ATTENTION CONTROL AND THE EFFECTS OF ONLINE TRAINING IN IMPROVING CONNECTED SPEECH PERCEPTION BY LEARNERS OF ENGLISH AS A SECOND LANGUAGE

by

Burcu Gökgöz-Kurt

Bachelor of Arts
Boğaziçi University, 2006

Master of Arts
Middle East Technical University, 2008

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Linguistics

College of Arts and Sciences

University of South Carolina

2016

Accepted by:

D. Eric Holt, Major Professor

Mila Tasseva-Kurktchieva, Major Professor

Amit Almor, Committee Member

James Dean Brown, Committee Member

Isabelle Darcy, Committee Member

Lacy Ford, Senior Vice Provost and Dean of Graduate Studies

DEDICATION

To my husband and best friend, Can, *my life*.

To my little ones, *the future.*

ACKNOWLEDGEMENTS

Writing a dissertation is hard work especially because it takes away from the precious time you could have spent with your loved ones, the living and the deceased. Therefore, you may end up sacrificing in many different ways. I could not have completed this work were it not for all the people in my life who have always stood beside me.

First, I would like to express my gratitude to my advisor, Dr. Eric Holt for his support and patience, which I needed the most throughout this long process. By being critical, flexible, prompt and understanding, he was guiding me through the path. He was always patient and nice even when pointing out to the worst mistakes that needed to be revised. This dissertation would not have been possible without your support and patience; thank you for being my advisor! My deepest gratitude goes to my advisor Dr. Mila Tasseva-Kurktchieva, who never hesitated to generously spend her time by brainstorming ideas and giving feedback to improve the earlier drafts of this dissertation. When I got stuck, she was there to help me out personally and academically even when she barely had time for her own work. This dissertation owes a lot to your insightful feedback, thank you so much for everything! I also would like to thank Dr. Amit Almor for pointing out to potential experimental fallacies in my design and answering my challenging questions. I am grateful for those light-bulb moments in my brain every time I came to see you! Also, I would like to thank Dr. James Dean Brown for creating time to help me with his expertise in this field. He was flexible enough to accommodate the time differences when I needed him online. He was always available to direct me through the challenging process of data analysis. Without his help, my results would be uninterpretable. Thank you for your time and guidance! Finally, thank you from the bottom of my heart to Dr. Isabelle Darcy for making this dissertation real by spending her hours online to brainstorm ideas even when she was on sabbatical overseas, answering questions after a long day of conference presentations, helping me have access to data and experimental designs of hers, and giving psychological support throughout this process. This dissertation could not have been possible without her. Thank you for accepting to be on my committee and being my mentor!

Other than my committee members, I also would like to thank Dr. Alexandra Rowe for allowing me to work for English Programs for Internationals (EPI). It was great to be a part of EPI, which always felt like home. Special thanks to Dr. Dan Fogerty for letting me use his speech lab, and his lab staff Dr. Gibbs and Courtney for helping me with the recordings. I also would like to extend my gratitude to my friends, Bill and Henning, at the Foreign Language Lab for supporting me during the most challenging stages of this process.

My friends and colleagues at EPI deserve the biggest gratitude in this process. Thank you Alisha, for checking the oddness of my stimuli anytime I showed up in your office and more importantly for being a great friend! Thank you to Regina and Bronia, my office mates for being so supportive of my work, sharing their resources, and for sharing my sorrow when I lost my loved ones back home. I also would like to extend my gratitude to Svetlana for her encouragement and help for me to apply for the grants, and

Timothy with his help with the arrangement of the data collection equipment. Special thanks to all other colleagues, especially Shinichi, Maria Sophocleous, Jeff, Janee, Stephanie, Cameron, Mr. Stiteler, Russ, Dr. Cheek, and Angi for their flexibility and support as well as students at EPI for their participation. You have truly been my family here in South Carolina.

Also, thanks to Linguistics Program at the University of South Carolina (USC) and the Fulbright Scholarship for financially supporting my Ph.D. studies. This dissertation has also been partially funded by Support to Promote Advancement of Research and Creativity (SPARC) grant at USC.

I also would like to thank my mom and dad for supporting me throughout my never-ending educational life. More importantly, my little ones, who came to our life in the middle and end of this long journey, taught me the value of time and patience once again. And of course…last but not the least, my husband and my best friend, Can, who deserves the biggest gratitude, has always been supportive of my work even if it meant taking over all my responsibilities in our family in addition to his own. I have always been excused for being stressful and irritable while he was the one who never complained. I am so lucky to have him in my life, and I hope I am finally back!

ABSTRACT

One of the aspects of L2 English phonology which poses a challenge for L2 learners is learning how to decode the language, especially as spoken by native speakers. This difficulty may be due to the way the native speakers speak by 'draw[ing] [the sounds] together' (Clarey & Dixson, 1963), which results in realization of consonants and vowels differently than when uttered in isolation. This process is referred to as *connected speech* (e.g., pronouncing 'want to' as [wɑnə], and 'going to' as [gʌnə]). The challenge in teaching and learning these forms is that they lack perceptual saliency, requiring extra attentional resources for learners to be able to recognize these forms in spoken language. Therefore, a better understanding of the role of attention in learning these forms is needed. While some studies find a relationship between attention control as a cognitive ability and L2 phonological processing (Darcy, Mora & Daidone, 2014; Safronova & Mora, 2012a), other studies have failed to confirm the existence of such a relationship (Darcy, Park & Yang, 2015). More importantly, to date, no study has examined attention control as it relates to L2 phonological gains, especially in learning a phonological aspect of L2 English other than individual segments as the target linguistic item. Therefore, the present study aimed to explore the effects of training in improving the connected speech perception of L2 learners as well as the relationship between attention control and learners' improvement in connected speech perception.

To do this, English as a Second Language learners, who were assigned to an experimental (n = 33) or a control group (n = 25), took a two-option forced-choice, pre- and post-test. The experimental group received online training on word-boundary palatalization as a connected speech phenomenon for three weeks while the control group did not. Word-boundary palatalization occurs in the transformation of [toʊld ju]~[toʊldʒʊ] 'told you' or [want ju]~[wantʃʊ] 'want you'. To measure attention control, all students were given a Speech-Based Attention Switching Task (Darcy, Mora & Daidone, 2014; Darcy & Mora, in press; Mora and Darcy, in press) and an Attention Network Test (ANT) (Fan, McCandliss, Sommer, Raz & Posner, 2002).

The findings reveal that learners both in the control and experimental groups improved their scores on the post-test; however, the improvement in the scores of the experimental group was significantly higher than those of the control group (p =.007). Furthermore, correlation analyses showed a significant negative correlation between the post-test scores and attention control, and the gain scores and attention control as measured by the Speech-Based Attention Control Task (p =.002 and p =.008, respectively) and the conflict effect of the Attention Network Test (p =.004 and p =.032, respectively). Additionally, overlapping results between the two attention control tasks were also found as revealed by the significant correlation between the shift-cost and conflict effect measures (p =.009).

Overall, the results indicated that L2 learners benefit from online training in improving performance scores on a perception test of word-boundary palatalization, which is promising for further studies of connected speech teaching and learning. The findings also reveal a significant relationship between learners' attention control and phonological learning, which shows the crucial role attention control plays in learning connected speech.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AC .... Attention Control
ANT .... Attention Network Test
ATI .... Aptitude-Treatment Interaction
CALL .... Computer Assisted Language Learning
CG .... Control Group
CLT .... Communicative Language Teaching
CS .... Connected Speech
EFL .... English as a Foreign Language
EG .... Experimental Group
ELF .... English as a Lingua Franca
ESL .... English as a Second Language
FFI .... Form-Focused Instruction
FLMP .... Fuzzy Logical Model of Speech Perception
IEP .... Intensive English Program
L1 .... First Language/Native Language
L2 .... Second Language
LFC .... Lingua Franca Core
SLA .... Second Language Acquisition

CHAPTER 1

INTRODUCTION

1.1 STATEMENT OF THE PROBLEM

Language teaching practices have recently been less and less dominated by traditional ways of teaching such as drills and grammar-translation, giving way to a more communicative and authentic approach (Ito, 2006a). Despite this emphasis on the communicative use of language in the last couple of decades, language learners still seem to have difficulty comprehending native speakers' speech. There might be various reasons for this such as the characteristics of the listener, the interlocutor or the text type, (see Rubin, 1994); however, the major problem is due to the gap between the language used in and outside of the classroom (Ur, 1987). Learners who are usually exposed to fully articulated "teacher talk" during their second language (L2) learning experiences in classroom contexts find it frustrating when they cannot understand authentic conversations among native speakers and highly proficient nonnative speakers of a language. This is especially true for those who learn L2 vocabulary and grammar in their home countries in a rather decontextualized way because upon their arrival in the host country, they usually have a "rude awakening" (Ur, 1987, p. 10) due to their claim that native speakers talk "too fast" (Gilbert, 1995, p. 97). What gives L2 learners such an impression is, in fact, the way native or highly proficient speakers articulate sounds by "draw[ing] [them] together" (Clarey & Dixson, 1963, p. 12), which is also referred to as *connected speech, reduction* or *sandhi variation* in the literature (Henrichsen, 1984).

Highly proficient speakers of languages usually use connected speech for the sake of efficiency (Rost, 2013). However, understanding connected speech is not as easy for L2 listeners, because in this type of phonologically reduced speech, L2 listeners are expected to determine where the word boundaries are. A large majority of L2 learners do not usually have extensive experience of the commonly used reduced speech forms in L2 English, and this may cause them to have breakdowns in communication in real life. In fact, this has been proven by previous research, which found a strong contribution of successful connected speech perception and comprehension to overall L2 aural processing (Joyce, 2011). Beyond communicative purposes, research on the listening comprehension section of the TOEFL (Test of English as a Foreign Language), which is a high-stakes English proficiency test for L2 learners of English, also demonstrates that "sandhi-variation" is one of the factors which affects test takers' success in the listening comprehension section (Kostin, 2004). In other words, being unable to perceive connected speech forms has been shown to bring about problems in communication, perception and listening comprehension in L2 English.

One way to overcome this difficulty is to immerse learners in the L2 context so that they familiarize themselves with features of connected speech by experiencing the language. However, learning by exposure requires relatively extensive periods of time, and, due to various individual differences, does not guarantee that the learner will notice the problematic structures. Another way is to have explicit, form-focused instruction on connected speech features. Given that the likelihood of L2 speakers to encounter this type of speech is not restricted by the speed or context of speaking, L2 learners need this type of knowledge to maximize smooth communication with fewer breakdowns. In fact, instruction on L2 connected speech has been previously shown to be relatively successful in facilitating L2 speech comprehension (Brown & Hilferty, 1986; Matsuzawa, 2006; Underwood and Wallace, 2012). However, these studies were either carried out in English as a foreign language (EFL), rather than English as a Second Language (ESL) setting. Exploring connected speech in the context of ESL learning is, in fact, crucial for ESL learners as they usually encounter this type of language use in everyday life, which may make them feel more motivated to learn it. Therefore, an investigation of connected speech in ESL settings may provide us with valuable input as regards to the effects of instruction where learners potentially have higher motivation and more need to learn such forms. Although there are studies which explored connected speech in ESL and even more so in EFL settings, most of those studies had certain limitations in their designs such as the assessment of a variety of connected speech phenomena using a single dictation test or providing contextual information, complex structures or lexicon in the dictation tests (see Ito, 2006a). Additionally, in almost all of these studies, researchers only looked at learning conditions such as context, age, or length of exposure, which can be controlled. However, there are various individual differences in cognitive abilities, which are also responsible for explaining the success of connected speech learning in classrooms.

Attention Control (AC) is one of these cognitive abilities whose role has been shown to play a crucial role in second language acquisition. Previous research has revealed that when there is need for controlled processing, adult bilinguals are especially more successful than their monolingual counterparts, which is also called the bilingual advantage in language processing (Bialystok, Craik, Klein & Viswanathan, 2004, p. 302).

Similarly, because there is a constant need for adult L2 learners to control their L2 during language processing, their attention control functions may accordingly be affected and even vary depending on their proficiency in the L2; therefore, investigating the role of attention in L2 learning is important to better understand such a relationship and is yet to be further explored. However, part of the difficulty in studying the ability of AC in individuals may be attributed to the difference in the interpretation and the study of attention in relation to other related constructs (Robinson, 1995; DeSchepper & Treisman, 1996) in addition to the disagreement on the type and amount of attention required for L2 learning to take place (Schmidt, 2001; Schwartz, 1993; VanPatten, 1994). Therefore, the question of whether learners with "better" attention control benefit more from explicit teaching is a question to be resolved (Schmidt, 2001). While there are quite a number of studies looking at the relationship between attention and the success in L2 syntax or vocabulary learning, fewer researchers examined the role of AC in relation to L2 phonological learning (Darcy, Mora & Daidone, 2014; Ito, 2006b; Schmidt, 2001). Moreover, almost all relevant research looks into how various cognitive skills relate to segmental aspects of phonological acquisition (MacKay, Meador & Flege, 2001). As for the tasks used to measure attention control, most studies employed non-language based attention control tasks except a few recent studies (Darcy, Mora & Daidone, 2014; Darcy & Mora, in press; Mora and Darcy, in press). Therefore, researchers have yet to explore the extent to which individual differences in AC explain post-instructional L2 phonological gains in domains other than segmental acquisition. Also, research on various aspects of connected speech comprehension has been carried out either by cognitive psychologists in highly controlled laboratory conditions, or by classroom teachers in classroom settings using dictation tests. The present research aims to show how assessment and training of connected speech perception of L2 English can be achieved by a combination of these two approaches.

1.2 THE PRESENT STUDY: PURPOSE STATEMENT & RESEARCH QUESTIONS

The purpose of this study is to investigate the effects of in-class, online phonological training on improving connected speech perception by ESL learners in relation to their attention control abilities. In other words, the study proceeds from two premises: (1) explicit, form-focused "instruction" may increase the chance of input to become intake (Schmidt, 1995; Robinson, 1995), and (2) there seems to be a link between L2 learning and learners' attention control as a cognitive ability (Guion & Pedersen, 2007; Segalowitz & Frenkiel-Fishman, 2005). Based on these premises, the study investigates whether learners from a variety of proficiency levels benefit from online training on palatalization in English. Additionally, the study also looks at how students' gains in perception are related to their attention control abilities as measured by two types of attention control tests.

More specifically, the present study aims to examine how English as a Second Language (ESL) learners' post-instructional gains in the perception of certain connected speech features relate to their attention control as guided by the following research questions:

RQ1: What are the effects of form-focused online training in improving connected speech perception (specifically word boundary palatalization) by ESL learners?

RQ2: What is the relationship between students' performance scores on the perception test (pre-test, post-test, and gain scores) and their attention control (AC) as measured by two types of tests:

a) Attention Network Test (ANT) (Fan, et al., 2002; Weaver, Bédard, & McAuliffe, 2009, 2013)

b) Speech-Based Attention Switching Task (Darcy & Mora, in press; Darcy, Mora & Daidone, 2014; Mora & Darcy, in press)

RQ3: What is the relationship between attention control scores as measured by an online Attention Network Test (ANT) and a Speech-Based Attention Switching Task?

In order to answer these questions, ESL students who received three weeks of in-class online training on the most commonly used English palatalization forms occurring across word boundaries (e.g., the transformation as in [kʊd jə] ~ [kʊdʒə] for 'could you') took a pre- and post-test. Learners also completed two types of experiments which measured their attention control: An Attention Network Test (ANT) (Fan et al., 2002; Weaver et al. 2009, 2013) and a Speech-Based Attention Switching Task (Darcy, Mora & Daidone, 2014), both of which were adapted extensively from the original tasks due to time limitations in classroom studies. The scores from the pre- and post-tests as well as the two AC tasks are used to discuss the effects of online training and the role of AC in relation to improvement in connected speech perception.

First, it is hypothesized that ESL learners in the experimental group will benefit from this online, form-focused training, and will improve more than the control group by the end of a three-week period. The content of the online training will aim to improve learners' perceptual abilities of the word-boundary palatalization by helping them better understand these forms as shown by previous instructional studies of connected speech learning (Ito, 2006a; Matsuzawa, 2006). It is also predicted that there will be a relationship between perception test scores of learners and their attention control abilities because connected speech forms lack perceptual saliency, so more attention is required to be able to perceive them (Henrichsen 1984, p. 106). Finally, it is predicted that there will be a correlation between the AC scores measured by the Speech-Based Attention Switching Task and the ANT as the attention measures both of these tasks use are based on the calculation of switching costs.

The present study adds to the growing body of research in the field of Second Language Acquisition (SLA), as well as to L2 psycholinguistics, in that it sets out to examine (a) whether in-class, online training may be helpful in improving L2 phonological perception, and how effectively learners can apply the rules they learn during training to novel contexts in perception, (b) whether and how attention control affects the processing and learning of L2 connected speech, (c) whether speech-based attention control tests similar to the one used in this study can be used as an alternative tool to measure attention control. This study also provides empirical evidence that ESL learners should be afforded the chance to become familiarized with connected speech in the classroom to improve their listening and possibly speaking skills outside of the classroom. It also encourages ESL teachers and curriculum designers to include connected speech aspects of L2 English in their curriculum.

1.3 DEFINITION OF TERMS

Attention Network Test (ANT) is a type of task designed to measure three functions of attention (alerting, orienting, and executive attention) in a single task originally developed by Fan et al. (2002). The version used in this study is an adapted shorter version of this task following certain criteria provided by Fan et al.'s (2002) java version of the task as well as the 10-minute version of the task as described in Weaver et al. (2009, 2013), both discussed in the sections to follow.

Connected Speech is a term used to describe processes such as reductions, minimizations in which changes in word forms occur in spontaneous speech according to certain phonotactic rules (Brown & Kondo-Brown, 2006; Joyce, 2013) as well as certain temporal and articulatory constraints (Hieke, 1984, p. 341; Hieke 1987, p. 41). Connected speech processes have been most recently classified into six main categories: linking, deletion, insertion, modification, reduction, and multiple process (see Alameen & Levis, 2015).

English as a Second Language (ESL) is a term used to describe English which is learned mostly in a country where the dominant language is English. ESL learners need to communicate in English to conduct their daily routine activities.

English as a Foreign Language (EFL) is used to describe English which is learned in a country where the dominant language is not English. EFL learners normally do not need to use English to live in an English-speaking environment at the time of learning.

Palatalization, as referred to in this study, is the coronal palatalization in English, and is defined as a phonological process by which word-final alveolar obstruents [t, d, s, z] are palatalized and become palato-alveolars [tʃ, dʒ, ʃ, ʒ] when they are followed by the palatal glide [j] (Chomsky & Halle, 1968, pp. 230-234).

Speech-Based Attention Switching Task, which has also been most recently referred to as *Speeded Set-Switching Task* is a novel attention control task originally created by Darcy, Mora & Daidone (2014; also see Safronova & Mora, 2012a, 2012b). In this task, the purpose is to link attention and phonological acquisition by using phonological dimensions rather than non-phonological ones as used in similar previous studies (Segalowitz & Frenkiel-Fishman, 2005). The task the present study uses has been adapted to fit in a shorter time by using the template of this novel task, but the dimensions used are only speech-based rather than being exclusively phonological. For example, while the original task uses nasality as a dimension, the present study only uses consonant/vowel status with no specific phonological aspect.

1.4 DISSERTATION OVERVIEW

The rest of this dissertation is organized as follows. The next chapter provides a detailed review of the literature defining and studying connected speech processes as well as discussing palatalization as a connected speech process. Chapter 3 presents the role of attention control as a cognitive ability in second language perception within a cognitive as well as a pedagogical framework. Chapter 4 provides a complete picture of the research design and the method used in this study. Chapter 5 presents the findings of the study, which are then discussed in Chapter 6 in light of previous literature.

CHAPTER 2

DEFINING AND TEACHING CONNECTED SPEECH IN SPOKEN AMERICAN ENGLISH

2.1 DEFINITIONS AND FUNCTIONS OF CONNECTED SPEECH

The term connected speech (Brown & Kondo-Brown, 2006) is used to refer to processes such as reductions, minimizations or full eliminations (Brown & Kondo-Brown, 2006) occurring across word boundaries following certain language-specific phonotactic rules (Joyce, 2013). This type of speech has also been referred to as *reductions*, *reduced forms of speech*, *sandhi-variation*, or *weak-forms* (Brown & Kondo-Brown, 2006, p. 5; Ito, 2006a, p. 17). However, until recently, there was not an all-inclusive term to refer to such processes. For example, the term “sandhi”, which is originally a Sanskrit word meaning “putting together”, is used to refer to variations such as assimilation, reduction and contraction in the context of English (Henrichsen, 1984, p. 105). Since it has been claimed that the term *reduction* may not include sentence stress and timing features of spoken language, following Brown and Kondo Brown (2006), I will use the term *connected speech* to refer to all aspects of such phenomena.

The term *connected speech* may involve changes, additions or eliminations to sounds and sound sequences. Stress and intonation patterns of English play a significant role in determining which sounds or sound sequences are to be deleted or modified. While function words, which usually do not bear stress, undergo deletion, content words and their stress bearing syllables are not usually eliminated in connected speech processes. For example, the word ‘and’ in the phrase ‘now and then’ is pronounced as [ən] because ‘and’ is a function word, and thus, it is pronounced in its weak form in connected speech. According to Hieke (1987), these changes in citation forms of words occur as a result of certain “temporal and articulatory” features of spontaneous speech among other reasons discussed later.

Alameen and Levis (2015) recently classified connected speech processes into six main categories. These are *linking*, *deletion*, *insertion*, *reduction*, *multiple processes* and *modification*. Their definition of *linking* is limited to describe “situations in which the ending sound of one word joins the initial sound of the next, but only when there is no change in the character of the segments”, e.g., pronouncing “some of” [sʌm əv] as if it is one word (p. 162). *Deletion* includes elisions as in pronouncing “call him” as [kɔl ɪm**]** by eliding the initial [h], and *contractions* as in pronouncing “he will” as “he’ll”. For *insertion*, Alameen and Levis (2015) give the example of Popeye’s statement of “I am what I am” realized as “I yam what I yam”, in which vowels are connected by glides at word boundaries (p. 163). *Reduction* involves vowel reductions in unstressed syllables and some consonant reductions as with the lack of release in /d/ sound in the phrase “bad boy” (p. 163). Under the category *multiple*, they mention commonly used lexical chunks undergoing several changes simultaneously. These include phrases such as “want to” pronounced as “*wanna*” or “going to” pronounced as “*gonna*”. Finally, the category of *modification* involves four sub-categories: *assimilation* (e.g., the assimilation of [n] to [m] before a bilabial stop in a phrase like “sun beam”); *flapping* (e.g., pronouncing the alveolar stop [t] as an alveolar flap in North American English in the phase “sit around”); *glottalization* (e.g., pronouncing the phrase “can’t make it” as [kænʔmekɪt] as a result of the [t] sound before the nasal [m] being pronounced with a specific glottal articulation); and finally *palatalization*, which involves pronouncing “that you” as [ðæʧʊ]. It should also be pointed out that the term *connected speech* is not usually used to describe processes occurring within words (Alameen & Levis, 2015). For example, the coalescent assimilation in the transformation of the word ‘face’ [feɪs]~ ‘facial’ [feɪʃəl] is similar to the modification in pronouncing “that you” as [ðæʧʊ]. However, while the former type of palatalization occurs within words, the latter occurs across word boundaries.

In word-boundary palatalization, which is the primary focus of this study, one phoneme seems to substitute for other phonemes (Alameen & Levis, 2015, p. 163). This type of palatalization is commonly referred to as ‘coronal palatalization’ in English and is defined as a phonological process in which word-final alveolar obstruents [t, d, s, z] are palatalized and become palato-alveolar consonants [tʃ, dʒ, ʃ, ʒ] when they are followed by the palatal glide [j] (Chomsky & Halle, 1968, pp. 230-234). An example of this process is pronouncing the phrase “told you” as [toʊldʒʊ] rather than [toʊld ju]. Coronal palatalization is further discussed in Chapter 2.

As revealed by the examples above, speakers choose to speak using connected speech simply to save time and energy. In speaking, there is the concept of efficiency, which essentially tolerates the maximum elision of language patterns, in an effort to minimize the number of phonological units (Rost, 2013). This is also known as “the principle of least effort” (Zipf, 1949), or “law of economy” (Clarey & Dixon, 1963), both of which explain why speakers are attracted to speaking with elisions, contractions and assimilations in their conversations. On the other hand, from a purely linguistic perspective, the primary function of the use of connected speech is, in fact, to maintain the rhythm of English by "compressing" unstressed sounds and syllables and making articulation easier (Shockey, 2003). While speakers favor connected speech for various reasons, this brings about a big challenge on the part of L2 listeners: the difficulty of keeping up with the message while listening to this "reduced" speech. In other words, the more reduced a message is, the more effort L2 listeners have to make in order to perceive and process the spoken text.

Other than the formal discussion of which term to use in literature and reasons for its use, there is no agreement, either, as to when connected speech is commonly used. While such processes have been claimed to occur in *fast, colloquial, casual, informal* and *relaxed speech* (Brown & Kondo-Brown, 2006, p. 5; Trager 1982; Weinstein, 1982), some others go even further to label them as "lazy", "sloppy", "careless", "slack", "slipshod", "substandard", "low-class" or "slovenly" (see Brown & Kondo Brown, 2006, pp. 5-6). Nevertheless, it has been shown that such stereotypes are not taken seriously in academic contexts, and more importantly, such descriptions have been proven incorrect since connected speech may occur in all registers, including academic and formal settings (Brown, 1977, pp. 2-3; Brown & Kondo-Brown, 2006, p. 5; Ito, 2006a; Rogerson, 2006). In fact, Underwood (1989) explains how difficult it is to draw the boundaries of formality/informality as follows:

> …for the language learner the division is not as neat… It frequently happens, for example, that a lecturer delivering a very formal lecture from a prepared set of notes switches to informal language when making an aside or recounting an anecdote as an illustration of a point just made. Or a person involved in describing a complicated phenomenon to a friend over coffee may switch in and out of

> formal and informal styles depending on whether he/she is describing the phenomenon or commenting on it. Between the extremes, there is a range of formality/informality depending on the social setting, the relative ages and status of the speaker and listener, their attitudes to each other and the topic, the extent to which they share the same background knowledge, and so on. (p. 14)

In this vein, it does not seem reasonable to be too restrictive in making claims regarding the contexts in which connected speech is used, which means that depriving L2 learners from exploring the features of connected speech might not be the best choice to make in a language classroom. In fact, connected speech could potentially help L2 learners feel socio-linguistically more advantaged, and even when considered as a marker of informality, it might help learners determine any “switching” between informal and formal use of language in spoken discourse (Underwood, 1989). Unfortunately, this aforementioned stereotypical “sub-standard” (Brown & Kondo-Brown, 2006, p. 5) or “informal-only” view of connected speech prevails among many teachers and listeners, which is, in fact, one of the reasons why teachers tend not to teach it and learners tend not to consider it a priority in their English learning experience (Brown, 1977, p. 3; Gilbert, 1995, p. 105).

2.2 MOTIVATION FOR TEACHING AND LEARNING CONNECTED SPEECH

Besides its perceived “substandard” status, there are various other reasons for this reluctance to practice connected speech in an L2 classroom. One is teachers’ lack of knowledge of these forms and of appropriate methods and techniques to teach them (Ito, 2006a). Teachers are not usually familiar with these structures, especially in an EFL context, or even if they are, the instruction is not “systematic” enough for learners to make generalizations (Rogerson, 2006, p. 91). These challenges are exacerbated by the lack of materials and of sufficient time to devote to teaching connected speech in classrooms simply because they are not included in the curriculum (Brown & Hilferty, 1986/2006; Henrichsen, 1984; Joyce, 2013; Rogerson, 2006; Underwood & Wallace, 2012). All these result in an absence of focus on these forms in EFL and even in ESL classrooms, despite their significant role in L2 listening as well as speaking.

An important motivation for learning and teaching connected speech is that unlike the common belief that regards connected speech as a part of informal language only (Weinstein, 1982, p. vvi, but also see Brown, 1977, p. 3 and Rogerson, 2006, p. 93), it, in fact, occurs in all registers including formal speech contexts (Brown, 1977, p. 2; Brown, 2006; Ito 2006a; Rosa, 2002), and has been shown to play a crucial role in L2 listening comprehension. A comprehensive study by Joyce (2011) looked at the relationship among linguistic knowledge, psycholinguistic sub-skills and L2 listening proficiency to investigate the factors that may help determine the L2 listening ability most. His findings suggested that knowledge of connected speech processes, phonological modification knowledge, as he calls it, was one of the two individual sub-skills "most closely related" to the "latent L2 Aural Processing" having a statistically significant contribution (r = 0.73) (p. 86). As an implication of his study, he encourages test designers to make "the ability to accurately process….reduced forms" a part of their goals in designing their testing tools by adding that this information could be used to "adjust the difficulty" level of a listening test item as an indicator of proficiency (pp. 87-88). In fact, this is also in line with Kostin's (2004) study which investigated the factors affecting the difficulty of the Test of English as a Foreign Language (TOEFL) dialogues and found "sandhi-variation” as one of the phonological variables causing L2 listeners to have difficulties in comprehension. Thus, we can claim that being acquainted with connected speech features is highly important in understanding native or highly proficient speakers of English in a variety of contexts, including high-stakes testing conditions.

Brown’s (2012) analogy of buildings explains, in a nutshell, why we should keep connected speech as a part of our curriculum in pronunciation teaching. As Brown puts it, teaching only phonemes in a classroom is “like giving the students a pile of bricks and expecting them to be able to put them together and make a building” (Brown, 2012, p. xi). Instead, he suggests teaching phonemes along with explanations as to how they change in context. This way, learners can recognize these phonemes not only in isolation but also when pronounced in real speech contexts. It gives learners a “pragmatic” advantage and enables learners “to adjust to various sorts of context constraints” (Brown, ibid.).

2.3 PREVIOUS STUDIES IN L2 CONNECTED SPEECH PERCEPTION AND COMPREHENSION

The teachability and the effects of instruction on connected speech comprehension and perception have been systematically investigated in several studies (Brown & Hilferty, 1986/2006; Crawford, 2005 & 2006; Ito, 2006a; Matsuzawa, 2006), and the findings showed an improvement in learners’ listening ability. However, before moving on to review these studies, the distinction between the terms comprehension and perception needs to be made clear. For the purposes of the present study, perception refers to the perceptual processing taking place during the initial stages of oral language comprehension according to the model provided by Anderson (1995 as cited in

Vandegrift & Goh, 2012, pp. 21-22). During perception, learners recognize the smaller parts of the incoming input such as phonemes, pauses or other acoustic information. Then, they use word segmentation skills to parse the stream of sounds into meaningful units, which is considered challenging especially in recognizing words in connected speech. In other words, listeners pay attention to the language itself without utilizing the meaning carried by them, and this "utilization" is what is called comprehension. In fact, according to Anderson's model, perception is needed for a successful comprehension to take place because the information gained in perception is normally used for comprehension. Therefore, it is important to understand the difference between these two terms to better evaluate the findings of the studies exploring connected speech.

One of the pioneering studies looking at connected speech by Brown and Hilferty (1986/2006) investigated the effects of four weeks of daily ten-minute instruction on connected speech in an EFL context. Their findings suggested an improvement in Chinese university students' connected speech abilities (*n* = 32) on a dictation and an integrated grammar test, but there was no improvement on the general listening comprehension test. However, since the dictation test was presented in context, it has been argued that this might have affected the reliability of the test, as listeners had a chance to guess from contextual information due to the "interconnected nature of the conversation" (Joyce, 2013, p. 81).

A more recent study by Matsuzawa (2006) showed that instruction on reduced forms can in fact lead to improvement in L2 listening ability in an EFL context. After a total of four hours of instruction over a month in understanding connected speech, Japanese businessmen (*n* = 20) took a post-test in the form of a cloze-test. The cloze-test used in this study used blank parentheses corresponding to the number of words in each sentence. In other words, the only difference of such a cloze-test from a dictation test is providing the respondents with the number of words in a sentence. Matsuzawa used such an approach to testing connected speech because he wanted to make sure that learners both understood "the sound of a reduced form and the grammatically correct meaning of an utterance" (p. 61). Therefore, in scoring the data, he did not count it correct if a respondent wrote down 'cats' instead of 'cat's' in the following sentence: "The cat's been sick since Monday". According to him, it shows that the listener did not really comprehend the meaning of the sentence (p. 61). The results of the pre- and post-tests showed a significant improvement on post-test scores. In fact, the validity of cloze-tests has been previously questioned. It is true that there is less burden on test-takers' working memory while taking a cloze-test as opposed to taking a dictation test. Test-takers do not need to keep the whole sentence in mind while it is being dictated, as they only need to focus on one or two missing words. In contrast, in a dictation test, they are expected to write down every single word of the dictated text. This means putting more burden on working memory when taking a dictation test as opposed to a cloze-test. However, in a cloze-test, since learners see the rest of the sentence and know how many words there are in each sentence, their likelihood of guessing the missing word(s) is higher (Joyce, 2013). Possibly to avoid such a problem, unlike the traditional cloze-tests, Matsuzawa used a different type of 'cloze-test', in which she gave the number of words in a sentence using blanks, but did not provide any of the non-target words (e.g., ____ _____ _____ ? for "Where are you?") This solution may help to a certain extent to address the problem of guessing words from context stated by Joyce (2013); however, before administering such a test, respondents should be informed about what exactly corresponds to a ‘word’ or whether a contracted form is counted as one or two words.

Underwood and Wallace (2012), who looked at both production and comprehension, also found a significant improvement in Japanese learners’ connected speech comprehension and their self-confidence in conversational ability following ten, weekly instructional periods. Although the findings showed significant improvement in both production and comprehension, there was no correlation between learners’ ability to comprehend reduced forms in a listening test and their production in a spontaneous peer conversation.

Similarly, Alameen (2014) looked at the effects of different instructional methods on the ability to perceive and produce linking as a connected speech phenomenon in L2 English. Her results indicated no significant improvement on the perceptual ability based on the results of the dictation test; however, significant values were reached in the ability to produce linking.

Limited previous research on the production aspect suggests that the production of connected speech features improves over time when such features are practiced in a traditional classroom context (Underwood & Wallace, 2012) as well as using computer assisted language learning (CALL) (Yang, Lin, & Chung, 2009). However, there seems to be a disagreement among researchers and language practitioners as to whether or not producing connected speech should be taught in classrooms. Norris (1993, 1995) suggests that the purpose for learners should be the recognition of connected speech features in order to communicate well, rather than imitating native speakers’ use of connected speech features in learners’ own speech. Brown (1977) also explicitly disapproves of teaching students to "*produce*" these "assimilated" or "elided" forms because "sophisticated students who have been taught to be *aware* of these forms will introduce them into their own speech in a natural context when they feel able to control them" (pp. 156-7). However, she finds "the failure" to understand these forms as "disastrous for any student who wants to be able to cope with a native English situation" (p. 157).[1]

As may be seen, the literature agrees on the prominence of teaching the perception and comprehension of connected speech features more than producing them, mainly because the primary goal of pronunciation teaching is accuracy in perception and comprehension, followed by production. Thus, although production of connected speech has been studied and been found to be helpful (Underwood & Wallace, 2012; Yang, Lin & Chung, 2009; Alameen, 2014), for the purposes of this study, the teaching of

[1] This can be criticized as not taking into consideration such approaches to English as the Lingua Franca Core (LFC) or English as a Lingua Franca (ELF), a set of prioritized codes sufficient for communication to take place among different L1 speakers of L2 English. ELF is considered to be a part of English as an International Language (EIL), which typically takes native speakers into the bigger picture. According to ELF approach, for instance, producing certain phonological features of English (e.g., voiceless interdental fricative, [θ]) is not necessary for L2 speakers of English to understand each other (Seidlhofer, 2005), and thus, such features should not be considered a priority in L2 English classrooms. However, it should be noted while some learners of L2 English do not need to follow native-speaker norms in learning English, there are others who are motivated to learn and speak native-like L2 English. In addition, although learning and teaching connected speech forms may not be considered a part of ELF, Jenkins (2000, 2002) suggests that one should improve their ability to understand these forms if they are expecting to have considerable contact with native speakers.

connected speech forms was limited to perception rather than production. Previous literature on connected speech comprehension has shown that learners benefit from instruction in making progress on improving their comprehension of connected speech forms, although this does not necessarily bring about similar immediate gains in general listening comprehension.

This study aims to contribute to the existing literature of connected speech perception in various ways. First, it explores a single aspect of connected speech perception (i.e., word-boundary palatalization) rather than investigating a variety of forms simultaneously, as most previous studies have done. This is especially important to have a more nuanced look at and to make stronger claims regarding different types of connected speech processes. Second, the present study employs a rather different methodology to assess connected speech by using a two-option forced-choice test, which will provide a fresh look at the assessment of connected speech perception. The use of this task is very important because the inclusion of a variety of student proficiency levels in this study meant widely ranging lexical knowledge, listening comprehension skills, note-taking and spelling abilities, all of which would be hard to control and would confound the results of the study if a dictation test were used. Therefore, in order to make it a more focused test of perception rather than a test involving multiple processing skills, a two-option forced-choice perception test was used, and to date, no known study has used such a method to assess connected speech perception. Finally, as this study involves multiple classrooms, instead of the instruction taking place in a classroom, the presentation and practice of the target forms are given via online training in computer labs, which may reveal interesting findings as to the use of technology in teaching connected speech forms. Taken together, the present study hopes to shed new light on understanding the learning and teaching of connected speech.

2.4 FORM-FOCUSED INSTRUCTION AND ATTENTION IN L2 LISTENING

A brief look at the history of pronunciation teaching reveals that the focus and the type of approach to pronunciation teaching (i.e., behavioral, cognitive, psychological) has shifted as new language teaching methods were introduced. While previously the Audiolingual Method viewed pronunciation as one of the most prominent components of the language learning process, Communicative Language Teaching (CLT) regarded it to be less important because according to this method, for pronunciation improvement, "input alone" would suffice. This lack of interest in pronunciation was understandable because the main emphasis in CLT was "communicative competence", and pronunciation ability was considered to be more of a linguistic ability and, thus, did not receive much attention until recently (Isaacs, 2009; Morley, 1994).

Another drawback of CLT in terms of pronunciation teaching is that CLT deemphasizes the role of focus on form and disapproves of learning through non-meaningful, mechanical exercises. In other words, repetition, drills and informative explanations on forms are usually frowned upon, as they are not typically performed in a meaningful context. Since teachers who are schooled in CLT consider many known methods of pronunciation teaching incompatible with CLT, they end up not being comfortable teaching pronunciation at all.

However, only in the last couple of decades, cognitive science has shown the role of cognitive variables in learning and teaching languages by emphasizing the facilitative roles form-focused instruction and attention play in L2 learning (Ellis, 2001; Schmidt, 1995). According to Ellis (2001), learners are more likely to recognize the target forms in the input when they are explicitly instructed on them. Other studies have also shown that explicit instruction on pronunciation is more likely to benefit those learners who pay more attention to form (Trofimovich & Gatbonton, 2006). Given that attention is essential for (phonological) learning to take place (Schmidt, 1990, 1995), the next step is finding the best ways to help learners pay attention to forms. One challenge for this in learning connected speech is the problem of "perceptual saliency", which refers to features of speech making "certain features of the input more comprehensible, and thus more liable to become intake" (Henrichsen 1984, p. 106). Since perceptual saliency plays a crucial role in determining the ease of learning certain L2 features, Henrichsen (1984) sees "reduced forms" at a disadvantage as such forms are not salient by nature. Later studies have also supported this notion, and shown that when forms are phonetically more "salient", they are more easily noticed regardless of the proficiency level or task effects. A study by Kim (1995) investigated the role of attention in understanding speech at different speech rates and found that listeners paid less attention to speech when it was read at a faster rate. Therefore, according to him, listeners should be "encouraged to move from a more lexical mode […] to a more syntactic mode" (p. 78) because this way, they would be able to comprehend connected speech processes occurring across word boundaries, which would otherwise go unnoticed (Ito, 2006a, p. 23). These findings indicate that training learners to "notice" particular forms might prove helpful not only for L2 learning in general, but also specifically learning connected speech forms (Kim, 1995). Therefore, the present study will follow form-focused instruction (FFI) in teaching the target forms to learners to increase the likelihood of noticing these less salient forms

(Schmidt, 1990, 1995). However, one thing to note is that the term *training*, as opposed to *instruction*, is used in the present study, as there is not student-teacher or student-student interaction similar to what would normally occur in a language classroom. The only off-line interaction is between the speaker in the training videos and the student watching and doing the exercises shown in the video.

The concept of attention is further discussed, but from a more cognitive approach later in Chapter 3. The last section of this chapter discusses the status of palatalization as a connected speech phenomenon in English, which is the target linguistic form being investigated in this study.

2.5 PALATALIZATION AS A CONNECTED SPEECH PHENOMENON

As word-boundary palatalization in English is the target content to be learned by the participants in this study, this section discusses various approaches to explaining palatalization as a connected speech phenomenon.

Kochetov (2011) describes palatalization as "a phonological process by which consonants acquire secondary palatal articulation or shift their primary place towards or close to the palatal region, [which] usually happens under the influence of an adjacent front vowel and or a palatal glide." In this palatalization process, segments form a complex, yet single, segment (Gussenhoven & Jacobs, 1998, pp. 180-182), which is especially important in learning and teaching the syllable structure and pronunciation of English. The region that is referred to as *palatal* is the region behind the alveolar ridge, marked as "corner" in Figure 2.1, behind the end of the hard palate. This is also the region where alveopalatal and palatal places of articulation are located (Keating, 1991, p. 32; Bateman, 2007, pp. 5-6).

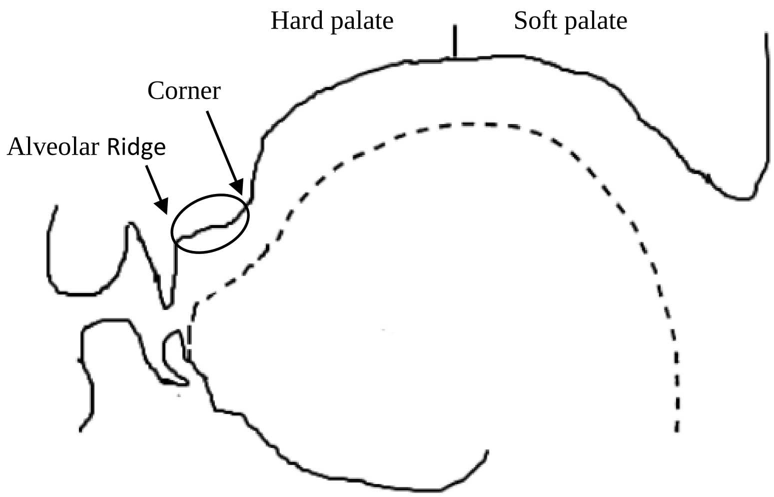


Figure 2.1. Overview of the palatal region of the vocal tract. Adapted from "Coronal Places of Articulation," by P. A. Keating, in C. Paradis & J.F. Prunet, (Eds.), *Phonetics and Phonology, The special status of coronals: Internal and external evidence 2* (p. 32), 1991, San Diego, USA. Copyright 1991 by Academic Press.

In this study, of the three types of palatalization processes in English described in Kochetov's (2011) work, the discussion is limited to coronal palatalization in English and other alternative approaches to interpretation of coronal palatalization within and across word boundaries.

Coronal palatalization in English describes a phonological process (Chomsky & Halle, 1968, pp. 230-234), in which word-final alveolar obstruents [t, d, s, z] are palatalized and become palato-alveolar consonants [tʃ, dʒ, ʃ, ʒ] when they are followed by the palatal glide [j]. To put it another way, coronal palatalization can be described as an alternation between [+anterior] coronals and [-anterior] coronals in the context of a palatal approximant glide resulting in tongue raising. It is also indicated that in coronal palatalization, stops tend to become affricates, and other consonants usually keep their manner of articulation (Bhat, 1978). This process originates from the fact that while [-

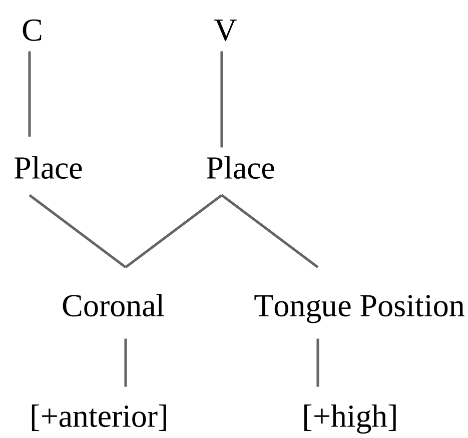
anterior] coronal sounds [tʃ, dʒ, ʃ, ʒ] in English are produced with a raised tongue position, the production of [+anterior] coronals [t, d, s, z] involves a lowered tongue position (Gimson, 1989, p. 176), which is displayed in (1) and (2) below. Kochetov considers this an ‘assimilatory’ process because the consonants undergoing palatalization change their place of articulation in a way to become phonologically closer to the place of articulation (i.e., hard palate) of the segment generating the palatalization. Furthermore, while all four sounds [t, d, s, z] change place of articulation, the alveolar stops [t, d] also become sibilant affricates through assibilation. Examples for within-word palatalization of the sounds [t, d, s, z] in English, which is not an optional process, and a feature geometric account of the process are given below in (1) and (2), respectively.

(1)	legisla[t]e	‘legislate’	legisla[ʧ]ure	‘legislature’
	resi[d]ue	‘residue’	resi[ʤ]ual	‘residual’
	pre[s]	‘press’	pre[ʃ]ure	‘pressure’
	clo[z]e	‘close’	clo[ʒ]ure	‘closure’

(2) Coronal Palatalization (adapted from Cavar 2003 adopting Lahiri & Evers, 1991)

Consonants become coronal [−anterior]:

C → palatalized / __ j



Kreidler (2008) explains the within-word palatalization shown in (1) as follows: “the /j/ of the suffix produces a change in the final consonant of the base word. The four apical consonants are replaced by the corresponding laminal consonants before /j/ and a vowel” (p. 247). Then, after the palatalization process occurs, this time the /j/ phoneme is dropped if the following vowel is unstressed, which is also called the ‘j-drop rule’ (p. 248).

While coronal palatalization occurs within words, as shown in (1) and (2), it can also occur across word boundaries when the word-final alveolar sounds [t, d, s, z] are followed by a word-initial [j] as exemplified in (3) below. As in the palatalization that occurs within words, the resulting sounds in word-boundary palatalization are the palato-alveolar consonants, [tʃ, dʒ, ʃ, ʒ]. However, unlike the lexical palatalization described in (1), this post-lexical palatalization occurs optionally.

(3)	[kʊd jə]	~	[kʊdʒə]	‘could you’
	[nɑt jɛt]	~	[nɑtʃɛt]	‘not yet’
	[ðɪs jir]	~	[ðɪʃir]	‘this year’
	[iz jərˈsɛlf]	~	[iʒərˈsɛlf]	‘ease yourself’

In this study, the focus is the type of post-lexical palatalization described in (3) above. A majority of linguists and language researchers have used the general term ‘palatalization’ to refer to processes that include but are not limited to the ones in (1) and (3). On the other hand, some researchers either used process-specific terminology (see Bateman, 2007 for a discussion), or they simply expressed their opinions on the irrationality of using the term ‘palatalization’ to describe the processes shown in (1) and

(3).[2] Other than the discussion of which term to use, there is an additional controversy as to whether lexical and post-lexical occurrences of 'palatalization' can be analyzed as two separate processes (Zsiga, 1995, 2000). In the remainder of this section, these different approaches to lexical and post-lexical palatalization are presented.

In her work, Zsiga (1995) looked at lexical and post-lexical palatalization of the alternation of [s] before [j] in American English to understand whether these two are distinct processes. Previous researchers (see Zsiga 1995, p. 282) had claimed that across word boundaries, palatalization could be gradient. Therefore, building on this assertion, she used acoustic and electropalatographic (EPG) methods to confirm this hypothesis. Her findings on the alternation of [s] before [j] showed that the palatalization process across word boundaries was gradient and variable while lexical palatalization was categorical. Categorical alternations do not normally vary in acoustic or articulatory ways. Therefore, she concluded that lexical and post-lexical palatalization are, in fact, distinct processes, and thus, they should be represented in different ways. She further indicates that the gestural overlap, which shows the variable and gradient nature of the process, may be a regular overlap that can be observed between any two sounds across word boundaries and, thus, may not need any specific palatalization rules. In other words, while categorical (lexical) palatalization triggers [-anterior] feature to spread across roots, in gradient (post-lexical) palatalization, there is just an overlap. As a result, according to Zsiga (1995), post-lexical palatalization would be better represented using gestural

[2] According to Shockey (2003), this is partly because the resulting sound is rarely truly palatal and that [j] sound may change a sound to a less palatal one; however, she also agrees that the term is 'well-established' and, thus, will still be used (p. 45).

overlap (Browman & Goldstein, 1986, 1990, 1992 as cited in Zsiga) because "the palatal constriction for [j] overlaps in time with the alveolar constriction for [s] when [s] and [j] are adjacent at a word boundary " (p. 294).

Maintaining that there was no phonological assimilation in what is traditionally claimed to be post-lexical palatalization in her previous studies, Zsiga (2000) aimed to challenge the arguments made against the analyses in her previous studies, this time using cross-linguistic evidence from Russian (see Zsiga, 1993, 1994, 1995). The main argument against her results from Zsiga (1995) was mostly related to her study design due to her limited number of the participants or some methodological problems such as the fit of the electropalates (Scobbi, 1995 as cited in Zsiga, 2000, p. 71). In this study, she addressed these concerns by including more contexts and participants, and by also adding Russian stimuli in order to show the difference between what she calls "an overlap" and "palatalization", as Russian palatalized consonants had a high potential to show this difference. This is because in Russian, the phoneme /s^j/ (as in /$des^j at$/ 'ten') is a phonemically palatalized consonant. According to Zsiga, "both Russian /s^j/ and English /s+j/ involve co-production of an alveolar fricative and palatal approximant; however, the English and Russian 'palatalized' fricatives sound very different" (p. 71). Therefore, Zsiga (2000) compared these two sounds by using spectral and temporal features of the Russian /s^j/ as evidence to show that while Russian /s^j/ is a result of a palatalization process, English /s+j/ is a result of gestural overlap.

The participants were L1 speakers of American English (*n* = 5) who were students of Russian, and L1 speakers of Russian, who were students of English (*n* = 5), although Zsiga (2000) clearly pointed out that no L2 data were reported in the study. Participants were given a set of sentence cards and were asked to read the sentences written on them three times and as naturally as possible. The sentences in English included target phrases such as 'press your point' or 'pressure point', exemplifying gestural overlap and true palatalization account of hers, respectively. Her Russian sentences also included target forms occurring in similar contexts in Russian sentences. Then, the recordings of the participants were acoustically analyzed to see if her gestural overlap account of [s + j] across word boundaries would be supported by the analyses.

Findings demonstrated that there was no statistical evidence indicating that word-boundary [s + j] sequences were forming a complete palatalization. Although for some speakers, the analyses revealed some kind of assimilation to certain degrees, only two of the 40 phrases showed complete assimilation. Her overall findings show that there is gradient palatalization in English, but not in Russian, and that the consonant clusters overlap more at word boundaries in English than in Russian. A partial overlap combined with blending brings about a gradient palatalization, which partially supports Zsiga's (1995) account of simple overlap.

The discussions show that there is need for more research to establish the status of post-lexical 'palatalization' as a process going beyond a simple overlap. Despite the discussion of Zsiga (2000) showing counter-evidence, the present project will adopt an account which considers word-boundary co-articulation as palatalization rather than mere overlap.

2.6 CHAPTER SUMMARY

This chapter provided a discussion of how to define and teach connected speech in light of previous literature. Some of the challenges of teaching connected speech as well as reasons why connected speech instruction receives less attention were also summarized. Because the present study uses online video-training following form-focused instruction to teach the target forms, the chapter also reviewed related theoretical frameworks. The last section of the chapter discussed palatalization within a linguistic framework as the specific type of connected speech form investigated in this study.

The following chapter will examine L2 listening from a cognitive perspective and discuss the role of attention control as a cognitive skill as it relates to L2 listening in general as well as in terms of L2 connected speech perception.

CHAPTER 3

COGNITIVE PROCESSES IN SECOND LANGUAGE LISTENING AND CONNECTED SPEECH PERCEPTION

3.1 BOTTOM-UP AND TOP-DOWN APPROACHES TO L2 LISTENING

Being able to understand connected speech is one of the major factors affecting successful L2 spoken text perception and comprehension (Campbell, et al., 2007, p. 5; Henrichsen, 1984; Ito, 2006a; Richards, 1983; Shockey, 2003). However, as noted at the outset, having learned the dictionary forms of words pronounced in isolation (Brown, 2011, p. 40), L2 learners are usually unable to understand authentic language regardless of the amount of formal instruction received for years. This is partly because “learners do not know how known words sound when they are put together in connected speech” even when they are familiar with the text and the vocabulary (Hagen, 2012, p. 2). Learners cannot recognize the way sounds, syllables, and words change in spoken discourse because this requires word recognition and lexical segmentation skills. In other words, in order to be able to perceive and decode spoken input successfully, listeners should be taught how to distinguish word boundaries in spoken words (Norris, 1993, p. 1; Kuo, 2012). Therefore, a better understanding of word recognition (Shockey, 2003) and lexical segmentation (Field, 2003; 2008c) is critical because an inability to segment and recognize words is one of the most commonly encountered problems hindering L2 listeners’ decoding and meaning building. A failure to understand connected speech is also considered to be a lexical segmentation problem (Field, 2003). In a study by Goh

(2000), who looked at the problems L2 listeners had during listening comprehension, five out of ten problems were perception problems, and two of these were directly related to connected speech. The first of these was being unable to chunk streams of speech, and the second problem indicated in the study was learners' inability to recognize the words they already knew. Both of these listening comprehension problems, which are related to connected speech perception, can be overcome through learning how to decode speech rather than merely using strategy training on how to use contextual information. To put it another way, if L2 learners wish to 'keep up with the acoustic blur' of naturally occurring spoken English (Brown, 1977), they may need more than a mere reliance on their ability to analyze only the most important aspects of the text by ignoring other specifics. In addition to this skill, they also should be taught how to cope with the specific phonological features that characterize connected speech because it does not seem possible to understand spoken texts without being able to decode the message (Norris, 1994; 1995; Al-jasser, 2008).

Equipping L2 listeners with such a skill is not an easy task because as Grosjean (1985) pointed out "word recognition is not a word-by-word, left-to-right process. Rather, the process is very much a feed-forward, feed-back operation, where there are constant adjustments being made to early and/or partial analyses" (as cited in Field, 2008c, p. 39). What Grosjean refers to in his article now corresponds to the widely debated top-down and bottom-up processing approaches to teaching L2 listening. Therefore, it is important to discuss these two processing approaches, which originally derive from cognitive psychology, within an SLA-oriented framework (Field, 2008c). Richard (2008) summarizes these two kinds of processing approaches to understanding spoken language as follows:

> Bottom-up processing refers to using the incoming input as the basis for understanding the message. Comprehension begins with the received data that is analyzed as successive levels of organization – sounds, words, clauses, sentences, texts– until meaning is derived. Comprehension is viewed as a process of decoding. Top-down processing, on the other hand, refers to the use of background knowledge in understanding the meaning of a message. Whereas bottom-up processing goes from language to meaning, top-down processing goes from meaning to language. (pp. 4-7)

In the history of L2 listening instruction, teaching in a way that is primarily bottom-up could be described as an era of "text-oriented instruction". In this type of listening instruction, there is a great emphasis on decoding and imitation; thus, meaning construction is considered to be built in an "incremental manner from individual sounds to words, to strings of words and, eventually, to a complete text." (Vandergrift & Goh, 2012, p. 6). However, what needs to be pointed out here is that the spoken texts employed in the 1950s and 1960s were, in fact, nothing but "traditionally written passages read aloud" (Vandergrift & Goh, 2012, p. 7), which may mean that they did not really represent the actual spoken language used. This is an important point to note because given the lexical and syntactic complexity of the written texts, learners might have needed a full deciphering of the text more than anything else (Vandergrift & Goh, 2012). However, this should not be taken as devaluing the contribution of bottom-up approaches to today's L2 listening instruction, in which more authentic spoken texts are used. In fact, this dissertation has set out to demonstrate the importance of bottom-up approaches in improving L2 listening perception. Specifically, as is clear from Richard's definitions above, understanding connected speech processes primarily requires bottom-up processing because it requires decoding the elements of spoken texts in the first place, which will be discussed in detail in the upcoming section.

Since the beginning of the 1970s, top-down approaches to L2 listening have been on the rise mainly because they go hand in hand with Communicative Language Teaching (CLT), whose focus has been on improving communicative competence. This approach mainly disfavors teaching and learning L2 listening in a bottom-up way. While meaning construction in bottom-up processing is realized from the smaller units to larger units, calling on linguistic knowledge, in top-down processing, the meaning building is realized in the opposite direction by drawing on background information, or schema. Top-down, or communication-oriented, L2 listening instruction aims to help learners in improving their "micro and macro skills" by underscoring the importance of metacognition and background knowledge (Vandergrift & Goh, 2012). This background knowledge may either be "previous knowledge about the topic of discourse, situational or contextual knowledge", or it could refer to the "knowledge in the form of 'schemata' or 'scripts'" (Richard, 2008). However, various researchers have challenged the adoption of a fully top-down processing approach to L2 listening instruction. While they suggest an interactive approach in which L2 listeners are ideally using their linguistic and non-linguistic knowledge rather than relying merely on one or the other for listening comprehension, some especially underscore the importance of being cautious about the use of contextual information to compensate for the inadequacy of decoding input. This is because the fundamental goal of L2 listeners should be hearing the actual spoken text and having less reliance on guessing from context (Wilson, 2003). Two major studies providing empirical evidence in favor of an interactive processing model are summarized below:

Tsui and Fullilove (1998)

Tsui and Fullilove's (1998) highly cited, long-term study found that successful Chinese learners of L2 English can be identified as those who can use their bottom-up rather than top-down processing because "schema" cannot always be adequate to make up for the lack of speech perception or word recognition (Buck, 2001; Rost, 2002). One of the items they gave as an example asked, "What saved the estate from burning down?" When such a question was asked, L2 listeners showed signs of creating a schema of fire fighters, which lead listeners to choose fire services as the correct answer. At first, in line with general schemata building expectations, the passage mentioned fire fighters and how they worked hard to extinguish the fire. However, later in the passage, it stated that what really saved the estate was the wind's blowing in the opposite direction. Less-skilled listeners were reported to score low on such items. This proves how relying primarily on context without decoding the input may mislead L2 listeners. The fact that more-skilled listeners were overall better able to use the top-down information in this study is not surprising. It is simply because more-skilled listeners have better decoding abilities and, thus, are able to make use of both top-down and bottom-up listening compensatorily.

Yi'an (1998)

Using an immediate retrospection verbal report procedure, Yi'an's study aimed to investigate how L2 listeners employed linguistic and non-linguistic knowledge sources when taking a multiple-choice test. In this study, the test takers were of intermediate proficiency and above. The findings showed a parallel activation of top-down and bottom-up listening procedures regardless of the proficiency level; however, he indicated that there were differences in the ways L2 listeners employed them during testing. Proposing to make a distinction between the "compensatory" versus "facilitating" functions of top-down information, Yi'an suggested that while less-skilled L2 listeners were more likely to employ the former function due to their less successful linguistic processing, more skilled L2 listeners were using facilitating functions of the top-down information. However, Yi'an points out the prominent role linguistic knowledge exerts in listening comprehension because (a) when listeners were only partially successful in their use of linguistic knowledge, the compensatory use of non-linguistic information did not always lead to correct comprehension of the listening passage, and (b) some listeners with partially successful linguistic processing allowed their non-linguistic knowledge to prevail over otherwise correct interpretations they had made based on linguistic knowledge.

These two studies summarized above mention the notion of "compensatory" functioning of top-down and bottom-up processing to make sense out of the input. This notion has its roots in Stanovich's (1980) Interactive Compensatory Hypothesis, which aims to explain how L1 readers process reading passages. According to this hypothesis, the more the readers are confident about the input, the less they will depend on non-linguistic information when they read and vice versa. In the case of L2 listeners, however, the challenge is that decoding might be a crucial problem for overall comprehension; thus, making use of compensatory strategies might not bring about success, as evidenced by research (Yi'an, 1998).

Attempting to demonstrate how linguistic and non-linguistic knowledge are employed by L2 listeners of varying proficiency levels, these studies contributed to literature to help revise the definition of a skilled L2 listener in a way that decoding abilities in addition to use of background knowledge could be used to describe a skilled L2 listener. Moreover, their findings can potentially encourage L2 researchers to reconsider the current position of bottom-up listening within L2 listening instruction and research in general.

3.2 SPOKEN WORD PROCESSING AND CONNECTED SPEECH PERCEPTION

One notion in Stanovich's hypothesis is the key in understanding the relationship between top-down and bottom-up processing in L2 listening: interactivity. As Field (2008a) indicates, both decoding (bottom-up) and meaning-building (top-down) are, in fact, interdependent in nature and this relationship is very much dependent on the way they interact rather than being a constant process, and the contribution of each processing skill in this interaction varies depending on the situation (pp. 134-135). Therefore, rather than studying them as alternative notions, they should be considered mutually dependent (Field, 2008b, p. 3).

This interaction between the bottom-up and top-down processing has in fact been studied for a couple of decades in the broader literature of speech perception and lexical access. Massaro's Fuzzy Logical Model of Speech Perception (FLMP) describes bottom-up processes as "mental operations that analyze the acoustic properties of a given speech stimulus" while top-down processes are those which "activate a set of potentially matching phonological representations [by using] information […] to select the best possible candidate from among the set of candidates activated by bottom-up processes" (as cited in Traxler, 2012, p. 69). Therefore, the integration of the information coming from bottom-up and top-down sources is seen as essential in speech perception and processing (Masssaro & Chen, 2008). The TRACE model, which is a top-down, interactive model of lexical access for visual and auditory input, also supports an interaction between different levels of input in order to activate the correct word form. According to the TRACE model, activation in the word-level will affect the letter-level processing at lower levels through interactive processing (Traxler, 2012, pp. 106-108). These recent models of speech perception and lexical access support the existence of interaction between bottom-up and top-down information in the activation and recognition of lexicon. Despite the differences they have, their view of non-serial word recognition by using information coming from different processing units support the interaction of bottom-up and top-down processing in word recognition and speech perception.

In the second language acquisition literature, cognitive processes involved in the realization of bottom-up and top-down processes in listening are illustrated by Vandergrift and Goh (2012) in Figure 3.1.

As Figure 3.1 shows, during decoding, listeners initially process the stream of speech or the input with very little involvement of prior knowledge. Although linguistic knowledge (i.e., phonology, lexicon, and syntax) is crucial at this stage, if such an approach is not combined with the top-down knowledge, it is not enough for comprehension.

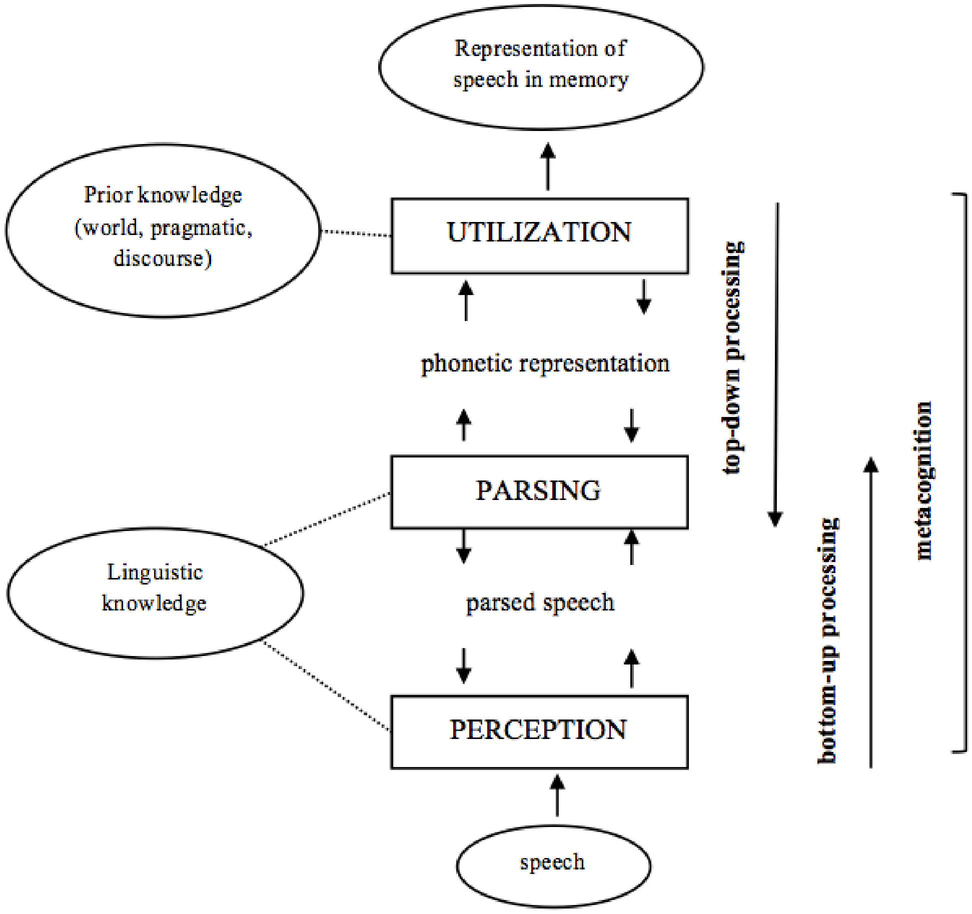


Figure 3.1. Cognitive processes and knowledge sources in L2 listening comprehension. Adapted from *Teaching and Learning Second Language Listening: Metacognition in action* (p. 27), by L. Vandergrift and C. Goh, 2012, New York, NY: Routledge. Copyright 2012 by Taylor and Francis. Reprinted with permission.

Similarly, the top-down processing component alone would not suffice for successful comprehension because without the necessary linguistic knowledge, successful decoding seems unlikely (Vandergrift & Goh, 2012). What this means for classroom teaching is that while teachers encourage the use of co-text, the linguistic environment of a word, and context in processing L2 aural input, they should not disregard the value of perceiving the speech signal itself. Therefore, in addition to possessing relevant strategies to use top-down information in aural text processing, learners should also be equipped with the necessary skills to be able to decode speech appropriately in order for meaning construction to occur.

However, since linguistic knowledge is the primary component of decoding, learners are faced with various challenges in developing L2 listening skills. This is mainly because listening is an online process in which the listener is to “decode the sounds of speech at a delay behind the speaker of as little as a quarter of a second, and this is about the length of a syllable” (Field, 2008a, p. 129). Viewed in this light, it can be claimed that a syllable, as well as sounds or phonemes, apparently play an indispensable role in speech perception and comprehension. This seems very much linked to the online nature of the listening activity, which is not usually fully controlled by the listener unlike reading, for instance, so listeners have to determine the boundaries of each syllable, word, or phrase, as they perceive speech. The primary challenge is that most spoken texts are not fully articulated and are more or less authentic; consequently, listeners usually need to cope with a variety of issues resulting from connected speech processes and stress assignment while listening to spoken texts. In this regard, the ability to segment words, sometimes even syllables and sounds, is crucial for decoding, but it can be a very challenging task for L2 listeners to recognize the word boundaries due to various reasons. One of these reasons is L1 transfer. The knowledge of the already existing L1 segmentation procedures that learners apply in their native languages is inevitably transferred to their L2 segmentation procedures. This happens more during the initial stages of L2 learning, and might pose problems when listening to L2 spoken texts, especially if the L1 and L2 word segmentation procedures and rhythm are too distinct. When allophonic and phonotactic structures of the native and the target languages differ widely, L2 listeners would have more difficulty in word recognition and lexical segmentation. Another related factor that makes word segmentation challenging for L2 listeners is L2 learners' unfamiliarity with the stress patterns in English. English has been commonly described as a stress-timed language; therefore, certain groups of words such as function words are naturally realized as weak syllables, and are not stressed except for emphasis or contrast. This phonological feature of English might add to the difficulty of decoding it if L2 listeners are not familiar with the stress patterns in English mainly due to the saliency issues. To summarize, the extent to which L2 listeners are able to utilize allophonic, prosodic, phonotactic and other cues might determine the success in L2 segmentation (Vandergrift & Goh, 2012). In addition, researchers also found factors such as frequency, density (the number of competing words), and the recent activation of the words, as having an effect in word activation processes in L2 listening (see Cross, 2009, pp. 34-35).

In light of the models and findings summarized so far, it becomes clear that despite the undeniable importance of making use of non-linguistic information in L2 listening comprehension, L2 listeners cannot accomplish successful comprehension of L2 spoken language by solely relying on their non-linguistic knowledge. However, it should be made unequivocal that rather than contrasting the contribution of these two knowledge sources (see Tsui & Fullilove , 1998; Yi'an, 1998), the present project aims to further explore the *perception* aspect of the entire process, which is very central to listening comprehension as evidenced in this section. In doing so, it hopes to draw attention to the prominence of listening for perception, rather than meaning building alone, which has been swept under the proverbial rug in recent years.

So far, Chapter 2 and 3 have (a) underscored the importance of teaching and learning connected speech to improve L2 listening skills, (b) established why and how connected speech comprehension is an important bottom-up skill that could help improve L2 listening perception especially through instructional methods, and finally (c) demonstrated how connected speech perception is viewed in cognitive psychology. The following two sections discuss the role that attention control plays in L2 phonological learning, and how this role of attention may be explained within an Aptitude-Treatment Interaction framework (Cronbach, 1953, 1957; Edwards & Cronbach, 1952; Snow, 1991).

3.3 INVESTIGATING THE ROLE OF ATTENTION CONTROL IN L2 PHONOLOGICAL LEARNING

3.3.1 A COGNITIVE APPROACH TO ATTENTION

In investigating individual differences, the contribution of cognitive psychology to the field of SLA is undeniable. It not only provides us with innovative methods to investigate established constructs in the field, but it also enables SLA researchers to understand the underlying reasons for learning outcomes and related findings by tapping into individuals' cognitive processing. Attention is one of these cognitive abilities whose role in L2 learning has been established in the field of SLA as well as cognitive psychology (Schmidt, 2001). Research by cognitive psychologists (Carlson & Dulany, 1985; Carr & Curran, 1994; Posner 1992, 1994) as well as SLA researchers approaching L2 learning from a more cognitive perspective maintain the idea that attention is crucial for learning to take place (Bialystok, 1994; Ellis, 1994; Robinson, 1995).

Despite the well-established status of attention as a key factor in learning, studies are yet to determine how exactly it relates to L2 phonological learning. To be able to explain the role attention plays in phonological learning, it is important to understand the way it is conceptualized in the SLA literature with a pedagogical as well as psycholinguistic focus on attention. Nonetheless, it should be noted that a thorough account of the theories of attention as well as different conceptualizations of attention in cognitive psychology or SLA is not relevant for the purposes of the present study (see Schmidt, 2001; Tomlin & Villa, 1994).

In their review study, Tomlin and Villa (1994) explained how cognitive scientists and L2 researchers viewed attention differently. While cognitive scientists have been more interested in the process of attention itself, L2 researchers have been interested in how attention is related to learning and teaching of languages (Segalowitz, 2010). On the other hand, it can be claimed that a conciliation has been reached between these two perspectives with the integration of psycholinguistic approaches to second language acquisition processes since then, especially increasingly in the last couple of decades.

As defined by Segalowitz & Frenkiel-Fishman (2005), attention is “the ability to shift focus of attention from one language-based attention directing function to another” (p. 646), and similar definitions have also been suggested (Eviatar, 1998; Talmy, 1996). Attention was found to affect L2 phonological learning (Francis, Baldwin & Nusbaum, 2000); thus, it can be argued that attention control is in fact needed in speech processing (Assmann & Summerfield, 1994; Gordon, Eberhardt, & Rueckl, 1993) as well as speech comprehension (Craik & McDowd, 1987; Wood & Cowan, 1995). It has also been suggested that phonological learning needs “intentionally focused attention” more than some other skills due to the fact that learners already have an established phonological system and that they filter the L2 speech they listen to. Therefore, they need to attend to new information in order to process the incoming message because focused attention to form in listening to L2 speech might help learners notice and, consequently, learn these forms (Schmidt, 2001).

Such focused attention requires control of attentional resources. Because controlling attention requires foregrounding and backgrounding of information (Segalowitz & Frenkiel-Fishman, 2005; Talmy, 1996), in speech processing, this means individuals will have to "bring relevant acoustic information to the foreground while keeping irrelevant information in the background" (Segalowitz & Frenkiel-Fishman, 2005, p. 645), which requires a considerable amount of attentional resources (Craik & McDowd, 1987; Wood & Cowan, 1995). More specifically, in terms of phonological AC, more efficient AC in the acquisition of L2 phonology has been shown to lead to better performance scores in processing acoustic-phonetic information in L2 speech (Darcy, Mora & Daidone, 2014; Guion & Pedersen, 2007; Mora & Gilabert, 2012; Mora & Darcy, 2013; Safronova & Mora, 2012a). However, there are also studies that looked at the relationship between phonological development and AC and reported "less clear", inconclusive, or non-significant results (Darcy, Park & Yang, 2015; Darcy & Mora, in press). Below is a review of some of these studies conducted using a similar type of method to measure AC as used in the present study, which is explained in detail later in Chapter 4. Still, it should be noted that each of the following studies is rather broad in scope, so for the purposes of this study, only the sections and findings which primarily deal with AC measures are reported below.

Safronova and Mora (2012a)

In their study, Safronova and Mora investigated the inter-subject variability in the perception of English /iː/ and /ɪ/ contrast by Spanish/Catalan EFL learners and whether and to what extent individual differences in AC predict their L2 vowel perception. They hypothesized that learners with better AC scores would be better at perceiving the vowel contrast in L2 English. To test this, 58 Spanish/Catalan speakers of L2 English took an AXB vowel discrimination task, which asked participants to determine whether the second stimulus was identical to either the first or the third one, and an attention control task. The attention control task was claimed to be novel in that it was designed to be the speech-based version of two previous similar tests which measure attention control using an alternating runs paradigm (see Rogers & Monsell, 1995; Segalowitz & Frenkiel-Fishman, 2005). The task required participants to switch their attentional resources between two alternating but predictable dimensions of voice quality (male or female) and segmental duration (long or short vowel). During the AC task, the screen was split into four brackets (not visible by any lines) and in each, stimuli were presented aurally in a clockwise manner. The participants only saw a picture a loudspeaker followed by the auditory stimulus and were then asked to press a certain key on a keyboard to answer the question on duration or voice quality. The dimensions, and thus, the questions of the top two and the bottom two rows were always no-shift conditions and the shift conditions were the vertical switch between them, which made the alternation predictable. The findings showed that AC scores explained about 31% of the unique variance in the results of L2 vowel perception scores (p <.001), and the researchers concluded that “greater AC may provide learners with an advantage in L2 target-like cue-weighting and may explain inter-learner variation in L2 phonological attainment".

Darcy, Mora & Daidone (2014)

The overall purpose of their study was to examine the role of attention and inhibition in L2 learners' phonological development. In order to investigate this, 16 L2 English learners of L1 Spanish and 18 L2 Spanish learners of L1 English took a speeded ABX categorization task targeting vowel and consonant contrasts to measure perception, and a delayed sentence repetition task to measure production. The tasks measured L2 learners' ability to perceive and produce certain phonemes presented in non-word stimuli. To measure the participants' attention control, the researchers used a novel speech-based attention-switching task. This task, which included two dimensions, asked participants to determine (a) whether the initial sound of the word they heard was a nasal, and (b) whether the word being spoken was a Spanish or an English non-word. As the task was designed in a way to require them to switch between two dimensions, reaction time difference between the shift and no-shift trials was operationalized as their measure of AC. A correlation of the shift-cost scores and the scores on ABX task showed that L2 English learners who scored better on ABX task had also shown to have more efficient attention control. However, no such relationship between the production and AC scores was reported for L2 Spanish learners.

Darcy and Mora (in press)

Using a very similar task to measure attention control, this time Darcy and Mora (in press) looked at "cognitive control" mechanism in phonological learning, which included attention control, phonological short-term memory (PSTM), and inhibition.

Their participants were two groups of late L2 English learners: 15 monolingual L1 Spanish speakers, and 30 Spanish and Catalan bilinguals, who were further split into balanced and unbalanced bilinguals. To measure the extent participants were able to diminish the effect of L1 influence in their perception, an ABX categorization task similar to the one used in Darcy, Mora and Daidone (2014) was used. The attention control task measured their attention switching skill across two dimensions: nasality and native language pronunciation, as in Darcy, Mora and Daidone's (2014) study. The overall results of the study showed PSTM and inhibition were related to phonological learning although there was variation across groups. However, despite their fine-grained analyses, the results of this study did not show a relationship between L2 learners' perception scores and their AC.

Mora and Darcy (in press)

The purpose of this study was to investigate the role of cognitive control (PSTM, attention and inhibitory control) as it relates to pronunciation accuracy measured by a production test. Pronunciation accuracy scores were calculated by acoustic analyses as well as through comprehensibility and accentedness ratings by native speakers. The participants included monolingual (L1 Spanish) and bilingual (Spanish and Catalan) speakers of L2 English. The attention control task used in this study was essentially the same as the one used in the previous study (Darcy & Mora, in press).

The results of the hierarchical regression analysis showed that monolingual L2 learners of the study with better attention control performed better in producing more target-like sounds in the production test. The results further showed that attention explained the variance in the spectral distance scores of the production test; however, the direction of the relationship between attention control and pronunciation accuracy was reverse in this case. In other words, the monolingual L2 learners with higher shift costs in the attention task were better in their performance. The authors attribute this result to the "non-unitary nature" of the concept of attention (Cohen et al., 2004; Tomlin & Villa, 1994 as cited in Mora & Darcy, in press). An attention control task as well as a delayed sentence repetition task may require different kinds of cognitive loads and processing. While in the attention control task, the learners were required to switch their attention, in the sentence repetition task they were to focus their attention to produce it correctly. Because attention switching and selective attention are triggered by different cognitive mechanisms (Fan et al., 2005 as cited in Mora & Darcy, in press), it may have been that learners with better "selective" attention did better on the production task while performing low on the task which required them to switch attention. Mora & Darcy also maintain that the reason why attention control could not predict the pronunciation accuracy scores for bilingual learners of L2 English could be that bilinguals are naturally good at switching between two languages in their daily lives, and this ability to switch might have transferred to their ability to perform naturally better at similar tasks that require them to switch, which was the case in the AC task they used in their study.

Darcy, Park and Yang (2015)

This comprehensive study aimed to investigate the predictors of L2 phonological acquisition in terms of cognitive skills in Korean learners of L2 English. For this purpose, L2 participants (a) took three phonological tests to determine their performance in L2 phonology, and (b) took cognitive tests to measure their attention control, processing speed and working memory (both in L1 and L2). The attention control task required the participants to switch between two dimensions, which were voice identity for indexical information (male or female voice) and lexical dimension (word or non-word). The overall results of this comprehensive study showed correlations between certain phonological tasks and cognitive skills of individuals; however, there was no significant correlation between the AC task scores and any of the L2 phonological acquisition tasks (see Darcy, Park & Yang, 2015 for further details).

The series of recent studies cited above are very valuable in that they present innovative ways to measure attention control specifically as it relates to L2 phonological acquisition by using speech-based attention control tasks. In addition to the Speech-Based Attention Switching Task discussed so far, the present study also uses a second task to measure attention control mechanisms in order to see whether conventional attentional network measurements would also yield comparable results regarding learners' attention control skills, which will provide an answer to the third research question in this study. For this purpose, an Attention Network Test (ANT) was determined to be used in this study as a second tool to measure learners' attention control (see Chapter 4 for more information about this task).

This task originates from Posner and Petersen's (1990) claim that attention can be examined in three related networks. These network functions are alerting, orienting and conflict (or executive control). Alerting can be defined as "achieving and maintaining an alert state" (Fan et al., 2002, p. 1). It is responsible for the capacity to increase attentiveness or response readiness to an upcoming stimulus. This readiness brings about various changes in the body following the presentation of a stimulus, which are necessary for higher performance on a given task (Fan et al., 2009).

Orienting, which is defined as “selection of information from sensory input” (Fan et al. 2002, p. 1), “involves rapid or slow shifting of attention among objects within a modality or among various sensory modalities, with three elementary operations: disengaging attention from its current focus, moving attention to the new target or modality, and engaging attention at the new target or modality” (Posner, Walker, Friedrich, & Rafal, 1984 as cited in Fan et al., 2009, p. 2). In behavioral studies, it is usually manipulated by using a cued signal asking for the location of the stimulus. It has been shown that there are differences in the regions of brain where activation occurs depending on whether the stimulus appears in a cued vs. uncued location (Rueda, Posner & Rothbart, 2004).

Finally, the conflict function is defined as “resolving conflict among responses” (Fan et al., 2002, p. 1). It is also referred to as ‘executive control function’ when explaining the conflict effect measured by the ANT. The executive control function mentioned in this study may be described as a function of the general attention control mechanism which has two components: conflict monitoring and conflict resolution (Ye & Zhou, 2009). It has been studied by using tasks which involve conflict resolution requiring respondents to pay attention to a weaker dimension of a stimulus rather than a stronger, conflicting dimension (Fan et al., 2002, 2009), a good example of which is the color Stroop Task (Bush, Luu, & Posner, 2000).

The ANT has been used previously in studies looking at attentional network mechanisms mostly in bilinguals rather than L2 second language learners. One such study by Yang, Yang and Kang (2014) examined the relationship between phonological awareness and executive function in Chinese-English bilingual children. They used the child-version of the ANT (Rueda et al., 2004) to measure three separate attentional networks: alerting, orienting and conflict. Their findings indicated a bidirectional relationship between phonological awareness and executive attention. Similarly, Costa, Hernández and Sebastián-Gallés (2008) used the ANT to determine whether and to what extent attention control affects speech performance in bilinguals. Their findings suggest that bilingual participants were faster and more efficient than monolinguals in their responses, suffering less interference, less switching costs from trials with incongruent flankers, indicating that bilinguals were more efficient in alerting and executive control network. However, since both studies reported above used life-long bilinguals rather than adult L2 speakers, the results should be interpreted cautiously as these two groups may have different attentional network profiles.

As mentioned earlier, considering the fact that language itself is seen as "an attention-directing system" (Segalowitz & Frenkiel-Fishman, 2005) and that L2 learning and processing require more attention control than L1 processing (Segalowitz, 2010; Slobin, 1996), it is conceivable that L2 learners' performances in phonological perception might be related to individual differences in their AC. To test this, both the ANT and the Speech-Based Attention-Switching task provide researchers of L2 speech perception with exciting opportunities to measure L2 learners' attention control in different but possibly overlapping ways. The findings of previous research on the issue are inconclusive as to the nature of the relationship between phonological AC as a cognitive ability and L2 speech perception. While learners with more efficient AC could possibly make more gains following an instructional intervention, there may also be no significant differences between the scores of L2 learners with more efficient AC and those with less efficient

AC. Therefore, the goal of this dissertation is to investigate these possibilities by looking at the post-instructional gains of L2 learners in connected speech perception and the results of the Speech-Based Attention Control Test and the ANT to explore any potential relationship among them.

3.3.2 AN APTITUDE-TREATMENT INTERACTION (ATI) APPROACH TO INTERPRETING THE ROLE OF ATTENTION CONTROL

It is commonly known that there is always variation in learners' gains in an L2 classroom in terms of their pace of learning and ultimate achievement. Learning conditions such as L1 background, context of learning, length of residence or age, as well as individual differences in cognitive abilities have been shown to influence learning experiences. While learning conditions might be controlled, this is not always true for cognitive abilities even after learning conditions are held constant (Darcy, Mora & Daidone, 2014). From a psycholinguistic point of view, researchers are interested in knowing how and to what extent such individual, cognitive skills might affect L2 learning. In an attempt to explain such an interaction, several researchers initiated an innovative paradigm in the early 1950s called the Aptitude-Treatment Interaction Hypothesis (ATI). The ATI paradigm has been formed following a series of research initiated by Edwards and Cronbach (1952) and Cronbach (1953, 1957), which suggested a way to show how individuals vary in terms of the gains they made following a certain type of treatment. So, the purpose was to show the interaction of aptitude and treatment on the learning outcome. Snow (1991) summarizes the basic purpose of the ATI paradigm as follows:

> ATI methodology is designed to take individual differences among treated persons into account systematically in treatment evaluation to assess the degree to which alternative treatments have different effects as a function of person characteristics and thus determine whether particular treatments can be chosen or adapted to fit particular persons optimally. (p. 205)

Understanding the term “aptitude” is crucial here. Cronbach and Snow (1977) define “aptitude” in simple terms as “any characteristic of a person that forecasts his probability of success under a given treatment” in which “treatment covers any manipulable variable” (p. 6). Later, Snow (1991, 1992) points out to the fact that the term “aptitude” as defined in ATI is not restricted to “intelligence” or “a set of differential abilities” as commonly assumed in the field of general psychology. It can be rather described as a combination of “complex personal characteristics identified before and during treatment that accounts for a person’s end state after a particular treatment” (1991, p. 205). In other words, aptitude may include a variety of individual differences that may affect the outcomes of the treatments in a certain way; however, these constituents of aptitude may not be one of the “conventionally defined aptitude constructs” which are usually considered as “correlates” of learning. In ATI, such individual differences are described as “propaedeutic”, meaning “they are needed as preparation to successful response to learning conditions” (Snow, 1989, p. 14). This means learners’ preexisting individual aptitudes play an important role for them to benefit from a certain type of instructional treatment over another; therefore, ATI suggests using a combination of various types of aptitude indicators of individual differences to explain the learning outcomes as there is claimed to be an interaction between treatment methods and aptitude. It should also be noted that ATI paradigm has also been extensively criticized for a variety of reasons. In her own work, Snow (1992) mentions some of these criticisms which deem ATI as “a fad” and “dead” or “not practical” when attempted to be applied in real life and she briefly addresses each of these criticisms and concludes that such criticisms are injudicious. However, a lengthy discussion of such arguments are beyond the scope of this study (for a review, see Kowollik, 2009).

Although studies testing the ATI paradigm usually apply more than one treatment method of instruction in their designs in order to see the effects of varying treatment methods, it should be noted that the present study does not attempt to compare two instructional methods or techniques. Nevertheless, this study attempts to examine the role of the individual variable attention control, as it relates to improvement in connected speech perception to show any existing relationship between the two. In this regard, the way the cognitive ability of attention control is treated in this study is presented within an Aptitude-Treatment Interaction framework without favoring a certain type of instructional method over another.

3.4 PURPOSE OF THE STUDY

In this chapter, it was shown that improving L2 listening and speech perception may benefit from learning which uses bottom-up processing. This is very important because learning and teaching connected speech is known to be a bigger part of bottom-up than top-down processing. Therefore, in learning these forms, attention to form (or focus on form) may benefit learners to a great extent because of the lack of saliency in these forms as discussed earlier (Henrichsen, 1984).

Previous studies which have looked at AC were different from the present study in two ways. First, none of those studies had an instructional or training component which measured gains in L2 phonological learning. Rather, they were using one-time discrimination tasks. Related to the same point, a second aspect in which the present study differs from the previous ones is that it looks at connected speech, which is an under-researched phonological dimension of L2 English. To date, to the best of my knowledge, no studies have investigated AC and L2 phonological learning together in the way the present study did.

Therefore, whether learners' attention control affects how much they benefit from training or instruction on connected speech is an interesting question to ask because answering such a question not only contributes to the general understanding of the role of individual differences in L2 phonological development, but it also helps us test new ways of measuring attention control, which may eventually lead to more established ways to measure AC specifically in L2 learners.

With regard to the perception of connected speech features in L2 English, the present study aims to investigate the following research questions:

RQ1: What are the effects of form-focused online training in improving connected speech perception (specifically word boundary palatalization) by ESL learners?

RQ2: What is the relationship between students' performance scores on the perception test (pre-test, post-test, and gain scores) and their attention control (AC) as measured by two types of tests:

a) Attention Network Test (ANT) (Fan, et al., 2002; Weaver, Bédard, & McAuliffe, 2009, 2013)

b) Speech-Based Attention Switching Task (Darcy & Mora, in press; Darcy, Mora & Daidone, 2014; Mora & Darcy, in press)

RQ3: What is the relationship between attention control scores as measured by an online Attention Network Test (ANT) and a Speech-Based Attention Switching Task?

As stated in Chapter 1, the present study makes the following predictions: (a) ESL learners will benefit from online, form-focused training as measured in pre- and post-test results, and (b) learners' performance scores on the perception tests will be higher for those with more efficient AC than those with less efficient AC, and (c) there will be a correlation between attention control measurements of the Speech-Based Attention Switching Task and the ANT.

3.5 CHAPTER SUMMARY

In light of the literature mainly in the intersection of cognitive psychology and SLA, the present chapter provided a discussion of bottom-up and top-down processing in L2 listening followed by relevant models of spoken word-recognition. Additionally, concepts such as attention and attention control have been defined and explained this time in a cognitive framework to build on the discussions of the related concepts (i.e., attention in SLA) and instructional approaches (i.e., FFI) explained in Chapter 2. Studies which first used speech-based attention control tasks similar to the one used in this study have also been summarized. The last section of the chapter discussed the ATI framework in order to show how the ATI paradigm has previously interpreted individual differences to predict instructional gains reported in the literature. The next chapter will present the research design and methods used in this study.

CHAPTER 4

METHOD

4.1 EXPERIMENTAL DESIGN

In order to test the research hypotheses presented in the previous chapter, the present study employs a quasi-experimental pre-test —treatment—post-test design. Students in the same proficiency level ESL classes have been pre-assigned to their sections by the program administrators based on their proficiency scores prior to their voluntary involvement in the study; however, in this study, students in each group have been selected as a whole and then randomly assigned to experimental and control groups, which might be considered cluster random sampling making it a quasi-experimental design (Mackey & Gass, 2005).

Students in all eight intact classes took a forced-choice pre-test (dependent variable) measuring their connected speech perceptual accuracy. The instructors were present during most of the sessions, but they did not participate or were not involved in any aspect of the instructional/training phase.

The experimental groups (four out of eight classes) received treatment (Independent Variable I) in the form of online connected speech training for three consecutive weeks in their computer labs during regular class session. They watched a series of movie clips in which content-related questions had been integrated by the researcher using various interactive video assessment tools. The control group did not receive any treatment on the target forms and only took a pre- and post-test.

Students in both groups also took two types of attention control tests (Independent Variable II) to see whether and how attention control performance of students is related to their scores in connected speech perception. The type of connected speech investigated in this research was the word-boundary palatalization that occurs optionally under certain phonological contexts in English. In addition, as the speech-based attention switching task used in this study is a relatively new task, to compare its findings with a rather conventional AC task, the ANT was used as a secondary measure of AC. Therefore, a correlation between the two different types of attention control tasks was also sought in order to shed new light on understanding the role of AC in phonological development.

4.2 PARTICIPANTS AND SETTING

To determine the sample size needed for analyses in a systematic way, a priori power analysis using the software package GPower, v3.1 was conducted (Faul, Erdfelder, Lang & Buchner, 2007). Estimating an effect size *f* of .40, and a .05 significance level, the analysis indicated that 16 participants in each group were needed to achieve statistical power of .80 for interaction effects for the RM ANOVA calculation. This was found to be achievable with the number of potential participants. A prior calculation of the required sample size was also employed for the correlation analyses to be run between the AC measures and perception test scores. Considering a medium effect size of .30, a conventional alpha ($\alpha = .05$), a two-tailed test, and a conventional power of .80, a total sample of 68 was needed.

The participants in this study initially included a total of 86 non-native speakers of English studying English as a Second Language (ESL) at the intensive ESL program hosted at a major research university in southeastern United States. Of those, 28 students were eliminated from the analysis for not taking the pre-, post-tests, any of the AC tasks or missing one of the training sessions due to tardiness or withdrawal from the program. As a result, a total of 58 L2 learners were included in the final analyses [Experimental Group (EG) = 33, Control Group (CG) = 25]. The number of participants at the beginning of the study was 86; therefore, the number at the beginning of data collection proved reasonable to achieve enough power. However, due to various reasons, the total number of students to be included in the final correlation analyses dropped more than anticipated at the end of the data collection ($n_{\text{initially}}$ =86, $n_{\text{included in the final analyses}}$ = 58), possibly causing loss of power.

A large majority of students enrolled in this program were international ESL learners whose purpose was to attain high enough L2 English proficiency scores to meet the requirements for matriculation into degree-seeking academic programs of colleges and universities in the U.S. Students were enrolled in a skills-based Speaking and Listening Class. There were six placement levels in the ESL program at the time of data collection. For the purposes of the study, the lowest and the highest proficiency levels (Level 1 and Level 6 in a six-level curriculum) were not included in the study as the target content would be either too hard or not challenging enough for them to study. Therefore, the remaining four proficiency levels (upper-elementary, lower intermediate, intermediate, and upper intermediate) were included in this study. The placement of the students was based on their beginning-of-term (if they were returning students) or end-of-term speaking oral interview and listening test scores, which were then weighted to determine the final placement score, and thus, the placement level of each student. At the time of data collection, the listening test used at this Intensive English Program (IEP) was College Board's Accuplacer ESL Listening Test, which is an online, computer-adaptive multiple-choice assessment tool. The conversations in the test questions typically "take place in academic environments such as lecture halls, study sessions, a computer lab, the library, and the gymnasium; and in everyday environments such as at home, shopping, at a restaurant, at a dentist's office, listening to the radio, reading the newspaper, and performing tasks at work." (College Board, 2015, p. 36). Students listen to the conversations and the related question, and then answer it by choosing one of the four response options. It is also possible to replay any of the answer choices, question or the conversation itself for two additional times (College Board, 2015, p. 36).

Two sections from each proficiency level were included in the study, and students in each section of a specific level were assigned to either Control Group or Experimental Group. So, the reported number of students in each level was based on this original placement of the students. However, it should be pointed out that since the primary focus of this study is understanding L2 perception rather than production, the scores from the online listening test were also used to determine the listening proficiency of the students to partial out any effects of the speaking test score in their respective placement level. As a result, two types of placement scores were used to analyze the data to explore the proficiency level variable: (a) the actual placement levels based on speaking and listening scores, and (b) their placement levels based on listening scores alone irrespective of their actual classroom assignments. The listening test score was calculated by taking an average of students' beginning-of-term and end-of-term listening scores to determine their specific listening proficiency levels.

The selection of ESL students in this study was also theoretically-driven because most of the studies investigating connected speech were carried out in an EFL setting (see Ito, 2006a), and those which examined AC and phonological development did not recruit ESL students with a variety of language backgrounds (Darcy, Yang & Park, 2015; Darcy & Mora, in press; Mora & Darcy, in press). Moreover, selecting ESL speakers currently learning English as participants, rather than those who already matriculated in colleges, was important to address the research questions in this study and to fill the gap in the literature, which calls for empirical studies testing gains in L2 phonological development (Mora & Darcy, in press). Therefore, the present study recruited students from various linguistic backgrounds actively learning ESL.

Demographic data collected for the student sample included age, gender, country of origin, native language as well as specific questions about students' previous and current L2 English learning experience (see Appendix B). The summary of the demographic and language background information pertaining to student participants are summarized in Table 4.1 below.

Of 33 students in the experimental group (Level 2: 7 students; Level 3: 7 students; Level 4: 9 students; Level 5: 10 students), 24 were males and 9 were females. Students' first languages included Arabic (n = 10), Mandarin Chinese (n = 11), Japanese (n = 8), Vietnamese (n = 2), Korean (n = 1), and Turkish (n = 1). In the control group, of 25 students (Level 2: 5 students; Level 3: 5 students; Level 4: 5 students; Level 5: 10 students), there were 16 males and 9 females.

Table 4.1

Demographic and Language Background Characteristics of Participant Groups

Measure	Experimental Group (*n* = 33)		Control Group (*n* = 25)	
	M	*SD*	*M*	*SD*
Age at testing (years)	22	6.43	23	6.76
Age of first exposure to L2 (years)	11	5.59	11	3.43
Length of residence in the U.S. (months)	5	4.53	6	5.24
L2 Instruction in home country (months)	97	4.07	92	3.7
L2 listening outside of class (hours)	1.44	.86	1.58	.86
Listening Proficiency Test Score [1]	74.15 (62%)	13.9	75.16 (63%)	12.43

[1] This score is an average of the beginning-of-term and the end-of-term listening test scores of the participants. The maximum score is 120.

The L1s represented were Arabic (*n* = 11), Mandarin Chinese (*n* = 6), Japanese (*n* = 5), Korean (*n* = 2), Nepali (*n* = 1). Table 4.1 also demonstrates the experimental and control groups were similar based on their listening proficiency scores alone.

Learners were also asked additional yes-no questions regarding their English language learning settings in the U.S. outside of their classroom. Table 4.2 below shows the number and the proportion of students who provided an affirmative response to the respective questions in each group.

In the control group, of 25 students, only 4% (*n* = 1) reported living with an American family, and 24% (*n* = 6) reported living with an American roommate while 32% (*n* = 8) of them stated that they used English as their means of communication at home. About 44% (*n* = 11) of the students stated that they watched American movies in English very frequently.

Table 4.2

A Summary of Learners' Responses to Questions about Their L2 Learning Setting

Measure	Experimental Group (n = 33)	Control Group (n = 25)
	Total (%)	Total (%)
Learners living/having lived with an American family	3 (9%)	1 (4%)
Learners living/having lived with an American roommate	6 (18%)	6 (24%)
Learners using English as a means of communication at home	9 (27%)	8 (32%)
Learners watching American movies in English on a frequent basis	27 (60%)	11 (44%)

In the experimental group, of 33 students, about 9% (n = 3) reported living with an American family, about 18% (n = 6) reported living with an American roommate, and about 27% (n = 9) stated that they used English as their means of communication at home. About 60% (n =20) reported watching American movies in English on a frequent basis. The information provided in Table 4.1, Table 4.2 as well as the information on L1 background, gender and the number of students in each proficiency level shows that Experimental and Control Groups were not very different from each other in their representation of similar groups of participants. It should, however, be noted that the total number of students in each group was not equal, resulting in unequal number of students in each proficiency level across groups.

Finally, to meet the ethical requirements in doing research with human subjects, at the beginning of the study, first, the purposes and the requirements of the study were explained to the participants. Required procedures such as getting (a) participants' informed consent (b) permission from the IEP to collect data, and assuring (c) the anonymity and the security of the data obtained were also addressed (see Appendix A).

Prior to data collection, necessary credentials have been obtained from the Institutional Review Board (IRB) (reference # 00044442) for approval and comply with all their rules and regulations including researcher's training and certification for research with human participants.

4.3 TARGET LINGUISTIC ITEMS BEING INVESTIGATED

The target linguistic item being investigated in this study was word-boundary palatalization found in naturally spoken English. There are four potential occurrences of such palatalization in English as summarized in Table 4.3.

Table 4.3

Target Linguistic Forms

	Type of Palatalization	Phrase Spelling	Citation Form Transcription	Reduced Form Transcription
Word-boundary Palatalization	[d+j]	…told you…	[toʊld ju]	[toʊldʒʊ]
	[t+j]	…want you…	[want ju]	[wantʃʊ]
	[s+j]	…kiss you…	[kɪs ju]	[kɪʃʊ]
	[z+j]	…turns yellow…	[tɜrnz jɛloʊ]	[tɜrnʒɛloʊ]

Each of these examples of palatalization occurs at varying degrees of frequency in naturally spoken English and, thus, has been proven to pose a challenge for L2 learners of English (Brown & Kondo-Brown, 2006; Henrichen, 1984; Joyce, 2011; Kostin, 2004). However, these forms, like many other aspects of pronunciation teaching, have not received enough attention in the language classroom despite the important role they play in successful language comprehension (Brown & Kondo-Brown, 2006). Moreover, since palatalization results in the reduction of certain sounds in each example seen in Table 4.3, these forms are naturally hard to notice during listening. This problem has been previously attributed to lack of “perceptual saliency” in connected speech forms (Henrichsen, 1984; Kim, 1995). In other words, successful perception of connected speech forms requires attention as these forms are hard to decode making these forms interesting to explore in relation to attention control. So, in addition to looking at whether training helps learners to improve their perception of connected speech forms, this study examines how learners’ attention control is related to their performance scores on the connected speech perception test.

4.4 TREATMENT: CONNECTED SPEECH PERCEPTION ONLINE TRAINING

In ESL classrooms, unlike most EFL contexts, class size is usually relatively very small. When there are multiple classes involved in a study, a variety of factors affect the learning taking place in a classroom. It seems less likely that all groups of learners will receive the same type of input, even in cases where materials and instructors are the same. One way to address this problem is to expose learners to the same type of input inside the classroom by having them receive online training on targeted forms. However, it is worth noting that the primary purpose of the present research is not to investigate the effects of video use in the classroom. A comparison of use and non-use of video in class in terms of its benefits is beyond the scope of this study, rather, video, in this study, was used as a medium to present students with the targeted input. There are several reasons why video was used in this study.

An important reason for preferring video for content presentation in this study was logistical. Since there were multiple classes involved in this study, there was a need to assure that the learners were exposed to the same type of input and assessment without taking up an extended amount of regular classroom period. This eventually enabled collection of data from a larger amount of ESL learners simultaneously.

In addition, use of video for content presentation can be an effective tool to grab students' attention and to activate or stimulate multiple senses in students with various learning habits given the effectiveness of the dual coding (i.e., the visual and the oral input) video offers (Houston, 2000; Sherman, 2004). Although specific inquiry into such use of video use in ESL classrooms has been limited, existing research suggests that inclusion of video in the field of education in general deserves consideration (see Berk, 2009). On the other hand, watching videos in classroom has also been criticized as being a passive activity, which may be avoided by integrating interactive video assessment tools in the videos. Many of these tools available online provide their users a variety of ways to present and assess the content learners are watching. While watching, learners are expected to focus more attention to comprehend the input, and possibly get feedback every time they answer a content-related question. This also helps them clarify and reinforce the input they are processing, which presumably enhances learning. These tools usually come with an analytics function which enables the user, in this case any faculty member or teacher, to check the student performances on quizzes embedded in videos. This could provide critical information on student success and effort with regard to their assignments by serving as a formative assessment method, and could help the instructor determine the concepts to review in class based on the scores (Edel-Malizia et al., 2015). Moreover, by working on these video recordings and answering embedded questions, learners take charge of their own learning, which is considered to be the first step towards autonomous learning (Benson, 2007; Little, 1996).

Given the potential benefits of video as a medium to convey new information for learners especially when enhanced with interactive video assessment tools, the present study adopted a similar approach by presenting its target content on connected speech perception in the form of online video episodes, which were prepared on a graphical design platform, and then were supported by an online interactive video assessment tool. The video series included embedded custom interactive assessment questions, which were then presented to students at various points of the video timeline. The students were also provided immediate feedback on their responses to questions with short explanations. The fact that students' understanding of the content was checked by their answering various types of embedded questions was important because these questions are an effective way to enhance learner's mental focus by keeping them active (Szpunar et al., 2013) while allowing them to work autonomously.

For these reasons, the researcher prepared a video script to be used in the training videos. This video script was prepared completely from scratch in the light of the existing pronunciation textbooks and some pronunciation teaching resources (Avery & Ehrlich, 1992; Brown, 2012; Celce-Murcia, Brinton & Goodwin, 2010; Cook, 2000; Grant, 2010; Weinstein, 2001). The content presented in the videos included a general introduction to the concept of connected speech followed by [t + j] and [d + j] transformation (Week 1), [s + j] and [z + j] transformation (Week 2), and a wrap-up, review and practice session (Week 3) including all four types of palatalization presented in Week 1 and Week 2. The script of the training video series can be found in Appendix D.

These training video series on connected speech were created using a form-focused instructional method in light of a presentation, practice and production (PPP) approach (see Richards & Rogers, 2001, p. 246). The content presentation in the training videos aimed to follow a sequence of three modes: "noticing", "awareness" and "practice" as described by Lyster (2007). The presentation of the content included focus on the rules, followed by example sentences asking students to repeat the same sentences. However, it should be pointed out that since there was no "instruction" taking place in a classroom environment, an actual realization of these stages cannot be strongly claimed. For example, the practice stage was encouraged, but as observed by the researcher during data collection in each group, there were still some students who did not actively repeat the practice sentences presented in the videos. Table 4.4 below summarizes the content presented in each of the three weeks (see Appendix D for details).

Table 4.4

Summary of Content for the 3-week Training on Palatalization

WEEK	CONTENT
WEEK 1	RAISING AWARENESS Introduction to Connected Speech PALATALIZATION (PART 1) Introduction to palatalization [t] and [d] followed by [j] Examples for these transformations Rule derivation Multiple-choice questions on content
WEEK 2	PALATALIZATION (PART 2) Introduction to palatalization [s] and [z] followed by [j] Examples for these transformations Rule derivation Multiple-choice questions on content
WEEK 3	REVIEW Review of Week 1 and Week 2 Content Sentence practice Multiple-choice questions on content

The movie clips were presented to learners in the experimental groups in their computer lab for about 20 minutes for three weeks, that is, a total of about 60 (3 X 20) minutes including the time spent for the assessment questions they were supposed to answer during watching. The audio of scripts used in the movie clips was recorded by a female native speaker of English with clear enunciation in a sound-proof room in the Speech Perception Lab located in the Department of Communication Sciences and Disorders/Speech Pathology at the university, in the presence of the researcher. After recordings were made, the researcher used a web-based animation software called Powtoon (2015) to create the animated videos to be used in the study. Powtoon is a powerful tool used by educators as well as business people to present content to people in an interesting way. During the animated video preparation, some of the sections in the script were found to be either redundant or lacking enough support for examples. Therefore, a couple of necessary changes and additional recordings were made to improve the original script of the movie clips before the final product was produced.

Once the movie clips were prepared, they were uploaded to EDpuzzle (2015) website, which is a web-based tool for interactive video assessment. This tool provides teachers and learners with a variety of interesting options to make the best use of videos in classroom. However, the present study only used the question integration function of the EDpuzzle online tool. In other words, the researcher integrated 5-6 multiple choice questions with one or more correct answer options to be selected by the learners during watching. Each time a question appeared, the video stopped streaming until the learners provided an answer, but there was also an option to skip the question. Once they answered the question, they were given feedback on their responses. Students' responses, however, were not recorded as this required registration, which would take up at least an additional 10-15 minutes of class-time, which was not favored in this particular EIP program. Screenshots of the final movie clips after integration of the questions can be found in Appendix J.

While the classes assigned to experimental group received training on connected speech perception, the classes in the control group followed their regular curriculum. In this EIP, ESL students are not required to buy any textbooks for their Speaking and Listening classes; however, on the curriculum guide, teaching connected speech is encouraged starting from Level 4 or 5. However, none of the teachers in eight intact classes reported teaching palatalization as a connected speech phenomenon in their classes after the study was finalized. Prior to or during data collection, teachers were not explicitly told the purposes and the specifics of the study, either.

4.5 INSTRUMENTATION

4.5.1 MEASURING IMPROVEMENT IN PERCEPTION OF PALATALIZATION: FORCED-CHOICE PERCEPTION TEST

In the assessment of connected speech perception, dictation tests have been the most commonly used and reliable method (Brown & Hilferty, 1986; 2001; Fountain & Nation, 2000; Ito, 2001; Joyce, 2013; Matsuzawa, 2006). Joyce's (2013) study is notable in creating a set of criteria for developing more reliable dictation tests in the light of various testing theories by making use of three measures "to safeguard the trait purity of the test" (p. 83). Previous studies have shown that when connected speech items are presented in context, learners are more likely to use it to decipher the dictated text. Therefore, the first measure Joyce (2013) suggests is using decontextualized sentences in the dictation task because that way the task would be testing learners decoding abilities rather than their abilities to guess the words from context. According to him, decontextualizing sentences is also useful for psychometric reasons. For instance, when material is presented in context, “due to the interconnected nature of the dictation, it would be difficult to remove material and replace it with alternative items” (p. 81). This is not helpful if the purpose is to develop tests that can be easily adapted. The second measure Joyce mentions is concerned with the length of the sentences. It is suggested that the researcher limit the number of words to seven in each sentence (Miller, 1956, cited in Joyce, 2013) because listeners with longer short-term memory span would be at an advantage when sentences are longer than seven words. Therefore, unless short-term memory is a variable in a study, the length of sentences should meet this criterion. The last measure Joyce (2013) mentions is about the lexical and syntactic difficulty of the sentences. As explained previously, CS forms are known to be less salient compared to fully articulated forms; therefore, if listeners are unable to decipher the language spoken, they tend to apply their knowledge of English sentence structure. For this reason, when creating sentences to be used in dictation tests, Joyce suggests using sentence structures only taught in elementary level ESL grammar books.

Although dictation tests are widely used in assessing the knowledge of connected speech processes, the present study did not use a dictation test to measure improvement in connected speech perception. The primary reason for this was the variety of student proficiency levels in the study, which meant having learners with widely ranging lexical knowledge, listening comprehension skills, note-taking and spelling abilities. All these factors were anticipated to be hard to control and to confound the results of the study due to the difficulty of creating dictation test items that would not decrease the validity and reliability of the test. Therefore, for the present study, I decided to develop a new forced-choice perception test with two response options to test the effects of training on perception of L2 English palatalization. However, the criteria set by Joyce (2013), which were summarized above, were used in forming the testing questions created for this study. Here is an example of the perception test response options used in the study:

(4a) They all **could use** [kʊdʒuz] any of these.

(4b) They all **could choose** [kʊd ʧuz] any of these.

ESL learners first heard either (4a), which was the target form pronounced in a palatalized, "connected" way or (4b), which was not the target form, but a combination of forms which sounded similar to the target form. Immediately after learners heard one sentence, the two options (4a & 4b) appeared on the screen for students to choose the sentence they heard. In order to have a closer look at the questions formed, the types of palatalized forms used in this study were summarized in Table 4.5. In addition to the summary of contrasting phrases and forms in Table 4.5, full sentences used in the study along with explanations for each contrasting sentence pair can be found in Appendix C.

Other than the ones listed in Table 4.5, an additional group of contrasting sentence pairs was also included in the perception test to see if the learners were able to distinguish these four types of palatalized forms when the forms were contrasted with each other. In other words, instead of pairing up a palatalized sound with a palatal sound, this time, both members of the pair were sounds palatalized at the word boundaries (see Table 4.6). Another difference between the contrasted pairs in Table 4.5 and Table 4.6 is that the latter limited the contrasts to either ‘you’ or ‘your’ instead of including other y-initial words (see Appendix C for a full list of sentence pairs and explanations).

Table 4.5

An Overview of the Contrasts Being Investigated

Phenomenon	Paired with	A list of the target contrasts used in study [1]	
		Target Form	Contrasting Form
[t] + [j] => [tʃ]	[dʒ]	haven’t used => [tʃuzd]	haven’t juiced => [ʤust]
(voiceless)	[ʃ]	thought your => [tʃɔr]	thought shore => [ʃɔr]
	[tʃ]	can’t yards => [tʃɑrdz]	can’t charts => [tʃɑrts]
	[tʃ]	not yet => [tʃɛt]	not chat => [tʃæt]
[d] + [j] => [dʒ] (voiced)	[tʃ]	could use => [dʒuz]	could choose => [tʃuz]
	[ʃ]	old you => [dʒʊ]	old shoe => [ʃu]
	[dʒ]	should you => [dʒʊ]	should June => [dʒun]
	[dʒ]	sold younger => [dʒʌŋgər]	sold jungle => [ʤʌŋgəl]
[s] + [j] => [ʃ] (voiceless)	[dʒ]	in case young => [ʃʌŋg]	in case junk => [dʒʌŋk]
	[tʃ]	this year => [ʃɪər]	this cheer => [tʃɪər]
	[ʃ]	miss your => [ʃɔr]	miss shore => [ʃɔr]
	[ʃ]	bless your => [ʃɔr]	bless sure => [ʃʊr]
[z] + [j] => [ʒ] (voiced)	[dʒ]	knows you =>[ʒʊ]	knows June => [dʒun]
	[tʃ]	has your => [ʒɔr]	has chore => [tʃɔr]
	[ʃ]	suppose yelling => [ʒɛlɪŋ]	suppose shelling => [ʃɛlɪŋ]
	[dʒ]	knows you did => [ʒʊ dɪd][2]	knows Judith => [ʤudɪθ][2]

Note. Underlined sections represent the transcribed forms.
[1] Target words may differ in cues such as vowel quality, vowel length, final consonant, and stress patterns
[2] This pair had to replace another pair not listed here which was found to be faulty after the piloting.

Table 4.6

Contrasted Pairings of Palatalized Sounds

Contrasted palatalized sounds		Target Form 1	Target Form 2
[tʃ]	[dʒ]	lent you => [lɛntʃʊ]	lend you=> [lɛndʒʊ]
[tʃ]	[ʃ]	walked your => [wɔktʃʊər]	walks your => [wɔkʃʊər]
[tʃ]	[ʒ]	sent your => [sɛntʃʊər]	sends your => [sɛndʒʊər]
[tʃ]	[ʃ]	typed your => [taɪptʃʊər]	types your => [taɪpʃʊər]
[dʒ]	[ʒ]	memorized your => [mɛməraɪzdʒʊər]	memorize your => [mɛməraɪʒʊər]
[ʃ]	[ʒ]	price you => [praɪʃʊ]	prize you => [praɪʒʊ]

Since previous literature in connected speech assessment mainly used dictation tests, there were specific challenges to preparing a forced-choice test for the assessment of palatalization. Part of the difficulty was coming up with similar sounding yet contrasting pairs of grammatical phrases while adhering to the criteria set by Joyce (2013). Another challenge was not being able to avoid stress shift when even a close-to perfect contrast was available. Below is an example pair of sentences, which were then eliminated from the study after the piloting phase:

(5a) I found you boxes to store this stuff.

(5b) I found shoe boxes to store this stuff.

The reason for the elimination of this pair from the study was the difference in stress assignment, which made the correct answer too obvious for the participants. In (5b), the word ‘shoe’ is stressed while the targeted contrasting syllable ‘you’ receives no stress. This difference in stress assignment lead students to figure out the right answer too easily, which was not desirable. In forming the sentence pairs for this tool, the researcher aimed to keep such problems to a minimum, but due to the nature of the task, despite such stress assignment difference in some questions, the sentence pairs were not replaced, instead, stress placement was kept as a variable to see if it would have an effect on item difficulty, which is discussed in Chapter 6. In addition to stress placement, other variables such as vowel quality and consonant contrast were also used to pair up the target forms of palatalized sounds, with similar sounding forms.

Following the criteria set by Joyce (2013), the sentence pairs were formed in a way to avoid context to help determine the right answer. This was another challenge in creating the perception test items in this study because while keeping the likelihood of the occurrence and authenticity of the task, the target items had to sound similar enough to challenge the students taking the test. Therefore, there were no contextual clues used in the response options and the authenticity and probability of the sentence meanings and usage were aimed to be kept as equally reasonable as possible. However, it should be noted that in some pairs one sentence may sound more likely to be heard than the other. This was something unavoidable due to the difficulty of coming up with equally likely sentences, but still sounding similar. However, this is not considered to be a big confound for the test because learners were explicitly pointed out multiple times both in writing and orally that all the sentences were grammatically possible and meaningful. Therefore, they were encouraged to listen and choose the sentence that they heard, not the one they found more grammatical or familiar.

Another criterion was the use of familiar grammatical structures as well as lower-level vocabulary in forming the test questions. Although the priority was given to finding similar-sounding words and phrases, frequency of words and phrases used in the tool was checked using two basic sources: the Frequency Dictionary of Contemporary American English (Davies & Gardner, 2010) and its extended online database version, the Corpus of Contemporary American English (COCA) (Davies, 2008-). All the target words used in the study were commonly used words found in the frequency list of Davies and Gardner (2010) and had a similar range of frequency. As for the grammatical structures used in the response options, it was assured that mostly basic sentence structure was used. Even in cases where some sentences were longer than others, the syntactic structures used to form them were simple enough for all proficiency levels to understand. Finally, the short-term memory issue mentioned by Joyce (2013) was not applicable to the test used in this study as the participants were not required to keep strings of words in their memory to note them down. However, a special criterion for keeping the number of words and syllables preceding and following the targeted contrasting words was used. Specifically, target contrasts were presented mostly in the middle of each response option. The number of syllables preceding and following the targeted sounds mostly ranged between 2 to 5. However, it should be noted that although the criteria suggested by Joyce (2013) were used in creating the question response options for the perception test, some issues regarding the use of familiar vocabulary in each pair or inclusion of longer sentences were not regarded as major problems as they would be in a dictation test and, thus, should not be considered as threats to proper analysis and interpretation of the results.

In light of the limited previous research and specifically formed measures provided by Joyce (2013) for the assessment of CS, the present study employed a new type of forced-choice test to measure gains in palatalization by ESL learners. The test included 16 main target items (16 sentence pairs x 2 times = 32), 6 additional paired within-group target items (6 x 2 = 12), 16 control item pairs (16 x 1= 16), 16 filler item pairs (16 x 2= 16) and 3 practice items (3 x 3= 9). The target item and filler item sentence pairs were presented twice for each correct answer while control item pairs were only presented once. This decision was made based on the results of the pilot testing mostly for the purposes of class period allocation.

The pilot testing was conducted using 6 students representing lower, intermediate and upper intermediate levels coming from four different native language backgrounds (Arabic, Chinese, Japanese and Turkish). Following the piloting phase (a) some of the items were replaced with new ones, (b) some technical issues with the presentation of the stimuli and software were addressed, (c) the length of the test was shortened by presenting control item pairs only once by excluding the target response as the correct option as it caused learners to guess the correct answers later on, and (d) instructions and response time limits were revised.

The trial and critical sentence pairs were then recorded by a male native speaker of American English with a Ph.D. in speech science working in the Speech Perception Lab at the university. The recordings were made using *Matlab* at a 16-bit depth and a sampling rate of 48 KHz in a sound-proof booth. The length of each sentence was set to 4000 milliseconds regardless of the number of syllables in a sentence. These sentences were then presented to students in each of the eight intact classes on Apple iPads using the stimuli presentation software, Paradigm (2007). The testing took place during regular class hour in computer labs. The presentation of the questions and answers were randomized for each individual. The test also included 3 practice sentence pairs, but each pair was presented three times rather than twice to make it clear to the students that it was possible to hear each sentence pair more than once with an aim to avoid strategic responding, which, in simple terms, means learners' determining the answer by just remembering their responses to the same question from its previous presentation. Apart from the online written instructions provided at the beginning of the perception test, learners were also given in-class written and oral instructions and explanations by the researcher before taking the test in order to make sure that students understood that it was a perception test rather than a test of grammatical accuracy. This was done because in the piloting, some students thought they were supposed to check the sentences for the accuracy. Following the instructions page, for the trial block and the test block, students first heard a sentence once on a blank screen with the dark sea-green background color. Immediately following, they were presented with two options in written form. Learners in all eight classes (4 control and 4 experimental groups) took the same test at two different times. It took approximately 25 minutes to complete the perception test.

When a newly devised test is used in research designs, reporting validity and reliability measures is highly crucial. Reliability is defined as "the consistency of the results and how sure readers can be of the replicability of the research" (Woodrow, 2014, p. 26). Despite being created by the author, the Cronbach's alpha value for this test was not calculated because each target stimulus and its contrasting pair used a variety of variables such as vowel quality or consonant contrast and, thus, had heterogeneous weights in the overall test, which directly impacts the internal consistency correlations among questions in a negative way. Even within groups of palatalized sounds, the contrasting pair may be assessing a distinct aspect than the other three pairs in the group.

Although this was a deliberate decision in the construction of the test, it makes the Cronbach's alpha calculation not suitable for this test. As an alternative, item difficulty analysis was determined to be conducted to reflect on the variety of question types (see Chapter 6).

Research instruments should also be valid, that is, "we want them to reflect what we believe they reflect and that they are meaningful in the sense that they have significance not only to the population that was tested, but, at least for most experimental research, to a broader, relevant population" (Mackey & Gass, 2005, p. 106). Items included in the instrument cover all four types of word-boundary palatalization in English, and can be claimed to bring a better understanding of the learning of this linguistic phenomenon, which assures the content validity. The content validity and face validity for this newly designed perception test were also checked throughout its preparation by a panel of five experts, which consisted of two faculty members at the university, three speaking and listening teachers experienced in teaching pronunciation as well as other native speakers of English. Revisions were made multiple times over a period of five months until the final version of the instrument was created. Validity was also confirmed after piloting and refining the instrument by deleting, replacing and adding new items based on expert and participant feedback and suggestions as well as native speaker intuitions naive to the purposes of the study. More importantly, it should also be noted that to avoid rote-learning, none of the target words or phrases given as examples in the training videos were used in the perception test.

4.5.2 MEASURING ATTENTION CONTROL

4.5.2.1 TASK 1: SPEECH-BASED ATTENTION SWITCHING TASK

In order to measure attention control, a speech-based attention switching task was used. This task is an auditory analog of the Dimension Change Sort Task (Bialystok & Martin, 2004) and a speech-based version of the alternating runs procedure (Rogers & Monsell, 1995; Segalowitz & Frenkiel-Fishman, 2005), in which participants are expected to inhibit their attention to a dimension which was previously selected, and then are asked to focus on a different dimension of the same stimulus for a second time (Darcy, Mora & Daidone, 2014; Darcy & Mora, in press; Mora & Darcy, in press; also see Safronova & Mora, 2012a & b; Safronova, 2013). This task was originally designed and used by Darcy and her colleagues in a series of studies in order to test attention control in relation to L2 phonology (Darcy et al. 2014; Darcy & Mora, in press; Mora & Darcy, in press). The present study adopted their task design to a large extent by replacing the dimensions in a way to answer the research questions in the present study and by also decreasing the total number of trials by half. Dimensions refer to the alternating questions and answers. In the present study, one dimension (question) was whether the person who uttered a one-syllable non-word in the recording was a male or a female while the other dimension (question) targeted whether a stimulus was consonant-initial or vowel-initial (Darcy, 2014 personal communication).

The stimuli in this task included 40 (+ 9 practice items) items, of which 20 were consonant-initial and 20 were vowel-initial one-syllable English non-words found in the ARC database (Rastle, Harrington & Coltheart, 2002). All non-words were monosyllabic with one of the following syllable structures: CVC, VCC, VCCC, CVCC, CVVC, VCVC, VVCC. However, it should be pointed out that the listeners did not see the words in written form, and when the non-words were pronounced, all the non-words sounded one-syllable (see Appendix E). A male and a female native speaker of English recorded the stimuli in a sound-proof booth. The task consisted of two main sections. In the training phase, participants answered 9 practice trials (with feedback), and the test phase included 40 trials with no feedback. The full list of stimuli used in the study may be found in Appendix E.

In the piloting phase, some lower level learners had needed additional guidance on vocabulary items such as consonant and vowel. Therefore, initial briefing on such words was provided prior to the test. At each trial, participants were expected to tell (a) whether the non-word they heard was consonant-initial or vowel-initial, and (b) whether it was spoken by a male or a female. Unlike the original task, instead of asking participants to respond by choosing from a "yes/no" option, the present study presented the participants with both of the options (male/female) or (consonant/vowel) along with the question (only in the practice trials) to make it easier for lower proficiency level ESL learners. Again, based on the piloting results and feedback from the participants, each time they were presented with a consonant-initial or vowel-initial question type, to help them remember what a vowel or a consonant was, right under the box, which says "vowel", all vowels in English were provided to the participants for the practice trials to familiarize the participants with the vocabulary. It should also be noted that the language used in the experiments were kept minimally complex. One such example is the replacing of the response options 'female' and 'male' with 'woman' and 'man' in the experiment after initial feedback from the piloting.

The stimuli were presented on 9.7-inch Apple iPads and administered with the commercial experimental stimuli presentation software Paradigm (2007). At each trial, after a fixation for 500 milliseconds, the auditory stimulus was presented for 4000 milliseconds, and was immediately followed by the question (for practice trials only) and answer choices. Participants chose one of the two answer choices (Man/Woman or Consonant/Vowel) by clicking one of the two response boxes appearing on the touchscreen (see Appendix K for screenshots of the task windows). Originally, this task was designed in a way so that the next stimuli would appear in neighboring quadrants in a clockwise rotation from top left to top right, bottom right, and then to bottom left to help participants understand the predictable nature of the presentation of the stimuli visually (Rogers & Monsell, 1995; Segalowitz & Frenkiel-Fishman, 2005). However, since in this study the data were collected in an ESL classroom setting, to make the task less complex for the participants, there were no visual blocks in the form of quadrants. Although auditory feedback mechanisms have been found to make participants pay more attention to accuracy (Safronova & Mora 2012a), the researcher decided to keep it to increase participant interest and motivation while limiting it to the first 9 trials in the practice phase. Also, before the training phase, both orally and in written form, the participants were explicitly informed that they had to respond as quickly as possible, and that they were being tested on how fast and accurate they were. In fact, as a part of the feedback provided to participants in the trial phase, when learners got the answer correct, the feedback said "Correct! Be faster!" to help them understand the importance of being fast in their responses. There were two kinds of trials in this test. One of them was a 'No-shift' trial, which presented the same type of question with the previous one, and the other one was a ‘Shift’ trial, which presented a different type of question from the previous one. All trials were arranged in a predictable ‘Shift, No-shift, Shift, No-shift’ sequence using 10 test blocks of trials. There was no randomization within blocks, but the blocks were randomized among themselves. Randomization within blocks was not possible as it would not keep the trial sequence. Shift trials required participants to switch their attention to a different speech dimension. Analyses were made based on the shift-cost measure, which was calculated by subtracting the mean RT for no-shift-trials from the mean RT for shift-trials for correct trials [Shift Cost = RTshift - RTnoshift/repeat]. Accuracy (in terms of error rate) on shift, no-shift and shift-cost measurements were also calculated. Figure 4.1 illustrates the trial sequence.

All 49 auditory tokens, 9 of which were trial items, were ordered in a way to make sure that the arrangement of the stimuli took perseverance into account. An example of a perseverative behavior occurs when a previously presented set of questions or responses determines the response in a set which requires a different type of response, also known as the “stuck-in-set” tendency (Milner, 1963). Therefore, in the arrangement of the presentation order of the stimuli, the purpose was to minimize perseverative errors in participants’ responses. In other words, even if participants were to make errors in their responses simply due to the difficulty of switching between different question and response types, the arrangement of the stimuli made sure that it was not due to perseverance. This template was created by Darcy (2014, personal communication) and generously shared with the researcher. For a full list of trial blocks and how this was realized, see Appendices F and G.

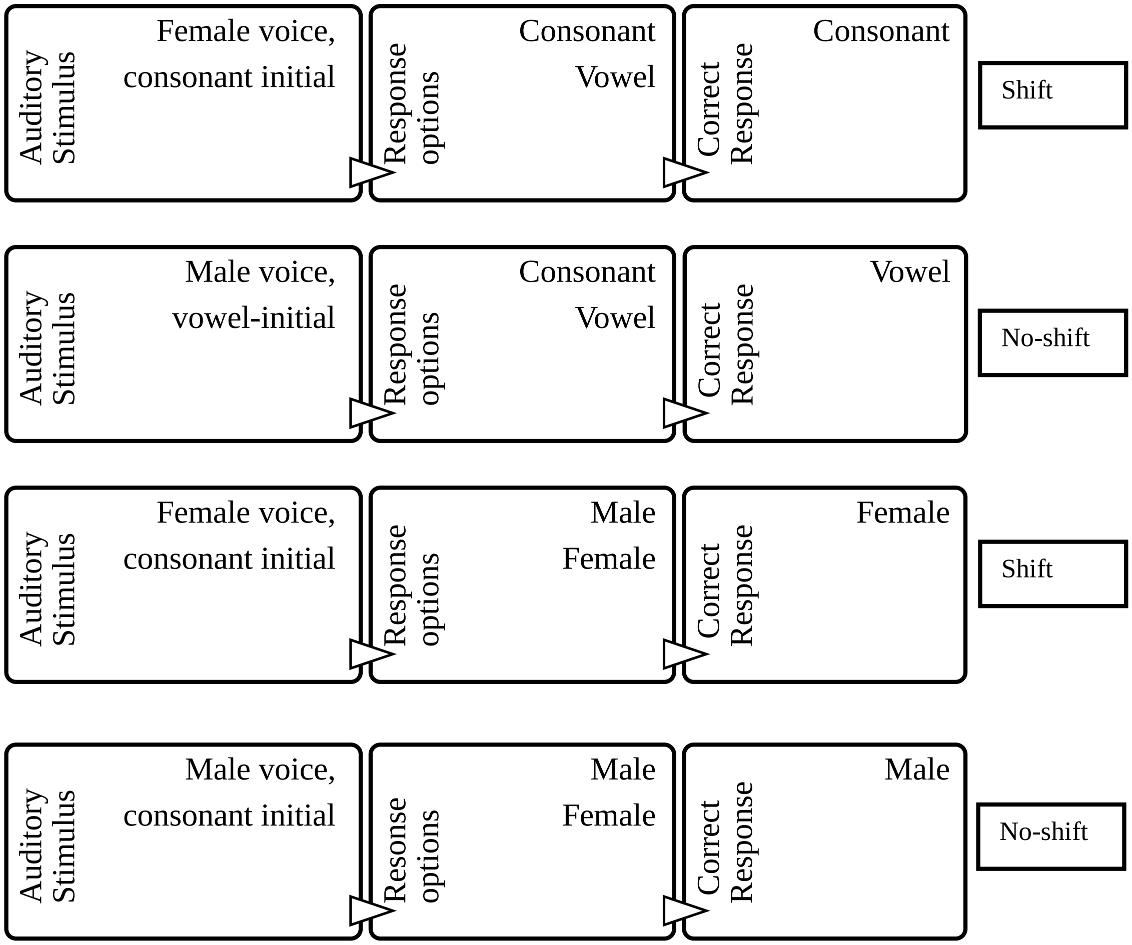


Figure 4.1. Sample trial squence for attention control task with correct response and trial type.

This attention control task is relatively new in that it measures an individual's attention control using speech-based dimensions. This is crucial especially given that the purpose is to investigate the relationship between AC and phonological acquisition because such a task measures AC in a more specific way. This novel task, to the best of my knowledge, has not been used to interpret AC as it relates to post-instructional gains in speech perception. Previous studies summarized in Chapter 3 using various versions of the same test only looked at perception and production of consonant and/or vowels and did not include an instructional component.

4.5.2.2 TASK 2: ATTENTION NETWORK TEST

In addition to the speech-based attention switching task, which is relatively new, an additional, traditional test to measure attention control was used. The purpose was to see whether there was a correlation between the findings of the two types of attention control tasks.

For this purpose, the present study used the Attention Network Test (Fan et al., 2002; Rueda et al., 2004; Weaver et al., 2009, 2013), which was originally designed to measure three different dimensions of attention, i.e., alerting, orienting and conflict, using a single task as explained earlier in Chapter 3 (Fan et al., 2002; Posner & Petersen, 1990).

The original ANT uses three flanker (neutral, congruent, incongruent), and four cue conditions (no, center, double or spatial-down & up). Each trial begins with a central fixation cross. The target figure is either a single arrow or a horizontal row of five arrows, presented above or below the fixation cross. The participants are to determine whether the middle arrow (or the single central arrow in the neutral flanker condition) is pointing to the right or left. While on congruent trials, the flanking arrows point in the same direction as the central arrow, on incongruent trials, the flankers point in the opposite direction, and on neutral trials, the central arrow appears alone. One of the four warning cue conditions (no, double, central and spatial) precedes each target. In the center cue condition, an asterisk appears where the fixation cross is located, and in the double cue condition, an asterisk is presented above and below the fixation cross. Finally, in the spatial cue condition, an asterisk appears in the location of the upcoming target, either above or below the fixation cross and in no cue condition, no asterisk precedes the target. Scores on three attentional network measurements (alerting, orienting and conflict) are calculated based on a number of specific mathematical formulas using the RT on correct trials, and their accuracy measurements (Fan et al., 2002).

Following the criteria followed by similar studies which shortened the task for various reasons, the present study also adapted the original task in a way to have a task which would not take more than 10-15 minutes due to time limitations in classroom settings. For this purpose, first, instead of investigating all three attentional network measures, this study only looked at orienting and conflict scores, which were more relevant for the purposes of this study because based on the descriptions of the alerting attentional network provided by Fan et al. (2002), the speech-based attention network task already used in this study does not seem to assess the alerting network specifically. Not including alerting network calculations enabled the researcher to exclude the double cue condition from the ANT. Second, the neutral flanker condition was removed leaving only congruent and incongruent flanker types. In fact, the shorter version of ANT created by Fan, the original designer of the task, also eliminates the neutral flanker because it is not used in any of attentional network calculations; thus, in the ANT version used in this study, the neutral flanker condition was also removed. This task is available on the website of the Sackler Institute for Developmental Psychobiology (www.sacklerinstitute.org/cornell/ assays_and_tools/ant/jin.fan). In addition to these two steps, time intervals within trials were also reduced. As a result, a task which took about 10 minutes for ESL learners was created.

In this task, at each trial, participants were presented with a row of five right-pointing or left pointing arrows. The participants' task was to determine the direction of the centrally presented arrow and respond by clicking the text boxes located at the bottom of the screen, which were marked as “left” or “right”. There were two types of flanker conditions (congruent and incongruent) and three cue conditions in this shortened version of the task (no cue, center cue and spatial cue). The whole session consisted of one 8-trial full-feedback practice block and one experimental block of 64 trials (4 warning types x 2 target locations x 2 target directions x 2 flanker conditions x 2 repetitions) (see Appendix I). As the number of stimuli was low, there were only two blocks (practice and test blocks), with a short break (3000 msec.) between them.

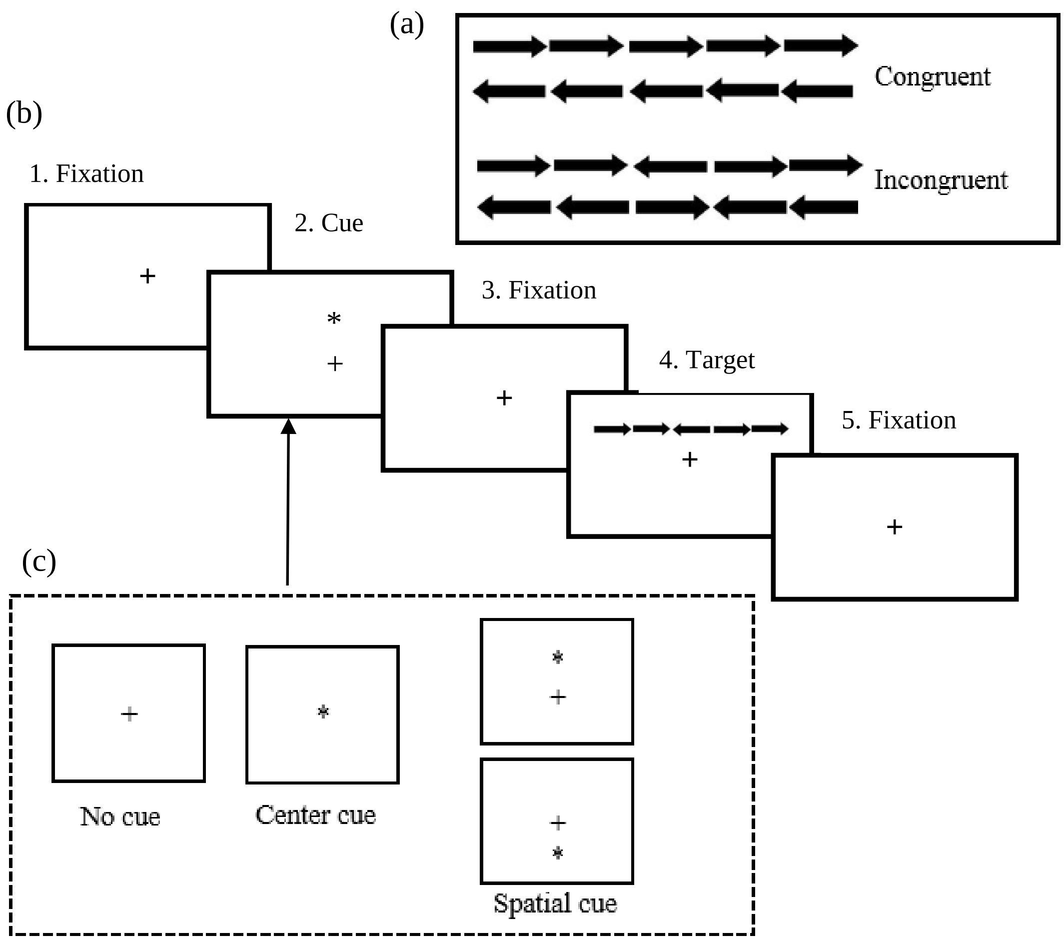


Figure 4.2. Experimental procedure. (a) Four types of target stimuli figures used in the experiment; (b) Sequence of a sample trial (c) Cue conditions.

The presentation of the stimuli was randomized. There were five events in each trial. The trial started out with a fixation period for 400-1000 milliseconds, immediately followed by a warning cue, which remained on the screen for 400 milliseconds. Third, following a short fixation period for 400 milliseconds, after the warning cue, the target and flankers were presented simultaneously and remained on the screen until the participants responded. Finally, there was a post-target fixation period for 2000 msec. The experimental procedure for the task is illustrated in Figure 4.2 above.

4.6 PROCEDURE

At the beginning of the study, the researcher explained the phases of the study in each class by giving the students information about what they were expected to do each week if they agreed to participate. In addition to oral explanations, each week, students were also presented with posters demonstrating the steps to be taken and other related guidelines. Since the researcher was also teaching in the same program at the time of data collection, students were encouraged to see her if they had additional questions or concerns. It was pointed out to the participants that their participation was voluntary and would not affect their grade for the class, and that the links to the training videos would be made available to the learners in the control groups after they took the post-test. Participants were not paid, but offered sweet snacks as a token of appreciation. Information sheet and consent forms were also collected from the participants (see Appendices A & B). Figure 4.3 below summarizes the procedure for the data collection.

Data were normally collected in computer labs. However, due to unexpected circumstances such as holidays falling on computer lab day for two of the control groups, the data collection was done in regular classrooms twice.

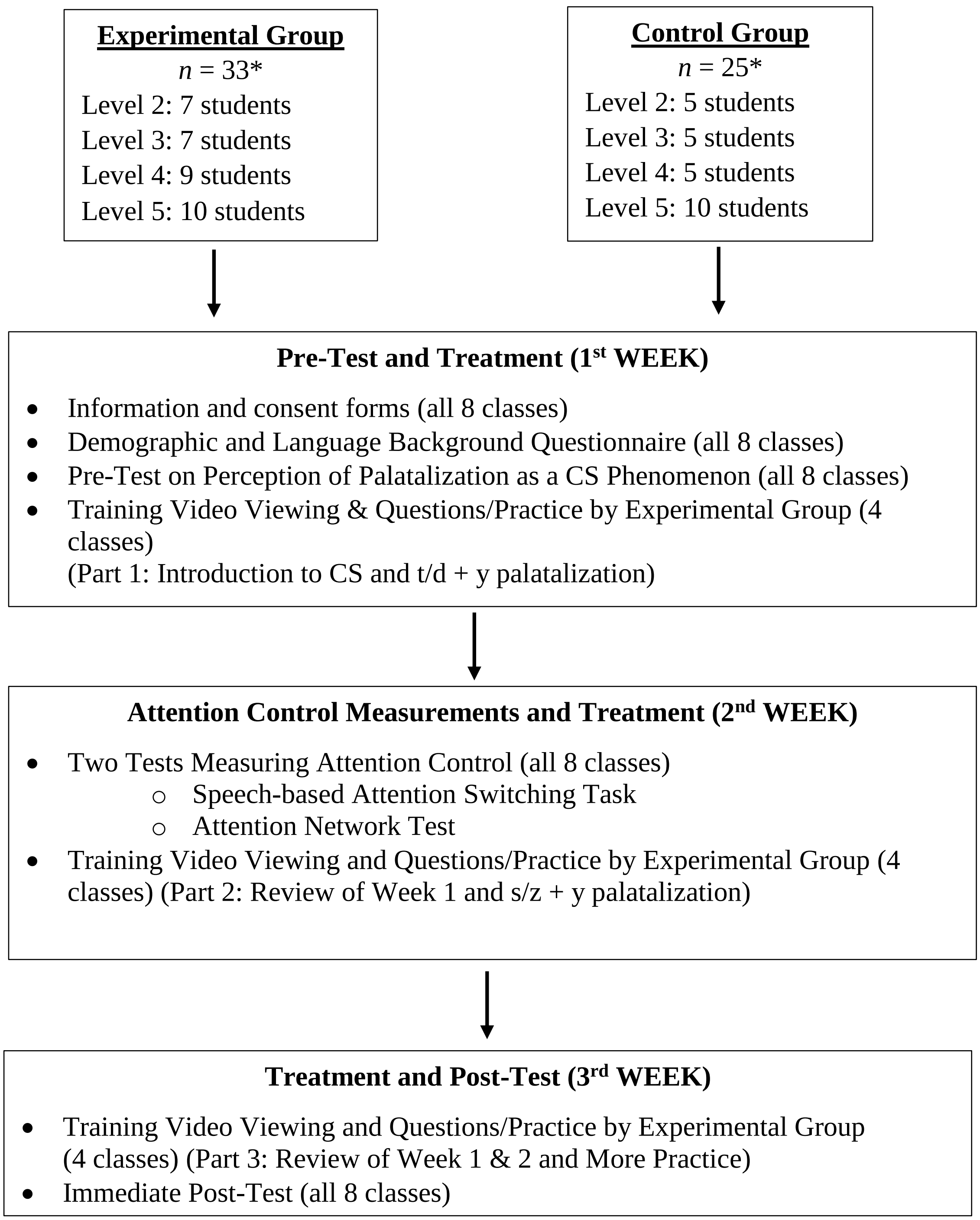


Figure 4.3. Graphical summary of the procedure.
*shows the number of the participants included in the final analyses.

Data collection began in the 5th week of a 9-week term to be able to collect data in eight intact classes during three full weeks. As all Speaking & Listening Classes were scheduled in the same time slot and only four days a week, the challenge was to visit each classroom once every week for three weeks. In total, the researcher made 24 visits to complete the data collection. Due to overlaps in computer lab times, the timing of the visits varied. While in some classes data were collected in the last 30 minutes of the class period, in some others it took place in the first 30 minutes or occasionally in the middle of the lab hour.

Once the students were informed about the study and their consents were gained, they completed the language background questionnaire. Then, they were given the pre-test using the commercial stimuli presentation software, Paradigm (2007), on 9.7-inch Apple iPads with headphones attached. Next, students, who were in one of the four classes assigned to experimental group, watched Part 1 of the training video series on a Macintosh computer in the computer labs. The training videos had to be watched on computers because during video viewing, the students were supposed to answer content-related questions presented through EDpuzzle, which worked better on computers than mobile devices. On the other hand, the control group only took the pre-test.

In Week 2, all classrooms took two types of attention control tasks in a single session. There was an option to take a break up to 3 minutes between tasks, but they could also proceed without stopping. On average, it took students 20 minutes to complete both attention control tasks. Similar to Week 1, the learners who ended up in the experimental group also watched the second part of the training video series while the control group members did not.

In Week 3, all students took the Post-Test. There was no Delayed Post-Test for this study as duration of each term was too short to give such a test at this IEP program.

The students assigned to the experimental group also took the final phase of the treatment (i.e., Part 3 of the training video series) prior to the Immediate Post-Test. Additionally, the control groups were provided with the links to the training videos once they took the Post-Test.

4.7 SCORING PROCEDURES

For the perception test, responses were either correct or incorrect, so the software used for data collection was set to assign '1' for correct and '0' for incorrect answers. A total score was calculated for each individual for two time points, the pre-test and the post-test. A gain score was also calculated for each individual by subtracting the pre-test score from the post-test score.

For the speech-based attention control task (AC Task 1), two measurements were reported. Learners' mean reaction time (RT) measurements for shift and no-shift trials were computed. This calculation only included correct trials. A mean shift cost was calculated by subtracting the mean RT for no-shift-trials from the mean RT for shift-trials [Shift Cost = RT_{shift} - $RT_{noshift/repeat}$]. Error rates on each of the shift, no-shift (repeat) and shift cost measurements were also calculated for each individual.

For the ANT (AC Task 2), individual scores were calculated across all flanker types and cue conditions. These calculations included (a) a mean RT and error rate for incongruent (across all cue conditions) and congruent (across all cue conditions), (b) a mean RT and error rate for center cue and spatial cue condition across both flanker types, (c) a mean RT and error rate for no cue, congruent; no cue, incongruent; center cue, congruent; center cue incongruent; spatial cue, congruent; spatial cue, incongruent. Then, a final calculation of two types of attention control measurements were obtained. The score for orienting for each individual was calculated by subtracting the mean RT for spatial cue condition (across both flanker types) from the mean RT for center cue condition (across both flanker types). The conflict score was calculated by subtracting the congruent trial types from the incongruent trials across all three cue conditions.

4.8 CHAPTER SUMMARY

This chapter provided a detailed account of the experimental design and methods of data collection. This included context, data collection tools and procedures, ethics of research and data scoring. The following chapter presents the statistical procedures used to analyze the data along with the findings to answer the research questions.

CHAPTER 5

RESULTS

5.1 INTRODUCTION

This chapter reports the results of the statistical analyses that were carried out in order to answer the proposed research questions. Table 5.1 lays out the research questions and variables in addition to statistical procedures used to answer these research questions.

The analyses for Research Question 1 were conducted using a two-way Repeated Measures Analysis of Variance (RM ANOVA) using learners' scores on the pre- and post-tests to examine the effects of training in connected speech perception. The independent variables (IVs) in this analysis were group assignment (EG, CG), and time (pre-test and post-test) while the dependent variable (DV) was the mean percentage score of the forced-choice perception test on the pre- and post-tests. To answer Research Question 2, which aimed to explore whether there was a relationship between learners' gains as measured by the pre- and post-tests and their attention control as measured by two different AC tasks, both parametric and non-parametric correlation analyses were carried out. Finally, the results of the correlation analysis between the two different AC tasks, which formed Research Question 3, were calculated by conducting separate correlation analyses. All statistical analyses were done using IBM Statistical Package for Social Sciences (SPSS), version 22.

Table 5.1

Research Questions, Variables and Analysis Methods

Research Question	Variables	Analysis Method
RQ1: What are the effects of form-focused online training in improving connected speech perception (specifically word boundary palatalization) by ESL learners?	IV1: Treatment (CG, EG) IV2: Time (Pre- & Post-Test) DV: Mean scores based on the scores from pre-test to post-test DV: Gains based on the scores from pre-test to post-test	Two-way repeated measures ANOVA
RQ2: What is the relationship between students' performance scores on the perception test (pre-test, post-test, and gain scores) and their attention control (AC) as measured by (a) Attention Network Test (ANT) and (b) Speech-Based Attention Switching Task?	V1: Pre- test, Post-Test and Gain Scores V2: AC accuracy scores and reaction time measurements of AC Task 1 V3: AC accuracy scores and reaction time measurements of AC Task 2	Pearson's product-moment correlation coefficient (PPMCC) Spearman's rank order correlation
RQ3: What is the relationship between attention control scores as measured by an online Attention Network Test (ANT) and a Speech-Based Attention Switching Task?	V1: AC accuracy scores and reaction time measurements of AC Task 1 V2: AC accuracy scores and reaction time measurements of AC Task 2	PPMCC

5.2 AN ANALYSIS OF THE PRE- AND POST-TEST SCORES

5.2.1 DESCRIPTIVE STATISTICS

Distribution of perception test scores was examined separately for pre- and post-tests. A summary including the means, standard deviations, skewness, and kurtosis values for each time point is provided in Table 5.2. The sample means suggest meaningful differences in the average scores for the two groups at two different times; however, a more formal test of the null hypothesis, $H_0 : \mu_{CG} = \mu_{EG}$ is needed.

Table 5.2

Descriptive Statistics for Scores on Pre- and Post-Tests by EG and CG

Group	Test[1]	*M*	*SD*	*Sk*	*SE*	*Ku*	*SE*
(1) Exp. Group	Pre-test	30.85	3.78	-.060	.409	-.797	.798
(*n* = 33)	Post-test	33.66	3.73	-.225	.409	-.802	.798
	Gains	2.81	3.15	-.450	.409	-.487	.798
(2) Control Group	Pre-test	32.76	4.21	.106	.464	-.976	.902
(*n* = 25)	Post-test	33.44	4.87	-.296	.464	-.904	.902
	Gains	.68	2.67	-.011	.464	-.568	.902
(3) Total	Pre-test	31.67	4.05	0.95	.314	-.707	.618
(*n* = 58)	Post-test	33.57	4.22	-.289	.314	-.734	.618
	Gains	1.9	.408	-.121	.314	-.737	.618

[1] The maximum score for the pre- and post-test is 44.

The hypothesis was determined to be tested by RM ANOVA, but it should be kept in mind that the validity of the resulting p-value normally depends on the assumptions of normality, homogeneity of variance and sphericity. Since there are two time points in the study, the sphericity assumption was not an issue for this RM ANOVA procedure as no values were calculated through Mauchley's Test of Sphericity.

Perception pre-test and post-test scores were analyzed for normality of distribution across two proficiency levels. The normality assumption appeared reasonable for both groups at both times through an examination of histograms, box-plots and scatterplots. They also had skewness values that were mostly close to zero, and the kurtosis values that were not greater than +1/-1. In addition, the Shapiro-Wilk test yielded nonsignificant results ($p > .05$) for each group at each time point. In addition, the assumption of homogeneity of variances has been met based on the non-significant results of Levene's test ($p > .05$) for each group at each time point.

5.2.2 RM ANOVA PROCEDURE

To examine the variability between groups before training, a one-way ANOVA was conducted on the participants' pre-test scores (n = 58), and the mean scores were found to be similar at the beginning of the data collection [F (1, 56) = 3.30, p = .074, η^2 = .06)]. Based on the information learners reported on the demographic and language background questionnaire, experimental and control groups were also compared on each of the variables presented in Table 4.1 (mean age, length of residence, age of onset, listening to English as well as exposure to English outside of class) using separate one-way ANOVAs. The results indicated no significant differences between groups on any of the variables (p > .05) before the treatment.

Having met the assumptions, a two-way RM ANOVA was conducted with an alpha level set to α = .05 for each effect to compare the main effects of test time and group as well as interaction between time and treatment. The within-groups factor was time of test (pre-test, post-test) and the between-groups factor was instructional condition (treatment group and control group). The results of the two-way RM ANOVA are summarized in Table 5.3 below. It should be noted that the RM ANOVA calculation was based on performance scores on target items only, and scores on control and filler items were not reported as a part of the analyses made in this work. On the perception test, there was a main effect for time, [F (1, 56) = 24.83, p =.000, η_p^2=.307], but no main effect of group was found [F (1, 56) = .681, p =.413, η_p^2=.012]. In addition, there was a statistically significant interaction effect of time and group [F (1, 56) = 7.743, p =.007, η_p^2=.121].

Table 5.3

Summary Table for the Two- way RM ANOVA (n=58)

Source	*SS*	*Df*	*MS*	*F*	*Partial* η^2	*Power*
Between-subjects effects						
Group	20.19	1	20.19	.681	.012	.128
Error	1659.12	56	29.63			
Within Subjects effects						
Time	104.31	1	104.31	24.83*	.307	.998
Time X Group	32.52	1	32.52	7.743**	.121	.781
Error	235.18	56	4.20			

* p = .000 ** p = .007

The findings indicate that a mere 1.2 % of the between-subjects variance was accounted for by the main effect of group. Additionally, the effect of time accounts for 30.7 % of the Time plus within-subjects error variance, while Time by Group interaction accounts for 12.1 % of the Time X Group plus within-subjects error variance. An interaction graph is provided in Figure 5.1 to show the relative position of the four sample means.

The size of the main and interaction effect can also be expressed using Cohen's effect sizes because as Larson-Hall (2010) expresses squared values such as the partial eta-squared, in fact, "lack directionality" (Rosenthal & DiMatteo, 2001 as cited in Larson-Hall, 2010, p. 116). According to her, they only show the "observed proportion of explain variance" (Kline, 2004, p. 100 as cited in Larson-Hall, 2010); therefore, she calls them percentage variance effect sizes (p. 116). The effect size she recommends using for ANOVA measures is Cohen's *f*. Thus, partial eta squared figures were also converted to Cohen's *f* for time main effect (f = .67) and time X group interaction (f =.37), and according to Cohen's (1998, 1992) guidelines for interpreting effect sizes the main effect of time was found to be large (f > .40), while the time X group interaction effect was determined to be medium (f < .40).

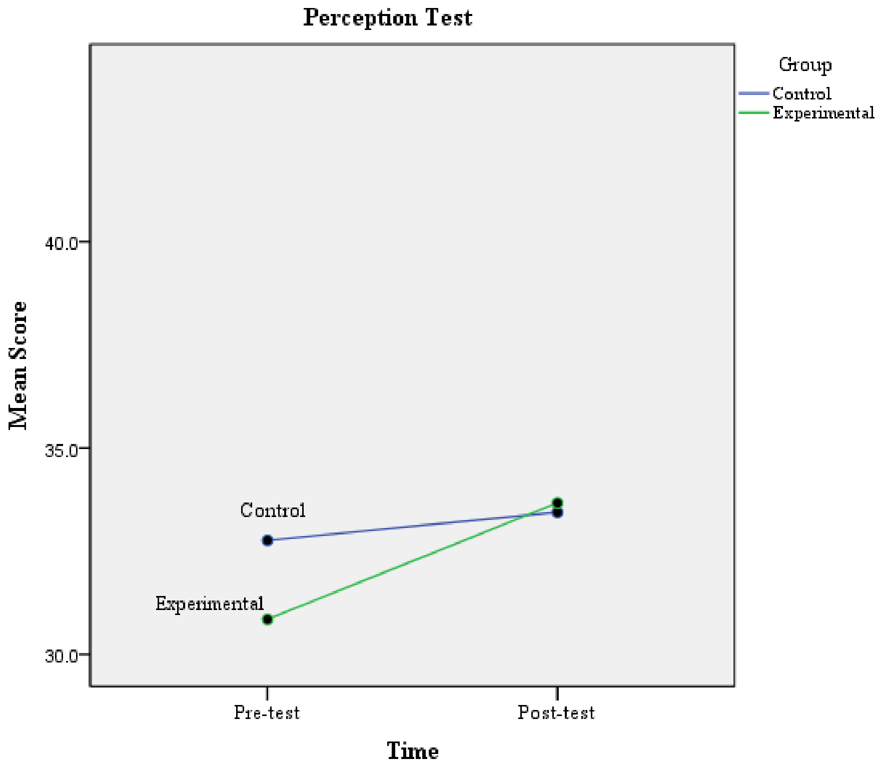


Figure 5.1. Interaction plot for Time X Group.

The conventions provided for partial eta-squared effect sizes by Huck (2012, p. 306) support Cohen's conventions in that the effect size is large for both the main effect of time, and medium for interaction effect in this study as calculated by RM ANOVA.

It should be noted that the two-way RM ANOVA findings reported above in Table 5.4 do not include the proficiency level variable within groups in the analysis. In order to see whether proficiency levels differed in their performance on the pre- and/or post-test, a series of separate (a) 2 X 2 X 2, (b) 2 X 2 X 3, and (c) 2 X 2 X 4 RM ANOVA analyses were conducted by separating learners into (a) two proficiency groups based on their listening test scores as upper and lower, (b) three proficiency groups based on their listening test scores as low, intermediate and upper, and (c) four proficiency groups based on their listening test scores as determined by the criteria set by the intensive language program they were enrolled in. However, none of the RM ANOVA procedures revealed any significant effect of proficiency as a measure to predict pre-test, post-test or gain scores. Furthermore, no significance level was reached when the same analyses were conducted based on their actual placement levels in the program, either. The results indicated no significant effect of proficiency as a between-subjects variable as revealed by the analyses, which may be attributable to the low number of students in each proficiency level after the elimination of the participants who missed either the pre-, post- or one of the AC tasks.

5.3 THE RELATIONSHIP BETWEEN PERCEPTION TEST SCORES AND ATTENTION CONTROL

5.3.1 AN ANALYSIS OF CORRELATION BETWEEN SPEECH-BASED ATTENTION SWITCHING TASK AND PRE-TEST, POST-TEST & GAIN SCORES

An attention control score for each individual was calculated using the Speech-Based Attention Switching Task (AC Task 1). This score was calculated for each individual using the scoring procedures explained earlier in Chapter 4.

Table 5.4

Descriptive Statistics for RT (msec) and Error Rate (%) for Shift and No-shift Trials

Trial Type	Group	RT	*SD*	*Sk*	*Ku*	Error rate	*SD*	*Sk*	*Ku*
Shift	Experimental (*n*=33)	1242	441.7	1.16[1]	.378[4]	.21	2.83	-1.98[1]	4.05[4]
	Control (*n*=25)	1307	464.6	1.20[2]	.359[5]	.20	1.97	-.617[2]	-.705[5]
	Total (*n* =58)	1270	448.9	1.15[3]	.259[6]	.21	2.48	-1.77[3]	3.91[6]
No-shift	Experimental (*n*=33)	1205	426.1	1.18[1]	.255[4]	.20	1.83	-.901[1]	-.104[4]
	Control (*n*=25)	1232	436.9	.991[2]	.089[5]	.22	2.52	-1.16[2]	1.12[5]
	Total (*n* =58)	1217	427.1	1.03[3]	.054[6]	.21	2.25	-1.07[3]	.754[6]

Note. RT= Reaction time

[1] SE =.409 [2] SE = .464 [3] SE = .314 [4] SE = .798 [5] SE = .902 [6] SE = .618

The descriptive statistics summarizing the mean reaction time (RT), error rate, standard deviation for shift and no-shift trials as well as the shift-cost measurements are presented in Table 5.4. The purpose of presenting the descriptive statistics in Table 5.4 above was to see whether the groups were similar enough to conduct the correlation analysis as a single group of respondents. The mean reaction time and accuracy scores for each group shows that they were relatively similar in their performance on the speech-based AC task.

In addition, to see whether groups differed in their attention control task performance, an RM ANOVA was conducted separately for a) mean accuracy scores, and (b) mean reaction time for shift and no-shift conditions as within subject variables. For the accuracy measurements, the *F* test demonstrated that there was no effect of group [F (1, 56) = .030, p = .862)], no effect of condition on accuracy [F (1, 56) = .035, p = .853)], nor any significant effect of interaction [F (1, 56) = .933, p = .338)]. This shows that both experimental and control groups performed similarly across shift and no-shift conditions. The RM ANOVA analysis conducted for RT measurements revealed that there was no effect of group [F (1, 56) = .159, p = .692)], but a significant effect of condition on reaction time [F (1, 56) = 22.24, p = .000), η_p^2 = .28], responses being faster on no-shift trials than responses on shift trials. There was no significant effect of interaction [F (1, 56) = 2.90, p = .094)] which indicates that both experimental and control groups performed similarly across shift and no-shift conditions. The results on mean accuracy and RT across groups show that there was no significant differences between groups on AC task. For the analysis of the AC task and perception test scores, the groups were not split into experimental and control groups. Therefore, assuring that there was no observed effects of group or interaction was crucial.

In addition, before computing the correlation analysis between the perception test scores and the attention control task findings, any effects of speed-accuracy trade-off in the data was also examined. In order to do this, the Pearson product-moment Correlation Coefficient was calculated between the accuracy and RT on shift trials to see if there was a positive correlation between those two variables.

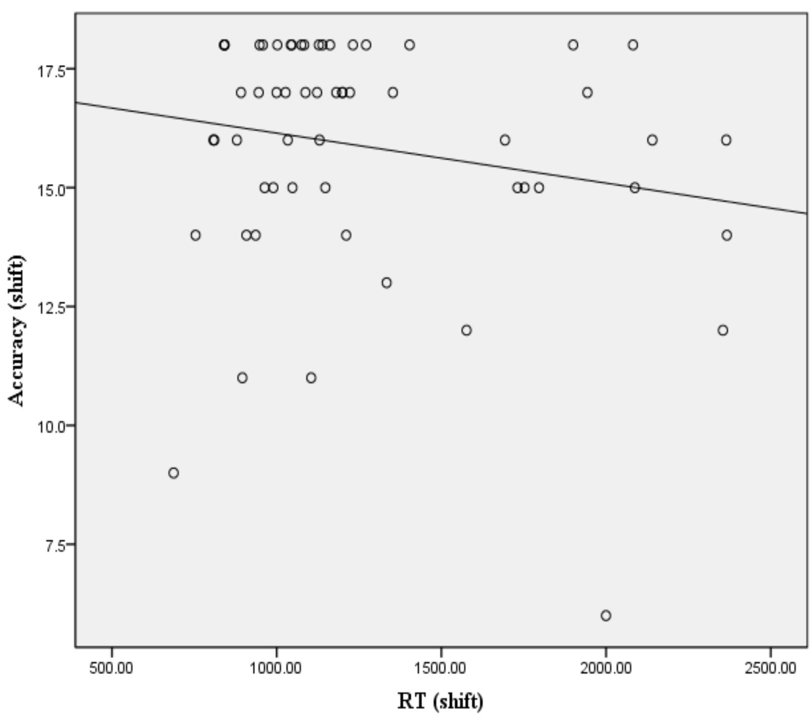


Figure 5.2. The relationship between RT and accuracy measures on shift trials.

As is illustrated in Figure 5.2, there was no positive correlation between the accuracy and reaction time measurements on shift trials, which indicates no sign of speed-accuracy trade-off in participants' responses (r = -.190, p = .153).

To answer our main research question of whether there was a correlation between learners' attention control scores on the Speech-Based Attention Switching Task and their scores on the connected speech perception test, the Pearson product-moment Correlation Coefficient (PPMCC) was determined to be computed. First, a shift-cost was calculated by subtracting the mean RT on no-shift trials from the mean RT on shift trials [$RT_{shift} - RT_{noshift}$] for accurate responses on the AC task (n = 58, M = 53.17, SD = 87.27, *Sk.* = .017, *Ku.* = .684, Shapiro-Wilk p =.55) and Figure 5.3 below shows the distribution of reaction time for all respondents.

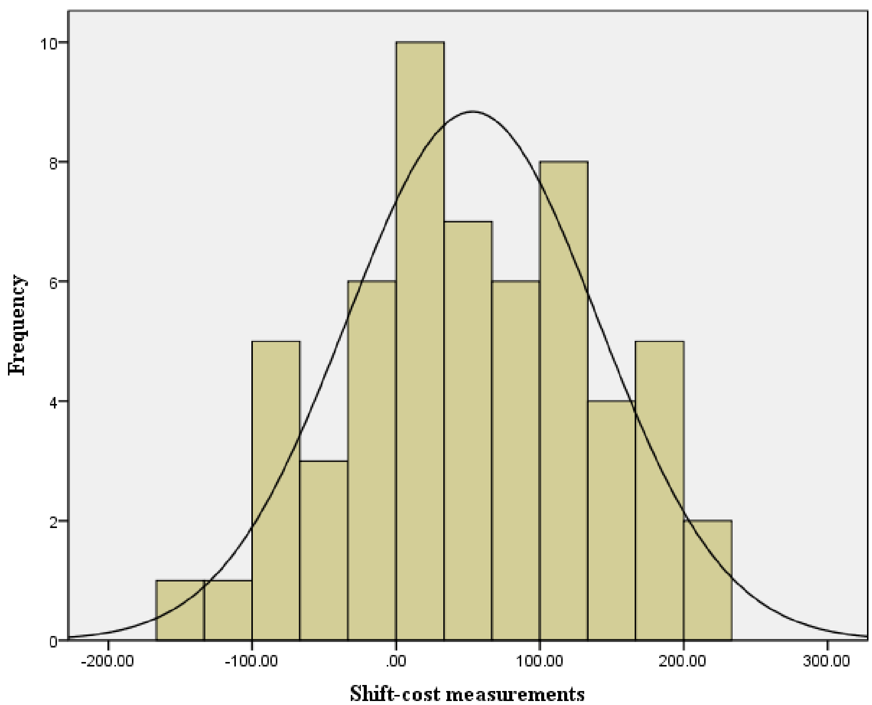


Figure 5.3. A histogram of shift-cost measurements (n = 58).

Assumptions associated with the Pearson's correlation analysis were examined and no violations were noted. The linearity assumption was determined to be normal via examination of a scatterplot. The normality assumption appeared to be tenable for each of the variables, which had skewness and kurtosis values not greater than -1/+1, and nonsignificant Shapiro-Wilk test for all variables (*p* >.05) confirmed this. Furthermore, an examination of the residual plot showed residuals appeared to be homoscedastic and normally distributed as no absolute value of studentized residual greater than 3 was detected.

Therefore, to investigate the relationship between learners' attention control and scores on the connected speech perception test, Pearson's correlation analyses were conducted between the learners' AC task shift-cost measurements and their scores on the Pre-test, Post-test, and Gain scores (Score $_{post\text{-}test}$ – Score $_{pre\text{-}test}$).

Table 5.5
Correlations between Shift-cost and Pre-test, Post-test & Gain Scores (n=58)

Variables	Shift-cost	Pre-test	Post-test	Gain Scores
Shift-cost	1.00			
Pre-test	-.16	1.00		
Post-test	-.39[a]	.73[c]	1.00	
Gain Scores	-.34[b]	.32[d]	.42[e]	1.00

[a] p =.002 [b] p =.008 [c] p =.000 [d] p = .014 [e] p = .001

The results revealed negative significant relationships between (a) the AC task shift-cost measurement and the post-test [r (58) = -.392, p = .002], and (b) the AC task shift-cost measurement and gain scores [r (58)= -.344, p = .008], both of which yielded medium effect sizes (Cohen, 1988), but no correlation was found between the pre-test and AC task performance (see Table 5.5). The scatterplots showing the linear relationships revealed by the correlation analyses are presented in Figure 5.4 below.

A post-hoc power analysis conducted using the *R* statistical software indicated that the sample size (n = 58), with two-tailed tests, and alpha, α = .05, yields a statistical power of .87, and .76 for each of the significant correlation analyses reported between shift-cost and post-test scores and shift-cost and gain scores, respectively.

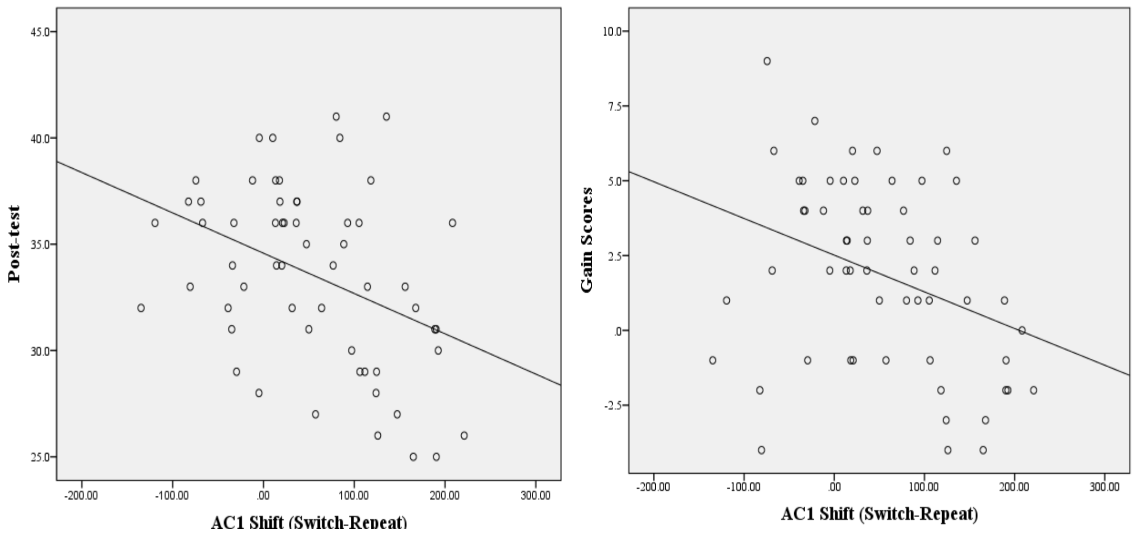


Figure 5.4. Scatterplots showing the linear relationships between the shift-cost measurement and post-test and gain scores.

An additional correlation analysis was also conducted between the performance scores on the perception test by experimental group and the shift-cost measure. Similar to the findings of the overall correlation analysis, the results revealed negative significant relationships between (a) the AC task shift-cost measurement and the post-test [r (33) = -.430, p = .012], and (b) the AC task shift-cost measurement and gain scores [r (33) = -.523, p = .002], but no correlation was found between the pre-test and AC task performance of the experimental group.

5.3.2 AN ANALYSIS OF CORRELATION BETWEEN ATTENTION NETWORK TEST AND PRE-TEST, POST-TEST AND GAIN SCORES

Parametric and non-parametric correlation analyses were determined to be run to investigate any relationship between learners' performance on the Attention Network Test and their scores on the perception test. Before presenting the correlation results, the mean RT, error rate (%) and standard deviations on both flanker types across all three cue types are presented in Table 5.6 and Figure 5.5. Table 5.6a summarizes RT data pooled from correct trials as a function of cue conditions and flanker types. Across all cue conditions, response time was longer for the incongruent flanker type than the congruent flanker type.

Table 5.6

Mean RTs (msec) and Error Rate (%) under Each Condition

	Warning Type		
Congruency	No-cue	Center cue	Spatial cue
(a) Mean RTs (msec) and standard deviations			
Congruent	657 (176)	694 (283)	644 (159)
Incongruent	760 (219)	754 (243)	771 (244)
(b) *Error Rate (%) and standard deviations*			
Congruent	.4 (1.06)	.3 (.88)	.3 (1.63)
Incongruent	.12 (1.89)	.12 (1.84)	.11 (3.59)

Note. Standard deviations are presented in brackets.

As is presented in Table 5.6 and graphically shown in Figure 5.5, the difference in the mean RT on congruent and incongruent conditions is the largest for spatial cue condition, followed by center cue and no cue conditions. Error rates also show a considerable difference for two different flanker conditions. Possible reasons for and consequences of such a difference in error rates are discussed in Chapter 6.

A set of cognitive subtractions, which were described earlier in Chapter 4 and are repeated below, was used to assess the efficiency of the two attentional mechanisms measured by the adapted, shorter version of the ANT used in the current study. In the calculation of orienting effect per respondent, the mean RT of spatial cue conditions was subtracted from the mean RT of center cue [$RT_{center\ cue}$ - $RT_{spatial\ cue}$].

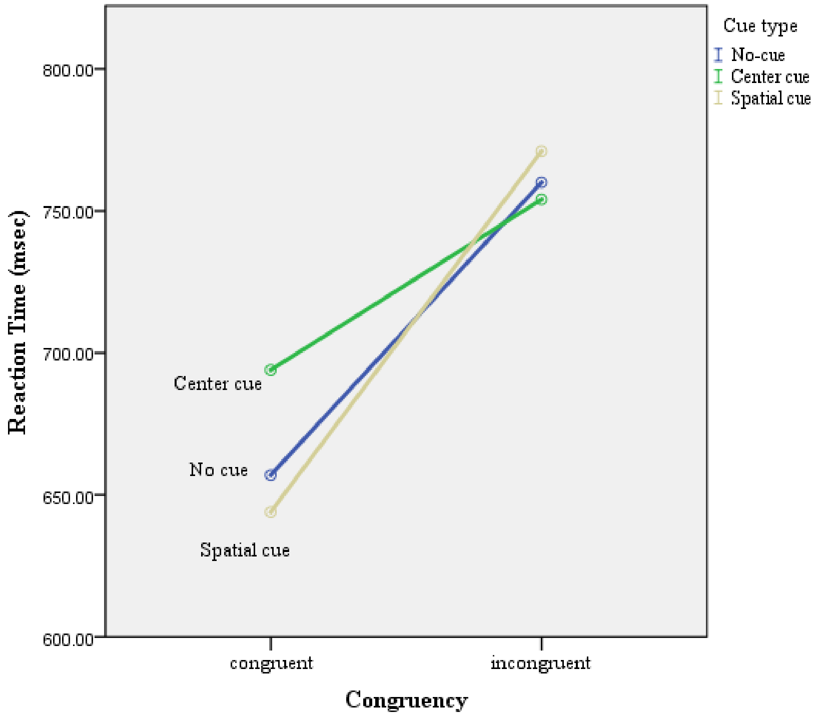


Figure 5.5 Overall mean RTs (msec) for the six conditions presented in the experiment.

The conflict (executive control) effect per subject was calculated by subtracting the mean RT on congruent flanker trials across all cue conditions from the mean RT on incongruent flanker conditions [$RT_{incongruent} - RT_{congruent}$].

The conflict effect (congruent vs. incongruent trials), which was also described as executive control in Chapter 4, was found to be significant indicating faster RTs for congruent (*M* = 773, *SD* = 310.6) than incongruent trials (*M* = 875, *SD* = 353.5); [*t* (57) = 9.77, *p* = .000]. The results of the t-test revealed that the orienting effect was not significant indicating that RTs on trials preceded by a spatial cue as compared to those preceded by a central cue were not significantly different [*t* (57) = -1.29, *p* = .203].

After the calculation of the conflict effect for each respondent, the Pearson's correlation analysis was conducted to examine the relationship between the conflict attentional network and learners' pre-test, post-test and gain scores on the connected speech perception test. The linearity assumption for conducting the correlation analysis was determined not to be violated via examination of a scatterplot. The distributional shape of conflict effects was also examined to determine the extent to which normality assumption was met as suggested by a bell-shaped distribution displayed by histogram, as well as tenable skewness (*Sk* = .311, *SE* = .314) and kurtosis (*Ku* = -.219, *SE* = .618) values closer to zero, and finally nonsignificant Shapiro-Wilk test (*S-W* = .962, *df* = 58, *p* >.067). Furthermore, an examination of the residual plot showed residuals appeared to be homoscedastic and normally distributed.

A summary of the Pearson's correlation analysis between the conflict effect and learners' scores on the Pre-test, Post-test, and Gain scores ($Score_{post\text{-}test} – Score_{pre\text{-}test}$) is presented in Table 5.7 below.

Table 5.7
Correlations between the Conflict Effect and Pre-test, Post-test and Gain Scores[*]

Variables	Conflict	Pre-test	Post-test	Gain Scores
Conflict	1.00			
Pre-test	-.18	1.00		
Post-test	-.37[a]	.73[c]	1.00	
Gain Scores	-.28[b]	.32[d]	.42[e]	1.00

[*] n = 58 [a] *p* =.004 [b] *p* =.032 [c] *p* =.000 [d] *p* = .014 [e] *p* = .001

The results revealed negative significant relationships between (a) the conflict effect and the post-test [*r* (58) = -.37, *p* = .004], and (b) the conflict effect and the gain scores [*r* (58) = -.28, *p* = .032], both of which yielding medium and small effect sizes (Cohen, 1988), but no significant correlation was found between the pre-test and the conflict effect [*r*

(58) = -.18, *p* = .181]. The scatterplots showing the linear relationships revealed by the correlation analyses are also presented in Figure 5.6.

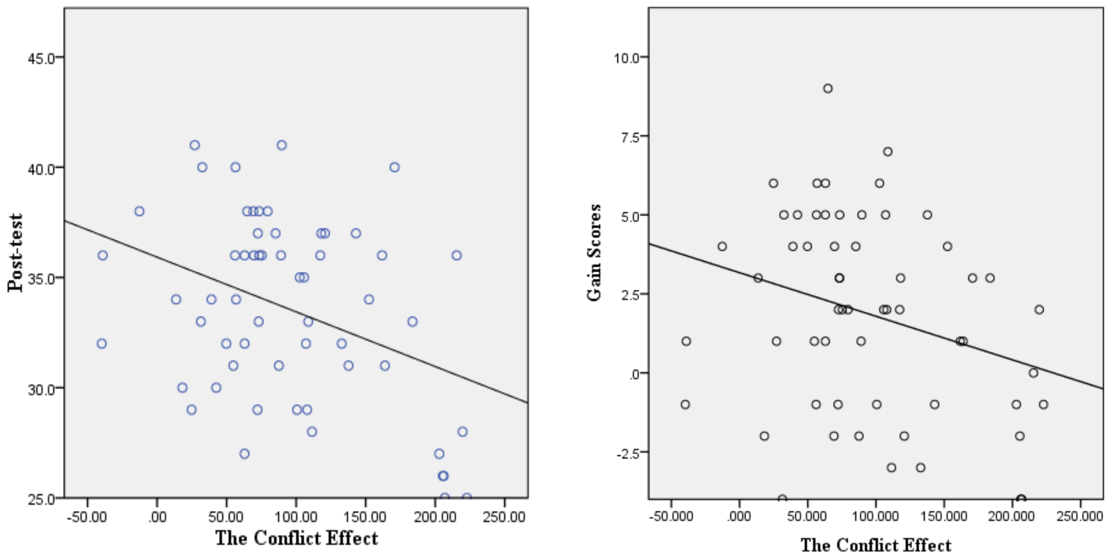


Figure 5.6. Scatterplots showing the linear relationships between the conflict effect and post-test and gain scores.

The findings of the post-hoc power analysis using the *R* statistical software indicated that with a sample size of 58 and alpha level set to α = .05 for this two-tailed test, the statistical power for correlation between the conflict effect and the post-test [*r* (58) = -.37, *p* = .004] was .83 while the power for the correlation between the conflict effect and gain scores [*r* (58) = -.28, *p* = .032], was found to be .58.

Next, in order to conduct a correlation analysis between the orienting effect and perception test scores, assumptions were checked. As mentioned earlier, the orienting effect refers to “selection of information from sensory input” (Fan et al. 2002, p. 1), which “involves […] disengaging attention from its current focus, moving attention to the new target or modality, and engaging attention at the new target or modality” and it is measured using cued signal asking for the location of the stimulus (Posner, Walker, Friedrich, & Rafal, 1984 as cited in Fan et al., 2009, p. 2). For the orienting data, an examination of the scatterplot revealed that the linearity assumption was not violated. However, the distributional shape of the orienting effect appeared to be non-normal as revealed by an examination of boxplot, histogram (Sk = -.097, SE = .314; Ku = 3.74, SE = .618), and significant Shapiro-Wilk test (S-W = .870 df = 58, p =.000). Therefore, it did not appear reasonable to proceed with the Pearson's correlation analysis. To investigate any potential relationship between the orienting effect and the perception test scores, first the extreme outliers were removed to see if removing the outliers would help the data to be considered normal. Then as an alternative to removing outliers, a non-parametric correlation, instead of Pearson product-moment correlation, was also conducted. The results of the procedures are described below.

The first method was the hierarchical removal of outliers, one at a time. As a result of removing a total of eight outliers in the upper and lower bounds, a nonsignificant Shapiro-Wilk test was achieved (S-W = .955, df = 50, p = .057). When the normality assumption was minimally met, first the Pearson's correlation analysis was conducted, but no significance has been reached between any of the perception test scores and the orienting effect [r (50) = .107, p = .459; r (50) = .079, p = .588, n = 50; r (50) = -.047, p = .747, for pre-test, post-test, gain scores, respectively]. Removal of even more outliers, again one at a time, to improve the normality did not change the findings of the correlation analysis in a positive way [r (46) = -.132, p = .381; r (46) = -.015, p = .924; r (46) = .138, p = .359, for pre-test, post-test, gain scores & orienting effect, respectively]. Since removal of outliers did not reveal any significant relationship, as a non-parametric alternative to Pearson's correlation, a Spearman's rank-order correlation was run, keeping the outliers; however, no significant relationship was reached by this procedure, either [r_s (58) = .057, p = .671; r_s (58) = .068, p = .613; r_s (58) = -069, p = .609]. The overall results indicate that there seems to be no linear relationship between the orienting as an attentional network and perception test performance of the learners in the current study.

So far, (a) the results of the RM ANOVA investigating the effects of online training on learners' connected speech perception and (b) the correlation analyses examining the existence of any linear relationships between learners' performance on the perception tests, and their attention control as measured by two types of tasks were presented. The findings reveal a significant effect of time, indicating that both experimental and control groups improved in their ability to perceive the target connected speech forms over time. However, the significant effect of time X group interaction further demonstrates that the experimental group improved more than the control group did, suggesting positive effects of treatment in improving the perception of connected speech forms. In addition, the findings of the correlation analyses showed that there was a negative correlation between the following: the post-test & the gain scores and shift-cost measurement; the post-test & the gains scores and the conflict effect.

The following section looks at to what extent the attention control mechanisms or attentional network properties of the learners were correlated as measured by two different types of AC tasks.

5.4 A CORRELATIVE ANALYSIS OF THE FINDINGS OF THE SPEECH-BASED ATTENTION SWITCHING-TASK AND ATTENTION NETWORK TEST

The tasks used to assess learners' AC measured their AC using different methods as detailed in Chapter 4. The results of the correlation analyses between AC measurements by two different types of AC tasks and the perception test scores revealed significant negative correlations of some sort for both AC tasks. These findings call for a further correlation analysis between these two types of AC tasks for two reasons. First, a significant correlation between any one of the perception scores and both of the AC tasks indicates the existence of a potential significant correlation between the AC scores measured by different AC tasks. Second, since the ANT is a more widely-used measure of attention control, it would be very interesting to see the extent to which the ANT and the Speech-Based Attention Switching Task agree in their assessments of learners' attention control mechanisms in this study.

In order to answer this question, two different correlation analyses were determined to be conducted based on the previous procedure of assumption checking for shift-cost measurements, conflict and orienting data in the previous sections of data analyses. While a Pearson's correlation coefficient was determined to be calculated to examine the relationship between the conflict effect and shift-cost measurements, the relationship between the orienting effect and the shift-cost measurements was examined by Spearman's rank-order correlation. The assumptions were checked in the previous sections for each of the variables and, thus, are not repeated here.

The results of the Pearson's correlation between the conflict effect of the ANT and the shift-cost measure of the Speech-Based Attention Switching task showed a positive significant relationship [r (58) = .38, p = .009]. According to Cohen's (1988) conventions, the correlation coefficient indicates a medium effect size. The findings of the post-hoc power analysis indicated that with a sample size of 58 and alpha level set to $\alpha = .05$ for this two-tailed test, the statistical power for correlation between the conflict effect and the shift-cost measurements [r (58) = .34, p = .009] was .75. Figure 5.7 below illustrates the linear relationship between the two variables.

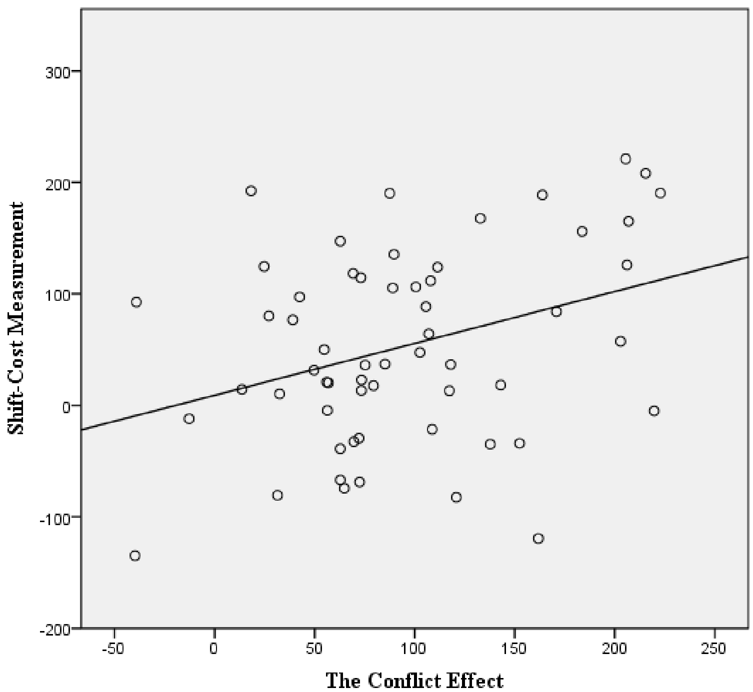


Figure 5.7. The relationship between two types of AC scores. The shift-cost measure was calculated by Speech-Based Attention Switching Task and the Conflict Effect was calculated by the ANT.

As for the relationship between the orienting effect of the ANT and the shift-cost measurement by the Speech-Based Attention-Switching Task, the results of the Spearman's rank-order correlation analysis revealed no significant correlation between the two AC scores [r_s (58) = -.70, p = .60]. The results are discussed in the following chapter.

5.5 CHAPTER SUMMARY

In this chapter, results of the experiments were summarized using descriptive and inferential statistical procedures. A significant effect of interaction between the group assignment and the time of testing was found as revealed by the results of the RM

ANOVA. Furthermore, the AC measurements of both AC tasks showed significant linear relationships with the post-test and gain scores with varying degrees of statistical and practical significance; however, no such relationship was discovered for the pre-test. Finally, a significant correlation was also found between learners' AC scores measured by the Speech-Based Attention Switching Task (shift-cost measurement) and the ANT (the conflict effect, but not the orienting effect).

CHAPTER 6
DISCUSSION

6.1 OVERVIEW

This chapter discusses the findings reported in Chapter 5 in reference to the three research questions that guided this study and in the light of relevant research. A section on the pedagogical implications of the study for SLA will follow, and then the dissertation concludes with limitations and suggestions for further research.

6.2 DISCUSSION OF RESEARCH QUESTION 1

RQ1: What are the effects of form-focused online training in improving connected speech perception (specifically word boundary palatalization) by ESL learners?

It was hypothesized that ESL learners in the experimental group would benefit from online, form-focused training and, thus, would improve more than the control group by the end of a three-week period. The content of the online training, which aimed to improve learners' perceptual abilities of the word-boundary palatalization, was predicted to help them better understand the target forms as shown by previous instructional studies of connected speech learning (Ito, 2006; Matsuzawa, 2006).

In order to investigate whether form-focused online training on connected speech leads to improvement in the perception of the target forms, a forced-choice perception test with two response options was given to the learners on a pre- and post-test basis. The findings showed the main effect of time was significant with a large effect size, indicating that both control and experimental groups improved regardless of the training; however, the interaction effect further revealed that the experimental group improved more than the control group did, revealing a medium effect size.

The findings were in line with previous research which suggested improvement when various connected speech forms were taught to learners (Brown & Hilferty, 1986; Matsuzawa, 2006; Underwood and Wallace, 2012). The fact that the present study employs a different approach to measuring improvement in perception of connected speech by using a forced-choice test instead of a dictation test makes the contribution of this study to the literature more valuable. This is because the use of this task helps better control learner variables such as vocabulary knowledge, note-taking or spelling abilities, and this means being able to measure the perceptual abilities with fewer confounding factors involved. Also, not all studies looking at connected speech perception found significant improvement. In a study by Alameen (2014) looking at linking in connected speech, learners took a dictation test before and after the treatment on a pre-, post- and delayed post-test basis. The results indicated no significant improvement in the perception of linking although there was improvement in production. She attributed lack of improvement in perception to several factors, one of which is the shorter amount of training time compared to previous studies that reported significant improvement. Additionally, she also claimed that as opposed to training students to perceive connected speech forms such as ‘going to’ vs. ‘*gonna*’, improving the ability to perceive linking may require a different perceptual competence which may be more challenging and time-consuming. I agree with Alameen (2014, p. 79) in that different types of connected speech forms may require different skills and varying periods of time to improve the ability to perceive them. Although finding non-overlapping results in perceptual improvement in connected speech may well be related to a variety of factors, such as the length and type of instruction, learner background, or the context of learning and testing, another way to look at this may be the way the perception of connected speech is measured. Instead of using a dictation test, using a test which would help learners to recognize or to focus more on perception alone may yield different results in showing perceptual improvement in connected speech forms of different types. In this respect, the use of a forced-choice test in this study may initiate the use of a new approach to assess and explore connected speech perception.

Despite reporting the relevant findings to answer the research questions, Chapter 5 does not present any data regarding the questions on which learners had relatively high or low mean accuracy scores. The classification of the questions based on accuracy scores may provide a more nuanced look at the different sub-types of palatalization investigated in the study. This may help us go beyond the overall findings of the study, which report improvement, and make stronger claims on the type of palatalization with which the learners had the most difficulty. For this purpose, an item facility/difficulty analysis of all the questions used in the perception test was conducted. Brown (1996) describes IF (item facility), which is also known as item difficulty (DIF), as "a statistical index used to examine the percentage of students who correctly answer a given item" (p. 64). In order to calculate IF, also commonly known as p-value in classical test theory, the sum of correct responses is divided by the total number of students taking the test. The p-value ranges from 0.00 to 1.00, and as the value increases, the item difficulty decreases.

The ideal difficulty for binary items (having two response options) is calculated to be halfway between the percentage of pure chance (50%) and 100%, and is calculated to be .75 [.50 + (1.00 - .5 / 2] (Thompson & Levitov, 1985 as cited in Matlock-Hetzel, 1997). Therefore, although there is no agreement on the exact values to be used in the literature, any item below P ≤ .25-.30 is considered to be difficult, and any item above P ≥ .85-.90 is generally considered easy. Table 6.1 below reports the questions which students found the easiest and the most difficult based on item facility (IF)/ item difficulty (DIF) analysis. In reporting the difficulty of the items, first DIF values and standard deviations under six different conditions were calculated (Pre/Post for Experimental and Control as well as Combined Pre/Post DIF values). Then, the items which were above the difficulty level of 80% and below the difficulty level of 50% across all scores were determined, as there were not many questions below the difficulty level of 30%. Table 6.1 reveals some interesting insights on the data and the type of the questions.

First, DIF calculations show that when the question is a non-target form, that is, a non-palatalized form paired with a palatalized form in a certain question, learners were much better in getting the correct answer. In fact, this is not considered to be surprising because the non-palatalized form of the pair is usually more easily discernable. An example of this is seen in questions ‘knows June’ and ‘knows Judith’ paired up with ‘knows you’ and ‘knows you did’ forms. When the former two were the correct options, respondents found it very easy to choose the correct answer. It may be due to the fact that each of the words ‘Judith’ and ‘June’ bears stress as opposed to the pronoun ‘you’ in ‘knows you did’ or ‘knows you’, and each has word-final consonants ([θ] and [n]) that are salient when pronounced.

Table 6.1
An Item Difficulty Analysis of the Most Frequent Correct and Incorrect Questions

Item	Control		Experimental		Total	
Target Items	Pre	Post	Pre	Post	Pre	Post
(a) The most difficult items						
typed your (vs. types your) [t] + [j] => [tʃ]	.4	.45	.40	.58	.43	.50
walks your (vs. walked your) [s] + [j] => [ʃ]	.4	.42	.44	.58	.41	.52
prize you (vs. price you) [z] + [j] => [ʒ]	.36	.45	.28	.55	.41	.43
can't yards (vs. can't charts) [t] + [j] => [tʃ]	.48	.27	.48	.33	.36	.40
sold younger (vs. sold jungle) [d] + [j] => [dʒ]	.16	.15	.28	.43	.16	.38
(b) The easiest items						
knows you did (vs. knows Judith) [z] + [j] => [ʒ]	.92	.85	.80	.82	.88	.81
this year (vs. this cheer) [s] + [j] => [ʃ]	.84	.91	.80	.82	.88	.81
miss your (vs. miss shore) [s] + [j] => [ʃ]	.84	.88	.84	.88	.86	.86
knows you (vs. knows June) [z] + [j] => [ʒ]	.96	.76	.92	.88	.85	.81
Non-target items						
(a) The most difficult item						
miss shore (vs. miss your)	.52	.55	.52	.48	.53	.50
(b) The easiest items						
knows June (vs. knows you)	.92	.97	.92	.97	.95	.95
knows Judith (vs. knows you did)	.88	.91	.88	.88	.90	.88
suppose shelling (vs. suppose yelling)	.90	.82	.92	.95	.86	.97
could choose (vs. could use)	.88	.85	.89	.79	.86	.84
sold jungle (vs. sold younger)	.92	.79	.84	.91	.84	.88
not chat (vs. not yet)	.84	.85	.95	.91	.84	.93
in case junk (vs. in case young)	.96	.73	.88	.85	.83	.86

Note. Difficulty classification: Difficult: < .30; Moderate: > .30 and < .84; Easy: > .85.

Another example of this is the question which contrasts the pair ‘in case junk workers’ and ‘in case young workers’. In this question, while the former option has a very high item p-value across almost all time points by all groups ($p > .80$), as seen in Table 6.1, the latter received a p-value of .60 or below in most time points by both groups.

Similarly, the pair ‘haven’t juiced’ vs. ‘haven’t used’ received relatively different item p-values by almost all groups at all time points ($p > .75$ and $p < .55$). This may evidence learners’ not having the knowledge of respective palatalization, and thus regardless of the sentence they heard, they may have thought that the correct answer was the sentence which sounded like the palatalized form (i.e., the non-target form), but not really the palatalized form of the pair. This is revealed by the pre-test scores, which seemed to have improved in the post-test scores.

Furthermore, the use of vowel quality as a means to select the correct answer was more challenging for learners, as revealed by the high error rate for the question which asked learners to tell the difference between ‘can’t yards’ vs. ‘can’t charts’. Although there is a contrast in the final consonant clusters, learners were more inclined to think that the correct answer would be the non-target form where there was no palatalization taking place across word boundaries.

Overall, the results indicate that when there is a stress shift or an additional consonant in the non-target form which paired with a palatalized, target form, the error rate decreased. However, when the discriminating variable or cue between the pairs is vowel quality rather than an additional consonant or stress, learners’ error rate increased, indicating that using vowel quality as a means to differentiate between two similar sounding forms was harder for learners.

On the other hand, DIF also demonstrates that the questions which the learners had the most difficulty with were the ones which paired two target, palatalized forms (e.g., ‘walks your’ vs. ‘walked your’). They were rather challenging for learners because (a) both forms sounded very much alike, and (b) learners had to have the knowledge of or ability to evaluate the palatalization rules in both options, as both underwent palatalization, unlike the other type of questions where palatalization was observed in only one of the options.

Despite the attempts to come up with sentences in which target and non-target forms would both equally make sense and sound equally authentic, the DIF analysis indicates that some sentences still made more sense semantically or seemed more reasonable due to factors such as semantic associations or contextual information. Below is an example of the response options on the test:

(6a) The store sold younger animals at a higher price.

(6b) The store sold jungle animals at a higher price.

While the non-target form (6b) received very high accuracy scores ($p > .80$ in almost all measurement points by both groups), the target form (6a) was at the bottom of the difficulty index, indicating its being a very hard question. There may be possible reasons for this. One is the fact that the phrase ‘jungle animals’ includes words that are semantically closer than the words in the phrase ‘younger animals’, thus making (6b) a better option for learners across all L1 groups regardless of the correct answer. The reason might also be attributable to the easiness of discerning the non-palatalized forms as in earlier examples.

An additional alternative explanation might be the L1 background of the students bringing down the correct mean score of this specific question because out of 30 Mandarin Chinese and Japanese students only 1 student provided the correct answer for the question when the answer was the target form (i.e., 6a). Since the contrast [r] vs. [l] has been shown to be especially problematic for Chinese and Japanese speakers for various reasons related to their L1 phonology (Nilsen & Nilsen, 2010, p. 109), the mean accuracy scores may have also been affected accordingly.

In fact, this brings about the discussion of the effect of L1 background on the results. The effects of L1 background on the results have been analyzed by grouping learners according to various criteria including country of origin, L1 background, Asian vs. non-Asian languages, Arabic vs. non-Arabic languages, and Arabic vs. Asian vs. all other languages and similar other grouping using a variety of statistical methods, but no significance indicating an effect of L1 background was reached. A further analysis was also conducted to see if the existence of some type of palatalization in the L1 would predict any of the findings. It might have been that learners who had palatalization in their L1 might have some type of metalinguistic awareness about the palatal sounds or processes. Using the data presented in Bateman's work (2007), languages with palatalization have been identified and then various statistical methods have been used to see if students from an L1 background in which palatalization of some type prevails would score better on the perception test. However, no evidence in support of this was found.

Finally, as the findings of the study revealed, the proficiency level of students did not seem to be a predictor of performance scores. As mentioned earlier, the proficiency level variable was analyzed in two different ways: one, using the original placement levels reported in this study, which was based on weighted scores on the speaking and listening tests, and two, using the average of the scores on the end-of-term and the beginning-of-term listening tests. However, the analyses revealed that proficiency level was not a variable that explained any of the scores on the perception test. A possible reason might be the low number of the remaining students included in the final analyses in each proficiency level. Although there were initially 8 to 15 students in each section of each proficiency level, a total of 28 students were eliminated, causing the number of students in each classroom to drop. This might have resulted in loss of enough statistical power to reveal proficiency level as a predictor to explain the findings.

To summarize, the examples presented above in Table 6.1 highlight various aspects regarding the types of challenges learners experienced in decoding palatalization in the target forms. It appears that the most important factor was learners' inefficient word segmentation skills. Instead of trying to segment the words and phrases, learners were relying on cues to eliminate the wrong answer options. These cues include lexical stress, vowel quality or consonant features. When the /l/ vs. /r/ distinction served as a cue to find the correct response, learners of certain L1s seemed to have scored lower due phonemes which do not map well onto each other between English and their L1 (Guion, Flege, Akahene-Yamada & Pruitt, 2000). However, in the absence of any strong cues, learners needed to decode the spoken words. When learners were unable to decode the palatalized form in a sentence, all they did was to guess, as revealed by the type of questions in which both options had some type of palatalization (e.g., ‘walks your dog’ and ‘walked your dog’).

Overall, results indicate that learners were able to improve their ability to perceive palatalization as a connected speech process when they received online training on it. It was only when two different types of palatalization were to be recognized that their performance scores did not show as much improvement. It may be concluded that learners may improve their connected speech perception if the right amount of input and instruction is provided. However, applying what they learn to novel contexts by recognizing those forms in spoken texts may require more time and practice.

6.3 DISCUSSION OF RESEARCH QUESTION 2

RQ2: What is the relationship between students’ performance scores on the perception test (pre-test, post-test, and gain scores) and their attention control (AC) as measured by (a) Attention Network Test (ANT) and (b) Speech-Based Attention Switching Task?

It was predicted that there would be a relationship between perception test scores of learners and their attention control abilities because connected speech forms lack perceptual saliency, so more attention is required to be able to perceive them (Henrichsen 1984, p. 106).

The Pearson’s correlation analyses showed that learners’ post-test and gain scores were negatively correlated with the shift cost measurements calculated by the Speech-Based Attention Switching Task. Findings of the previous studies looking at the relationship between AC and segmental discrimination abilities are not consistent. While Darcy and Mora (in press) and Mora and Darcy (in press) failed to find a significant relationship between the two, Darcy, Mora and Daidone (2014) and Safronova and Mora (2012a) found significant relationship between AC and phonological skills. The findings of these specific studies are crucial as they used the original version of the Speech-Based Attention Switching Task used in this study. The findings of the present study are in line with those studies which found a correlation between the ability to shift between different dimensions and the ability to discriminate certain segments in L2. Two things the present study and previous studies cited above differed were that (a) the present study had an instructional component to it, and that (b) it investigated connected speech perception rather than segmental acquisition in L2 phonology. Therefore, the presence of a significant negative correlation may also be interpreted as an indication that higher attention control mechanisms may be related to higher instructional gains, which is also evidenced by previous research (Schmidt, 1990, 1995; Trofimovich & Gatbonton, 2006).

The second test used to measure attention control in this study was the Attention Network Test. As revealed by the findings of the correlation analyses, there was a negative relationship between the post-test, gain scores and the conflict effect, but not the orienting effect. The results may be interpreted in various ways.

First, the ANT was relatively complicated for the learners to understand as observed in the question-answer sessions taking place prior to data collection as well the pilot testing. Especially the lower proficiency level learners in the study had difficulty understanding the instructions of the study despite the detailed oral and written explanations provided. This is revealed by the findings in two different ways. First, the error rates for the incongruent trials for most of the students in lower levels were very high. This means when the central arrow did not agree with the rest of the arrows, students tended to make significantly more mistakes in determining the way the central arrow was pointing. A closer look at the error rates revealed a potential underlying factor: some learners in the lower levels were clearly unaware that they were supposed to check the direction of the center arrow, to which they were not paying careful attention. Rather, respondents were checking which way the majority of the arrows were pointing. Despite this, the results showed a significant relationship between the conflict effect ($RT_{congruent} - RT_{incongruent}$) probably because this issue was mostly limited to the performance of the lower proficiency level students.

On the other hand, the analyses of various correlation methods did not prove any relationship between the orienting effect and any of the perception test scores. This may again be explained by the task complexity issues. As mentioned earlier, the orienting effect is calculated by using the RT and accuracy measures on the spatial cue and the central cue. One possible problem with this may be that ESL learners might not have been aware that the asterisks presented above and below the fixation cross cued the appearance of the upcoming stimuli. Even if they realized this after a while on their own, this might have affected the final numbers, which eventually affects the spatial cue mean RT used in the calculation of the orienting effect. This may be the reason why no significant level of correlation was reached between the orienting effect and the perception test scores.

The overall findings of the study are promising in that they are in support of previous research suggesting a relationship between attention and L2 phonological development (Schmidt, 1990, 1995; Trofimovich & Gatbonton, 2006). The present study confirms these findings in two additional ways by using two different attention control tasks, both of which revealed the role of attention control as it related to L2 connected speech learning.

6.4 DISCUSSION OF RESEARCH QUESTION 3

RQ3: What is the relationship between attention control scores as measured by an online Attention Network Test (ANT) and a Speech-Based Attention Switching Task?

It was predicted that there would be a correlation between the AC scores measured by the Speech-Based Attention Switching Task and the ANT as the attention measures both of these tasks use are based on the calculation of switching costs.

The findings of the correlation analysis showed a significant relationship between the shift-cost measures and the conflict effect of the two tasks while no such relationship was present between shift-cost and orienting. The findings are very important in showing how the speech-based attention task results overlap with the results of the ANT to a large extent. This is very promising for future studies which wish to assess the role of attention control in L2 phonological acquisition studies using a speech-based task.

Although the ways the two tasks were designed seemed very different from each other, the fact that the shift-cost and the conflict effect of the speech-based attention task and the ANT, respectively, were found to be separately correlated with the post-test and the gain scores as summarized in the previous section indicated a potential relationship between the two AC tasks, which were later confirmed by separate correlation analysis. This in fact may not be surprising when we consider how the shift-cost and the conflict effect are calculated. The shift cost is calculated by subtracting the mean reaction time on no-shift trials from the mean RT on shift trials. In other words, the calculation is done to measure an individual's ability to switch from one dimension to another one. Similarly, the conflict effect of the ANT is calculated by subtracting the mean RT on congruent trials from the mean RT on incongruent trials. In each of these flanker conditions, like in the previous AC task, the respondents are being measured for how long it takes for them to respond to congruent as opposed to incongruent trials. It is normally expected to take them shorter to respond to the congruent trials, which is true for the no-shift condition in the speech-based AC task. As is seen, despite the difference in design and method, there are similarities in the way they measure attention control. It may be due to the fact that both tasks can be considered switching tasks, calculating attention control based on shift-cost calculations. Such tasks have been claimed to be "good candidates for the core measures of executive control" which is a term used to refer to a variety of cognitive abilities such as inhibition or monitoring (see Wager, Jonides & Smith, 2006). The reason why the orienting effect was not correlated with the shift-cost measures of the speech-based task may be, then, attributable to the fact that it did not really measure a component of executive control functions.

Overall, the findings are very promising in showing a relationship between the novel speech-based task and the ANT in that it proves that L2 phonological development and attention control can be examined using new types of language-based AC tasks.

6.5 LIMITATIONS AND FURTHER RESEARCH

With respect to the current research, the first thing to be kept in mind is that the variety of L1 backgrounds in the study might have impacted the results. The way an L1 phonology maps onto the features of the L2 phonology directly affects learners' interlanguage and how much learners benefit from instruction. It may have been that a learner from an L1 background in which all the target palato-alveolar sounds exist will inevitably be more likely to improve more and faster in perception of palatalization given this metalinguistic advantage. L1 syllable structure and the presence of other types of connected speech phenomena could also be potential sources of L1 transfer in learning word-boundary palatalization in English. Therefore, further studies of this type could explore similar aspects of connected speech using learners with the same L1 background and manipulating relevant L1 phonological features.

Another limitation of this research is the high number of students who dropped out by the end of the data collection, and thus were excluded from the final analyses. This is commonly observed in classroom research involving multiple stages of data collection, and one way to avoid this could be having higher number of students in each group or motivating learners by providing them with some kind of incentives to take part in the study.

Additionally, while the researcher attempted to minimize the use of cues in determining the correct option in the perception test, the results indicated that there were multiple cues such as stress assignment, semantic associations or additional consonants which made the correct response option too easily discernable for learners. Therefore, further studies may wish to improve the response options by minimizing or ideally eliminating all the cues which are cueing the response options rather than assessing learners' decoding skills. An alternative way could be including a dictation test or a cloze-test in the study design to compare its results to the multiple-choice test.

The proficiency level variable in this study did not seem to predict the gains in connected speech learning; however, as mentioned earlier, this might be due to the final number of students used in the analyses. Therefore, further studies may use more participants in each of the proficiency levels as it may reveal interesting findings as to whether there is a difference in the way learners in different proficiency levels benefit from instruction on connected speech.

Furthermore, unlike the experimental group, the control group in this study did not receive any instruction nor were they exposed to the target connected speech forms. Future replications of the study might consider exposing the control group to target connected speech forms without providing any explicit instruction on these forms. That way, it might be possible to investigate the effects of instruction versus mere exposure to target forms. Then, stronger claims regarding the sources of improvement can be made.

Although the way the ANT task was designed and presented to the students was intended to minimize any confusions regarding the expectations of the task, the accuracy scores, in particular, evidenced that this was not fully achieved. Testing learners individually rather than as a whole classroom by also adding the double cue condition to the task design to calculate the alerting network might be considered for further studies. Assessing learners' AC may yield healthier results if learners are tested individually without distractions of the classroom environment, making sure that each individual knows what they are expected to do to complete the AC task.

Finally, further studies may investigate different types of connected speech processes in relation to other cognitive abilities such as short-term working memory or inhibition. This could provide a better picture of the role of cognitive abilities in L2 phonological learning and help researchers understand the factors underlying L2 phonological acquisition.

6.6 IMPLICATIONS OF THE STUDY FOR L2 PEDAGOGY

A major finding of the present study shows that online training might be an effective way to present content to students when there is need to assure that the same type of input is presented to all students. However, the role of the embedded questions in keeping the students alert during video viewing should not be underestimated. The use of online tools for embedding questions in videos might help teachers realize that watching videos in classrooms does not have to be a passive activity. In teaching L2 phonology, videos might be extremely helpful as they may help learners visualize, for instance, the pronunciation of sounds in a certain L2.

The present study also shows that it is possible to teach connected speech processes to learners. More specifically, the fact that learners' perception of word-boundary palatalization improved by the end of a three-week instructional period is very encouraging in teaching and learning of connected speech forms. The take-away from this finding for textbook writers, curriculum designers and ESL teachers is that connected speech forms are in fact 'teachable' and must be a part of learners' training in L2 English phonology. As Brown (2012) points out, without teaching learners how to put together the bricks (i.e., individual sounds) into a building (i.e., connected speech), knowledge of sounds alone is only helpful in a limited way. Teaching the way the sounds change in spoken language, on the other hand, helps them learn the language they need outside of the classroom, as well as for high-stakes testing situations, as revealed by previous research (see Kostin, 2004).

The findings also suggest that attention is linked to L2 perceptual development in L2 connected speech. Although attention is a cognitive ability of an individual, there may be things that teachers can do to help learners pay more attention to the content presented in class to maximize learning. One way could be to be employ a variety of teaching methods in teaching L2 phonology. Instead of using, say, a communicative approach to teaching a certain phonological aspect to students, integrating other methods such as form-focused instruction in teaching may enhance the learning process. This is because FFI enables learners to focus on the rules or forms by drawing their attention to a specific aspect of the form. This is especially helpful if the target form is a less salient form, like connected speech processes, in which learners need extra guidance in deciphering the language. This is also related to the use of top-down or bottom-up approaches in teaching listening skills. Depending on the purposes of the lesson, instead of preferring activities which favor top-down processing, techniques requiring bottom-up processing, such as dictation tests, should be a part of the classroom learning.

6.7 CONCLUSIONS

The present study set out to explore the effects of online training on L2 phonological development, more specifically L2 connected speech perception on a pre- and post-test basis. In addition, it also sought to examine the relationship between attention control and L2 phonological development. The findings showed that learners in the experimental group improved their performance scores on the perception test more than the learners in the control group. The findings also revealed that more efficient attention control was related to L2 phonological learning. Although no causal relationship can be claimed based on the analyses, this relationship may indicate that learners with better attention control were able to better focus on the distinct properties of the L2 phonology, which in turn might have yielded higher instructional gains. Finally, the positive correlation between the two different types of AC tasks may indicate that the novel speech-based task might be a promising tool to measure AC in language-based studies.

These findings shed new light on our understanding of L2 phonological learning as it relates to attention control. It has been shown that teaching and learning connected speech via online training is possible, so curriculum developers, textbook writers and teachers should consider inclusion of various connected speech forms in classroom teaching as a way to help L2 learners of English improve their listening and speaking abilities. Additionally attention control has been found to be related to performance scores on the perception test of connected speech forms. Therefore, it can be claimed that some degree of attention is needed for connected speech learning to take place. Finally, the possibility of measuring attention control using a speech-based or phonology-based task is very promising for L2 psycholinguistics, and may serve as a model for further research.

REFERENCES

Alameen, G. (2014). *The effectiveness of linking instruction on NNS speech perception and production* (Doctoral dissertation). Iowa State University.

Alameen, G., & Levis, J. M. (2015). Connected Speech. In M. Reed & J. Levis (Eds.), *The Handbook of English Pronunciation*, (pp. 159-174). Wiley Blackwell.

Al-jasser, F. (2008). The effect of teaching English phonotactics on the lexical segmentation of English as a foreign language. *System*, *36*(1), 94-106.

Assmann, P. F., & Summerfield, Q. (1994). The contribution of waveform interactions to the perception of concurrent vowels. *Journal of the Acoustical Society of America*, *95*, 471–484.

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

Bateman, N. (2007). *A crosslinguistic investigation of palatalization* (Doctoral dissertation). University of California, San Diego.

Benson, P. (2007). Autonomy in language teaching and learning. *Language Teaching*, *40* (01), 21-40.

Berk, R. A. (2009). Multimedia teaching with video clips: TV, movies, YouTube, and mtvU in the college classroom. *International Journal of Technology in Teaching and Learning*, *5*(1), 1–21.

Bhat, D. N. (1978). A general study of palatalization. In J. S. Greenberg, (Ed.), *Universals of human language*, *Volume 2: Phonology*, (pp. 47-92). Stanford UP, Bloomington, CA.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition*, *16*(02), 157-168.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science*, *7*(3), 325-339.

Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and aging*, *19*(2), 290.

Brown, G. (1977). *Listening to spoken English*. London: Longman

Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall.

Brown, J. D. (2006). Authentic communication: Whyzit importan'ta teach reduced forms. In *Proceedings of the 5th Annual JALT Pan-SIG Conference*, *13* (Vol. 24).

Brown, J. D. (2012). *New ways in teaching connected speech*. TESOL International Association Press. Alexandria, VA.

Brown, J. D., & Hilferty, A. (1986). The effectiveness of teaching reduced forms of listening comprehension. *RELC Journal*, *17*(2), 59-70.

Brown, J. D., & Hilferty, A. (2006). The effectiveness of teaching reduced forms of listening comprehension. In J.D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers*, (pp. 51-58).

(Reprinted from The effectiveness of teaching reduced forms of listening comprehension. *RELC Journal*, *17*(2), 59-70.)

Brown, J. D., & Kondo-Brown, K. (2006). Introducing connected speech. In J.D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers*, (pp. 1-15). National Foreign Language Resource Center, University of Hawai'i at Mānoa.

Buck, G. (2001). *Assessing listening*. Cambridge University Press. Cambridge, UK.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in the anterior cingulate cortex. *Trends in Cognitive Science, 4(6)*, 215–222.

Campbell, D. F., McDonnell, C., Meinardi, M., & Richardson, B. (2007). The need for a speech corpus. *ReCALL*, *19*(01), 3-20.

Carlson, R. A., & Dulany, D. E. (1985). Conscious attention and abstraction in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 45.

Carr, T. H., & Curran, T. (1994). Cognitive factors in learning about structured sequences. *Studies in Second Language Acquisition*, *16*(02), 205-230.

Cavar, M.E. (2004). *Palatalization in Polish* (Doctoral dissertation). University of Potsdam, Germany.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching Pronunciation: A Course Book and Reference Guide*. 2nd edition. Cambridge. CUP.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Clarey, M. E., & Dixson, R. J. (1963). *Pronunciation exercises in English*. Prentice Hall.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Newbury Park, CA: Sage.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

College Board. (2015). *ACCUPLACER Program Manual*. New York: College Board.

Cook, Ann. (1991). *American Accent Training: A Guide to Speaking and Pronouncing Colloquial American English for Everyone Who Speaks English as a Second Language.* NY: Barron's Educational Series.

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*(1), 59-86.

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 474–479.

Crawford, M. (2005). Bottom-up skills and L2 listening pedagogy: A study on the teaching of reduced forms in English. *Journal of the English Literary Society of Hakodate*, *44*, 23-40.

Cronbach, L. J. (1953). Correlation between persons as a research tool. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and Research*, (pp. 476-388). Oxford, England: Ronald Press.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671-684.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.

Cross, J. (2009). Diagnosing the process, text, and intrusion problems responsible for L2 listeners' decoding errors. *Asian EFL Journal*, *11*(2), 31-53.

Darcy, I., Mora, J. C., and Daidone, D. (2014). Attention control and inhibition influence phonological development in a second language. *Concordia Working Papers in Applied Linguistics, Volume 5,* March 2014, 115-129. Concordia University, Montreal, Canada.

Darcy, I. & Mora, J. C. (in press). Executive control and phonological processing in Spanish learners of English: the effect of learning in a monolingual or a bilingual context. In G. Granena, D.O., Jackson & Y. Yilmaz (Eds.), *Cognitive Individual Differences in L2 Processing and Acquisition.* John Benjamins (Bilingual Processing and Acquisition series).

Darcy, I., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences, 40*, 63-72.

Davies, M. (2008-) *The Corpus of Contemporary American English: 520 million words, 1990-present*. Available at http://corpus.byu.edu/coca/.

Davies, M., & Gardner, D. (2010). *A frequency dictionary of contemporary American English: Word sketches, collocates and thematic lists*. Routledge.

DeSchepper, B., & Treisman, A. (1996). Visual memory for novel shapes: Implicit coding without attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 27-47

Edel-Malizia, S., Brautigam, B., Bittner, K., & Blackstock, D. (2015). Investigating Interactive Video Assessment Tools for the Flipped Classroom. Paper selected for presentation at the *International Conference on E-Learning 201*5. Retrieved from https://www.researchgate.net/publication/281492488_Investigating_Interactive_Video_Assessment_Tools_for_the_Flipped_Classroom

EDpuzzle [Web-based video assessment tool]. (2015). Accessible at https://edpuzzle.com/

Edwards, A. L., & Cronbach, L. J. (1952). Experimental design for research in psychotherapy. *Journal of Clinical Psychology, 8*(1), 51-59.

Ellis, N. C. (1994). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. *AILA Review, 11,* 37-57.

Ellis, R. (Ed.). (2001). *Form-focused instruction and second language learning: Language learning monograph*. Vol. 51. Wiley-Blackwell.

Eviatar, Z. (1998). Attention as a psychological entity and its effects on language and communication. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of neurolinguistics* (pp. 275–287). New York: Academic Press.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*(3), 340-347.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Field, J. (2003). Promoting perception: lexical segmentation in L2 listening. *ELT journal*, *57*(4), 325-334.

Field, J. (2008a). *Listening in the language classroom*. Cambridge: CUP.

Field, J. (2008b). Guest editor's introduction emergent and divergent: A view of second language listening research. *System*, *36*(1), 2-9.

Field, J. (2008c). Revising segmentation hypotheses in first and second language listening. *System, 36,* 35-51.

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics, 62,* 1668-1680.

Gilbert, J. (1995). Pronunciation practice as an aid to listening comprehension. In D. Mendelsohn & J. Rubin (Eds.), *A guide for the teaching of second language listening,* (pp. 97-111). San Diego, CA: Dominie Press.

Gimson, A.C. (1989). *An Introduction to the Pronunciation of English*. London: Edward Arnold.

Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, *28*(1), 55-75.

Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, *25*, 1–42.

Grant, L. (2001). *Well said: Pronunciation for clear communication*. 3rd edition. Heinle & Heinle Publishers.

Guion, S. G. & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O. S. Bohn & M. Munro (Eds.), *Language Experience in Second Language Speech Learning*, (pp. 57-77). Amsterdam: John Benjamins.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, *107*(5), 2711-2724.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *The Quarterly Journal of Experimental Psychology*, *56*(7), 1213-1236.

Gussenhoven, C., & Jacobs, H. (1998). *Understanding phonology*, *2nd ed.* Routledge.

Hagen, S. A. (2000). *Sound advice: A basis for listening*. New York: Pearson Education.

Henrichsen, L. E. (1984). Sandhi-variation: A filter of input for learners of ESL. *Language Learning*, *34*(3), 103-123.

Hieke, A. E. (1984). Linking as a marker of fluent speech. *Language and Speech, 27*(4), 343-354.

Hieke, A. E. (1987). Absorption and fluency in native and nonnative casual speech in English. In A. James and J. Leather (Eds.), *Sound Patterns in Second Language Acquisition* (pp. 41-58). Dordrecht, the Netherlands, Foris Publications.

Houston, C. (2000). Video usage and active learning strategies among community college faculty members. *Community College Journal of Research and Practice, 24,* 341-357.

Huck, S. W. (2012). *Reading Statistics and Research* (6th edition). Pearson Education Publishing: Boston, MA.

Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Canada Journal*, *27*(1), 1-12.

Ito, Y. (2001). Effect of reduced forms on ESL learners' input-intake process. *Second Language Studies*, *20*(1), 99-124.

Ito, Y. (2006a). The significance of reduced forms in L2 pedagogy. In J.D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers*, (pp. 17-25). National Foreign Language Resource Center, University of Hawai'i at Mānoa.

Ito, Y. (2006b). The comprehension of English Reduced Forms by Second Language Learners and Its Effect on Input-Intake Process. In J.D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers*, (pp. 67-81). National Foreign Language Resource Center, University of Hawai'i at Mānoa.

Jenkins, J. (2000). *The phonology of English as an international language: New models, new norms, new goals*. Oxford UP.

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

Joyce, P. (2013). The Development and Validation of a Second Language Listening Reduced Forms Test. *Iranian Journal of Language Testing*, *3*(2).

Keating, P. (1991). Coronal places of articulation. In C. Paradis & J.F. Prunet, (Eds.), *Phonetics and Phonology*, *The special status of coronals: Internal and external evidence 2* (pp. 79-100). San Diego: Academic Press.

Kim, Hae-Young. (1995). Intake from the speech stream: Speech elements that L2 learners attend to. In Schmidt, R (Ed.) *Attention and awareness in foreign language learning*, (pp. 65-84). National Foreign Language Resource Center, University of Hawai'i at Mānoa.

Kochetov, Alexei. (2011). Palatalization. M. van Oostendorp, C. Ewen, B. Hume, & K. Rice. (Eds.) *The Blackwell Companion to Phonology, Phonological Interfaces, Volume 4*, (pp. 1666-1690). Oxford: Wiley Blackwell.

Kostin, I. (2004). Exploring item characteristics that are related to the difficulty of TOEFL dialogue items. *Research Report Educational Testing Service.* Princeton, NJ.

Kowollik, V. (2009*). The interaction between general cognitive ability and training structure: Using meta-analysis to investigate aptitude-treatment interactions in the training literature (*Doctoral dissertation). University of Oklahoma, OK.

Kuo, F. L. (2012). Factors Affecting Chinese EFL Learners' Spoken Word Recognition. *NCUE Journal of Humanities*, *Vol. 6*, 1-14.

Kreidler, C. W. (2008). *The pronunciation of English: A course book*. John Wiley & Sons.

Lahiri, A. & V. Evers. (1991). Palatalization and Coronality. In C. Paradis & J.F. Prunet, (Eds.), *Phonetics and Phonology*, *The special status of coronals: Internal and external evidence 2* (pp. 79-100). San Diego: Academic Press.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS.* Routledge: New York.

Little, D. (1996). Freedom to learn and compulsion to interact: promoting learner autonomy through the use of information systems and information technologies. In R. Pemberton, E. S. Li, W.W. Or & H. D. Pierson (Eds.) *Taking control: Autonomy in language learning*, Vol 1 (pp. 203-218). Hong Kong UP.

MacKay, I. R., Meador, D., & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica*, *58*(1-2), 103-125.

Mackey, A., & Gass, S. M. (2013). *Second language research: Methodology and design*. Routledge.

Massaro, D. W., & Chen, T. H. (2008). The motor theory of speech perception revisited. *Psychonomic bulletin & review*, *15*(2), 453-457.

Matsuzawa, T. (2006). Comprehension of English reduced forms by Japanese business people and the effectiveness of instruction. In J.D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers*, (pp. 59-66). National Foreign Language Resource Center, University of Hawai'i at Mānoa.

Matlock-Hetzel, S. (1997, January). Basic concepts in item and test analysis. Paper presented at *the Annual Meeting of the Southwest Educational Research Association*, Austin, TX. Retrieved from http://files.eric.ed.gov/fulltext/ED406441.pdf

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.

Milner B. (1963) Effects of different brain lesions on card sorting. *Archives of Neurology*, *9*,100–10.

Mora, J.C. & Darcy, I. (2013). Inhibition and Phonological Processing in a Second Language. Paper presented at *the EUROSLA 2013 Conference*, Amsterdam, the Netherlands.

Mora, J. C. & Darcy, I. (in press). The relationship between cognitive control and pronunciation in a second language. In Isaacs, T. & Trofimovich, P. (ed.) *Interfaces in Second Language Pronunciation Assessment: Interdisciplinary Perspectives.* Multilingual Matters

Mora, J.C. & Gilabert (2012). Individual factors in utterance and perceived fluency: some empirical issues. *Workshop Fluent Speech: Combining cognitive and educational approaches*. University of Utrecht. Utrecht, the Netherlands.

Morley, J. (1994). *Pronunciation Pedagogy and Theory: New Views, New Directions*. TESOL. Alexandria, VA.

Nilsen, D. L., & Nilsen, A. P. (2010). *Pronunciation contrasts in English*. Waveland Press.

Norris, R. W. (1993). Teaching reduced forms: An aid for improving lower-level students' listening skills. *Fukoka Women's Junior College Studies, 46*, 49-56.

Norris, R. W. (1995). Teaching reduced forms: Putting the horse before the cart. In *English Teaching Forum, 33*(3), 47-50.

O'brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, *29*(04), 557-581.

Perception Research Systems. (2007). *Paradigm Stimulus Presentation*. Retrieved from http://www.paradigmexperiments.com

Posner, M. I. (1992). Attention as a cognitive and neural system. *Current Directions in Psychological Science, 7*, 11-14.

Posner, M. I. (1994). Attention: the mechanisms of consciousness. *Proceedings of the National Academy of Sciences*, *91*(16), 7398-7403.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Reviews in Neuroscience, 13*, 25-42.

PowToon [Web-based animation software]. (2015). Accessible at https://www.powtoon.com/

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology*, *55A*, 1339-1362.

Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL quarterly*, *17*(2), 219-240.

Richards, J. C. & Rogers, T. S. (2001). *Approaches and Methods in Language Teaching.* 2nd ed. New York: Cambridge UP.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning*, *45*, 283-331.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of experimental psychology: General*, *124*(2), 207.

Rogerson, M. (2006). Don'cha know? A survey of ESL teachers' perspectives on reduced forms instruction. In J.D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers*, (pp. 85-97). National Foreign Language Resource Center, University of Hawai'i at Mānoa.

Rosa, M. (2002). Don't cha know? A survey of ESL teachers' perspectives on reduced forms instruction. *Second Language Studies*, *21*(1), 49-78.

Rost, M. (2002). Teaching and researching: listening. London, Longman.

Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, *78*(2), 199-221.

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*(8), 1029-1040.

Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2011). Attentional control and self-regulation. In K.D. Vohs & R.F. Baumeister (Eds.). *Handbook of self-regulation: Research, theory, and applications*, *2nd edition*, (pp. 284-299). Guilford Press.

Safronova, E. & Mora, J. C. (2011) Acoustic and phonological memory in L2 vowel perception. *35th AEDEAN Conference*. Barcelona, Spain.

Safronova, E. & Mora, J.C. (2012a) Attention control in L2 phonological acquisition. *30th AESLA Applied Linguistics Conference.* University of Lleida, Spain.

Safronova, E. & Mora, J. C. (2012b) Acoustic memory, phonological memory and attention control in L2 speech perception. *AAAL 2012 Annual Conference*. Boston, MA.

Safronova, E. (2013) Attention-shifting skill and L2 vowel perception. *New Sounds 2013 (International Symposium on the Acquisition of Second Language Speech).* Concordia University, Montreal, Canada.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied linguistics*, *11*(2), 129-158.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.) *Attention and* *awareness in foreign language learning*, (pp. 1-63). National Foreign Language Resource Center, University of Hawai'i at Mānoa.

Schmidt, R. (2001). "Attention." In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Taylor & Francis.

Segalowitz, N. and Frenkiel-Fishman, S. (2005) Attention control and ability level in a complex cognitive skill: attention-shifting and second language proficiency. *Memory and Cognition*, 33, 644-653.

Seidlhofer, B. (2005). English as a lingua franca. *ELT journal*, *59*(4), 339.

Sherman, J. (2003).*Using authentic video in the language classroom.* Cambridge: Cambridge University Press.

Shockey, L. (2003). *Sound patterns of spoken English*. John Wiley & Sons.

Schwartz, B. D. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition*, 15, 147- 163.

Snow, R. E. (1989). Aptitude treatment interaction as a framework for research on individual differences in learning. In P. L. Ackerman, R. J. Sternberg, & R. G. Glasser (Eds.), *Learning and individual differences* (pp. 13-59). New York: Freedman.

Snow, R. E. (1991). Aptitude-treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology*, *59*(2), 205.

Snow, R. E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational Psychologist*, *27*(1), 5-32.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading research quarterly*, 32-71.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, *110* (16), 6313-6317.

Talmy, L. (1996). The windowing of attention. In M. Shibatani & S. A. Thompson (Eds.), *Grammatical constructions* (pp. 235–288). Oxford: Oxford University Press

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*(02), 183-203.

Trager, E. C. (1983). *PD's in Depth*. Prentice Hall.

Traxler, M. J. (2012). *Introduction to psycholinguistics: Understanding language science*. John Wiley & Sons.

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *The Modern Language Journal*, *90*(4), 519-535.

Tsui, A. B., & Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied linguistics*, *19*(4), 432-451.

Underwood, P. R., & Wallace, W. (2012). The effects of instruction in reduced forms on the performance of low-proficiency EFL university students. *The Asian EFL Journal*, *14*(4), 1-24.

Ur, P. (1987) *Teaching Listening Comprehension*. Cambridge, England. Cambridge University Press.

Vandergrift, L., & Goh, C. C. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.

VanPatten, B. (1994). Evaluating the role of consciousness in second language acquisition: Terms, linguistic features & research methodology. *AILA Review*, *11*, 27-36.

Wager, T. D., Jonides, J., & Smith, E. E. (2006). Individual differences in multiple types of shifting attention. *Memory & cognition*, *34*(8), 1730-1743.

Weaver, B., Bédard, M., McAuliffe, J., & Parkkari, M. (2009). Using the Attention Network Test to predict driving test scores. *Accident Analysis & Prevention*, *41*(1), 76-83.

Weaver, B., Bédard, M., & McAuliffe, J. (2013). Evaluation of a 10-minute Version of the Attention Network Test. *The Clinical Neuropsychologist*, *27*(8), 1281-1299.

Weinstein, N. J. (1982). *Whaddaya say?* Culver City, CA: English Language Services

Weinstein, N. J. (2001). *Whaddaya say? Guided practice in relaxed speech* (2nd ed.) London: Longman.

Wilson, M. (2003). Discovery listening—improving perceptual processing. *ELT journal*, *57*(4), 335-343.

Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: how frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 255.

Woodrow, L. (2014). *Writing about Quantitative Research in Applied Linguistics*. Palgrave Macmillan. UK.

Yang, J. C., Lin, Y. L., & Chung, C. I. (2009, August). Developing a Film-based Learning System with English Verbal Reduced Forms for Supporting English Listening Comprehension. In *Proceedings of the 4th International Conference on E-Learning and Games (Edutainment2009)*. Banff, Canada. *Lecture Notes in Computer Science: Learning by Playing, 5670*, 74-84. (EI)

Yang, H., Yang, S., & Kang, C. (2014). The relationship between phonological awareness and executive attention in Chinese-English bilingual children. *Cognitive Development*, *30,* 65-80.

Ye, Z., & Zhou, X. (2009). Executive control in language processing. *Neuroscience & Biobehavioral Reviews, 33*(8), 1168-1177.

Yi'an, W. (1998). What do tests of listening comprehension test?-A retrospection study of EFL test-takers performing a multiple-choice task. *Language testing*, *15*(1), 21-44.

Zipt, G. K. (1949). Human behavior and the principle of least effort. *Addison Wealey*.

Zsiga, E. C. (1993). *Features, gestures, and the temporal aspects of phonological organization* (Doctoral dissertation).Yale University.

Zsiga, E. C. (1994). Acoustic evidence for gestural overlap in consonant sequences, *Journal of Phonetics*, 22, 121-140.

Zsiga, E. C. (1995). An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English. In B. Connell & A. Arvaniti (Eds.)

Phonology and phonetic evidence, papers in laboratory phonology (pp. 282-302). Cambridge: Cambridge University Press.

Zsiga, E. C. (2000). Phonetic alignment constraints: consonant overlap and palatalization in English and Russian. *Journal of phonetics*, *28*(1), 69-102.

APPENDIX A

INFORMED CONSENT FORM

ABOUT THIS STUDY

TITLE OF THE STUDY: Attention Control and the Effects of Online Training in Improving Connected Speech Perception by Learners of English as a Second Language

INVESTIGATOR : Burcu Gökgöz-Kurt, Ph.D. Student, Linguistics Program, U. of South Carolina

- **What is this study about?**

This is a study about teaching and learning connected speech and the effects of individual differences on learning related forms. You were selected as a possible participant in this study because you have indicated that you are an ESL learner enrolled in a Speaking & Listening Class in the program. Your participation is completely voluntary and you are free to choose not to participate without any negative consequences or stop participating at any time.

- **What am I expected to do if I agree to participate?**

You will do the following: (a) take a pre-test, a post-test, and a two-part test measuring your attention control over three weeks. For the pre- and post-tests, for each question, you will listen to sentences on iPads through headphones, and then will be asked to choose one of the two sentences you listened to. Each test should take about 20-25 minutes, (b) may participate in connected speech perception training which will take 15-20 minutes during three weeks in your speaking & listening classes. In addition to these, you will also be asked to answer questions about your age, nationality and language learning background. Please note that you do not have to answer all questions of the questionnaire and tests in order to participate.

- **What are the benefits of participating in this study?**

There are no direct benefits or risks of participating in the study; however, you may receive an online training on the connected speech perception in ESL English, and this may help you better understand the use of English especially by native speakers. The major benefit of this study is to further our understanding of the effects of instruction on connected speech perception and how it relates to attention.

- **Should I be worried about the confidentiality of records?**

All information gathered will remain confidential. All materials will be entered and saved on researcher's personal computer and protected by a password. If the results are published, your identity will remain confidential.

- **What if I have a question?**

Taking part in the study is your decision. You may also quit participating in the study at any time or decide not to answer any question you are not comfortable answering. Participation, non-participation or withdrawal will not affect your grade in any way. We will be happy to answer any questions you have about the study. You may contact me, Burcu Gokgoz Kurt, at (gokgozku@email.sc.edu) or my faculty advisor, Dr. Eric Holt at (holt@sc.edu) if you have study-related questions or problems. If you have any questions about your rights as a research participant, you may contact the Office of Research Compliance at the University of XXX at XXX-XX-XX. Prior to the study, necessary credentials have also been obtained from to the Institutional Review Board (IRB) with the approval number 00044442.

- **Signature Text**

I have read the contents of this consent form and have been encouraged to ask questions. I have received answers to my questions. I give my consent to participate in this study, although I have been told that I may withdraw at any time without negative consequences. I have read and understand that this is a voluntary task; it will NOT affect my grade at EPI and all the information below will be kept confidential (safe).

☐ **CHECK √ THIS BOX IF YOU UNDERSTAND AND WANT TO PARTICIPATE IN THE STUDY.**

STUDENT'S SIGNATURE: ____________________

DATE: ____________________

APPENDIX B

DEMOGRAPHIC AND LANGUAGE BACKGROUND QUESTIONNAIRE

Today's Date ___________________

1. PLEASE PICK A NICKNAME OR NUMBER THAT YOU WILL REMEMBER LATER ON!!! THEN, WRITE IT IN THE BOX. THIS IS IMPORTANT.

Nickname or Number
____

2. Speaking & Listening Class and Level. Circle one please:

SL2a (Instructor X))	SL2b (Instructor X)	SL3a (Instructor X)
SL3b (Instructor X)	SL4a (Instructor X)	SL4b (Instructor X)
SL5b (Instructor X)	SL5c (Instructor X)	

3. Age: ________

4. Gender: ☐ Male ☐ Female

5. Do you have any hearing problems (as told by a doctor)? ☐ Yes ☐ No

6. Where are you from?

____

7. How old were you when you FIRST started learning English in your home country?

____

8. How long (how many years) did you study English for in your home country?

____

9. How long have you been studying English for in the U.S.A. (in total)?

____

10. Please answer the following question.

Outside of EPI, I spend about _______ minutes/ _______ hour (s) listening to English every day (listening to people, TV, radio, music).

11. Read the following statements and circle YES or NO.

YES	NO	I live/have lived an American roommate.
YES	NO	I live/have lived with an American family for a while.
YES	NO	At home, my roommate and I speak English all the time.
YES	NO	I spend a lot of time watching (American) movies in English in my free time.

APPENDIX C

MULTIPLE FORCED-CHOICE PRE- AND POST- PERCEPTION TEST

Phenomenon 0: Practice items		Explanation of Contrast	Type of Stimuli
P1	What **do you want to** [wʌɾəjə wɑnə] do now? What **are you going to** [wʌɾəjə gʌnə] do now?	Want to / Going to This item is repeated three times with varying correct answers (a, b, and a again) to show students that this repetition is something they should expect. This is to avoid strategic responding.	Practice
P2	Did you talk to **(h)er** this morning? Did you talk to **Earl** this morning?	h-deletion This item is repeated three times with varying correct answers (a, b, and a again) to show students that this repetition is something they should expect.	Practice
P3	She has got **to hurry**. She has got **a worry**.	Juncture This item is repeated three times with varying correct answers (a, b, and a again) to show students that this repetition is something they should expect.	Practice
Phenomenon 1: [t] + [j] => [tʃ] (voiceless)		**Explanation of Contrast**	**Type of Stimuli**
1	They still haven'**t‿used** [tʃuzd] the ones you gave them. They still haven't **juiced** [ʤust] the ones you gave them.	[tʃ] vs. [dʒ] Explanation: a consonant contrast with a final consonant cluster difference	Critical Stimuli

2	They always though**t your** lines [tʃɔr, tʃʊər, tʃər] were unforgettable. They always thought **shore**lines [ʃɔr, ʃoʊr] were unforgettable.	[tʃ] vs. [ʃ] Explanation: consonant contrast with a possible vowel difference	Critical Stimuli
3	So, why can'**t yards** [tʃɑrdz] be very big, again? So, why can't **charts** [tʃɑrts] be very big, again?	[tʃ] and [tʃ] in both cases Explanation: same consonants with a contrast in the final consonant cluster	Critical Stimuli
4	He likes to sing, but no**t yet** [tʃɛt] here at home. He likes to sing, but not **chat** [tʃæt] here at home.	[tʃ] and [tʃ] in both cases Explanation: same consonants yet a contrast in the middle vowel	Critical Stimuli
Phenomenon 2: [d]+]j] => [dʒ] (voiced)		**Explanation of Contrast**	**Type of Stimuli**
5	They all coul**d use** [dʒuz] any of these. They all could **choose** [tʃuz] any of these.	[dʒ] vs. [tʃ] Explanation: a consonant contrast	Critical Stimuli
6	The 40-year ol**d you** [dʒu, dʒʊ, dʒə] would be worn out by now. The 40-year old **shoe** [ʃu] would be worn out by now.	[dʒ] vs. [ʃ] Explanation: a consonant contrast with a possible vowel difference	Critical Stimuli

7	What else shoul**d you** [dʒu, dʒʊ, dʒə] not do for lunch? What else should **June** [dʒʊn] not do for lunch?	[dʒ] and [dʒ] in both cases Explanation: same consonants, but an additional final consonant	Critical Stimuli
8	The store sol**d younger** [dʒʌŋgər] animals at a higher price. The store sold **jungle** [dʒʌŋgəl] animals at a higher price.	[dʒ] and [dʒ] in both cases Explanation: same consonants with a final consonant difference	Critical Stimuli
Phenomenon 3: [s] + [j] => [ʃ] (voiceless)		**Explanation of Contrast**	**Type of Stimuli**
9	It is just in ca**se young** [ʃʌŋg] workers won't be here then. It is just in case **junk** [dʒʌŋk] workers won't be here then.	[ʃ] vs. [dʒ] Explanation: a consonant contrast with a final consonant difference	Critical Stimuli
10	He thought that this **year** [ʃɪər] was a bit different. He thought that this **cheer** [tʃɪər] was a bit different.	[ʃ] vs. [tʃ] Explanation: a consonant contrast	Critical Stimuli
11	It's not fair to **miss your** [ʃɔr, ʃʊər, ʃər] jobs when you are sick. It's not fair to miss **shore** [ʃɔr, ʃoʊr] jobs when you are sick.	[ʃ] and [ʃ] in both cases Explanation: same consonants with a possible vowel contrast	Critical Stimuli

12	They will bless‿**your** [ʃɔr, ʃʊər, ʃər] hands for being helpful. They will bless **sure** [ʃʊr] hands for being helpful.	[ʃ] and [ʃ] in both cases Explanation: same consonants with a vowel difference	Critical Stimuli
Phenomenon 4: [z] + [j] => [ʒ] (voiced)		Explanation of Contrast	**Type of Stimuli**
13	She already know**s‿you** [ʒu, ʒʊ, ʒə] never got here on time. She already knows **June** [dʒun] never got here on time.	[ʒ] vs. [dʒ] Explanation: a consonant contrast with an additional final consonant	Critical Stimuli
14	He always ha**s‿your** [ʒɔr, ʒʊər, ʒər] lists pinned on the board. He always has **chore** [tʃɔr] lists pinned on the board.	[ʒ] vs. [tʃ] Explanation: a consonant contrast with a possible vowel reduction	Critical Stimuli
15	I suppo**se‿yelling** [ʒɛlɪŋ] is very important. I suppose **shelling** [ʃɛlɪŋ] is very important.	[ʒ] vs. [ʃ] Explanation: a consonant contrast	Critical Stimuli
16	She already know**s‿you did** [ʒʊ dɪd] very well. She already knows **Judith** [ʤudɪθ] very well.	[ʒ] vs. [ʃ] Explanation: a consonant contrast, and a final consonant difference (and a prosody difference)	Critical Stimuli (had to replace a faulty item)

Contrasted Palatalized Sounds in Each Pair		Explanation of Contrasts	Type of Stimuli
17	They sure len**t you** [tʃu, tʃʊ, tʃə] a lot of money. They sure len**d you** [dʒu, dʒʊ, dʒə] a lot of money.	[tʃ] vs [dʒ]	Critical Item
18	Then she walke**d your [**tʃɔr, tʃʊər, tʃər] dog in the yard. Then she walk**s your** [ʃɔr, ʃʊər, ʃər] dog in the yard.	[tʃ] vs. [ʃ]	Critical Item
19	She then sen**t your [**tʃɔr, tʃʊər, tʃər] clothes to them. She then send**s your** [ʒɔr, ʒʊər, ʒər] clothes to them.	[tʃ] vs. [ʒ]	Critical Item
20	She sure type**d your [**dʒɔr, dʒʊər, dʒər] letters very fast. She sure type**s your** [ʃɔr, ʃʊər, ʃər] letters very fast.	[dʒ] vs. [ʃ]	Critical Item
21	They all memorize**d your [**dʒɔr, dʒʊər, dʒər] names very quickly. They all memori**ze your** [ʒɔr, ʒʊər, ʒər] names very quickly.	[dʒ] vs. [ʒ]	Critical Item

22	This is the lowest pri**ce you** [ʃu, ʃʊ, ʃə] will get. This is the lowest pri**ze you** [ʒu, ʒʊ, ʒə] will get.	[ʃ] vs. [ʒ]	Critical Item
Control Items **These items were paired with all the target sentences in the first 4 phenomena**		**Explanation of Contrast**	**Type of Stimuli**
23	They still haven'**t‿used** [tʃuzd] the ones you gave them. They still haven't **seen** the ones you gave them.	Control Item for Question 1	Control
24	They always though**t‿your** lines [tʃɔr, tʃʊər, tʃər] were unforgettable. They always thought **their** lines were unforgettable.	Control Item for Question 2	Control
25	So, why can'**t‿yards** [tʃɑrdz] be very big, again? So, why can't **stocks** be very big, again?	Control Item for Question 3	Control
26	He likes to sing, but no**t‿yet** [tʃɛt] here at home. He likes to talk, but not **right** here at home.	Control Item for Question 4	Control
27	They all coul**d‿use** [dʒuz] any of these. They all could **pass** any of these.	Control Item for Question 5	Control

28	The 40-year ol**d‿you** [dʒu, dʒʊ, dʒə] would be worn out by now. The 40-year old **shirt** would be worn out by now.	Control Item for Question 6	Control
29	What else shoul**d‿you** [dʒu, dʒʊ, dʒə] not do for lunch? What else should we not do for lunch?	Control Item for Question 7	Control
30	The store sol**d‿younger** [dʒʌŋgər] animals at a higher price. The store sold **larger** animals at a higher price.	Control Item for Question 8	Control
31	It is just in ca**se‿young** [ʃʌŋg] workers won't be here then. It is just in case **those** workers won't be here then.	Control Item for Question 9	Control
32	He thought that thi**s‿year** [ʃɪər] was a bit different. He thought that this **team** was a bit different.	Control Item for Question 10	Control
33	It's not fair to **miss‿your** [ʃɔr, ʃʊər, ʃər] jobs when you are sick. It is not fair to miss **these** jobs when you are sick.	Control Item for Question 11	Control

34	They will bless‿**your** [ʃɔr, ʃʊər, ʃər] hands for being helpful. They will bless **giving** hands for being helpful.	Control Item for Question 12	Control
35	She already knows‿**you** [ʒu, ʒʊ, ʒə] never got here on time. She already knows **they** never got here on time.	Control Item for Question 13	Control
36	He always has‿**your** [ʒɔr, ʒʊər, ʒər] lists pinned on the board. He always has **store** lists pinned on the board.	Control Item for Question 14	Control
37	I suppose‿**yelling** [ʒɛlɪŋ] is very important. I suppose **cleaning** is very important.	Control Item for Question 15	Control
38	She already knows‿**you did** [ʒʊ dɪd] very well. She already knows **she did** very well.	Control Item for Question 16	Control
Filler Items		**Explanation of Contrast**	**Type of Stimuli**
39	These are the **last cheers** you will see from her	Sound contrast	Filler

	These are the **last tears** you will see from her.		
40	This is his **bike** you saw yesterday. This is his **kite** you saw yesterday.	Sound contrast	Filler
41	We are all **busy** at the moment. We are all **crazy** at the moment.	Sound contrast	Filler
42	Do you have **a false** schedule? Do you have **a fall** schedule?	Juncture	Filler
43	I know we **loan** a lot. I know we'**ll own** a lot.	Juncture	Filler
44	What do you need (**h)im** for? What do you need **it** for?	h-deletion	Filler
45	I'm **taking a nice** cold shower. I am **taking an ice** cold shower.	Juncture	Filler
46	Is this the **night rain** I am hearing? Is this the **night train** I am hearing?	Juncture	Filler
47	Oh no**, this guy** is falling. Oh no, **the sky** is falling	Juncture	Filler
48	The **stuff he knows** can lead to problems. The **stuffy nose** can lead to problems.	Juncture	Filler
49	**Some mothers** I have seen were confused. **Some others** I have seen were confused.	Juncture	Filler

50	He sure has **got a cold** today. He sure has **got to go** today.	Juncture	Filler
51	She already **made a mistake** today. She already **made him a steak** today.	Juncture	Filler
52	Could you find the **ripe pears** please? Could you find the **right pairs** please?	Juncture	Filler
53	Who likes a **gray day** all the time? Who likes a **grade A** all the time?	Juncture	Filler
54	She should **have an aim** as you said. She should **have a name** as you said.	Juncture	Filler

APPENDIX D

CONNECTED SPEECH PERCEPTION TRAINING VIDEO SERIES TRANSCRIPT

In the following script, the pronunciation of words were not phonetically transcribed not to complicate the task for the recorder. The phrases and sentences in italics indicate that they person who recorded it read it in a 'connected' way.

Week 1: Connected Speech in American English and Transforming [t/d + j] sounds

Hello everyone! Sometimes when you are listening to Americans talking, they may sound too fast to you. You may know a lot of vocabulary words and how to pronounce them, but you may just have difficulty understanding what they say.
For example, you hear someone saying '*I wanna go'*, but when you write it down, the full form is actually 'I want to go.' If you have never heard 'want to' pronounced as '*wanna*', the first time you hear it, you may not understand it.

This is quite normal because when advanced or native speakers speak at a natural pace, the words come closer to each other, and speakers may connect, reduce or change the pronunciation of certain sounds. For example, you may hear 'want to' pronounced as '*wanna*', or 'going to' pronounced as '*gonna*'. So, it is important to understand that Americans do not always actually talk fast, but that they connect their words as they talk. This means they blend the end of one word with the beginning of the next. They do this not only because it saves time and energy but it also makes the music of the English language, which is called its rhythm.

Here are some examples:

Please listen carefully:

Would you tell him what you saw?

Did you understand what I just said?

Now, this time, I will say the same sentence in a careful way and more slowly.

Would you tell him what you saw?

Did you understand what I just said?

The first time I said it, you might not have understood the whole sentence because I said it in a more natural way following the musicality and the rhythm of English. Maybe it was harder for you to understand the sentence. Now, let's see what made it hard for you to understand the question "Would you tell him what you saw?' when it was pronounced as '*Wouldja tell'em whatju saw?'*

In the first part, instead of saying 'Would You' I said '*Wouldja*',

In reading the pronoun 'him', in 'tell him', I deleted the 'h' sound and said '*tell'em*' instead of "tell him".

When I said 'what you saw', this time I said '*whatchu saw'*.

Here is another example:

Whaddaya want'em to do?

It could read as "What do you want them to do?"

Let's practice again "*Whaddaya want'em to do*?"

Today, we will talk about what kind of transformation occurs when a word ending in a 'd' or 't' sound is followed by a 'y' sound in an unstressed syllable. It sounds complicated, but when you see the examples, you will understand it every easily.

First, we will discuss the transformation of 'd+y' sounds. Here, we are talking about 'd' as the final consonant of a word and 'y' as the first consonant of the next word. When they come in contact, they affect each other in the way they are pronounced. These two sounds come together and form the [dʒ] sound. In fact you have heard this sound in many other words in English. Some of these are ‘j and in judge’, or ‘j as in jar’ or ‘j an in Jessica’. Please repeat after me: j, j, j. j. Judge, jar, Jessica.

Typically, at a normal speed, especially in casual speech, Americans would not say ‘Would you rather stay home?’ Instead, they connect their speech, in other words, they blend the end of one word with the beginning of the next and say ‘*Wouldja*’, as in ‘*Wouldja rather stay home*? So, here Americans say ‘would’ and ‘you’ together. The ‘d’ sound at the end of the word ‘would’ and ‘y’ sound at the beginning of the word ‘you’ end up as ‘*wouldja’*. Here are more examples. Please repeat after you hear the sentences and phrases.

‘Could you?’ may be pronounced as ‘*Couldja’*

Couldja open the window?

‘Did you?’ may be pronounced as ‘*Didja’*

Did you do your homework?

‘Hide your’ may be pronounced as *‘hidejar’*

Hide your books.

Can you notice what happened to the d+y sound? They ended up forming a new sound, which is [dʒ]. This transformation is not only limited to ‘you’ or ‘your’ or ‘yours’. Here is another example:

‘They offered yoga classes every day.’

……which may be pronounced as ‘*They offeredyoga classes every day.’*

Here because the word ‘yoga’ begins with a ‘y’ sound, when the word ‘offered’ that comes before it ends in a ‘d’ sound, it is still pronounced as [dʒ]. So, it sounds like: ‘*They offeredyoga classes every day.’* However, it is important to note that not all words beginning with a ‘y’ sound do that. Some words are stressed all the time, so they do not transform. In other words, we should not expect all the words starting with a ‘y’ sound to transform when a t-final word comes before them. For example, in the following sentence, we do not connect the sounds ‘y’ and ‘d’ in that way: “She poured yeast into the bowl”. Here is an interesting example:

‘She had university exams’, which may be pronounced as ‘*She hadjuniversity exams.’*

In this sentence, we do not see any ‘y’ letters, but the ‘u’ sound at the beginning of the word ‘university’ in fact sounds like ‘y’ when it is pronounced. As you know, in English, sometimes the spelling and the pronunciation of words might seem different. So, it sounds like: *‘She hadjuniversity exams.’*

Now let’s talk about what kind of transformation occurs when a word ending in a ‘t’ sound is followed by a ‘y’ sound in an unstressed syllable. This is actually a very similar transformation to d+y, which we just talked about. The main difference is the resulting sound. In this case, the resulting sound is ‘ch’. In fact, it is a very common sound in English. Here are some example: ‘ch’ as in ‘China’, or ‘ch’ as in ‘chips’ or ‘ch’ as in ‘cheap’ and so on. Please repeat after me: ‘ch ch ch ch ch’. China, chips, cheap.

In natural speech, especially in casual speech, Americans would not say ‘Who hit your arm?’ Instead, they will connect their speech and say ‘*Who hitchour arm?*’

So, here we say ‘hit’ and ‘your’ together. The ‘t’ sound at the end of the word ‘hit’ and ‘y’ sound at the beginning of the word ‘your’ ended up sounding like *‘hitchour’,* *"Who hit your arm?"*

Here are some more examples. Please repeat after you hear the sentences,

'Can't you?' may be pronounced as *'Can'tchu'*

Can'tchu do it?

'Let you' may be pronounced as *'Letchu'*

I will letchu know.

'don't you' may be pronounced as '*dontchu'*

Why don'tchu join us?

Can you notice what happened to the t+y sound? They ended up forming a new sound, which is 'ch'.

However, it is important to note that not all words beginning with a 'y' sound do that. For example, in the following sentence, we do not connect the sounds 'y' and 't' in that way: 'He felt yawning was rude'. This transformation is not only limited to 'you' or 'your' or 'yours' or some y-initial words. Here is another interesting example: 'They've got unique questions to ask.' Which may be pronounced as '*They've gotchunique questions to ask.*' In this sentence, we do not see any 'y' letter, but the 'u' sound at the beginning of the word 'unique' in fact sounds like 'y' when it is pronounced. So, it sounds like this: '*They've got unique questions to ask.'*

I hope the questions and the explanations have been helpful for you to understand the transformation that occurs when t- and d-final words are followed by 'y'-initial words in English in unstressed syllables. It is hard to cover all the aspects of connected speech but we will cover them to better understand the naturally spoken American English. See you next week!

Week 2: Connected Speech s+j z+j

Transforming s + y sounds

Hello guys! Welcome to another session on connected speech in English. Today, we will talk about what kind of transformation occurs when a word ending in a ‘s’ and ‘z’ sound is followed by a ‘y’ sound in an unstressed syllable. It sounds complicated, but when you see the examples, you will understand it every easily.

First, we will discuss ‘s+y’ transformation. Here we talk about ‘s’ as the final consonant of a word and ‘y’ as the first consonant of the next word. When they come in contact, they affect each other in the way they are pronounced. In fact, together they form a ‘sh’ sound, which is a very common sound in English. This is the ‘sh’ sound as in ‘ship’ or ‘sh’ as in ‘shark’, or ‘sh’ as in ‘sheep’ and so on. Please repeat after me please ‘Sh, Sh, Sh, Sh, Sh, Sh...’, Sheep, Shark, Ship.

Typically, at a normal speed, especially in casual speech, Americans would not say ‘Will you pass your plate?’ Instead, they connect their speech, that is, they blend the end of one word with the beginning of the next and say *‘passyour’*, as in *‘Will you pass your plate?’*. So, here Americans say ‘pass’ and ‘your’ together. The ‘s’ sound at the end of the word ‘pass’ and ‘y’ sound at the beginning of the word ‘your’ end up as ‘*passsjour’*. So, ‘pass your’ may become ‘*passyar’, ‘Will you passyour plate?’*

Here are some more examples. Please repeat after you hear the sentences and phrases.

‘kiss your’ may be pronounced as ‘*kisssjar’*

Kiss your son every day.

‘dress yourself’ may be pronounced as ‘*dress shorself’*

You need to dress yourself tonight.

‘Yes you are’ may be pronounced as *‘yes-shu-ar’*

Yes you are a good friend.

‘Pass your’ may be pronounced as *‘pas-shar’*

Did you pass your test?

‘Guess your’ may be pronounced as *‘guess-shor’*

Let me guess your age.

Can you notice what happened to the s+y sound? They ended up forming a new sound, which is ‘sh’. This transformation is not only limited to ‘you’ or ‘your’ or ‘yours’ or ‘yourself’.

Here is another example:

Then she puts yoghurt in it.

may be pronounced as ‘*Then she putshoghurt in it.’*

Here because the word ‘yoghurt’ begins with a ‘y’ sound, when the word ‘puts’ that comes before it ends in a ‘s’ sound, you may hear ‘*putsyoghurt’* instead of ‘puts yoghurt’. So, it sounds like this: ‘*Then she putshoghurt in it.’*

However, it is important to note that not all words beginning with a ‘y’ sound do that. Some words are stressed, so they do not transform. In other words, we should not expect all y-initial words to transform when a s-final word comes before them. For example, in the following sentence, we do not connect the sounds ‘y’ and ‘s’ that way: ‘He thinks Yiddish is an interesting language.’

Now let’s talk about the transformation, which occurs when a word ending in a ‘z’ sound is followed by a ‘y’ sound in an unstressed syllable. This is actually similar to the ‘s+y’ transformation, which we just talked about. The main difference is the resulting sound. When a word ending in a ‘z’ sound is followed by a ‘y’ sound, they may form a [ʒ] sound as in the first letter of the word ‘genre’. This word is one of the rare examples where we see this sound at the beginning of a word. We do not usually hear this sound at the beginning of the sounds, but we hear it in the middle of the words such as the [ʒ] as in ‘vision’, [ʒ] as in ‘casual’ or [ʒ] as in ‘usual’ and so on. Repeat after me please: [ʒ, ʒ, ʒ, ʒ] as in vision, casual, usual’. Typically, at a normal speed, especially in casual speech, Americans would not say ‘She says you are okay’ most Americans connect their speech, that is, they blend the end of one word with the beginning of the next and say ‘*She sayzjar okay.’* So, here we say ‘says’ and ‘you’ together. The letter ‘s’ at the end of the word ‘says’ actually sounds like ‘z’. Try saying it ‘says’, ‘says’. Can you hear the ‘z’ sound at the end? So, when you write it, it is ‘says’, but when you read it out loud, it is actually ‘say[z]’. In other words, we cannot see it in writing but there is a ‘z’ sound at the end of the word. This ‘z’ sound which is written as ‘s’ at the end of the word ‘says’ and ‘y’ sound at the beginning of the word ‘you’ end up sounding as ‘*sayzyou*’. It sounds like *‘She sayz you’r okay.’*

Here are some more examples. Please repeat after you hear the sentences and phrases.

‘as you’ may be pronounced as *‘azyou’*

Do it as you like.

‘please your’ may be pronounced as *‘pleazeyar’*

You need to please your customers.

‘Freeze your’ may be pronounced as ‘*freezyar’*

You’ll freeze your toes.

Can you notice what happened to the z+y sounds? They ended up forming a new sound, which is [ʒ].

Here is another example:

'How was yesterday's party?'

may be pronounced as *'How wasyesterday's party?'*

Pronounce 'was': 'was', 'was'. Do you hear the 'z' sound at the end of the word 'was'. The 's' letter at the end of the word 'was' actually sounds like 'z'. So, we cannot see it in writing but there is a 'z' sound at the end of the word 'was'. Now, let's go back to our sentence. Here because the word 'yesterday' begins with a 'y' sound when the word that comes before it ends in a 'z' sound, it is still pronounced as [ʒ]. However, it is important to note that not all words beginning with a 'y' sound do that. In other words, we should not expect all y-initial words to transform when a z-final word comes before them. For example, in the following sentence, we do not connect the sounds 'y' and 'z' that way:

'Then, she adds yam to the dough'.

I hope the questions and explanations have been helpful for you to understand the transformation that occurs when s- and z-final words are followed by 'y'-initial words in English in unstressed syllables. See you next week!

Week 3: Review and Wrap-up

Hello guys! Welcome to our last session on connected speech. Therefore, this week, we will review the topics we have covered so far. When the consonants 's, z, t, d' are followed by a [j] sound in an unstressed syllable, the two sounds may combine to form a palatalized consonant. Here is what happens.

So, when a ‘d’ sound at the end of a word is followed by a ‘y’ sound in the next word, as in the first example, the resulting sound is [dʒ] as in ‘*couldja’, ‘couldja open the door’.* When a ‘t’ sound at the end of a word is followed by a [j] sound in the next word as in the second example, they form a [tʃ] sound as in *‘hitjar’, ‘who hit your arm?*

When an ‘s’ sound at the end of a word is followed by a ‘y’ sound in the next word as in the third example, they form a [ʃ] sound as in ‘*passyar’, ‘passjar plate’.*

Finally, when a ‘z’ sound at the end of a word is followed by a ‘y’ sound in the next word as in the last example, they form a [ʒ] sound as in ‘*sayzyou’, ‘She saysyou’re ok’.*

Here are more examples for you to practice from previous weeks. First we will see examples of d+y sounds transforming into a [dʒ] sound. Repeat after me please, it is [dʒ], [dʒ], [dʒ]. Some of these are ‘j and in judge’, or ‘j as in jar’ or ‘j an in Jessica’ and so on.

Here are some examples, please repeat after me:

‘Did you?’ may be pronounced as ‘*Didja’*

Did you do your homework?

‘Hide your’ may be pronounced as *‘hidejar’*

Hide your books.

‘offered yoga classes’ may be pronounced as ‘*offeredyoga classes’*

They offered yoga classes every day.

‘had university’ may be pronounced as *‘haduniversity’*

She had university exams.

Did you see how a ‘d’ sound followed by a ‘y’ sound in an unstressed syllable transformed into a [dʒ] sound?

Now, let's remember how a 't' sound followed by a 'y' sound combine to form a [tʃ] sound. Repeat after me please, it is [tʃ], [tʃ], [tʃ]. 'Ch' as in 'China', or 'ch' as in 'chips' or 'ch' as in 'cheap' and so on. Here are some examples, please repeat after you hear the sentences.

'Can't you?' may be pronounced as *'Can'tchu'*

Can'tju do it?

'Let you' may be pronounced as *'Letchu'*

I will letchu know.

'don't you' may be pronounced as *'dontchu'*

Why don'tchu join us?

'got yoghurt' may be pronounced as '*gotchoghurt'*

They've gotchoghurt drinks.

'got unique' may be pronounced as '*gotunique'*

They've got unique questions to ask.

Now we have seen the transformation of t+y sounds, it is time to review the s+y transformation. As you may remember from last week, this is the 'sh' sound in 'ship' or 'shark', or 'ship' and so on. Repeat after me please 'Sh, Sh, Sh, Sh' Sheep, Shark, Ship. Here are some examples, please repeat after me:

'kiss your' may be pronounced as *'kissshjar'*

Kiss your son every day.

'dress yourself' may be pronounced as *'dress shorself'*

You need to dress yourself tonight.

'Yes you are' may be pronounced as *'yes-shu-ar'*

Yes you are a good friend.

‘Pass your’ may be pronounced as *‘passhar’*

Did you pass your test?

‘Guess your’ may be pronounced as ‘*guess-shor’*

Let me guess your age.

‘puts yoghurt’ may be pronounced as ‘*putshoghurt’*

Then she putshoghurt in it.

Now, let’s practice z+y transformation. Remember, in almost all examples, you will see an ‘s’ sound at the end of the words instead of a ‘z’ sound, but the pronunciation of those ‘s’ sounds is in fact a [z]. When it comes in contact with a ‘y’ sound, it forms a [ʒ] sound. We do not usually hear this sound at the beginning of the sounds in English, but we hear it in the middle of the words such as the [ʒ] as in ‘vision’, [ʒ] as in ‘casual’ or [ʒ] as in ‘usual’ and so on. Repeat after me please: [ʒ, ʒ, ʒ, ʒ] as in vision, casual, usual’. Here are some examples. Please repeat after you hear each example:

‘as you’ may be pronounced as *‘azyou’*

Do it as you like.

‘please your’ may be pronounced as ‘*pleazeyar’*

You need to please your customers.

‘Freeze your’ may be pronounced as *‘freezyar’*

You’ll freeze your toes.

‘was yesterday’ may be pronounced as *‘wasyesterday’*

How wasyesterday’s party?

However, it is important to point out the fact that the amount of transformation will depend on many different factors. It depends on the speaker, style, and formality. In other words, some speakers may use it more while others may not use it as often. In more informal contexts, you may hear these forms more than in very formal contexts. When the speech rate gets faster and faster, you may hear these forms more. So, in fact, it all depends on a variety of factors.

I hope you enjoyed this video series and have already begun to better understand English speakers when you are listening to them or communicating with them. You do not have to speak exactly the same way native English speakers do, but as long as they understand you clearly, there are no problems. However, it is important for you to listen and understand when speakers speak because otherwise there might be a breakdown of communication. So, your primary purpose should be to better 'understand' the forms of connected speech, and through time and practice, learning these forms may help you speak better, too.

I wish you good luck with your English learning adventure! Bye!

APPENDIX E

SPEECH-BASED ATTENTION-SWITCHING TASK STIMULI LIST [1]

1. abes
2. ags
3. alds
4. ards
5. arls
6. arls Repeated
7. aves
8. baft
9. baft Repeated
10. beft
11. duft
12. duud
13. eaps
14. eeds
15. eens
16. ekes
17. epth
18. erf
19. erts
20. fost
21. guft
22. iked
23. kaft
24. kax
25. kest
26. konk
27. merb
28. moft
29. mubs
30. nied
31. noud
32. oals
33. olts
34. orns
35. orns
36. pold
37. pued
38. pued Repeated
39. seft
40. sest
41. talt
42. teep
43. ubs
44. umbs
45. utch
46. uts
47. zect
48. zect Repeated
49. zemp

[1] Four non-words marked as 'repeated' had to be read twice due to unforeseen experimental and technical problems as they had to replace the original words.

APPENDIX F

SPEECH-BASED ATTENTION-SWITCHING TASK BLOCK DESIGN

(Darcy et al. 2014; Darcy and Mora, in press; Darcy 2014 personal communication)

Block design was adapted from Darcy, 2014 (personal communication).

VF= vowel-initial spoken by a female

CF= consonant-initial spoken by a female

VM= vowel-initial spoken by a male

CM= consonant-initial spoken by a male

MC	Neutral	Practice block
MV	no-shift	
MV	Shift	
MC	no-shift	
MV	Shift	
FC	no-shift	
FC	Shift	
FV	no-shift	
FC	Shift	
FV	no-shift	Block1
MV	Shift	
FV	no-shift	
FC	Shift	
MC	no-shift	Block 2
FC	Shift	
MC	no-shift	
MV	Shift	
FV	no-shift	Block 3
MV	Shift	
FC	no-shift	
FC	Shift	
MC	no-shift	Block 4
FC	Shift	
MV	no-shift	
MV	Shift	

MV	no-shift	Block 5
MV	shift	
FV	no-shift	
FC	shift	
FC	no-shift	Block 6
FC	shift	
MC	no-shift	
MV	shift	
FV	no-shift	Block 7
MV	shift	
MC	no-shift	
MV	shift	
MC	no-shift	Block 8
FC	shift	
FV	no-shift	
FC	shift	
FV	no-shift	Block 9
MV	shift	
FV	no-shift	
FC	shift	
MC	no-shift	Block 10
FC	shift	
MC	no-shift	
MV	shift	

APPENDIX G

SPEECH-BASED ATTENTION-SWITCHING TASK: ASSIGNMENT OF STIMULI IN BLOCKS TAKING PERSEVERATION INTO ACCOUNT

(Template adopted from Darcy, 2014 personal communication)

Green= 'Yes' answers, Red= 'No' Answers

Dimension : Voice Quality
Question : Male?

	Condition			
consonant > gender	Consonant?			
	consonant Female	Vowel Male	vowel female	consonant male
	Male?			
	Female consonant	male vowel	Male vowel	female consonant
Within category (answer) perseverance	wrong	wrong	wrong	wrong
Across category (question) perseverance	wrong	wrong	wrong	wrong

gender > consonant	Male?			
	Female consonant	male vowel	female vowel	male consonant
	Consonant?			
	consonant Female	vowel male	consonant female	vowel male
Within category (answer) perseverance	wrong	wrong	wrong	wrong
Across category (question) perseverance	wrong	wrong	wrong	wrong

APPENDIX H

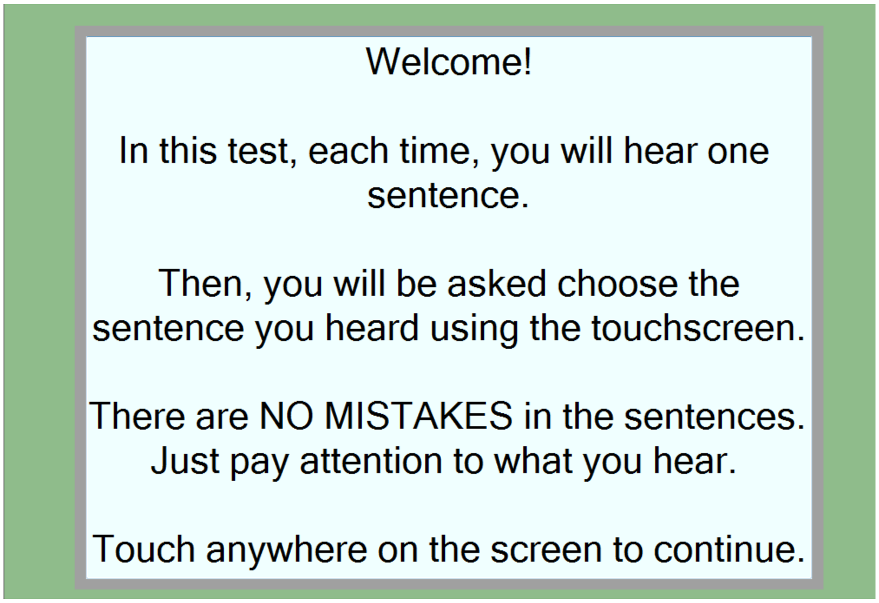
ATTENTION NETWORK TEST STIMULI DESIGN

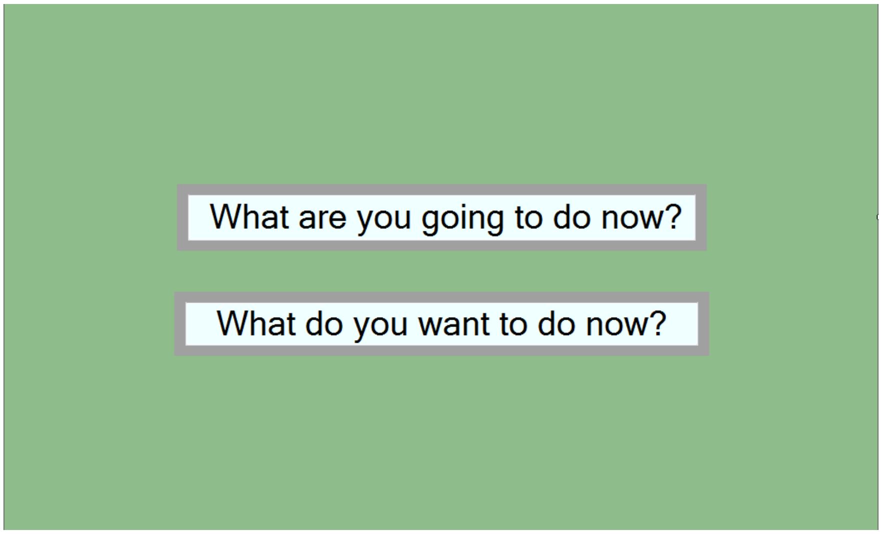
There are 64 trials in total (32 X 2)

Condition	Warning Type	Target Type	Flanker Type	Target Direction
1	No	UP	congruent	Left
1	No	UP	congruent	Right
1	No	DOWN	congruent	Left
1	No	DOWN	congruent	Right
1	No	UP	incongruent	Left
1	No	UP	incongruent	Right
1	No	DOWN	incongruent	Left
1	No	DOWN	incongruent	Right
2	Center	UP	congruent	Left
2	Center	UP	congruent	Right
2	Center	DOWN	congruent	Left
2	Center	DOWN	congruent	Right
2	Center	UP	incongruent	Left
2	Center	UP	incongruent	Right
2	Center	DOWN	incongruent	Left
2	Center	DOWN	incongruent	Right
3	Up	UP	congruent	Left
3	Up	UP	congruent	Right
3	Up	DOWN	congruent	Left
3	Up	DOWN	congruent	Right
3	Up	UP	incongruent	Left
3	Up	UP	incongruent	Right
3	Up	DOWN	incongruent	Left
3	Up	DOWN	incongruent	Right
4	Down	UP	congruent	Left
4	Down	UP	congruent	Right
4	Down	DOWN	congruent	Left
4	Down	DOWN	congruent	Right
4	Down	UP	incongruent	Left
4	Down	UP	incongruent	Right
4	Down	DOWN	incongruent	Left
4	Down	DOWN	incongruent	Right

APPENDIX I

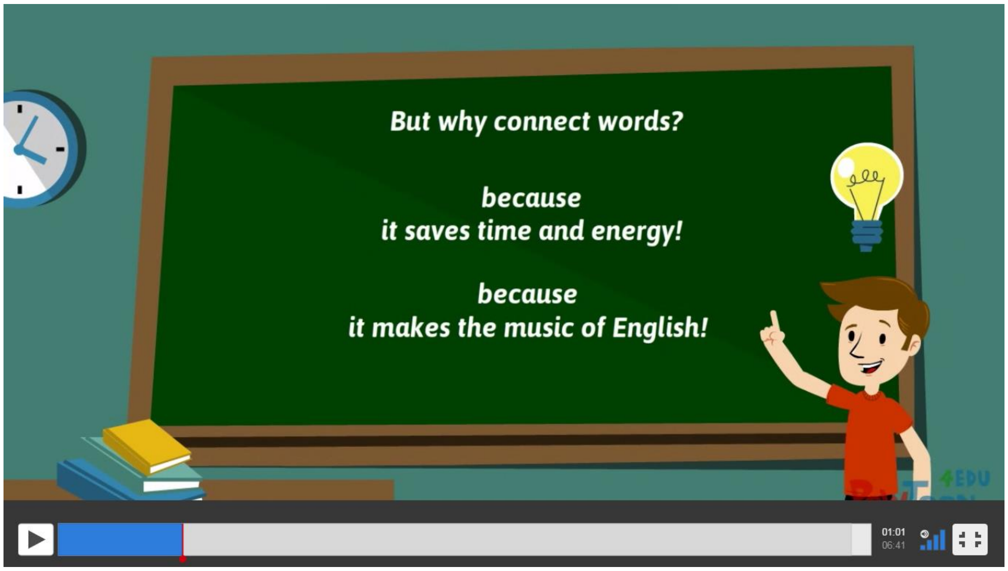
SCREENSHOTS OF THE PERCEPTION TEST

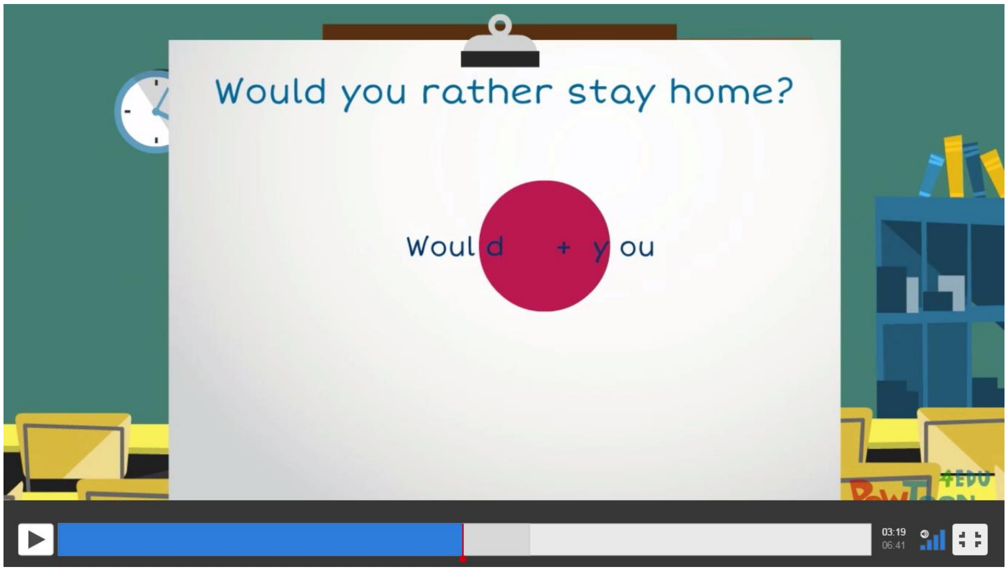


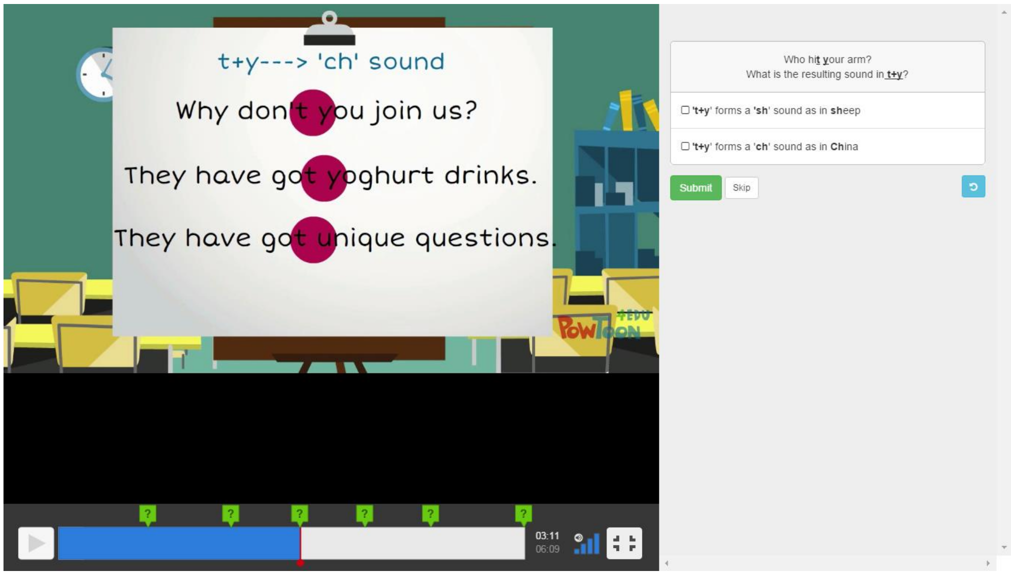


APPENDIX J

SCREENSHOTS OF THE TRAINING VIDEO SERIES

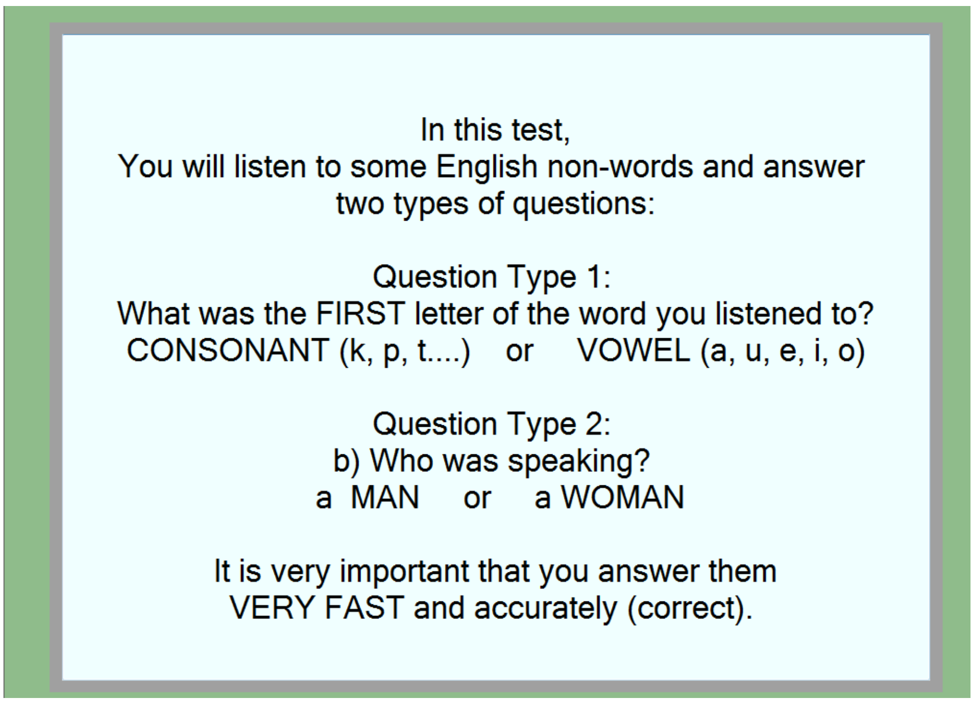


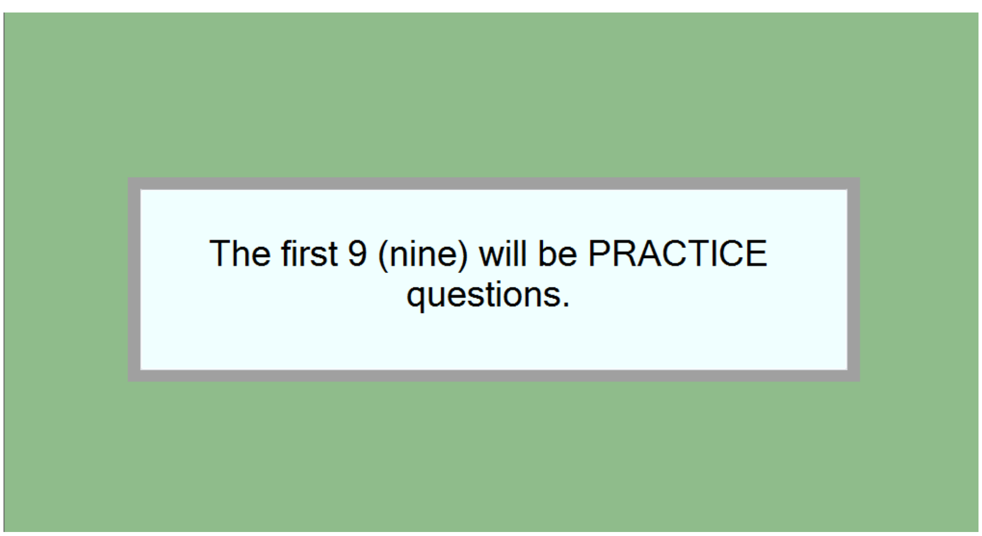

Would you rather stay home?
Would + you
03:19
06:41


t+y---> 'ch' sound
Why don't you join us?
They have got yoghurt drinks.
They have got unique questions.
03:11
06:09
Who hit your arm?
What is the resulting sound in t+y?
't+y' forms a 'sh' sound as in sheep
't+y' forms a 'ch' sound as in China
Submit
Skip

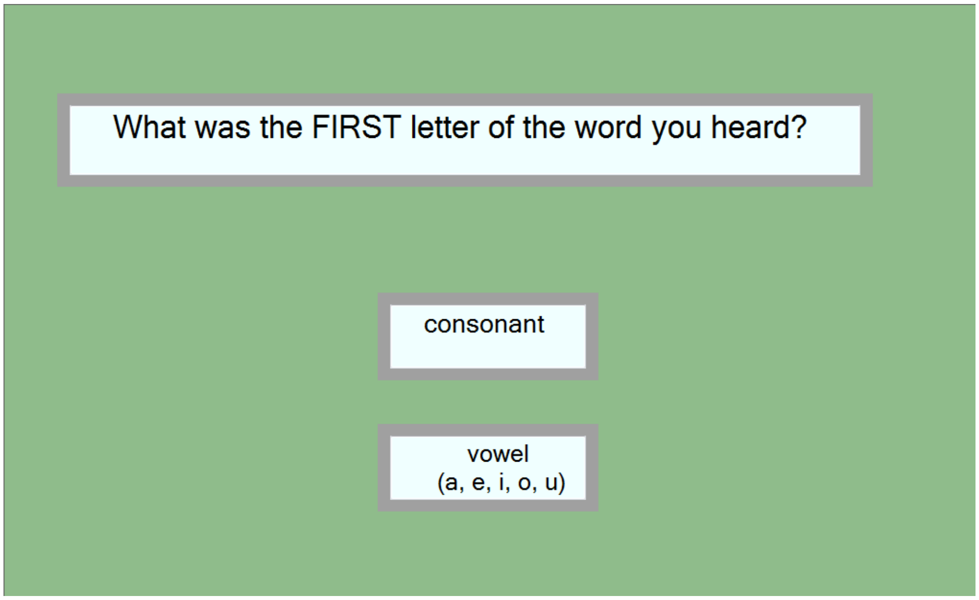
APPENDIX K

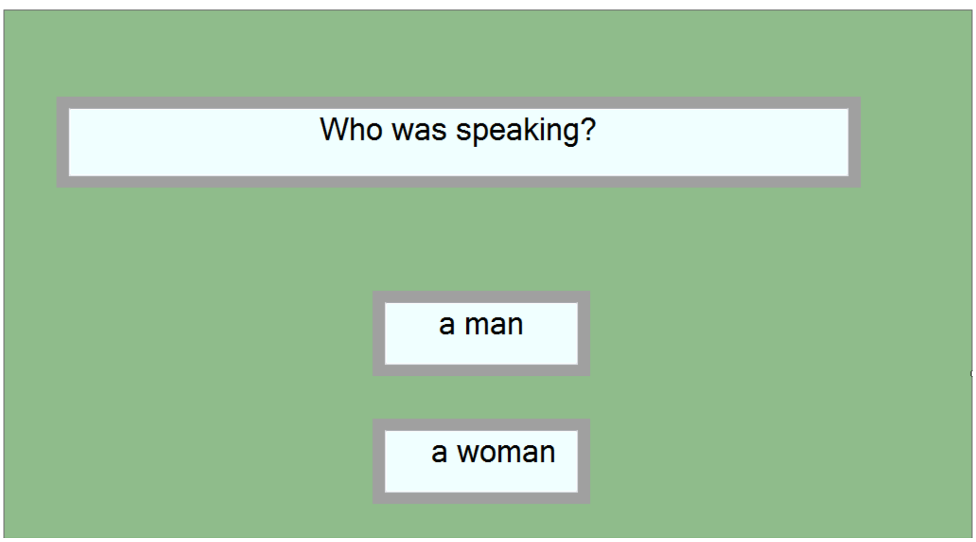
SCREENSHOTS OF THE SPEECH-BASED ATTENTION-SWITCHING TASK



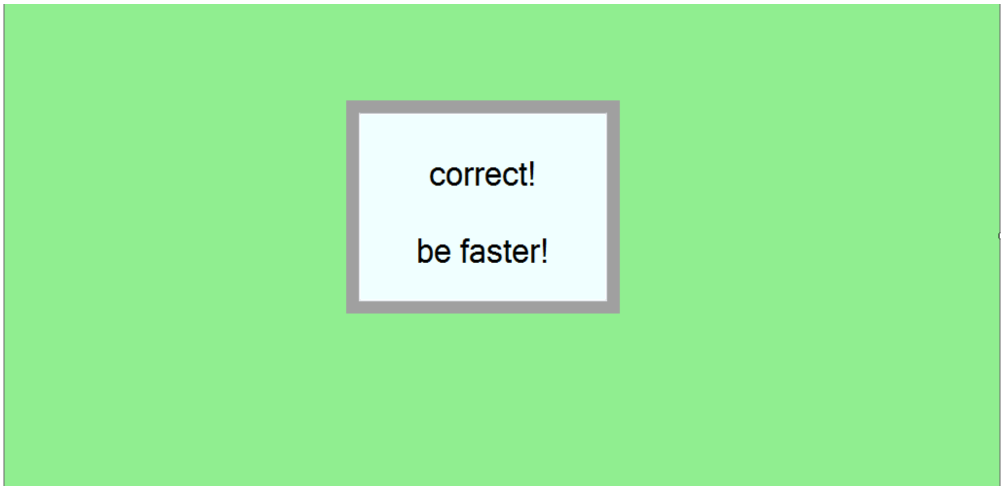


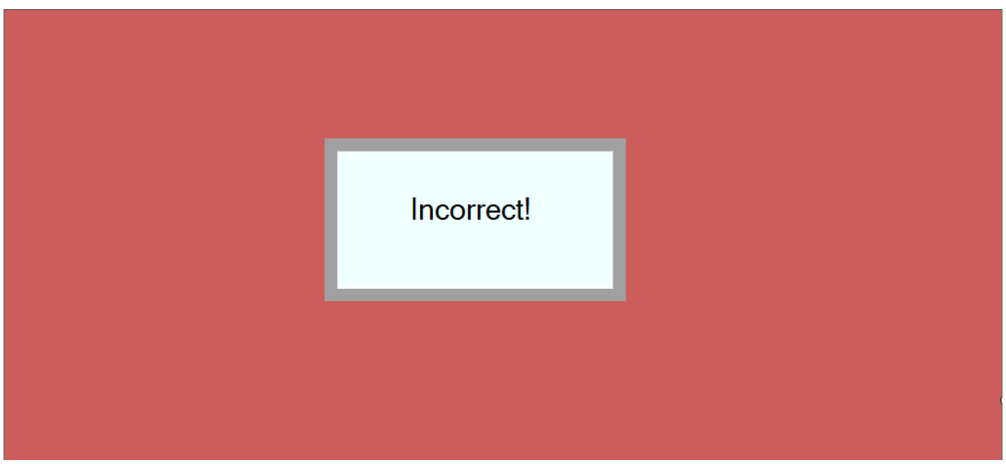
(Screenshots of the practice trial with questions and hints)



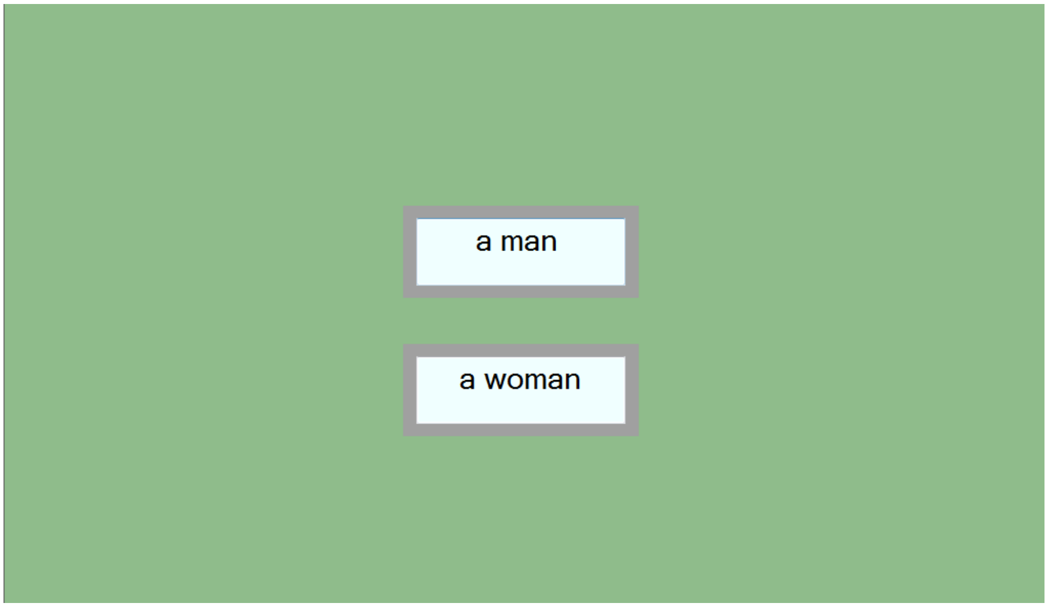


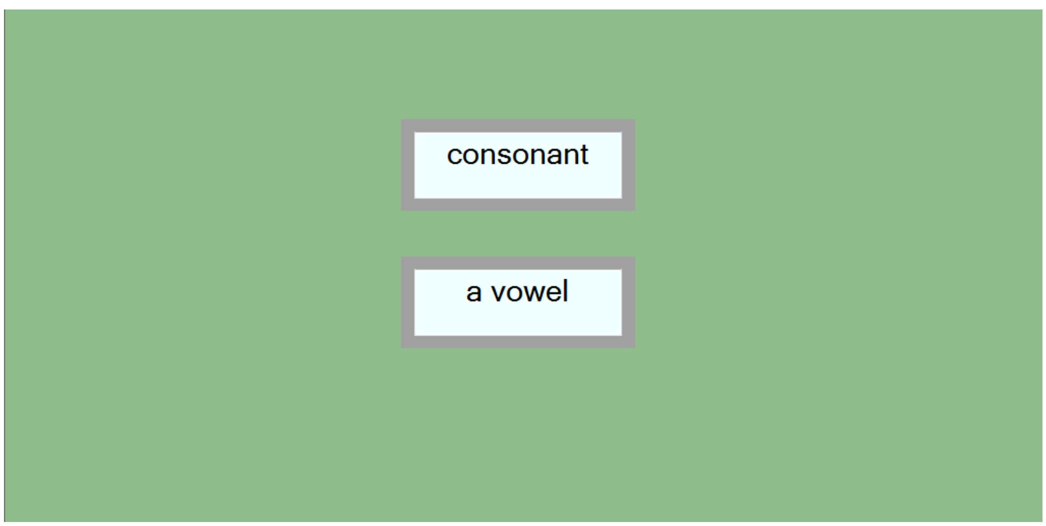
(Screenshots of feedback for practice trials)





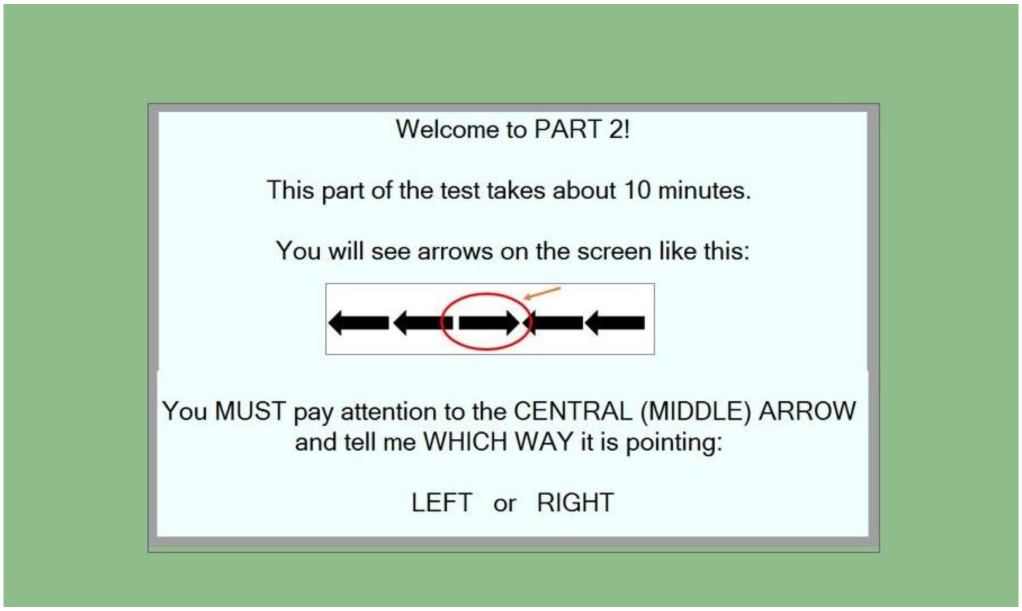
(Screenshots of the experimental trial with no questions, hints or feedback)

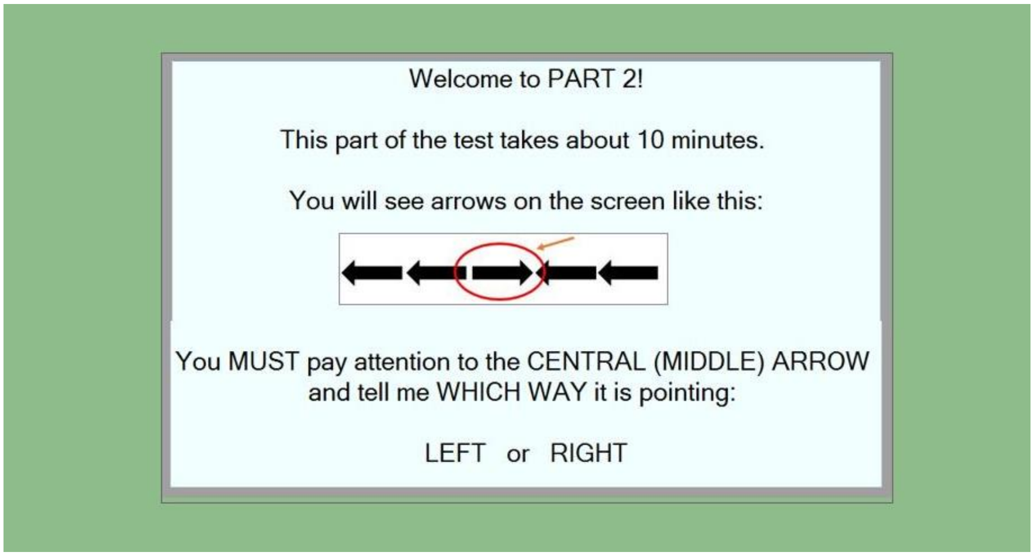


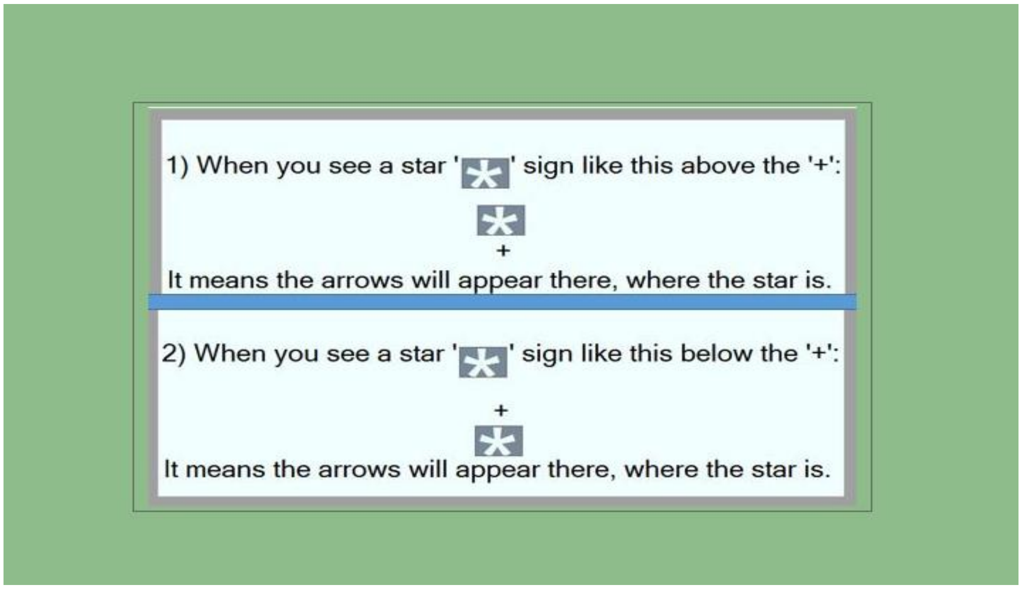


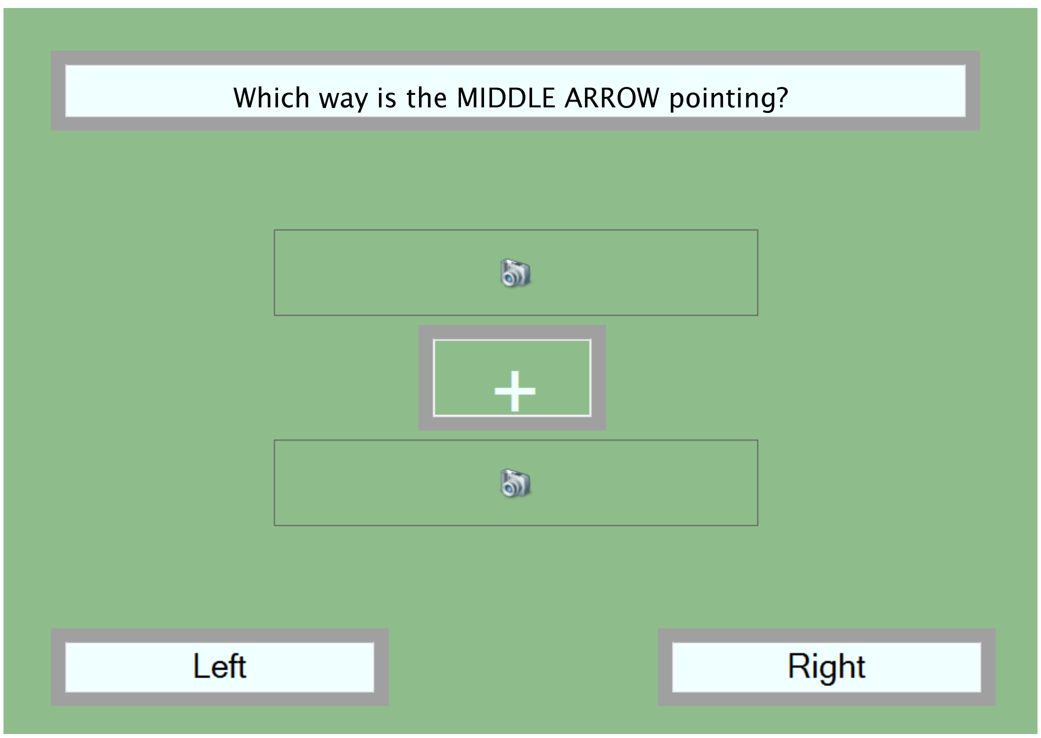
APPENDIX L

SCREENSHOTS OF THE ATTENTION NETWORK TEST









Note. Any of the four following sets of arrows appeared in the top or the bottom box with a camera icon in it, depending on the correct anwer of the trial type.