gave them a piece of paper with numbers on it for participants to write “S” (i.e., same) or “D” (i.e., different) for each pair. They were also instructed to listen to the pairs of words as many times as they wanted. After they finished this task, they were asked to continue with the next activity, which consisted on pasting the three-word sets in the TTS application, choosing any of the voices, setting the speed at 70% and deciding which word in the set contained a different vowel sound. Participants had also a piece of paper with numbers on it to write down the ‘odd’ word of each set. The next activity consisted of having the participants write three sentences with any of the words used in the previous activity in the Word document, paste them into VoiceText and listen to them using the three different voices at the speed rate they wanted. Finally, participants had to complete the gapped story by pasting it into the application and minimizing it to complete the gaps in the document. They were instructed to listen to the story for as many times as they wanted until they completed all gaps, using the voice they wanted. After completing all activities, participants were asked to save all changes in the Word documents, compress them and e-mailed them to the researcher.

In sum, the activities carried out by the TTS group were designed with the purpose of enhancing participants' exposure to the target vowel to promote its acquisition in both perception and production.

Experimental Treatment for the non-TTS Group

Participants belonging to the non-TTS group read and completed the same stories and activities as those adopted in the TTS group. The only difference was that the aural input was provided exclusively by the researcher, a fluent speaker of English. The activities were printed, photocopied, and distributed among the participants so that they could each have a hard copy of the story and associated tasks. Throughout each lesson, the researcher read the story aloud three times and had the participants read and listen to it at the same time. In this group, students answered the comprehension questions on their worksheets and they checked their answers with the researcher. For the following three listening and discrimination tasks, the researcher read the words aloud at a slow pace. In the sentence writing task, the researcher picked some of the participants' sentences and read them aloud for the whole class. Finally, in the activity with the gapped story, the researcher read the text as many times as necessary until all participants completed the gaps in their worksheets. It is important to mention that no feedback was provided to the participants' answers. They were told that at the end of the treatment, the researcher would e-mail them the correct answers for each lesson.

The activities conducted by the non-TTS group attempted to emulate the environment that characterized the TTS-based instruction without the use of the TTS.

Lessons in the Control Group

Each lesson in the control group consisted of three activities about different videos available on YouTube.com (one video per lesson). Each video was related to the week lesson they were studying in the regular classes of English (See Table 8 for the topics discussed in each lesson). Every class started with a warm-up activity about the topic that was going to be dealt with in the video. The participants were then asked to watch the video on the computer for a first time to get an idea of the content of the video. After that, they were given six or seven comprehension and vocabulary questions and were then asked to watch the video a second time to answer these questions. Once questions were answered, the researcher asked for volunteers to discuss their answers orally, followed by immediate feedback by the researcher.

Lesson	Topic
Lesson 1	Successful people and entrepreneurs
Lesson 2	Social etiquette
Lesson 3	Different places in the world
Lesson 4	Learning a foreign language

Table 8: Topics discussed in the lessons with the control group.

Analyses

The following section describes the scoring procedures carried out in data obtained in the perception tasks, followed by a detailed description of the randomization, scoring and sound normalization procedures conducted to the speech samples obtained from the two oral production tasks.

Scoring Procedure of Perceptual Tasks

Participants' perception of the target sound in the forced-identification task (FIT) and discrimination task (DT) was automatically coded as 0 or 1 for each token. That is, every correct answer was coded as 1 and every incorrect answer was coded as 0. The maximum score for each task was 10. Recall that in the FIT, participants listened to 20 stimuli, which were categorized as belonging to either "sheep" or "ship". Only those that were to be identified as "ship" were used for the analyses, i.e., 10. In the DT, participants listened to 10 pairs of pseudowords, i.e., 5 with the same vowel and 5 with different vowels. All 10 answers were used for the analyses.

Randomization and Scoring Procedure of Production tasks

In order to prepare the sound files that contained the tokens elicited in the word-list read-aloud task (WLT) and passage read-aloud task (PT) for later rating, the researcher used Praat and Audacity. Recall that the WLT had distractors, which had to be deleted from every sound file generated in each of the tasks for each participant, in every testing instance. For this task, the researcher edited every sound file in Audacity by deleting the distractors from the sound files and by inserting regular silence intervals of 2 seconds after every token. The amplitude of every token was then normalized so that they had similar "volume". It was not necessary to randomize the order of the tokens elicited in this task, since the researcher previously prepared eight different randomization templates for this task to have participants read them in different orders.

The randomization procedure of the tokens elicited in the PT was performed differently because of the nature of the task. Because this instrument consisted of passage reading, it was not possible to have tokens randomized from before participants read the text. The sound files generated from this task were edited using Praat first. All tokens from this task were first marked using the textgrids available in the program. In Praat, the user can create levels below the spectrograms to insert text. These levels are called textgrids. Figure 7 displays a screenshot of Praat that shows, from top to bottom, the oscillogram, the spectrogram and the textgrid with the marked tokens.

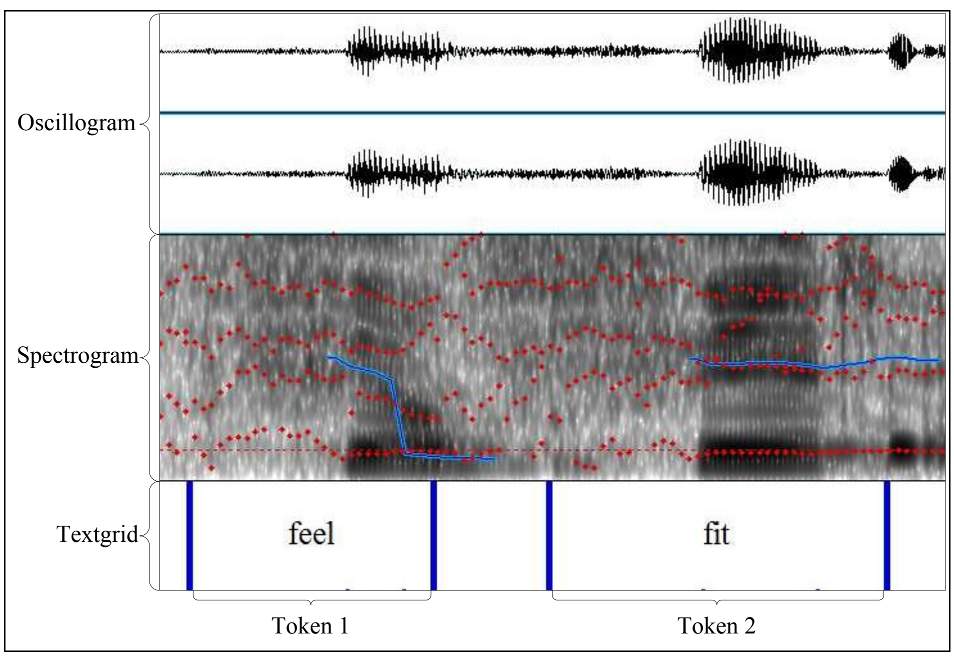

Figure 7: Marking of tokens of PT in Praat.

After all tokens were marked, the researcher used a Praat script to extract each marked token to individual sound files, which were then concatenated in different orders by means of another Praat script. This script created a sound file for each participant with the 9 tokens elicited in this task. The sound files generated by Praat were then edited in Audacity by inserting regular silence intervals of 2 seconds after every token. These tokens were also normalized in amplitude.

2660 tokens were obtained by means of the oral production tasks, i.e., 19 tokens per participant in each of the three tests (pre-, immediate and delayed post-test). One participant from the control group did not attend to the last interview, so it was not possible to obtain her data.

The rating of all tokens was carried out by a non-expert native speaker of English (NS). The NS received a 20-minute training session, in which she was instructed to assign a score of 1 to every target-like instance of /ɪ/ in every token, or a score of 0 to non-target-like instances. She was also instructed to focus her attention on the pronunciation of the vowel of every word and not to pay attention to the pronunciation of the surrounding consonants. The rater had several Excel spreadsheets with the name codes of every participant and the words they pronounced in randomized order. In order to play the sound files, the NS used Express Scribe, a transcription application that allows the user to listen to the sound files and manage the application in the background while writing a document.

The NS rated the tokens in randomized order during 3 sessions of 45 minutes each, spaced by three-day intervals. All rating sessions were carried out with the researcher present in case of any doubt.

In order to calculate inter-rater reliability, a proficient non-native speaker of English rated 10% of the speech samples, which were extracted from all three groups across all tests. The ratings of the tokens included in this 10% were then compared to the same tokens rated by the NS using the Kappa statistic, which was found to be .98 (p = 0.02).

Acoustic Analyses

The researcher conducted acoustic analyses of all the tokens elicited in both oral production tasks, i.e., WLT and PT, using Praat in order to obtain the first three formant values of each token, i.e., F1, F2 and F3. Formants are defined as the resonances in the vocal tract or in the oral cavity (Menke, 2010). In the spectrogram, these resonances are seen as dark bands that concentrate in different ranges of frequencies (see Figure 8). According to Menke, F1 is associated with tongue height, whereas F2 is associated with tongue backness. F3 is also associated to tongue backness to some extent, but usually F1 and F2 are the frequencies taken into account to describe vowels. F3 values were used in this study to allow normalization of the formant frequencies of vowels (see forthcoming discussion). Higher F1 values are associated with a lower tongue position (e.g., /ɪ/ has higher F1 values when compared to /i/ because it is articulated in a lower point in the mouth), and higher F2 values are associated with a more fronted tongue position (e.g., /i/ has higher F2 values than /ɪ/ because it is articulated closer to the palate, i.e., at a higher point in the mouth) (Menke, 2010).

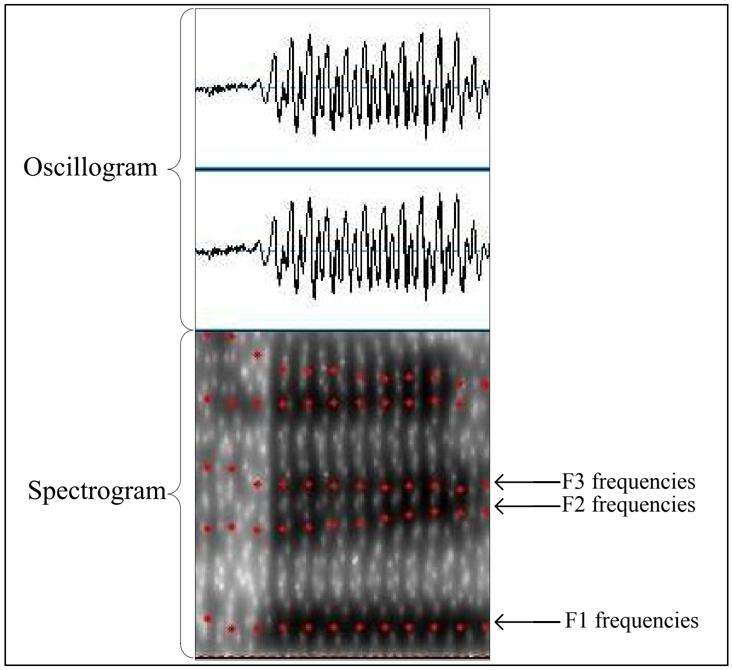

Figure 8: Formants in the spectrogram.

In order to obtain the formant values of vowels only, it was necessary to do so in a two-step process. First, the researcher placed the cursor at the beginning of the vowel, i.e., when the dark bands started and no traces of the preceding sounds were heard) and dragged it until the end of the vowel, i.e., when the dark bands ended and no traces of the following sounds where heard. Second, once the vowel was selected, the researcher put the cursor in the middle point of the selection to obtain the formant values of each vowel at that specific point. See Figure 9 for a visual representation of this procedure with a sample of the word “miss”.

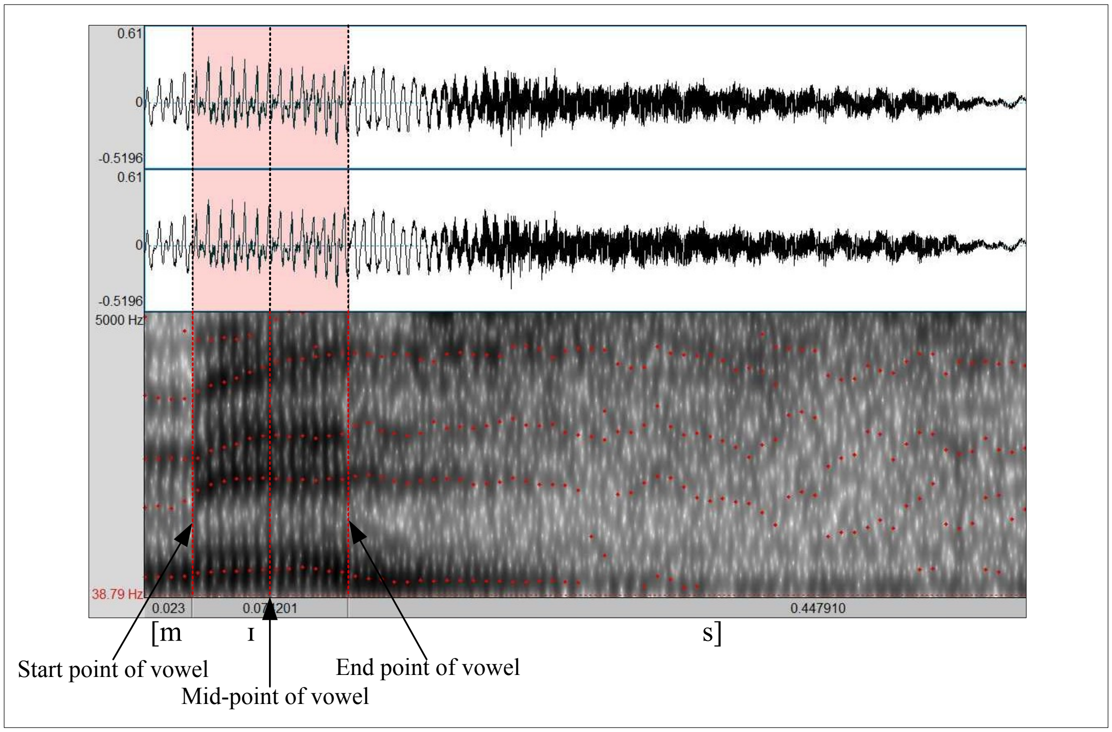

Figure 9: Procedure of the acoustic analysis of /ɪ/.

After the mid-point of the vowel was identified, the researcher used a Praat script to obtain the F1, F2 and F3 values of each token. Since formant frequencies are dependent upon physiological and anatomical characteristics of the vocal tract (Adank, Smits & van Houte, 2004) and therefore, they vary from speaker to speaker, it was necessary to convert them from Hertz to Barks –another measure of frequency– and normalize all formant values. For this purpose, the researcher used NORM (Thomas & Kendall, 2007), which is an on-line application that uses F1, F2, and F3 to make the above mentioned conversion to obtain Z scores: Z_1, Z_2 and Z_3, which correspond to the Bark-converted values of F1, F2 and F3 respectively. The application then calculates the difference between Z_3 and Z_1 (i.e., Z_3- Z_1) to obtain a value for tongue height, and then the difference between Z_3 and Z_2 (i.e., Z_3- Z_2) to obtain a value for tongue backness. These last two scores, i.e., Z_3- Z_1 and Z_3- Z_2 were used for later analyses and plotting.

Operationalization of Variables

Perception

Recall that perception in this study was measured through two tasks, i.e., the forced-identification and the discrimination tasks. Therefore, perception in this study is defined as the ability to identify the target sound as such and to differentiate it when it is heard in one of the pairs of pseudowords, as in the discrimination task. The improvement in these abilities is defined as higher scores in each of the perception tasks when groups are compared to each other in each post-test.

Production

Production was also measured by means of two tasks, i.e., the word-list read-aloud task and the passage read-aloud task. In this study, correct oral production is defined as the ability to appropriately pronounce the target sound in two different contexts as rated by the NS. The improvements in pronunciation is defined as higher scores in both tasks according to the NS rating, and both Z_3-Z_1 and Z_3-Z_2 scores.

This chapter described the methodology adopted to conduct the present study. In the next chapter, the results obtained by the groups in the four tasks in the pre-, immediate and delayed post-tests will be presented.

CHAPTER 4

RESULTS

This chapter presents the results of the study, which were analyzed by means of descriptive as well as inferential statistics. The first part contains the results from the pre-test, and the second part presents the results for the research questions posed in Chapter 2.

Pre-test

In order to know whether groups had any difference in performance at the outset of the experiment, the researcher conducted eight one-way ANOVAs to the scores obtained in both perception tasks and the scores obtained in the production tasks using the PASW Statistics package. Recall that for each production task, there were three sets of scores (i.e., NS ratings plus Z_3-Z_1 scores for tongue height and Z_3-Z_2 scores for tongue backness/frontness).

The one-way ANOVA conducted to the scores obtained in the forced-identification task yielded no significant differences between groups [$F(2, 44) = 1.39$, $p = .25$]. Similarly, no significant differences were found between groups for the mean scores obtained in the discrimination task [$F(2, 44) = 1.57$, $p = .21$].

The means obtained by the NS rating in the word-list read-aloud task revealed no significant differences between groups [$F(2, 44) = 4.76$, $p = .62$]. No significant differences were found for the Z_3-Z_1 scores between groups [$F(2, 44) = .77$, $p = .46$] or Z_3-Z_2 scores [$F(2, 44) = .02$, $p = .98$] in this task.

The one-way ANOVA conducted to the NS rating scores obtained by all subjects in the passage read-aloud task revealed no significant differences between groups [$F(2, 44) = .54$, $p = .58$], nor did the Z_3-Z_1 scores [$F(2, 44) = 1.36$, $p = .26$] or Z_3-Z_2 scores [$F(2, 44) = .00$, $p = .99$].

These results indicate that the three groups included in the study were very similar with respect to the perception and production of /ɪ/ in all tasks at the outset of the experiment.

Research Question 1

In order to know whether the TTS group showed a greater improvement in the perception of English /ɪ/ than the non-TTS and control groups after the experimental treatment, several analyses were carried out on the perception scores obtained both in the forced-identification task and the discrimination task. In the first part, I will present the descriptive statistics results, followed by the second part where I will provide the inferential statistics results for each task.

Descriptive statistics results

For the forced-identification task, the mean for the TTS group in the pre-test was 6.47 (SD=3.06), while in the immediate and delayed post-test, the mean scores were higher (M=7.94, SD= 2.72 and M=8.47, SD=2.26 respectively).

The non-TTS group had a mean score of 7.94 (SD=1.69) in the pre-test. The means in the immediate post-test were slightly lower (M=7.75, SD=2.23), while in the delayed-post test, the mean scores were higher than in the pre-test and immediate post-test (M= 8.12, SD=2.36).

The control group had a mean of 6.50 (*SD*=3.50) in the pre-test for the forced-identification task. In the immediate post-test, this score was slightly lower and higher in the delayed post-test (*M*=6.21, *SD*= 3.82 and *M*=8.15, *SD*=2.85 respectively).

Figure 10 shows a graph with the mean scores obtained in the forced-identification task by all three groups in all testing instances, together with the standard deviations, signalled by the T-shaped lines on the bars.

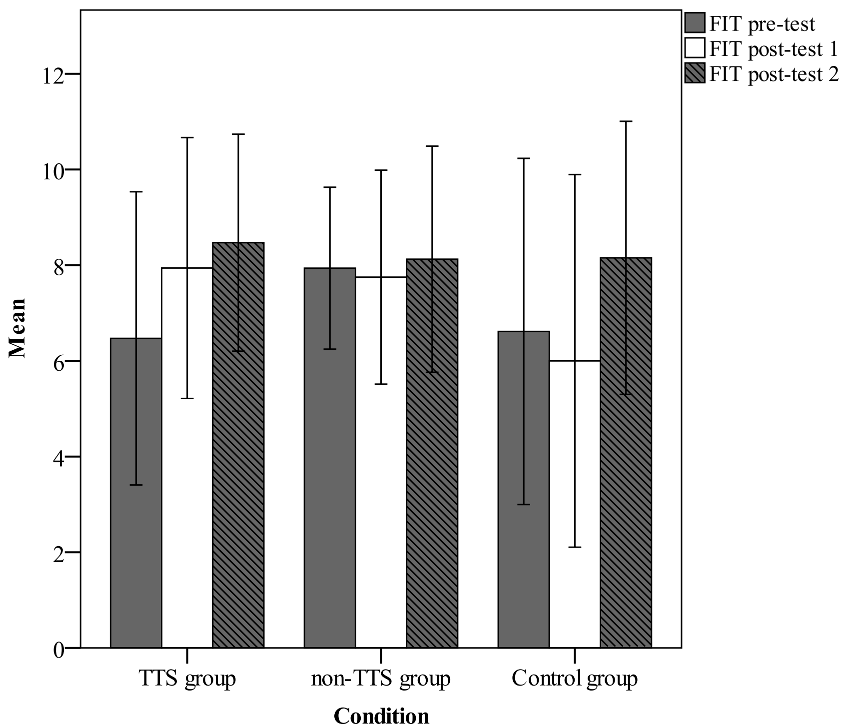

Figure 10: Mean scores and standard deviations of the forced-identification task.

In the second perception task, i.e., the discrimination task, the TTS group had a mean score of 7.29 (*SD*=3.06) in the pre-test, while in the immediate and delayed post-tests, the means were 8.65 (*SD*=1.96) and 7.58 (*SD*=1.46) respectively.

The mean score of the non-TTS group in the pre-test was 7.44 (*SD*=1.26), while in the immediate post-test, the mean was 8.44 (*SD*=1.26). In the delayed post-test, the mean of the non-TTS group was 7.56 (*SD*=1.59).

The control group had an initial mean score of 6.50 (*SD*=3.5) in the discrimination task, and a mean of 8 (*SD*=1.3) in the immediate post-test. In the delayed post-test, their mean score was 7.46 (*SD*=1.76).

Figure 11 shows a bar graph with the means obtained by the three groups in the discrimination task together with their standard deviations.

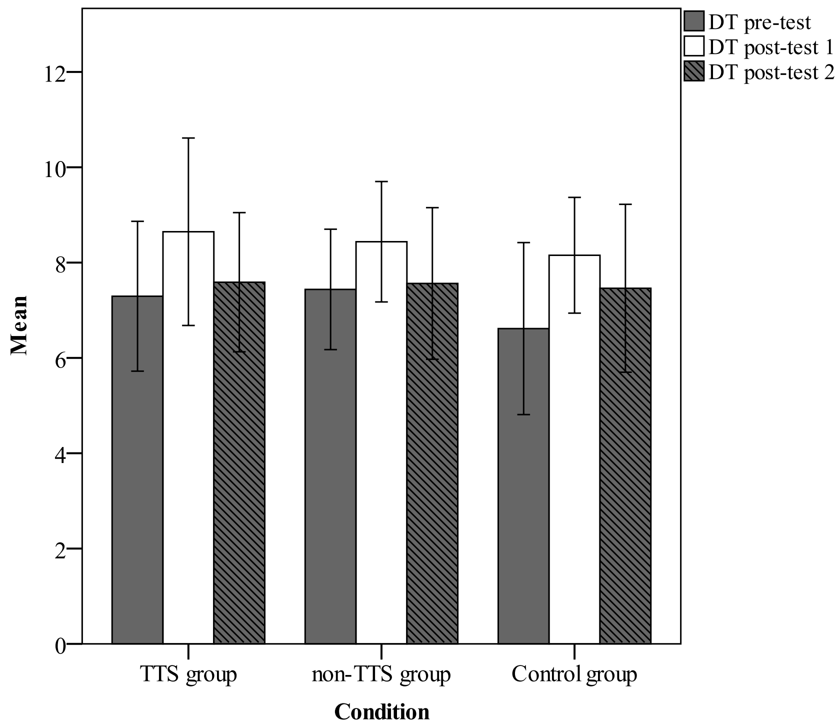

Figure 11: Means and standard deviations of the discrimination task.

These descriptive results indicate that the TTS group was the only one that showed an improvement in their ability to identify the target sound over time, whereas the other two groups showed a decrease in performance followed by an improvement –as showed by the forced-identification task results (shown in Figure 10). In terms of their ability to differentiate the target sound with /i/, as measured with the discrimination task, all groups showed a similar performance pattern, i.e., an improvement in the immediate post-test followed by a decrease in the delayed post-test (shown in Figure 11).

Inferential statistics results

In order to find out whether there were significant differences between groups in all testing instances in the perception tasks, two mixed factorial ANOVAs were carried out on the scores obtained by means of the forced-identification task and the discrimination task. Two factors were used for these analyses. The within-subject factor was Time, with three levels (i.e., pre-test, immediate post-test and delayed post-test) and the between-subject factor was Group, with three levels as well (TTS group, non-TTS group and control group). The alpha level was adjusted to p =.006 to fit the number of tests conducted to the data (i.e., eight tests in total) and avoid declaring statistically significant differences when not appropriate. Effect sizes were also calculated using eta squared (η^2). According to Cohen (1988), $\eta^2 = .01$ is a small effect size, $\eta^2 = .06$ is a moderate effect, whereas $\eta^2 = .14$ or greater is a large effect size. It is important to mention that the ANOVAs excluded all data from the participant belonging to the control group who missed the delayed post-test from all analyses; therefore, the data from 46 participants was used (TTS group: 17 participants; non-TTS: 16 participants; control group: 13 participants).

The ANOVA conducted to the scores obtained in the forced-identification task revealed no significant effects for Time [$F(2, 86) = 3.10$, $p = .05$, η^2=.06] or for the Time X Group interaction [$F(2, 86) = 1.21$, $p = .30$, η^2=.05]. The mixed factorial ANOVA conducted to the mean scores obtained in the discrimination task revealed a significant effect of Time [$F(2, 86) = 9.03$, $p < .001$, η^2=.17], but no significant effects for the interaction Time X Group were found [$F(2, 86) = .273$, p =.89, η^2=.01]. Post hoc pairwise comparisons of the level of Time using the Bonferroni correction revealed that there was a statistically significant difference between the scores obtained by all participants in the pre-test (*M*=7.11, SD=1.56) and those obtained in the immediate post-test (*M*=8.38, *SD*=1.55) for the discrimination task.

In sum, these results indicate that there were no statistically significant differences between the groups in terms of their ability to identify and differentiate English /ɪ/ as shown by the scores obtained both in the forced-identification and discrimination tasks.

Research Question 2

In order to know whether the TTS group showed a greater improvement in the pronunciation of English /ɪ/ than the non-TTS and control groups after the experimental treatment, several analyses were carried out on the scores obtained both in the word-list read-aloud task and the passage read-aloud task. As with research question 1, in the first part I will present descriptive statistics results, while the second part will consist of inferential statistics results for each task.

Descriptive statistics results

NS ratings

In the word-list read-aloud task, the mean of the TTS group according to the NS ratings was 5.41 (*SD*=2.34) in the pre-test, whereas in the immediate post-test the mean was 5.76 (*SD*=2.19). In the delayed post-test, the TTS group obtained a mean of 7.82 (*SD*=2.27).

In the same task, the non-TTS group obtained a mean of 5.94 (*SD*=3.23) in the pre-test. In the immediate and delayed post-test the means were 6.06 (*SD*=2.93) and 7.50 (*SD*=3.22) respectively.

The mean score of the control group in the pre-test of this task was 4.86 (*SD*=3.48), while in the immediate post-test, the mean score was 4.21 (*SD*=2.51). The mean of this group in the delayed post-test was 6.46 (*SD*=2.66).

Figure 12 shows a graph with the mean scores obtained in the passage read-aloud task by the three groups in all testing instances, together with the standard deviations, signalled by the T-shaped lines on the bars.

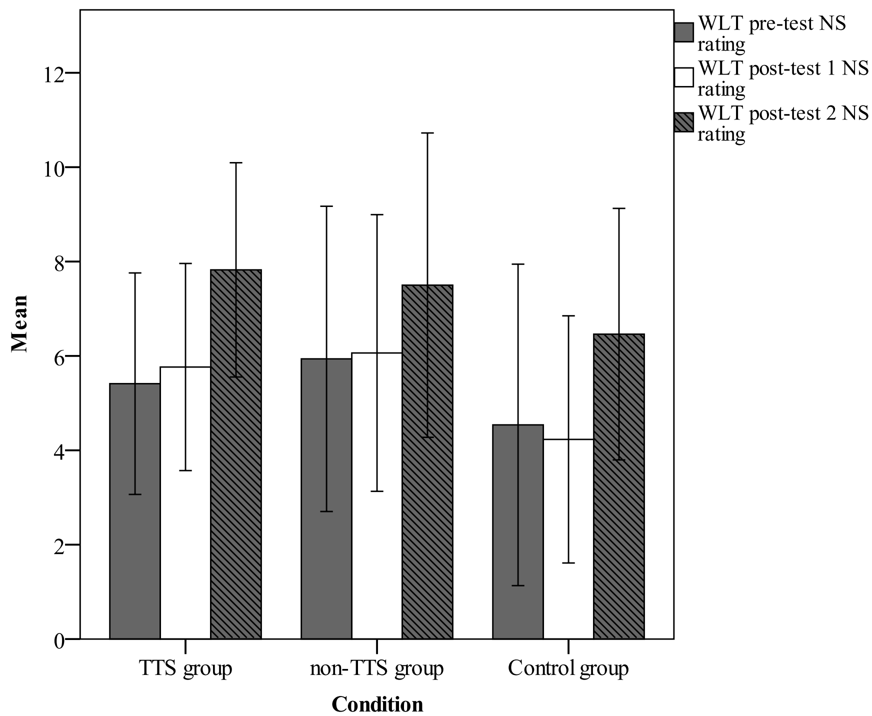

Figure 12: Means and standard deviations of the word-list read-aloud task.

According to the NS ratings of the passage read-aloud task, the mean for the TTS group in the pre-test was 5.53 (*SD*=1.80), while in the immediate and delayed post-test, the mean scores were 7.35 (*SD*= 1.65) and 4.17 (*SD*=2.26) respectively.

The non-TTS group had a mean score of 6.06 (*SD*=2.26) in the pre-test for this task. Their mean scores for the immediate and delayed post-tests were 5.06 (*SD*=1.94) and 3.06 (*SD*=2.51) respectively.

The mean score of the control group in the pre-test for this task was 6.21 (*SD*=1.71), while in the immediate post-test, their mean was 5.71 (*SD*=2.01). In the delayed post-test, the control group had a mean of 2.79 (*SD*=2.08).

Figure 13 shows a graph with the mean scores obtained in the passage read-aloud task by the three groups in all testing instances, together with the standard deviations, signalled by the T-shaped lines on the bars.

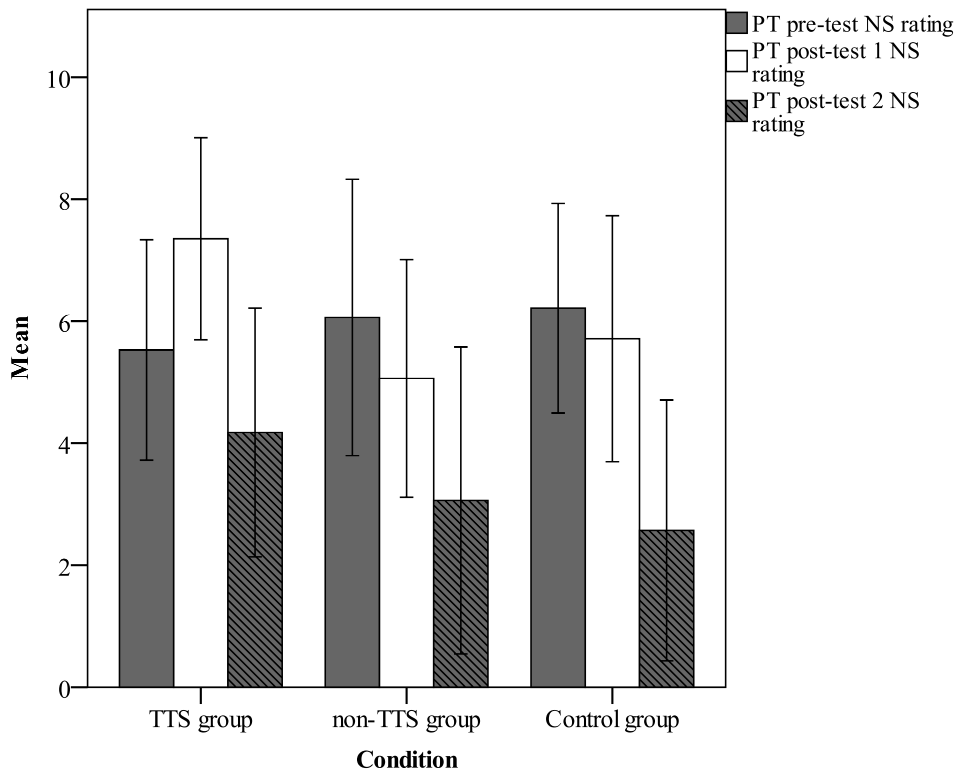

Figure 13: Means and standard deviations of the passage read-aloud task.

The results from the word-list read-aloud task show that the TTS and non-TTS groups improved their production of /ɪ/ overtime, whereas the control group showed a decrease in performance followed by an improvement. However, group performance in the passage read-aloud task was different. Despite the fact that the TTS group was the only one that showed an improvement in the immediate post-test, all three groups had a poorer performance in the delayed post-test.

Z_3-Z_1 and Z_3-Z_2 scores

The normalized formant values, i.e., Z scores, were also analyzed descriptively. It is important to mention that the decline in both Z_3-Z_1 and Z_3-Z_2 scores can be regarded as improvement from a higher and more fronted articulation of /ɪ/ to a lower and more retracted one. Therefore, lower Z_3-Z_1 values are associated with a decrease in tongue height, while lower Z_3-Z_2 values are associated with a more retracted tongue position. In the following paragraphs, the results of the word-list read-aloud task will be displayed first, followed by the results obtained in the passage read-aloud task.

In the word-list read-aloud task, the TTS group had a mean score of 11.20 (*SD*=.52) for the Z_3-Z_1 scores and a mean of 1.45 (*SD*=.39) for the Z_3-Z_2 scores in the pre-test. In the immediate post-test, the mean for Z_3-Z_1 was 11.10 (*SD*=.54), while that of Z_3-Z_2 was 1.54 (*SD*=.43). Finally, in the delayed post-test, their mean for the Z_3-Z_1 scores was 11.04 (*SD*=.52) and for the Z_3-Z_2 scores, 1.53 (*SD*=.35).

In the pre-test, the non-TTS group had a mean of 11.45 (*SD*=.74) for the Z_3-Z_1 values, and 1.46 (*SD*=.32) for the Z_3-Z_2 scores. In the immediate post-test, the mean for the Z_3-Z_1 scores was 11.35 (*SD*=.74), while the mean for the Z_3-Z_2 scores was 1.43 (*SD*=.32). The means in the delayed post-test were 11.30 (*SD*=.85) and 1.54 (*SD*=.33) for the Z_3-Z_1 and Z_3-Z_2 values respectively.

In the pre-test for the word-list read-aloud task, the control group had mean scores of 11.34 (*SD*=.59) and 1.47 (*SD*=.27) for the Z_3-Z_1 and Z_3-Z_2 values respectively. In the immediate post-test, the mean for Z_3-Z_1 was 11.37 (*SD*=.44), while that of Z_3-Z_2 was 1.77 (*SD*=.66). In the delayed post-test, their mean for the Z_3-Z_1 scores was 11.54 (*SD*=.64) and for the Z_3-Z_2 scores, 1.50 (*SD*=.35).

The descriptive results of the Z_3-Z_1 and Z_3-Z_2 scores obtained in the word-list read-aloud task indicate that the TTS group had a better performance in terms of tongue height and backness in the immediate post-test, whereas in the delayed they only improved in terms of tongue height. The non-TTS group improved in both tongue height and backness over time, whereas the control group had a poorer performance in terms of tongue height in both post-tests. In terms of tongue backness, the control group improved in the immediate post-test, but these gains were not retained in the delayed post-test. Figure 14 shows a plot summarizing the results of the Z_3-Z_1 and Z_3-Z_2 mean scores obtained by the three groups in the word-list read-aloud task. This plot represents the articulatory space, so the scores in the left side show tongue height, and those at the bottom show tongue backness/frontness. In order to have a point of reference to interpret the values obtained by the groups in this task, the Z scores obtained by a native speaker of English in this task have been included as reference.

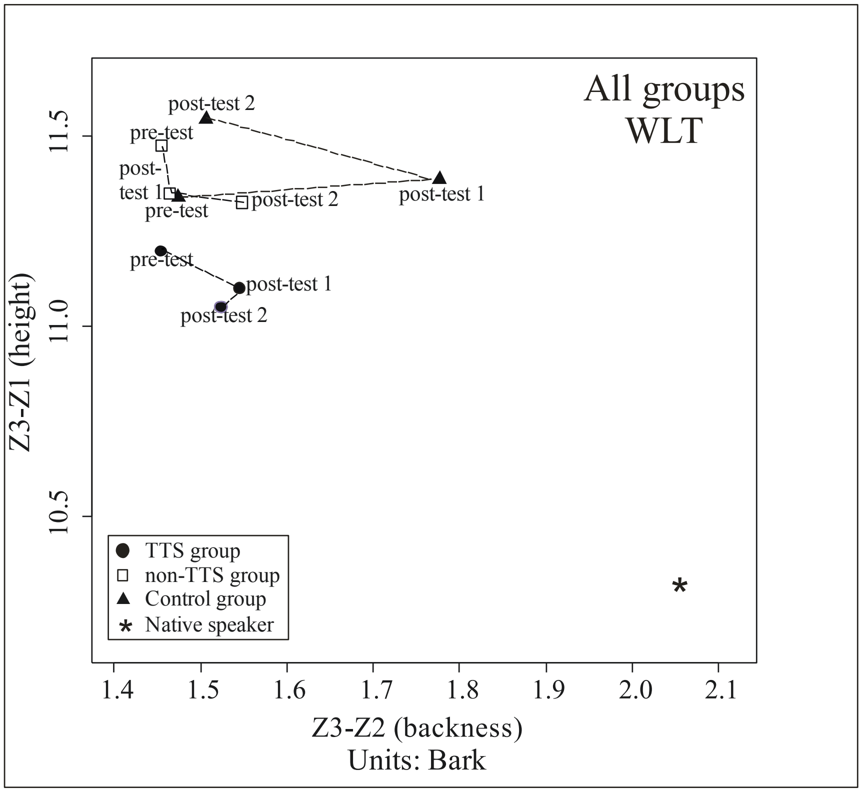

Figure 14: Plot of Z_3-Z_1 and Z_3-Z_2 mean scores in the word-list read-aloud task.

In the passage read-aloud task, the TTS group had a mean score of 11.03 (*SD*=.43) for the Z_3-Z_1 scores and a mean of 2.49 (*SD*=.33) for the Z_3-Z_2 scores in the pre-test. In the immediate post-test, the mean for Z_3-Z_1 was 10.82 (*SD*=.36), while that of Z_3-Z_2 was 1.50 (*SD*=.35). Their mean for the Z_3-Z_1 scores in the delayed post-test was 11.02 (*SD*=.43) and for the Z_3-Z_2 scores, 1.60 (*SD*=.28).

The non-TTS group had a mean of 11.29 (*SD*=.60) for the Z_3-Z_1 values, and 2.50 (*SD*=.32) for the Z_3-Z_2 scores in the pre-test. In the immediate post-test, the mean for the Z_3-Z_1 scores was 11.23 (*SD*=.62), while the mean for the Z_3-Z_2 scores was 1.36 (*SD*=.30). In the delayed post-test, the means were 11.35 (*SD*=.60) and 1.51 (*SD*=.33) for the Z_3-Z_1 and Z_3-Z_2 values respectively.

In the pre-test, the control group had mean scores of 11.31 (*SD*=.51) and 2.51 (*SD*=.35) for the Z_3-Z_1 and Z_3-Z_2 values respectively in the passage read-aloud task. In the immediate post-test, the mean for Z_3-Z_1 was 11.30 (*SD*=.51), while that of Z_3-Z_2 scores was 1.77 (*SD*=.82). Their mean for the Z_3-Z_1 scores in the delayed post-test was 11.39 (*SD*=.42) and for the Z_3-Z_2 scores, 1.76 (*SD*=.54).

For the passage read-aloud task, the descriptive results of the Z_3-Z_1 and Z_3-Z_2 scores show that the TTS group had a better performance in terms of tongue height in the immediate post-test, whereas in the delayed they only improved in terms of tongue backness. The non-TTS group improved in terms of tongue height from the pre-test to the immediate post-test, but these gains were not retained in the delayed post-test. In terms of backness, the non-TTS group showed a better performance only in the delayed post-test. Finally, the control group only showed an improvement of tongue backness overtime. Figure 15 shows a plot summarizing the results of the Z_3-Z_1 and Z_3-Z_2 mean scores obtained by the three groups in the passage read-aloud task. As in figure 14, the Z scores obtained by a native speaker of English in this task have been included as reference.

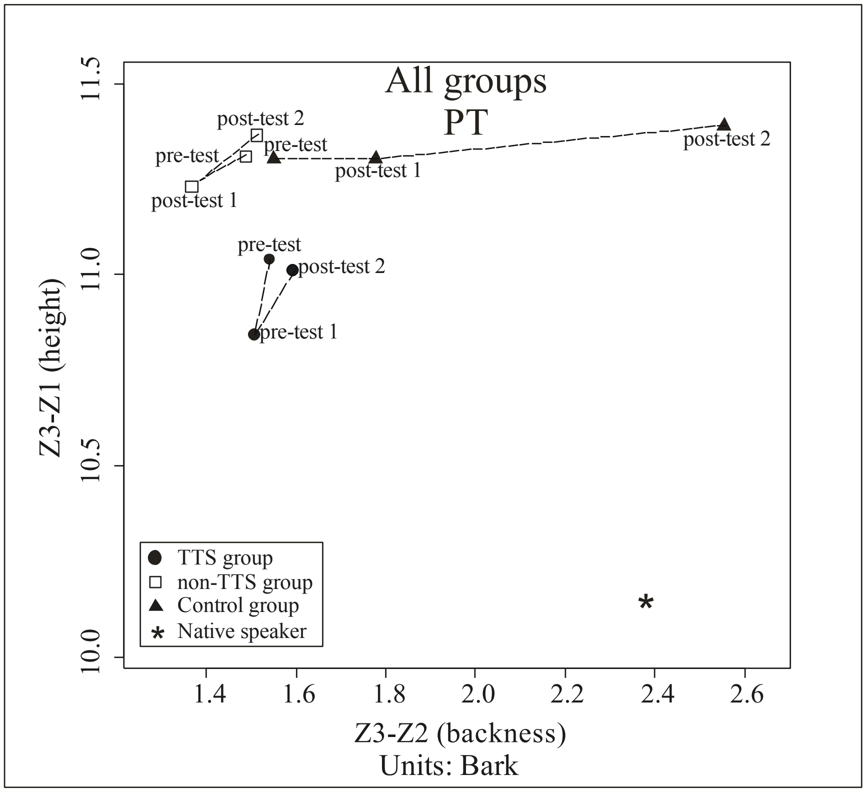

Figure 15: Plot of Z_3-Z_1 and Z_3-Z_2 mean scores in the passage read-aloud task.

Inferential statistics results

NS ratings

To find out whether there were significant differences between groups in all testing instances in the pronunciation of /ɪ/ in both tasks according to the NS rating, two mixed factorial ANOVAs were carried out on the scores obtained in the word-list read-aloud task and the passage read-aloud task. As in the perception tasks analyses, two factors were used. The between-subject factor was Time, with three levels (i.e., pre-test, immediate post-test and delayed post-test) and the within-subject factor was Group, with three levels as well (TTS group, non-TTS group and control group). As mentioned at the outset of the Results section, the alpha level was adjusted to p = .006.

For the word-list read-aloud task, the mixed factorial ANOVA revealed a significant effect of the factor Time [$F(2, 86) = 17.32$, $p < .001$, $\eta^2=.28$], but no significant effects of the Time X Group interaction [$F(2, 86) = .36$, $p =.83$, $\eta^2=.01$]. Post hoc pairwise comparisons for the levels of time using the Bonferroni correction revealed that there was a significant difference between the scores obtained by all participants in the pre-test (M=5.43, SD=2.99) and the delayed post-test (M=7.33, SD=2.74), and between the immediate post-test (M=5.40, SD=2.62) and the delayed post-test.

The ANOVA conducted to the scores obtained in the passage read-aloud task revealed a significant effect for Time [$F(2, 86) = 44.38$, $p < .001$, $\eta^2=.45$] and for the Time X Group interaction [$F(2, 86) = 4.27$, $p = .003$, $\eta^2=.08$]. Bonferroni-adjusted post hoc pairwise comparisons for the levels of time revealed that there was a statistically significant difference between the scores obtained by all participants in the pre-test (M=5.91, SD=1.93) and those obtained in the delayed post-test (M=3.39, SD=2.26). For the Time X Group interaction, Bonferroni-adjusted post hoc pairwise comparisons showed that there was a significant difference between the mean scores obtained by the TTS group (M=7.35, SD=1.65) and the non-TTS group (M=5.06, SD=1.94) in the immediate post-test.

In sum, these results indicate that there were no statistically significant differences between the groups in terms of production of /ɪ/ in the word-list read-aloud task. However, in the passage read-aloud task, the TTS group significantly outperformed the non-TTS group in the production of /ɪ/ in the immediate post-test.

Z_3-Z_1 and Z_3-Z_2 scores

The ANOVA conducted to the mean Z_3-Z_1 scores obtained in the word-list read-aloud task showed no significant effects for Time [$F(2, 86) = .71$, $p = .49$, $\eta^2=.01$] or for the Time X Group interaction [$F(2, 86) = 1.58$, $p = .18$, $\eta^2=.06$]. Similarly, no effects for Time [$F(2, 86) = 2.59$, $p = .08$, $\eta^2=.05$] or for the Time X Group interaction [$F(2, 86) = 1.83$, $p = .12$, $\eta^2=.07$] were found in the mean Z_3-Z_2 scores obtained in this task.

For the mean Z_3-Z_1 scores of the passage read-aloud task, the ANOVA also revealed no significant effects for Time [$F(2, 86) = 2.49$, $p = .08$, $\eta^2=.05$] or for the Time X Group interaction [$F(2, 86) = .50$, $p = .73$, $\eta^2=.02$]. A significant effect for Time was found in the ANOVA conducted to the Z_3-Z_2 scores obtained in this task [$F(2, 86) = 90.64$, $p < .001$, $\eta^2=.66$]. Bonferroni-adjusted post hoc pairwise comparisons for the levels of time revealed significant differences between Z_3-Z_2 scores obtained by all participants in the pre-test (M=2.50, SD=.05) and immediate post-test (M=1.56, SD=.07), and between those obtained in the pre-test and delayed post-test (M=1.62, SD=.05). Finally, no significant effects were found for the Time X Group interaction in the analyses [$F(2, 86) = 1.22$, $p = .30$, $\eta^2=.01$].

To summarize, these results indicate that the three groups did not perform significantly differently from each other in production in terms of tongue height and backness in both post-tests.

In this chapter, the results obtained from the perception and production tasks of the study were presented. In the next chapter, the results of the study and their implications will be discussed in relation to the research questions posed in Chapter 2 and the related hypotheses. Their relationship to the relevant theory and previous studies will also be addressed.

CHAPTER 5

DISCUSSION

As mentioned at the outset of this thesis, the assumption that learners have difficulties with the acquisition of L2 phonology because they perceptually filter L2 sounds in the aural input through their L1 is widely accepted in the field of SLA. However, L2 learners sometimes fail to perceive the differences between L1 and L2 segments and, as a result, they are unable to form new categories in their phonological repertoires. In this context, the concept of perceptual salience is relevant because learners might be able to establish differences between L2 and L1 sounds if they are perceptually prominent in the L2 input. Some researchers (e.g., Chapelle, 2003; Collins et al., 2009) suggest that multimedia environments in which learners are exposed to texts, images and sounds are beneficial because the language input can be manipulated so that learners become aware of the challenging forms. Consequently, this may have an effect on the acquisition process. The aim of this study was to explore the extent to which pedagogical instruction using TTS technology as a means to enhance the target form in the aural input may assist learners in the L2 acquisition of /ɪ/ by Hispanophone learners of English. In Chapter 2, two research questions were proposed: (a) will a group receiving TTS-based instruction show a greater improvement in the perception of English /ɪ/ than a group receiving instruction without a TTS and a control group? and (b) will a group receiving TTS-based instruction show a greater improvement in the production of English /ɪ/ than a group receiving instruction without a TTS and a control group? The hypotheses formulated predicted that the TTS group would significantly outperform the non-TTS group and the control group both in terms of perception and production. In the following discussion, I will start by summarizing the results presented in the previous chapter and I will then address these results with respect to the hypotheses proposed and in relation to the relevant theory and previous studies. The limitations of the study together with suggestions for future research and its implications for ESL teaching will also be addressed.

Summary of Results

Perception of English /ɪ/

In the forced-identification task, the results revealed that only the TTS group had a better performance in both post-tests when compared to the results obtained by the other groups. However, this performance was not significantly different from that of the non-TTS and the control group. Interestingly, the results of the delayed post-test showed that all three groups had a better performance than in previous testing instances, i.e., pre-test and immediate post-test.

In the discrimination task, all three groups behaved similarly, i.e., all participants showed an improvement from the pre-test to the immediate post-test and a decline in the delayed post-test. The comparisons between groups yielded no significant differences, but all groups performed significantly better in the immediate post-test than in the pre-test.

Production of English /ɪ/

The results obtained from the NS rating of the word-list read-aloud task revealed that learners in the TTS group improved their pronunciation of the target sound over time. However, their performance in the post-tests did not significantly differ from the other groups. Significant improvements were found for the scores obtained by all participants in the immediate post-test and delayed post-test when compared to their performance in the pre-test. The results of the normalized formant values revealed that when pronouncing the target sound, the TTS group improved in terms of tongue height and frontness from the pre-test to the immediate post-test, and continued to improve in terms of tongue height –but not in terms of frontness– from the immediate to the delayed post-test. However, no significant differences were found in the post-tests for the scores for tongue height and frontness between groups.

The results obtained from the NS rating of the passage read-aloud task showed that the TTS group significantly outperformed the non-TTS group in the immediate post-test. However, in the delayed post-test, no significant differences between groups were found. The results of the scores for tongue height and tongue frontness revealed that the TTS group outperformed the other groups only in terms of tongue height in the post-tests; however, their performance was not statistically different from the other groups. Significant improvements were found in terms of tongue frontness for the scores obtained by all participants in the delayed post-test when compared to those obtained in the pre-test.

Discussion of Results

Despite the fact that the TTS group showed an improvement in one of the perception tasks, the results indicate that the proposed hypothesis was not borne out. That is, the TTS group did not significantly outperform the other groups in terms of perception. However, in pronunciation, results were not so straightforward. According to the NS ratings, the TTS group did significantly outperform the non-TTS group in one of the pronunciation tasks; therefore, the proposed hypothesis was partially borne out. However, these results should be taken with caution for two reasons. First, the effect size of the experimental treatment was smaller than the effect size of the Time factor, and second, the TTS group did not significantly outperform the control group. This could indicate that the input provided by the TTS may be more effective than teacher input to assist learners with the acquisition of L2 pronunciation, but this does not seem to be entirely true: the control group, which was not subjected to the experimental treatment, was not significantly different from the TTS group.

In some of the tasks (forced-identification, word-list read-aloud [NS ratings] and passage read-aloud [scores for tongue frontness]), all participants showed a better performance in the delayed post-test. These results suggest that it may be possible for L2 learners to overcome the so-called "L1 filter" in phonology acquisition, and therefore perceive and produce L2 sounds with higher accuracy. However, these results also suggest that this improvement cannot be attributed to the experimental treatments to which groups were subjected, but they may be the results of learners' experience with the L2. This experience might have led them to better identify and pronounce the target sound and behave similarly in the delayed-post test. If this were the case, the results would resemble those of Flege et al. (1997), in which experienced learners encountered less difficulties to perceive and produce English vowels. Recall that the learners in this study attended to their regular English lessons during eight hours a week and had accumulated 56 hours of instruction when the delayed post-test took place, suggesting that their ability to identify and pronounce the target sound might have improved as a direct consequence of their experience with the L2.

From these results, it is also possible to infer that input manipulated by means of the TTS did not play such an important role and, therefore, it did not significantly influence learners' performance on these tasks. Despite the fact that the aural input was made prominent by the TTS application, it seems that learners did not focus their entire attention in the target forms during the experimental treatment. This goes along with what Ellis (2006) states about perceptual salience. He states that this prominence is associated with a more subjective experience with a stimulus, so what might be salient for an individual might not be salient for others. If a stimulus is not prominent, it seems logical to infer that learners will not pay attention to it. Researchers such as Schmidt (1990, 1993) and Leow (2007) suggest that it is not possible to learn an L2 feature if there is no attention involved in the process. Learners in the TTS group might have failed to focus their attention on the target sound and perceive its phonetic properties during the treatment, despite the fact that the input to which these groups were exposed was highlighted, and that most of the tasks carried out were intended to improve their ability to perceive the target sound. Previous studies in grammar acquisition have shown that learners who notice L2 features can actually acquire them (e.g., Leow, 1997, 2000; Rosa & O'Neill, 1999). In order to make learners notice the features of /ɪ/, they may need more instruction time with more focused noticing activities in which students are clearly shown the target form. The fact of only increasing the input and making it available to the learner does not seem to be enough for learning.

Another explanation for these results relates to the notion of form-focused instruction. Recall that learners in the experimental groups were not taught how the target sound differed from other confounding sounds such as /i/, or how it was articulated. The fact of explicitly directing learners' attention to the features of the target sound could have helped them improve their ability to perceive and pronounce English /ɪ/, as previous studies in L2 phonology have shown (e.g., Cenoz & García Lecumberri, 1999; García Pérez, 2003; Saito & Lyster, in press). This idea is also supported by some applied linguists (e.g., Celce-Murcia et al., 1996), who suggest that learners should be provided with description and articulatory features of foreign sounds so that they can better acquire a new L2 phonology.

An interesting trend that is present in all the results is the improvement in performance of the TTS group from the pre-test to the immediate post-test, which was not always the case for the non-TTS group and control group. Although results do not completely support the hypotheses proposed for both perception and production as they were not always significant, it may be possible to conclude that exposure to enhanced input via TTS could have a significantly positive influence on learners' acquisition of L2 phonology if the length of the treatment was extended. Recall that participants attended four 45-minute experimental lessons spaced by one-week intervals, which might have been insufficient to help them improve their perception and pronunciation.

In addition, the results obtained in this study resemble those found in previous TTS-based literature involving L2 phonological acquisition (Hincks, 2002; Kiliçkaya, 2008), in which L2 learners' working with this application showed an improvement in their performance in pronunciation, but not in a significant manner.

Finally, it is worth to mention that groups did not perform equally in the perception and production tasks (e.g., the TTS group did not have the same performance in the word-list read-aloud task as in the passage read-aloud task). This suggests that it may not be possible to make generalizations about learners' performance from a single task, i.e., more tasks seem to be needed in order to comprehensively investigate a given phenomenon.

Limitations of the Study and Suggestions for Future Research

This study encountered a number of limitations that will need further consideration in future research. One of these limitations relates to the sampling of participants. Recall that they were students in intact classes. Because of logistical constraints, it is often impossible for researchers to carry out studies in instructional settings in which participants are randomly assigned to different experimental conditions. However, this seems to be critical for experimental research, since most of the statistical tests used to detect differences, such as ANOVAs, work on the assumption that participants are randomly assigned to experimental groups (Field, 2009). When this assumption is violated, the results obtained may deviate from reality.

Another evident limitation of the current study is that the rating of the speech samples was carried out by one NS of English. Although inter-rater reliability was calculated by having a second rater evaluate 10% of the speech samples and that agreement was almost perfect, the NS might have experienced listening fatigue, as other studies have shown (e.g., Derwing, Munro & Wiebe, 1998). This could represent a serious threat to validity, since the ratings might not accurately reflect participants' performance in pronunciation tasks. To overcome this problem, researchers may want to consider having several raters evaluate different parts of the data across an extended period of time.

A third limitation of the current study that might have had an influence on the results was the length of the experimental treatment. As mentioned earlier, such a short treatment might have failed to influence learners' acquisition of /ɪ/ and made it difficult to detect learning that might have taken place. Previous studies in L2 phonology have shown that a considerable amount of hours of focused instruction are needed to help learners acquire challenging forms. For example, it has previously been suggested that Japanese learners are able to acquire English /ɹ/ and /l/, two challenging phonemes, only after extended treatments including focused instruction (e.g., 15 hours in Bradlow, Pisoni, Yamada, & Tokhura, 1997). Therefore, it is possible to hypothesize that linguistic forms that pose considerable difficulties for L2 learners, such as English /ɪ/ for Spanish speakers, may require more hours of instructional intervention. Future researchers may want to consider having learners exposed to extended experimental treatments to capture its effects and, in addition, to provide a longer time interval when conducting delayed post-tests.

A further limitation of this study is that for perception and pronunciation tasks, it included instances of /ɪ/ only in monosyllabic words in a CVC or CCVC environment. Although this was a conscious decision that allowed the researcher to isolate the phenomenon in order to create obligatory contexts for /ɪ/, if disyllabic or multisyllabic words containing /ɪ/ had been included, the results of the study could have been more generalizable to other phonetic environments and thus provide a clearer picture of the perception and production of English /ɪ/. Future researchers might want to consider including a varied range of prosodic contexts in which /ɪ/ occurs.

Implications and Concluding Remarks

Despite the fact that many of the tests did not yield significant differences between groups, the results obtained were consistently positive for the TTS group from the pre-test to the immediate post-test, i.e., this group outperformed the other groups in most of the tasks, albeit in a non-significant manner. Accordingly, the results of the current study may have several implications for ESL teaching, provided that suggestions made in the previous section be considered. First, it can be proposed that TTS could be integrated into ESL pedagogy as a way to expose learners to aural input in a more personalized manner. Learners can work individually with the application by having access to both aural and visual input without deviating their attention to external interferences such as interruptions from peers or noises in the learning environment. At the same time, learners can manipulate the input at their convenience by adjusting or changing certain features in the application, such as the speed rate and the synthetic voices.

TTS applications could also serve as an alternative provider of input for ESL learners. In many instructional settings, such as the one in this study, learners are exposed to one main source of input: the teacher. The use of TTS can make instruction more varied and dynamic, with different forms (in both quantity and quality) of input. The fact of being exposed to varied forms of input may have an impact in phonology acquisition. In fact, previous studies have shown that exposure to highly-variable input can help L2 learners establish robust L2 phonetic categories (e.g., with Japanese leaners; see Bradlow et al., 1997; Lively, Pisoni, & Logan, 1992; Lively, Pisoni, Yamada, Tokhura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991).

However, when implementing the use of TTS in the ESL classroom, teachers might want to consider having learners work on focused activities that target the relevant sounds and also provide learners with clear phoneme descriptions, since the results of this study have shown that mere input flood may not be enough to acquire an L2 feature.

Teachers could certainly benefit from the use of TTS well. This application has the advantage of taking on repetitive tasks, which is something that might not be appealing to them. With a TTS application, learners can listen to texts as many times as they want. However, ESL instructors might not feel comfortable reading a text many times during a lesson to make his/her students aware of different linguistic features. In addition, TTS can help teachers save time to find listening material for listening activities, since any written text can be entered into the application and then reused in aural activities. This could be a tremendous advantage in those instructional setting in which ESL teachers do not have enough time to dedicate to material preparation or where English is not spoken outside of the classroom.

The purpose of the study presented here was to investigate the effectiveness of TTS in the acquisition of English /ɪ/ by Spanish learners of English in an instructional setting. Despite the fact that the findings were not as anticipated and that there is still much research to be undertaken in this matter, it is hoped that this study will contribute to the field of computer-assisted second language learning and help fill the existing research gap in the literature regarding the controlled use of TTS and its influence on L2 phonological acquisition in instructional settings.

REFERENCES

Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, 116(5), 3099-3107.

Audacity (Version 1.3-12-beta) [Computer software]. Boston, MA: Free Software Foundation Inc.

Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation.* Oxford: Oxford University Press.

Best, C. (1993). Learning to perceive the sound pattern of English. *Haskins Laboratories Status Report on Speech Research,114,* 31–80.

Best, C. (1995). A direct realistic perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 167–200). Timonium, MD: York Press.

Bion, R. A. H., Escudero, P., Rauber, A. S., & Baptista, B. O. (2006). Category formation and the role of spectral quality in the perception and production of English front vowels. *Proceedings of INTERSPEECH 2006, USA,* 2913–2916. Retrieved from http://www.nupffale.ufsc.br/rauber/Interspeech%20article_2006.pdf

Bradlow, A. R., Pisoni, D. B., Yamada, R. A. & Tokhura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects on perceptual learning on speech production. *Journal of the Acoustic Society of America, 101*(4), 2299-2310.

Brown, C. (1998). The role of the L1 grammar in the acquisition of segmental structure. *Second Language Research*, *14*, 139–193.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 4–63). Oxford: Blackwell.

Cardoso, W. (2007). The variable development of English word-final stops by Brazilian Portuguese speakers: A stochastic optimality theoretic account. *Language Variation and Change*, *19*(3), 219–248.

Carlisle, R. S. (1998). The acquisition of onsets in a markedness relationship: A longitudinal study. *Studies in Second Language Acquisition, 20*, 245–260.

Carlson, R. (1995). Models of speech synthesis. *Proceedings of the National Academy of Science of the United States of America, USA, 92*, 9932–9937. Retrieved from http://www.jstor.org/pss/2368593

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages.* New York, NY: Cambridge University Press.

Cenoz, J., & García Lecumberri, L. (1999). The effect of training on the discrimination of English vowels. *International Review of Applied Linguistics in Language Teaching, 37*(4)*,* 261–275.

Chang, R. (2011). *365 ESL short stories.* Retrieved form http://www.eslfast.com/.

Chapelle, C. (1998). Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning and Technology, 2*(1), 21–39.

Chapelle, C. (2001). Innovative language learning: Achieving the vision. *ReCALL, 13*(1), 3–14.

Chapelle, C. (2003). *English language and technology*. Philadelphia, PA: John Benjamins.

Chapelle, C. (2007). Technology and second language acquisition. In N. Markee (Ed.), *Annual Review of Applied Linguistics* (pp. 98–114). Cambridge: Cambridge University Press.

Chapelle, C. (2009). The relationship between second language acquisition theory and computer-assisted language learning. *The Modern Language Journal, 93*, 741–753.

Cobb, T. (n.d.). Web Vocabprofile [Computer software]. Retrieved from http:// www.lextutor. ca/vp/eng/

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some input on the easy/difficult grammar question: An empirical study. *The Modern Language Journal, 93,* 336–353.

Czaykowska-Higgins, E., & Dobrovolsky, M. (2009). Phonology: The function and patterning of sounds. In W. O'Grady & J. Archibald (Eds.), *An introduction to contemporary linguistic analysis* (pp. 59-108). Toronto, ON: Pearson Education Canada.

De Lacy, P. (2006). *Markedness: Reduction and preservation in phonology.* Cambridge: Cambridge University Press.

Derwing, T., & Munro, M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*, 393–410.

Deutmert, A. (2003). Markedness and salience in language contact and second-language acquisition: Evidence from a non-canonical contact language. *Language Sciences, 25*, 561–613.

Dulay, H., Burt, M., & Krashen, S. (1982). *Language two.* Oxford University Press: Oxford.

Dutoit, T., & Cerňak, M. (2005). TTSBOX: A matlab toolbox for teaching text-to-speech synthesis. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05), USA, 5,* 537–540. Retrieved from http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1416359.

Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27*(2), 315–330.

Eckman, F. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition, 13,* 23–41.

Eckman. F. (2004). From phonemic difference to constraint rankings. *Studies in Second Language Acquisition, 26*, 513–149.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics, 27,* 1–24.

Ellis, N. C. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 state. In P.

Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 372–405). London: Routledge.

Escudero, P. (2000). *Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish teNS/lax vowels by Spanish speakers* (Master's thesis). Retrieved from http://www.fon.hum.uva.nl/paola/.

Escudero, P. (2006). The phonological and phonetic development of new vowel contrasts in Spanish learners of English. In B. O. Baptista & M. A. Watkins (Eds.), *English with a Latin beat: Studies in Portuguese/Spanish-English interphonology* (pp. 41–55). Amsterdam: John Benjamins.

Field, A. (2009). *Discovering statistics using SPSS.* London: SAGE Publications Ltd.

Finch, D., & Ortiz, H. (1982). *A course in English phonetics for Spanish speakers.* Bristol: Heinemann Educational Books.

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. 233–277). Timonium, MD: York Press.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske & A. Rohde (Eds.), *An integrated view of language development: papers in honor of Henning Wode* (pp. 217–244). Trier: Wissenschaftlicher Verlag.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in* *language comprehension and production* (pp. 319–355). Berlin: Mouton de Gruyter.

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

García Pérez, G. M. (2003). *Training Spanish speakers in the perception and production of English vowels* (Doctoral dissertation). Retrieved from Pro-Quest (Document ID 932345391).

Gatbonton, E., & Trofimovich, P. (2008). The ethnic group affiliation and L2 proficiency link: Empirical evidence. *Language Awareness, 17*, 229–248.

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning, 51*(1), 1–50.

Hallé, P. A., & Best, C. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stops + /l/ clusters. *Journal of the Acoustical Society of America, 121*(5), 2899–2914.

Hancin-Bhatt, B. (1994). Segment transfer: A consequence of a dynamic system. *Second Language Research*, *10*, 241–269.

Handley, Z. (2009). Is text-to-speech synthesis ready for use in computer-assisted language learning? *Speech Communication, 5*, 906–919.

Hincks, R. (2002). Speech synthesis for teaching lexical stress. *Proceedings of Fonetik-Speech, Music and Hearing: Quarterly Progress Status Reports, 44*(1), 153–156.

Hualde, J. I. (2005). *The sounds of Spanish.* New York: Cambridge University Press.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics, 23*(1), 83–103.

Julie (Version M16) [Computer software]. Santa Clara, CA: Neospeech Inc.

Kate (Version M16) [Computer software]. Santa Clara, CA: Neospeech Inc.

Kiliçkaya, F. (2008). Improving pronunciation via accent reduction and text-to-speech software. In T. Koyama, J. Noguchi, Y. Yoshinari, and A. Iwasaki (Eds.), *Proceedings of the WorldCALL 2008 Conference, Japan, 1,* 135–137. Retrieved from http://www.j-let.org/~wcf/proceedings/proceedings.pdf

King, R.D. (1967). Functional load and sound change. *Language, 43,* 831–852.

Kirstein, M. (2006). *Universalizing universal design: Applying text-to-speech technology to English language learners' process writing* (Doctoral dissertation). Retrieved from Pro-Quest (Document ID1188883621).

Ladefoged, P. (2005). *A course in English phonetics.* Boston, MA: Heinle & Heinle.

Lee, F. (1969). Reading machine: From text to speech. *Institute of Electrical and Electronics Engineers - Transactions Audio and Electroacoustics, USA, 17*(4), 275–282.

Lenneberg, E. H. (1967). *Biological foundations of language.* New York, NY: Wiley.

Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning*, *47*(3), 467–505.

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in Second Language Acquisition, 22*(4), 557–584.

Leow, R. P. (2007). Input in the L2 classroom. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 21–50). New York, NY: Cambridge University Press.

Levis, J., & Cortes, V. (2008). Minimal pairs in spoken corpora: Implications for pronunciation assessment and teaching. In C. Chapelle, Y. R. Chung, & J. Xu (Eds.) *Towards adaptive CALL: Natural language processing for diagnostic language assessment* (pp. 197–208). Ames, IA: Iowa State University.

Lively, S. E., Pisoni, D. B., & Logan, J. S. (1992). Some effects of training Japanese listeners to identify English /r/ and /l/. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 175 - 196). Tokyo: Ohmsha.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tokhura, Y. & Yamada, T. (1994). Training Japanese listeners to identify Englihs /r/ and /l/. III. Long-term memory retention of new phonetic categories. *Journal of the Acoustic Society of America, 96*(4), 2076-2087.

Logan, J. S., Lively, S. E. & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustic Society of America, 89*(2), 874-886.

McCandliss, B. D., Fiez, J. E., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Test of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience, 2*(2), 89–108.

Major, R. C. (1986). The ontogeny model: Evidence from L2 acquisition of Spanish r. *Language Learning, 36*, 453–504.

Major, R. C. (2001). *Foreign accent.* Amsterdam: Benjamins.

Menke, M. (2010). Examination of the Spanish vowels produced by Spanish-English bilingual children. *Southwest Journal of Applied Linguistics 28*(2), 98–135.

Morrison, G. (2008). L1-Spanish speakers' acquisition of the English /i/-/ɪ/ contrast: Duration-based perception is not the initial developmental stage. *Language and Speech, 51*(4), 285–315.

Morrison, G. (2009). L1-Spanish speakers' acquisition of the English /i/-/ɪ/ contrast II: Perception of vowel inherent spectral change. *Language and Speech, 52*(4), 437–462.

Munro, M., & Derwing, T. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System, 34*, 520–531.

Nation, I. S. P., & Heatley, A. (1994). Range: A program for the analysis of vocabulary in texts [Computer software]. Retrieved from http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

Odisho, E. (1992). *A comparative study of English and Spanish vowel systems: Theoretical and practical implications for teaching pronunciation* (Report No.

ED352836). Retrieved from Education Resources Information Center website http://www.eric.ed.gov/ERICDocs/data/

PASW Statistics (Version 18). Chicago, IL: SPSS Inc.

Paul (Version M16) [Computer software]. Santa Clara, CA: Neospeech Inc.

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms.* Princeton, NJ: Princeton University Press.

PowerPoint (Version 2007) [Computer software]. Redmond, WA: Microsoft Corporation.

Praat (Version 5.2.04) [Computer software]. Amsterdam: Boersma & Weeninck.

Proctor, C. P., Dalton, B., & Grisham, D. L. (2007). Scaffolding English language learners and struggling readers in a universal literacy environment with embedded strategy instruction and vocabulary support. *Journal of Literacy Research, 39*(1), 71–93.

Rauber, A. S., Escudero, P., Bion, R. A. H., & Baptista, B. O. (2005). The interrelation between perception and production of English vowels by native speakers of Brazilian Portuguese. *Proceedings of INTERSPEECH 2005, Portugal,* 2913–2916. Retrieved from http://www.nupffale.ufsc.br/rauber/Interspeech%20article_2005.pdf

Richards, J. C., & Schmidt, R. (2002). *Longman dictionary of language teaching and applied linguistics.* Harlow: Pearson Education.

Roach, P. (2000). *English Phonetics and Phonology, a practical course.* Cambridge: Cambridge University Press.

Rosa, E., & O'Neill, M. D. (1999). Explicitness, intake, and the issue of awareness. *Studies in Second Language Acquisition, 21*(4), 511–556.

Saito, K., & Lyster, R. (in press). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning.*

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 35–78.

Schmidt, R. W. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, *13*, 206–226.

Šef, T., & Gams, M. (2003). Speaker (GOROVEC): A complete Slovenian text-to-speech system. *International Journal of Speech Technology, 6*, 277–287.

Sisson, C. (2007). *Text-to-speech in vocabulary acquisition and student knowledge models: A classroom study using the REAP intelligent tutoring system* (Report No. CMU-LTI-07-009). Retrieved from Carnegie Mellon University, Language Technologies Institute website http://www.lti.cs.cmu.edu/Research/tech-reports.html

Stockwell, R. P., & Bowen, J. D. (1965). *The sounds of English and Spanish.* Chicago: The University of Chicago Press.

Stratil, M., Burkhardt, D., Jarrat, P., & Yandle, J. (1987). Computer-aided language learning with speech synthesis: User reactions. *Programmed learning and Educational Technology, 24*(4), 309–316.

Stratil, M., Weston, G., & Burkhardt, D. (1987). Exploration of foreign language speech synthesis. *Literary and Linguistic Computing, 2*(2), 116–119.

Thomas, E. R., & Tyler, K. (2007). NORM: The vowel normalization and plotting suite. [Online Resource: http://ncslaap.lib.ncsu.edu/tools/norm/].

Visual Basic (Version 6.3) [Computer Software]. Redmond, WA: Microsoft Corporation.

VoiceText (Version 3.11.1.0) [Computer Software]. Seoul: Voiceware Corporation Ltd.

APPENDIX A

Consent Form (Spanish)

A través de este documento manifiesto que acepto participar en un estudio que llevará a cabo Fernanda Soler (fono: 1-514-6527088; e-mail: f_soleru@education.concordia.ca) estudiante del programa Master of Arts in Applied Linguistics del Departamento de Educación, Concordia University, Montreal, Canadá.

A. PROPÓSITO

Se me ha informado que el propósito de esta investigación es el estudio de un nuevo enfoque pedagógico para el aprendizaje del inglés.

B. PROCEDIMIENTOS

Se me ha informado que se me pedirá completar el cuestionario que se adjunta y participar en tres entrevistas en donde mis respuestas serán grabadas (una antes y dos después del periodo de instrucción). También se me ha informado que se me pedirá asistir a cuatro sesiones de instrucción que durarán aproximadamente 45 minutos cada una, durante cuatro semanas consecutivas, que estarán incluidas dentro de mis horas de clases de la asignatura de Lengua Inglesa I.

C. CONDICIONES DE PARTICIPACIÓN

- Entiendo que puedo renunciar a participar en esta investigación y que mi disentimiento no tendrá ninguna consecuencia negativa.
- Entiendo que mi participación en este estudio es confidencial, es decir, la investigadora conocerá mi identidad, pero no la dará a conocer.

- Entiendo que los datos de este estudio pueden ser publicados o presentados en alguna conferencia científica y que los datos se darán a conocer de tal forma que mi identidad no será revelada.
- Entiendo que mi participación en este estudio no involucra ningún riesgo para mi persona, sino que puede representar un beneficio para mi aprendizaje del inglés.
- Entiendo que seré entrevistado en tres ocasiones y que mis respuestas serán grabadas.
- Entiendo que recibiré una pequeña compensación material cuando concluya mi participación en la investigación y que se me entregará cuando asista a la última entrevista.

HE LEÍDO CUIDADOSAMENTE LO ANTERIOR Y ENTIENDO COMPLETAMENTE ESTE DOCUMENTO DE CONSENTIMIENTO, A TRAVÉS DEL CUAL ACEPTO PARTICIPAR EN ESTE ESTUDIO.

FECHA: 11 de abril de 2011.

NOMBRE: ____

FIRMA: ____

Si en algún momento usted tiene alguna pregunta sobre sus derechos como participante de este estudio, por favor contacte a Adela Reid, miembro del Comité de Ética de Concordia University, al 1-514-8482424, anexo 7481 o al e-mail areid@alcor.concordia.ca

APPENDIX B

Language Background Questionnaire (Spanish)

Por favor, lee atentamente las siguientes preguntas y completa con letra clara toda información que se te pide a continuación:

NOMBRE: ________________________ EDAD: ______ GÉNERO____________

SECCIÓN EN LENGUA INGLESA I: _____ FECHA: 1 de abril de 2011.

1. ¿Cuál es tu lengua materna?____

2. ¿Qué idiomas estás estudiando actualmente en la carrera?______________________

3. ¿Hablas algún otro idioma? ¿Cuál(es)? ___________________________________

4. Durante tu educación básica, ¿asististe a un colegio bilingüe inglés/español?

   Sí ____ No ____

5. Durante tu educación básica, ¿cuántos años tuviste clases de inglés? ______ años

6. Durante tu educación media, ¿asististe a un colegio bilingüe inglés/español?

   Sí ____ No ____

7. Durante tu educación media, ¿cuántos años tuviste clases de inglés? ______ años

8. ¿Has tomado cursos de inglés fuera del colegio o la universidad (en algún instituto o escuela de idiomas, por ejemplo)? Si la respuesta es sí, responde la pregunta 9; si no, continúa con la pregunta 10.

   Sí ____ No ____

9. ¿Cuántos años has estudiado inglés fuera del colegio o la universidad? ______ años

10. Haciendo un cálculo promedio, ¿cuántas horas a la semana estás expuesto al inglés fuera de tus clases en la universidad? Esto incluye horas de televisión, de música, de interacción con otras personas en inglés, etc. ______ horas a la semana.

11. Evaluando tus habilidades lingüísticas con el idioma inglés, ¿Cómo las calificarías en una escala de 1 a 7, en donde 1 es muy básico y 7 muy avanzado? Encierra tus respuestas en un círculo.

Reading skills (habilidades de lectura):

1 2 3 4 5 6 7

Writing skills (habilidades de escritura):

1 2 3 4 5 6 7

Listening skills (habilidades auditivas):

1 2 3 4 5 6 7

Speaking skills (habiliadades de habla):

1 2 3 4 5 6 7

APPENDIX C

Word-list Read-aloud Task

Por favor, lee las siguientes palabras en voz alta. Deja que pase un segundo antes de leer la palabra que sigue y no leas los números.

1. bill
2. room
3. hill
4. short
5. hit
6. belt
7. kill
8. plumb
9. miss
10. bright
11. pick
12. jump
13. with
14. man
15. sit
16. more
17. thick
18. rude
19. wish
20. lake

APPENDIX D

Passage Read-aloud Task

Por favor, lee el siguiente texto en voz alta. No te preocupes si no conoces algunas palabras, sólo dilas como tú crees que se pronuncian. Lee a un ritmo normal.

How to Save Petrol and the Environment

We are all aware of the need to protect the environment but sometimes it's difficult for us to do the right thing. But did you know that making a few simple changes to your driving habits will not only do less damage to our world but will also save you money and could even represent a big advantage for your health? Here, we suggest some questions you might want to ask yourself and also give you a list of steps to help you.

Do you really need to take the car?

Try making more use of public transport. If the service is frequent and reliable you'll soon get used to using buses and trains. In fact, for shorter journeys, why not take the opportunity to get into shape, feel fit and go on foot?

Share the journey

How often do you see cars with just one occupant with the driver making the same journey as others who live nearby? Why not car share and half the cost of the trip? There are several websites where people can swap details and make arrangements to meet up.

Kill bad habits

When you must use the car, plan your journey so you can go to all the places you need to visit rather taking the car out again and again. If you get caught in a traffic jam, switch off the engine when you're stationary for a long time. Try not to brake too sharply or accelerate too quickly as this will lead to you using up more fuel. On cold mornings, don't warm up the engine before you start your journey. When you next put fuel in your car, think about whether you really need to fill up the tank. All that extra weight will put more pressure on the engine.

Servicing

Make sure you carry out basic maintenance like checking the tyre pressure each week. Keep your car regularly serviced so that it runs as efficiently as possible. Read the manual of your car to know when you should carry out each maintenance. If you have a problem, fix it as soon as possible.

Steps like these will save you money and help you do your bit to protect the environment. Let’s make a deal. You will certainly teach others to protect the environment as well!

APPENDIX E

Short Stories used in the Experiment

Story 1

Sara's Day

Sara Smith, a Pasadena resident, went shopping. She is 30, and has lived at 3037 North Foothill Street since 1992. Sara has been married to John for seven years. They have two children; Bob is five years old, and Nancy is three. Sara owns a 1995 four-door green Toyota. Yesterday at 9 a.m., Sara got into her car and drove to Barget, a department store a mile away.

Barget was having a holiday sale. Sara bought a four-slice toaster for $29.95 plus tax. The regular price was $39.95. She paid by check. On her way home, Sara stopped at Milk Plus to buy a gallon of nonfat milk. The milk was $3.50. Sara got 50 cents back in change.

Sara arrived home at 10 a.m., John and the kids were still sleeping. She woke them up and then made a hot and nutritious breakfast for everyone.

Story 2

Book Him

A man, who was accused of failing to return more than 700 children's books to five different libraries in the county, was released from jail yesterday, after a book publisher agreed to post his bond of $1,000. The publisher said, "There's a story here. This is a man who loves books. He just can't let go of them. He hasn't stolen a single book. So what's the crime? We think that Mister Barush has a story to tell. We plan to publish his story."

When asked why he didn't return the books, Mister Barush said, "Well, how could I? They became family to me. I was afraid to return them, because I knew that kids or dogs would get hold of these books and chew them up, throw them around, rip the pages, spill soda on them, get jam and jelly on them, and drown them in the toilet."

He continued, "Books are people, too! They talk to you, they take care of you, and they enrich you with wisdom and humor and love. A book is my guest in my home. How could I kick it out? I repaired torn pages. I dusted them with a soft clean cloth. I turned their pages so they could breathe and get some fresh air.

"Every week I reorganized them on their shelves so they could meet new friends. My books were happy books. You could tell just by looking at them. Now they're all back in the library, on the lower shelves, on the floors, at the mercy of all those runny-nosed kids. I can hear them calling me! I need to rescue them. Excuse me. I have to go now."

Story 3

Bill the bird

When I was a little child, my family and I used to live by the sea. In our house, we used to have a little bird, called Bill. He used to sing beautifully, and his singing would make us feel very happy.

I remember I always cleaned his cage, because my brother pretended not to hear my mother when she asked him to do it. He wouldn't even feed him. Only my parents and I would give the little bird its food and water to drink.

During the day, Bill was always outside. We would hang its cage in the patio, so he could enjoy the sunny days. But at nights, we would keep him inside the house, because nights got very chilly, and we were afraid that cold might kill him.

In the mornings, I used to wake up when I heard Bill singing, but one day I didn't hear him anymore. I missed his singing, and I knew that there was something wrong because everything was still. With surprising speed, I jumped from my bed and went to reach my mother. I asked her what was wrong, why I didn't hear Bill anymore. I'm sure she saw fear in my eyes, because I could tell from her face. She told me that Bill had passed away the night before because he was a bit old and weak. I suddenly started crying. I felt so bad because my little pet wasn't there anymore. Now, I don't cry anymore, but I miss his happy songs when I wake up every morning.

Story 4

His Stomach Is Growling

Derek was working on his computer doing some homework for school. Everything was still in Derek’s house. Suddenly, his stomach growled loudly. He didn’t feel all that hungry, but he thought that the “squeaky wheel” should get some grease. He went over to the refrigerator, and he opened the freezer door. It contained six empty ice cube trays! He shut it, and he opened the bottom door. He scrutinized the shelves: milk, butter, canned-fish, ketchup, mustard, a piece of ham, and some cheese slices. He opened the vegetable bin. Nothing in there, except some red onions, half a head of green cabbage, and a little bit of lettuce.

His refrigerator contained some food, but it was food that you would eat only if you were starving. He was reminded of his stint in the army. His buddies and he always joked about how bad the food rations were.

Derek went back to his desk and resumed using his computer. His stomach growled again, but he ignored it. He would wait until he was really hungry. Then he would walk down to the fried chicken place five minutes away and get some finger-licking food.

He never realized that the growl came from his dog lying under the table!

APPENDIX F

Sample Lesson TTS Group

Name: ________________________________

Activity 1

INSTRUCTIONS:

1. Please, write your name above.
2. Paste the following story in VoiceText.
3. Set the **speed at 90** at all times and do not change it throughout the activity.
4. Listen to the story **three times**, using Julie, Kate and Paul's voices.
5. Pay attention to the story because you will have to answer five comprehension questions at the end.
6. After pasting the story in the application, **save changes** and close this document.

Bill the bird.

When I was a little child, my family and I used to live by the sea. In our house, we used to have a little bird, called Bill. He used to sing beautifully, and his singing would make us feel very happy.

I remember I always cleaned his cage, because my brother pretended not to hear my mother when she asked him to do it. He wouldn't even feed him. Only my parents and I would give the little bird its food and water to drink.

During the day, Bill was always outside. We would hang its cage in the patio, so he could enjoy the sunny days. But at nights, we would keep him inside the house, because nights got very chilly, and we were afraid that cold might kill him.

In the mornings, I used to wake up when I heard Bill singing, but one day I didn't hear him anymore. I missed his singing, and I knew that there was something wrong because everything was still. With surprising speed, I jumped from my bed and went to reach my mother. I asked her what was wrong, why I didn't hear Bill anymore. I'm sure she saw fear in my eyes, because I could tell from her face. She told me that Bill had passed away the night before because he was a bit old and weak. I suddenly started crying. I felt so bad because my little pet wasn't there anymore. Now, I don't cry anymore, but I miss his happy songs when I wake up every morning.

Name: ___________________________________

Activity 2

INSTRUCTIONS:

1. Please, write you name above.
2. **Without looking at VoiceText**, please answer the following comprehension questions in relation to the story we just listened to. We'll check the answers after you're done.
3. Please, **save changes** before closing the document!

1. Where did the boy and his family live when he was a little child?

2. What was the name of his pet?

3. Why would they hang his cage in the patio?

4. Why didn't the boy hear his bird anymore?

5. What does he miss from his pet?

Name: ________________________________

Activity 3

The following words were extracted from the story we just read.

INSTRUCTIONS:

1. Please write your name above.
2. Paste the following word list in VoiceText.
3. Set the **speed at 70** at all times and do not change it throughout the activity.
4. **Listen to the list three times**, each time with a different voice (Julie, Kate & Paul).
5. Pay attention to the words because another activity will follow.
6. After pasting the words in the application, **save changes** and close this document.

speed - face - drink - feed - weak - make - hang - keep - kill - day - cage - feel - miss - cleaned - bit - sing - asked - still

Name: _________________________________

Activity 4

Now, we will work on the words we just listened to.

INSTRUCTIONS:

1. Please, write you name above.
2. Paste the pairs of words in the box below in VoiceText.
3. Set the **speed at 70** at the beginning. You may change it later if you want.
4. Listen to the **list three times**, each time with a different voice (Julie, Kate & Paul).
5. Please, **indicate below if you changed the speed** during this activity.
6. Before closing the document, please **save changes.**

   1. make - day
   2. sing - feel
   3. drink – cleaned
   4. hang - asked
   5. kill - feed
   6. still - keep
   7. bit – miss
   8. cage - face
   9. speed - weak

Activity 5

INSTRUCTIONS:

1. Please, do not delete the words from the last activity in VoiceText
2. Choose one of the voices (either Kate, Paul or Julie) and listen to the words one more time.
3. Set the **speed at 70.**
4. Pay attention to the **vowel sound** of each pair of words and decide whether they sound the same or differently.
5. Write an "S" next to the numbers in your paper if the **vowel** of the each pair sounds the same or "D" if they sound differently (Keep the paper to hand it in).

Name: ________________________________

Activity 6

INSTRUCTIONS:

1. Please enter your name above.
2. Paste the words below in VoiceText.
3. Set the **speed at 70** at least once. You may change it after if you want.
4. Choose one of the voices (either Kate, Paul or Julie) and listen to the sets of words.
5. In each set, there is a word that has **a different vowel sound**. Decide which one it is and write it down next to the numbers you wrote (Keep the paper to hand it in).
6. Remember to **save changes** before closing the document.

1. make - day - keep
2. sing - feel - feed
3. drink - cleaned - kill
4. hang - asked - face
5. still - keep - miss
6. speed - weak - bit

Name: ___________________________________

Activity 7

INSTRUCTIONS:

1. Please write you name above.
2. Choose **three words** from the previous activity (they are still pasted in VoiceText).
3. Write **one new** sentence with each word. Do not use sentences from the initial text.
4. Paste the sentences in VoiceText and listen to them using the three different voices (Julie, Kate and Paul).
5. You can set the speed at the rate you want. Before closing, **save changes**.

Sentence 1

Sentence 2

Sentence 3

Name: ___________________________________

Activity 8

INSTRUCTIONS:

1. Please write your name above.
2. Paste the following in VoiceText:

Now, I'm going to start reading the story for you to fill the gaps. Are you ready? OK. I'll start now.

3. Open the file "Activity 1" and paste the story again in VoiceText (any speaker) right after the sentences above.
4. You can use the speed you want.
5. Press play and minimize the window.
6. Listen to the story and complete the gaps in the text below.
7. Listen to the story **until you complete all the gaps**.
8. **Once you complete all the gaps**, check your answers by looking at the missing words in VoiceText.
9. Before closing, please **save changes**!

When I was a little child, my family and I used to ____. In our house, we used to have a little bird, called Bill. ____ beautifully, and his singing would make us feel very happy. I remember I ____ cage, because my brother pretended not to hear my mother when she asked him to do it. He wouldn't ____. Only my parents and I would give the little bird its food and ____.

During the day, Bill was always outside. We would ____ in the patio, so he could enjoy the sunny days. But at nights, ____ him inside the house, because nights got very chilly, and we were afraid that ____ him.

In the mornings, I used to wake up when I heard Bill singing, but one day I ____ anymore. I missed his singing, and I knew that there was something wrong because ____. With ____, I jumped from my bed and ____ my mother. I asked her what was wrong, why I didn't hear Bill anymore. I'm sure ____ in my eyes, because I could tell from her face. She told me that Bill had passed away the night before because he was a ____. I suddenly started crying… I felt so bad because ____ wasn't there anymore. Now, I don't cry anymore, but I ____ songs when I wake up every morning.