Perception of American English vowels by adult Spanish-English bilingual listeners

Paula Garcia

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy under the Executive Committee

of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2014

ABSTRACT

Perception of American English vowels by adult Spanish-English bilingual listeners

Paula Bibiana Garcia Cardona

Cross-linguistic studies have demonstrated that learning of a second language (L2) is influenced by the phonological system of the native language (L1), with L2 learners forming mental representations of new, non-native sounds by a process of assimilation to familiar native sounds (Best, 1995; Flege, 1995). Adult sequential successive bilingual Spanish-English speakers may be specifically challenged in perceiving and acquiring American English (AE) vowel contrasts that are signaled by multiple cues not phonemically relevant in their native language. Much of the existing research on vowel perception in L1-Spanish adults has focused on the AE vowel contrast /i/ vs. /ɪ/, as in *sheep* vs. *ship*, because discrimination errors between these two vowels are common (Escudero, 2000; Morrison, 2006; 2008; 2009). However, other vowel contrasts /ʌ/-/ɑ/ (as in *hut* vs. *hot*) have also been reported to present perceptual challenges for native Spanish-speaking learners of English (Flege, Munro & Mackay, 1995; Escudero & Chládková, 2010). It is assumed that such perceptual issues contribute to poor performance in second language acquisition and processing, and have implications for access to employment and academic opportunities for a large and growing immigrant population in the United States (Labor Employment and Training Administration Report, 2005).

The aim of this study is to implement electrophysiological and behavioral methods to further elucidate the perceptual and processing abilities of L1-Spanish adult learners of English, while examining less-studied AE vowel contrasts /ʌ/-/ɑ/, and to evaluate whether specific properties of these speech sounds, such as spectral and duration differences, contribute directly to difficulties encountered in L2 acquisition. More specifically, in this study we will examine response accuracy and reaction time, as well as Mismatch Negativity (MMN) and P300 Event-Related Potentials in two listening conditions: *natural vowel duration*, where target vowel sounds are presented naturalistically, and *neutral vowel duration*, in which speech sound discrimination is possible based on spectral cues alone. Event Related Potentials (ERPs – MMN and P300) are neurophysiological indices that can reflect native and non-native mental phonological representations. Findings from the pilot study that utilized natural and neutralized duration speech sounds revealed behavioral and neurophysiological differences between Spanish-English bilingual listeners and native English speakers responses to natural AE vowel contrasts. This raised a question of whether adult Spanish-English bilinguals relied on speech cues in a similar fashion to native English speakers when perceiving these AE vowel contrasts. It is understood that language-specific use of speech cues (e.g. spectral and durational) helps to distinguish between perceptually similar speech sounds. Therefore, it was assumed that removal of duration distinctions between the target vowels would reveal any underlying differences in the processing mechanism and how much L1-Spanish listeners rely on durational cues to perceive subtle differences between vowel pairs.

Findings from this dissertation study indicate that adult sequential Spanish-English bilingual listeners (Study group) showed indices of discrimination and identification of AE vowel /ɑ/ but not /ʌ/ at the attentional level, when both spectral and durational information about the vowels was perceptually available in the natural vowel duration condition, but also when duration was neutralized leaving only spectral cues available to distinguish the vowels. The current findings show that Spanish-English bilinguals may use spectral and durational cues, like native English speakers, to perceive the English vowel contrast /ɑ/-/ʌ/. However, this cannot be described as an "end state" in the sense of Escudero (2005), since the neurophysiological evidence shows that these L2 learners are able to reach native-like discrimination only when they recruit attentional and cognitive resources to facilitate the perceptual process.

Table of Contents

Table of Figures

Acknowledgments

I wish above all to thank my advisor, Karen Froud, for teaching and mentoring me all these years with such patience, wisdom, and encouragement. She has been instrumental both to this project and more broadly to my development as a scholar and a person. Over the past several years, Karen has inspired, challenged, and supported me, always at the right time and always in the right way. Working with Karen, in short, has been a wonderful learning experience.

I am deeply grateful to Laura Sanchez for having shared this, and many other learning experiences with me along these years.

I would like to express my very great appreciation to Dr. Erika Levy for her valuable and constructive suggestions during the planning and development of this research work. Her willingness to give her time so generously has been very much appreciated.

I am particularly grateful for the assistance given by Dr. Nishi, who generously answered multiple emails and phone calls to generate optimal speech stimuli, and also provided me with invaluable feedback on this research work.

I would like to thank my other committee members, Dr. Hansun Waring, and Dr. Robert Remez, for their helpful comments throughout the dissertation project and for challenging me to think about my work from the perspectives of multiple disciplines.

I am also grateful to the Biobehavioral Sciences department at Teachers College, for the institutional support I have received during my graduate years. In particular, I would like to thank Dr. Peter Gordon, Dr. John Saxman, and especially, Maria LaMadrid, and the Office of International Services for all the generous fellowships that have allowed me complete this program and to do the research presented in this dissertation.

My work and life as a graduate student has been greatly enriched by supportive colleagues and a wonderful community of collaborators. I am grateful to each of the members of the Neurocognition of Language Laboratory. Trey Avery, Reem Khamis-Dakwar, Lisa Levinson, Heather Green, Lauren Godwin, Felicidad Garcia, Chaille Maddox, Guannan Shen, and especially to Dayna Moya.

This research work would have not been possible without each of the wonderful people that endured two sessions of EEG data collection. I am deeply thankful to their participation and enthusiasm.

Finally, I am both thankful for and grateful to my family, and friends. Their support and friendship over these years has been invaluable to me. It would be impossible to name everyone here, but I would especially like to thank my parents, Gloria Esperanza and Raul, my sister Ana Maria, and my brother Raul. Finally, I most want to thank Carlos, whose patience and love have supported me through the challenges.

1. Introduction

Learning English in this global world is an indispensable skill, needed to access information and technology, to understand the intricacies of world politics and conflicts, and to participate in academic, work, and social environments. Mastering a second language (L2) becomes even more important when immigrating to a foreign country. According to the U.S. Census Bureau report (2010) in the American Community Survey, 20.6% of individuals over the age of 5 years living in the United States speak a language other than English at home, and 12.8% of those speak Spanish. The Hispanic population is the largest minority population in the United States and accounts for 47% of the total 40 million immigrants who are currently living in the U.S. Eighty-two percent of foreign-born Hispanics over the age of 18 years, who arrived after 2000 in the USA, reported they do not speak English well (Pew Hispanic Center, 2012).

The above statistics suggests that a large part of the U.S. Hispanic community has little confidence in its ability to communicate in the country's dominant language (English). They likely encounter difficulties communicating with native English speakers and other L2-English users, in daily life situations, and especially in linguistically demanding environments such as school or work. The US Department of Labor Employment and Training Administration Report (2005) stated that two in five foreign-born Hispanics experience language and cultural barriers to success, and are more likely to live in poverty than other segments of the population.

A study conducted by Kochhar (2005) indicated that foreign-born Hispanics who do not speak English are likely to have lower-income occupations (e.g., farming, serving, production, and construction) compared to those who speak the language fluently. The consequences of not mastering English are certainly detrimental for Spanish language-dominant Hispanics in the USA. Those consequences may include fewer opportunities for academic and professional qualification, restricted participation in social and community life, and limited access to qualified jobs for which they are otherwise qualified. In order to contribute to improving the language competence of this adult population, it is important to better understand the learning processes that occur during the process of learning a second language in adulthood. Adult learners of a non-native language differ significantly from children in their abilities to acquire linguistic competence. A body of evidence indicates that, as age increases, the ability to perceive and produce non-native speech sounds decreases (Best, McRoberts & Goodell, 2001; Flege, 1995, 2003; Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola & Nelson, 2008). The effective use of language requires acquiring knowledge and the ability to use that knowledge at different linguistic levels including morphology (the internal structure and formation of words), syntax (rules governing sentence structure), semantics (the meaning of language), vocabulary (the words in a language) and phonology (the systematic organization of sounds in a language). The proposed study is concerned primarily with the acquisition of L2 phonological knowledge.

Phonology is concerned with how the sounds of a language are systematically organized in a *phonological system*. The phonological system in every language constitutes an abstract representation of the speech sound classes and their relationships to convey meaning. All language sounds can be classified as *phonemes* (Raphael et al., 2011), referring to the smallest unit of sound that signals a difference in meaning, and *allophones* referring to alternative productions of the sound in a language without change in the meaning of the word. For example, in English, the phoneme /p/, is produced differently in the words *pin* [p^h] and *spin* [p], but if *pin* was produced using [p], it would still be *pin*, with a "non-native" sound for an English speaker. Therefore, the English phoneme /p/ has allophonic variants, including [p^h] and [p].

Phonemes and their allophones are specific to different languages. That is, in Spanish, [d] and [ð] are allophones of the phoneme /d/. For example, *dedo (finger)* is pronounced /dedo/ and *lado (side)* is pronounced /laðo/; however, in English /d/ and /ð/ are two separate phonemes that signal different meaning, in words such as *doze* /doʊz/ and *those* /ðoʊz/. The term *phone* is used to refer to the actual production of each member of a phoneme class (e.g. [p^h] and [p], as they are produced in the words /pin/ and /spin/ respectively, are phones). In addition to the abstract representation of the speech sounds in the phonological system as phonemes, another level of organization of the speech sounds corresponds to the phonetic categories. At this level, the listener assigns speech sounds to categories according to the acoustic or articulatory features of the speech signal.

The production of human speech sounds is a complex process that is the result of air flowing through the phonatory system. The air vibrates when reaching the vocal folds and resonates at the nasal and oral cavities. In the oral cavity, the position, and the size of the articulators (i.e., lips, teeth, tongue, alveolar ridge, palate, etc.) filter the sound waves creating specific resonating frequencies with acoustic characteristics that carry information and meaning for the listener. The speech signal travels to the listener's ears as pressure waves containing information that reveals the relative position of the articulators in the oral cavity, the length of the speech segment, the exact time of the release of bursts of air at the lips, and the identity of the speaker, among other cues that are crucial for the listener to decode the speaker's intended meaning.

The speech signal is characterized by its variability; it varies according to the talker's individual characteristics (e.g., vocal tract size and shape, gender, age, language, and dialect), the linguistic context (i.e., other phonemes that surround the target sound), and the speech conditions (e.g., background noise, speech rate). Even though it is heard as a linear stream, the speech signal is not linear and the sounds are overlapping, which makes it difficult to determine when one sound starts, when it ends, and how each sound affects its neighboring speech sounds. The listener's task is to extract the relevant (invariant) information from the speech signal and map it onto his/her mental representations. This is possible through the decoding of auditory-phonetic cues contained in the speech signal.

Speech auditory-phonetic cues inform the listener about the nature and identity of speech sounds. These cues are often redundant, and may have a primary or secondary status (i.e., more or less needed) in a given language for the identification of a speech sound (Nittrouer, 2002). Important cues that are known to be useful for discriminating between phonemes that are native to Spanish and English include spectral and durational cues (Escudero, 2000; 2005; Fox, Flege & Munro, 1995). Spectral cues refer to changes in signal properties caused by the relative position of the articulators in the vocal tract. The positions of articulators, and the constrictions that form in the oral cavity, cause the air to resonate in different frequencies. These resonance frequencies are represented as formants (i.e., bands of energy) in a spectrogram. Each formant frequency reflects the resonance of the vocal tract when the articulators are held in specific positions relative to one another (see figure 1). Formant 1 (F1) reflects resonant frequencies associated with tongue height in the oral cavity, while Formant 2 (F2) changes with front-back position of the tongue along the oral cavity. Resonance frequencies in Formant 3 (F3) are also associated with changes in the place of articulation. Changes from vowel to consonant or consonant to vowel are signaled by *formant transitions* – that is, rapid temporal shifts in resonant frequencies. Formants 1 and 2 in isolation have been shown to provide sufficient auditory-phonetic information for a listener to perceive vowel sounds (Fox & Jacewicz, 2009), so these two formants will provide a focus for this study.

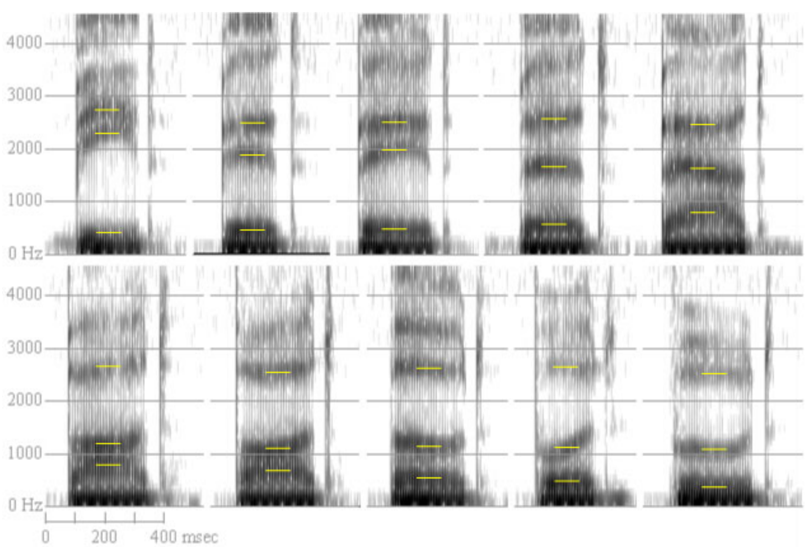

Figure 1. Wideband spectrogram of American English vowels.

Each bar represents a vowel. Top row, left to right: [i, ɪ, eɪ, ɛ, æ]. Bottom row, left to right: [ɑ, ɔ, o, ʊ, u]. Each yellow line represents a formant frequency. Reprinted from Rob Hagiwara's Monthly Mistery Webzone, R. Hagiwara. Retreived June 12, 2013, from http://home.cc.umanitoba.ca/~robh/howto.html#formants. Copyright 2009 by Rob Hagiwara. Reprinted with permission.

Durational cues inform the listeners about duration of a phoneme and the relative duration of a speech sound (i.e., phone) in specific linguistic contexts. Duration may be a primary or a secondary cue in different languages. For example, duration is a primary cue for Japanese vowels because length differences correlated with differences in meaning (Kinoshita, Behne & Arai, 2002). That is that in Japanese, each vowel has both short and long instantiations that signal meaning differences between words. For example, the contrasting words *obasan/*obasan "aunt" vs. *Obaasan/*obaasan "grandmother", or *tsuki* /tsuki/ "moon" vs. *tsūki* /tsuuki/ "airflow" are minimal pairs that differ phonetically only with respect to vowel duration (Vince, 1987). In English, duration is used as a secondary cue (Hillenbrand, Clark, & Houde, 2000). That is, duration does not play a primary contrastive role, but offers supporting cues to assist with disambiguating English speech sounds. For example, the vowels in English words *beat* /bit/ and *bit* /bɪt/ differ primarily in their spectral characteristics (i.e., formant frequencies). However, there is a subtle duration difference that can also aid a listener in distinguishing between these two phonemes.

In Spanish, unlike Japanese and English, durational cues are not used to signal meaning changes in words (Cebrian, 2006). That is, duration values in Spanish speech sounds do not provide listeners with relevant information about the identity of a specific sound. However, similar to English, spectral information is highly relevant for signaling differences between Spanish speech sounds. Other speech cues that are useful in perception of speech in different languages include voice onset time (VOT: a difference in the length of time taken to initiate vocal fold vibration following the release of a stop consonant), and tones (systematic frequency differences that signal lexical distinctions in some languages) (see Raphael et al., 2011).

From auditory-phonetic cues, listeners obtain relevant acoustic information that is mapped onto mental representations of the sounds. Due to the variability in the speech signal, listeners continually adjust their perceptual systems in order to detect and weigh the relevant acoustic cues that distinguish speech sounds from one another, ignoring information that does not contribute to this differentiation. Speech sounds are systematically represented in the phonological system, and the organization of the speech sounds provides meaning to the language. This organization differs from one language to the other. During the first months of life, infants are able to discriminate between many speech sounds, including some that are not a part of their native-language phonology. However, by the end of the first year of life, sensitivity to native contrasts increases, while sensitivity to non-native contrasts decreases (Jusczyk, 1997; Bosch & Sebastián-Gallés, 2003). Therefore, infants become specialized in extracting and weighting the cues in the phonemes of their native language (Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson, 2006).

Typically developing children use most of these cues of their native language by the age of three, without any formal language teaching. The mechanisms associated with language learning have motivated research from different approaches. Approaches such as Behaviorism (Skinner, 1957) and Nativism (Chomsky, 1957) have provided valuable insight into the innate capacities and learning abilities of humans. Behaviorist approaches argue that language is a learned behavior that could not happen without environmental input. The Nativist approach holds that humans are endowed with an innate faculty of language, sometimes referred to as Universal Grammar (UG), that interacts with environmental input in highly constrained and specified ways. Research evidence has demonstrated that the interplay between innate capacities, learning abilities, and environmental input plays a fundamental role in language acquisition (Kuhl, 2009; Saffran, Werker, & Werner, 2006).

The same theoretical approaches have been applied to the second language acquisition process. Here, I will briefly outline three perspectives on second language acquisition that vary with respect to the availability of UG for the second language learner. First, the Full Access position holds that L2 learners can acquire L2 features even if they are not part of the learner's L1 (Schwartz & Sprouse, 1996). Second, the Partial Access position claims that L2 learners only acquire properties of L2 that are part of their L1 (Bley-Vroman, 1990). Last, the No Access position argues that second language learning does not involve access to universal grammar principles (Meisel, 2001). Partial access and no access views argue that full access to the universal grammar is not possible in L2 learners due to ageing out of critical/sensitive periods (the concept of critical/sensitive period and its implications for first and second language learning will be discussed in chapter 2). One of the most influential authors from the nativist point of view is Krashen, who explains the process of second language acquisition from a nativist approach. According to Krashen, there is an acquisition system and a learning system that are at work in second language learning. The acquisition involves subconscious processes that are similar to those implemented during first language acquisition in children. This system requires meaningful interaction in the target language and communication experiences where speakers are not focusing on formal production but on the communicative act. On the other hand, conscious learning also plays a role, and formal instruction as well as error correction provides the learner with knowledge about the language, such as grammar rules, through the explicit learning system (Krashen, 1988).

By contrast, cognitive approaches (e.g., Ellis, 1994; Long, 1996) and socio-cultural approaches (e.g., Vygotsky, 1978) to second language learning agree on the importance of meaningful interactions. Long (1996) argues that interaction facilitates language acquisition connecting language input, inner learning capacities, and what the learner can produce, in meaningful ways. In this sense, through interaction, learners obtain appropriate input and feedback (Gass, 1997; Long, 1996; Pica, 1994) that allow them to adjust their linguistic output (Swain, 1995).

As can readily be seen from this brief overview of theoretical frameworks, very different perceptual challenges are involved in learning a second language in adulthood, compared to the task of L1 acquisition in childhood. Unlike children, who (under conditions of typical development) perceive the sounds of their L1 effortlessly, adults need to learn to perceive a different set of speech sounds and the relevant speech cues for new language's sound system. The task of L2 speech sound acquisition also necessarily involves analysis of acoustic stimuli so that cues in the speech signal can be used to establish organizational parameters of the new phonological system. This process occurs under the influence of the already established and specialized native phonological system. At least during initial learning stages, the L1 phonological system serves as a filter through which the adult listener classifies and assigns the new sounds to native phonological categories (i.e., groups of speech sounds that share certain characteristics) (Flege et al., 1995; Best, 1995; Best & Tyler, 2007). This influence makes it easier to acquire certain speech sound distinctions in the second language phonological system, whereas other contrasts are more difficult to acquire, depending on properties of the L1 system (Polka, 1991).

In addition to the effect of L1 on the perception of L2 speech contrasts, the sequence of L1 and L2 acquisition has been found to play a role in the second language learning process. The term most widely used to refer to children and adults who have learned more than one language in the course of their lives is *bilinguals.* Butler and Hakuta (2004) define bilinguals as individuals with communicative skills in two or more languages, with different degrees of proficiency, and who interact with speakers of the specific languages in a community. Researchers in the field have distinguished between the terms (1) *simultaneous bilinguals,* referring to those bilinguals who have learned one or more languages simultaneously during childhood (Meisel, 1989), and (2) *sequential/successive bilinguals*, referring to bilinguals who have been introduced to other languages later in life after the acquisition of L1. For sequential bilinguals the first language has been already established when the second language is introduced. In addition, Konhert & Bates (2002) apply the term *early sequential bilinguals* to those who learn L2 during childhood sequential to L1-only exposure in their early childhood. If their first exposure to the L2 happens after the age of three, children are generally considered sequential bilinguals (Genesee, Paradis & Crago, 2004). This study will focus on the L2 vowel perception characteristics of sequential bilinguals.

Research in adult cross-language speech perception has focused on how listeners perceive the speech sounds of a different language, and what factors make this task harder or easier for various listener groups. Lexical factors play a facilitative part in speech perception and production: for example, research has shown that Japanese learners of English find it difficult to perceive and produce the English consonant contrast /ɹ/ and /l/, which is a phonemic contrast (i.e., signals a meaning distinction) in English but not in Japanese (Logan, Lively & Pisoni, 1991; Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura & Yamada, 1994; Strange & Dittmann, 1984). Phonological cues also play a part in cross-language vowel perception. For example, American English (AE) vowels are distinguished by spectral information (such as formant shifts) and durational information (temporal distinctions); but those cues are not allocated equal importance by L1 English listeners (Hillenbrand, Clark & Houde, 2000). It has been shown that spectral changes are more important than durational information for L1-English listeners to identify their vowels (Hillenbrand et al., 2000).

Behavioral studies looking into adult L1 Spanish-speaking users of L2 English have shown that this population has difficulty perceiving English vowel contrasts that are signaled by multiple cues that are not phonemically relevant in their native language. Most of this research has focused on the English vowel contrast /i/-/ɪ/ as found in 'beat' and 'bit'. (Bohn, 1995; Escudero, 2000; Flege 1997; Fox et al., 1995; Morrison, 2006, 2008, 2009). Such studies have demonstrated that L1-Spanish listeners of English weigh spectral and duration cues differently from native L1-English speakers when perceiving and producing the contrast, at least during some early learning stages (Escudero, 2000). L1 Spanish listeners use spectral cues to distinguish between the five vowels, and durational cues are not phonemically used in Spanish. In addition, the spectral values between Spanish and some AE vowels are very similar. Therefore, adult Spanish listeners who learn AE vowels may have difficulty in acquiring AE speech sound contrasts that rely primarily on spectral durational cues (Escudero, 2000). However, other vowel contrasts such as /ʌ/-/ɑ/ (as found in 'hot' and 'hut') has also been reported to present perceptual challenges for Spanish listeners (Escudero & Chládková, 2010; Flege et al., 1995).

To conclude, multiple factors have been found to influence the perception of non-native speech sounds. Some of these factors are related to the listener: L1 background (Iverson & Evans, 2007; Iverson & Evans, 2009), age of the learner (Flege et al., 1995), or experience with L2 (Flege, Bohn, & Jang, 1997). Other factors are related to the characteristics of speech sounds, including relative perceptual difficulty (Polka, 1991; Strange & Dittman, 1984), or relevant cues for identification (Escudero & Boersma, 2004; Holt & Lotto, 2006; Morrison, 2006, 2008, 2009). In addition, adults who learn a second language in their L1 environment are usually enrolled in language classes taught by non-native, L2 English teachers, who may provide distorted L2 phonetic models (Best & Tyler, 2007), and inadvertently contribute to difficulties with L2 phoneme perception when learners eventually attempt communication with native speakers of the L2. Adults who learn a second language in an environment where L2 is dominant, by contrast, are likely exposed to more native phonetic models and that is thought to improve their perception and production towards a more native-like status (Jia, Strange, Wo, Collado, & Guan, 2006). However, many adults arrive in the L2 environment having already received instruction in the L2, or having had exposure to native input.

There is a great deal of research on the cognitive, social, economic, and emotional factors involved in adult second language learning. However, there is still a gap between existing behavioral investigations and our understanding of the brain mechanisms underpinning these overt outcomes. Many questions remain unanswered, such as those related to the process of perceiving the sounds of a new language. For example: In the process of learning new speech sounds, how does the brain support the perceptual changes that occur in second language learners from hearing foreign-language speech sounds for the first time over to completing the development of new phonological representations? Why do people who have received formal language training, and perform proficiently in some second language tasks, still have trouble perceiving and producing certain foreign phonemes accurately?

Knowing that speech is rapid, unfolding in the order of milliseconds, it is fundamental to look at what is happening in the brain (where speech is decoded), in a very fine-grained temporal dimension. A neuroscientific approach that implements electrophysiological measures enables an examination, at the level of cortical brain responses, of the phonological representations of foreign vowels in adults. Electrophysiological methodologies, such as event related potentials (ERP), can provide evidence about the neural responses that are correlated with processing non-native phonemes in adulthood, and can provide indirect evidence about how listeners weigh L2 phonemic cues in real time, with millisecond precision.

Electroencephalography (EEG) is a method for indexing brain functions by recording the electrical activity generated by large populations of neurons as summed voltages at the scalp. It is a non-invasive procedure that can be applied repeatedly in individuals, and carries minimal risk. The high temporal resolution (millisecond precision) of the neurophysiological method makes it very suitable for examining rapid responses to auditory speech stimuli that occur during specific speech decoding tasks. Although its spatial resolution is not as precise as that provided by functional magnetic resonance imaging (fMRI) methods, it is possible to identify signal generators of specific components recorded through EEG (Luck, 2005). The ERP method is a means for examining the brain's time-locked responses to specific cognitive events (Handy, 2005). Event-related Potentials (ERPs) are derived from the continuously recorded EEG, by averaging together the time-locked or synchronized electrical activity with responses to multiple instances of a cognitive event, such as a sound being heard, or a word being recognized. This averaging process enhances the signal-to-noise ratio of recordings and removes non-phase-locked activations (such as eye blinks and heart beats) so that (in principle) only activation related to the event of interest is represented in the averaged data. ERP components are characterized by simultaneous multi-dimensional measures of polarity (negative or positive voltage deflection), amplitude, latency, and scalp distribution. Several ERP components have been identified as reflecting activation associated with particular cognitive processes. The mismatch negativity and the P300 are two of the most common ERPs implemented to study non-native speech processing. In the section below, I provide brief descriptions of these components, and discuss how they are typically elicited and interpreted.

The mismatch negativity (MMN) is a negative voltage deflection that has been shown to reflect unconscious processes of change-detection in auditory stimuli. This makes it appropriate for use as an index of central auditory stimulus representation (Näätänen, 1995). The MMN is elicited by any discriminable auditory stimuli. Typically, the listener is presented with a series of several and equal auditory stimuli (i.e., the standard stimuli) that, at some points of the series, are replaced by a number of different stimuli (i.e., the deviant stimuli). This is known as the 'oddball paradigm'.

Generators of the MMN component have been localized bilaterally to auditory and frontal cortex (Alho, 1995; Rinne et al., 2000). The negative-going voltage deflection usually peaks at 150 to 250 milliseconds (ms) after onset of a deviant stimulus (Näätänen, Paavilainen, Rinne & Alho, 2007). The magnitude of the deviance determines the peak amplitude (measured in microvolts μV) and peak latency (measured in milliseconds, ms). The more pronounced the deviance, the greater the amplitude peak (voltage difference between a pre-stimulus baseline and the largest positive-going peak of the ERP) and the shorter the peak latency (time from stimulus onset to the point of maximum positive amplitude within the time window) (Näätänen, 1992; Pakarinen, Takegata, Rinne, Huotilainen, & Näätänen, 2007). The MMN is usually presented as a peak in the *difference wave.* The difference wave is obtained after subtracting the response elicited by standards, from that elicited by the deviants.

The P300 component is a positive voltage deflection, which can be elicited through the oddball paradigm only if the listener is actively engaged in the task of detecting the deviant stimuli. This component is considered an index of conscious cognitive processing because it is observed during conscious discrimination tasks. Discrimination generates a relatively large, positive-going waveform with a modal latency of about 300-800 ms (Toscano, McMurray, Dennhardt & Luck, 2010) when elicited using auditory stimuli in adults (Polich & Kok, 1995). Similar to the MMN, the P300 can be elicited using the oddball paradigm. In this case, elicitation of the P300 using the oddball paradigm requires the listener's attention to each stimulus in the series of standard and deviant stimuli. Attentional resources are allocated towards deviant stimuli more than towards standard stimuli, resulting in amplitude changes in the P300 time window. The amplitude of the P300 is larger over parietal sites, varies with the improbability of the deviants, and is also larger when the subject dedicates more effort to the task. P300 amplitudes are smaller when the subject is hesitant with respect to whether a given stimulus is a standard or a deviant (Polich, 2007; Polich & Kok, 1995). P300 latency is shorter over frontal areas and longer around parietal ones.

This study seeks to utilize neurophysiological indices of adult bilingual brain organization at the level of the phonological system representation, as a means of identifying results of exposure to English as a second language. The study examines brain responses from sequential bilingual, L1-Spanish/L2-English listeners to AE vowel contrasts that are signaled by spectral and duration cues (e.g., /i/-/ɪ/). One question is whether this population perceives these "difficult" vowel contrasts, and if they do, how native-like their perception is compared to that of native English speakers. This investigation aims to provide some insights about possible underlying factors that prevent some sequential bilingual, L1-Spanish/L2-English listeners from becoming fully native-like perceivers of L2. In addition, the study aims to contribute to the cross-language research field by providing behavioral and neurophysiological indices of the perceptual mechanisms involved in acquiring the speech sound system of a second language in adulthood. Findings could be translated into pedagogically meaningful information which could allow curriculum designers and teachers to propose new ways to help adult second language learners develop their full learning potential from the very start of the L2 learning process.

This dissertation is organized as follows: Chapter 2 focuses on speech perception in adult L2 learners, and provides an analysis of two relevant models of L2 adult speech perception: the Speech Language Model (SLM), the Perceptual Assimilation Model (PAM & PAM-L2), and the Second Language Linguistic Perception Model (L2LP). It also reviews how L2 adult learners perceive non-native speech cues, presenting behavioral experimental evidence on L1-Spanish adults' performances in perceptual tasks involving such cues. Chapter 3 concentrates on how the brain processes speech cues in L1 and L2 listeners, reviewing the relevance of the specific event-related potential (ERP) technique that is implemented in the study. Chapter 4 describes the proposed study, including the research questions, hypothesis, and its methodology. Chapter 5 contains the results and the statistical analysis of the experimental data. Chapter 6 presents the discussion of the findings, the limitation of the study, as well as future directions.

2. Speech Perception In Adult L2 Learners

The phoneme is the basic unit of sound in a language, and is the minimal contrastive unit in phonology. Each language uses only a subset of phonemes, far fewer than the number of sounds that can be produced by the human vocal tract. Infants undergo an auditory perceptual reorganization during their first year of life (Eimas, 1978), going from a general ability to discriminate virtually any pair of phonemes (language universal) to an enhanced sensitivity, through a kind of sensory narrowing, that permits them discriminate their the sounds of their native language (language specific). At the same time as this perceptual enhancement is occurring, the ability to discriminate the phonemes of non-native sound systems decreases (Werker & Tees, 1984). This perceptual reorganization - consisting of specification of the native speech sounds, and becoming highly sensitive to specific native speech cues - results in the formation of a native phonological system that will influence the acquisition of L2 speech sounds later in life. The auditory perceptual system remains flexible enough to allow the incorporation of new non-native sound categories (Flege et al, 1995; Kuhl, 2000). However, this flexibility differs between children and adults. There is evidence showing that young children can achieve native-like perception and production of non-native sounds (Williams, 1979; Flege, Yeni-Komshian, & Liu, 1999), while adult learners may not reach native-like levels at least to perceive and produce some L2 sounds (Werker & Tees, 1984; Flege et al., 1995; Best, 1995; Strange, 1995).

There is variability in adult perception of non-native speech sounds related to multiple factors such as intrinsic characteristics of the speech signal, the features of specific languages, the age of the listeners, length of exposure to the non-native sounds, and use of L1 and L2 on a daily basis for example. Even with so much variation, there is agreement on the fact that for most adults it is difficult to perceive and produce some L2 phoneme contrasts that are not present in their L1 (Flege et al., 1995; Strange, 1995).

This chapter presents an overview of current thinking among scholars who seek to explain non-native speech perception in adults. First, three of the currently proposed models of speech perception from the current experimental literature are described and compared - namely, the Perceptual Assimilation Model (PAM: Best 1995), PAM-L2 (Best & Tyler, 2007), and the Speech Learning Model (SLM: Flege, 1995). Then, a summary is provided of some of the most relevant factors that have been identified as influencing the perception of L2 sounds in adults. Next, I present a comparison between English and Spanish vowel inventories, focusing on the main perceptual challenges faced by adult Spanish listeners when learning AE vowels. Finally, this chapter reviews some of the most relevant behavioral and neurophysiological studies that have provided insight into the cognitive processes and neural indices of L2 vowel perception in adults.

2.1 Three Models of Speech Perception

A number of researchers in the areas of speech perception and linguistics have sought to describe and predict how adult listeners (naïve and experienced) progress through the process of learning non-native speech sounds. Such efforts have led to the development of models that offer insights into specific processes and outcomes depending on how listeners perceive non-native sounds. In this section, three of those models are described and compared, by examining their similarities and differences, and by showing how the models have been guiding researchers to test hypothesis about non-native speech perception.

The most widely accepted speech perception models are the Perceptual Assimilation Model (Best, 1995), with its extension, the Perceptual Assimilation Model L2 (Best & Tyler, 2007), the Speech Learning Model (Flege, 1995), and the Second Language Linguistic Perception (Escudero, 2005). In general, these models propose that learning L2 speech sounds is influenced by the L1 sound system. All models posit that younger learners have a perceptual advantage over older ones, and that relative similarities and dissimilarities between L1 and L2 determine listeners' perceptions of L2 sounds. The models differ with respect to the nature of the object of speech perception (gestures-articulatory movements versus acoustic-phonetic information). Hypotheses derived from these models have been tested in a number of studies and have been shown to provide insights into various factors involved in L2 adult speech perception and production.

2.1.1. The Perceptual Assimilation Model (PAM). The Perceptual Assimilation model proposed by Best (1995) is based on the Ecological Theory of Perception/ Direct Realist View (Gibson, 1966, 1979; Fowler, 1986) and the Articulatory Phonology framework (Browman & Goldstein, 1989). The Direct Realist view holds that perceivers obtain temporal and spatial information from the environment through integrated perceptual systems, without innate knowledge or acquired mental associations. This means that listeners perceive cues about the position and movement of the articulators in the oral cavity as the speaker is producing the sounds. Regarding first language acquisition, the model assumes that what is being perceived is the acoustic reflexes of specific articulatory gestures produced by the speakers' vocal tracts. Infants extract information from tongue, lips, jaw position and movement during speech production. The perceptual mechanisms involved are the integrated perceptual systems including vision, hearing, and proprioception. Cognitive mechanisms or mental representations are not needed because the environmental stimulation is rich in direct information for perceivers to engage their perceptual systems in active exploration. Infants perceive articulatory gestures with no linguistic content, and later on, detect high-level invariant features (relevant articulatory information) that signal meaning in their native language. There is an increased efficiency in detection of critical differences in native speech sounds that listeners continue to refine throughout life. However, the model does not address specific mechanisms for detection and extraction of the relevant speech features.

The PAM approach to adult non-native speech perception describes how adult naïve listeners (listeners who have no experience with the target language) assimilate non-native sound contrasts and predicts discrimination accuracy based on this assimilation. Following the Direct Realist View, the PAM proposes that adults do not have mental representations or mental mappings for sound perception. The basic premise of this model is that non-native speech sounds tend to be perceived according to their similarities to and dissimilarities from native-language speech sounds. What adults directly perceive are complex articulatory events (e.g. position and movements of the tongue and lips) from which they detect invariants. The native phonological system is the perceptual mechanism that is also in place for perceiving non-native language features, since the new sounds are compared and classified according to the native language categories. In this perceptual process non-native listeners assimilate new sounds to the native sounds that are perceived as most similar. Table 1 shows the possible assimilation patterns predicted by PAM to be available for non-native listeners' assimilations of new language sounds, and the predictions of this model for discrimination accuracy. These patterns are also described below.

In a Two-Category (TC) assimilation pattern, each of a pair of two non-native sounds is assimilated to two different native phoneme categories. Discrimination of both L2 sounds is expected to be excellent for these contrasts. For instance, Spanish listeners may assimilate English vowels /u/-/i/ to their congener Spanish categories /u/-/i/. In a Single-Category (SC) assimilation pattern, two L2 sounds are equally assimilable to one native phoneme category. In consequence, poor discrimination is expected. For example, Spanish listeners perceiving English vowel contrast /i/-/ɪ/ (as in *sheep-ship*) may assimilate both members of the non-native English contrast to their native Spanish vowel /i/ (Escudero & Boersma, 2004). In the category goodness difference (CG) pattern, two non-native sounds are assimilated to the same native phoneme category but differ in the goodness of fit to the category. This means that one of the non-native sounds is a poor exemplar of the native category. For this pattern, the discrimination is expected to range from good to moderate. For example, Greek listeners may assimilate the Southern British English contrast /æ/-/ʌ/ (as in *cap-cup*) to the Greek /a/ category, but perceiving one of the English vowels /æ/ or /ʌ/ as more similar to the Greek /a/ than the other (Lengeris, 2009). In the uncategorized-categorized (UC) pattern, only one of the two non-native sounds can be assimilated to a native category. Very good discrimination is expected. Aoyama (2003) found that Japanese listeners assimilate English /m/ to their Japanese /m/ category, and do not categorize the English /ŋ/ in final-syllable position (as in *thing*). The Uncategorizable (UU) pattern is evident when neither non-native sounds are assimilated to a native sound category, even though they are heard as speech. Aoyama (2003) also reported that Japanese listeners do not categorize syllable-final English nasal contrasts /n/-/ŋ/. Finally, in the Non-assimilated contrast (NA) pattern, both non-native speech sounds are heard as non-speech sounds, as in the case of English speakers who do not hear Zulu clicks as speech sounds (Best, McRoberts & Sithole, 1988). Even though PAM does not suggest a learning process or mechanism to extract high-order invariants from the articulatory features in either L1 or L2, it assumes that exposure to L2 may contribute to a reorganization of the assimilation patterns in which listeners split their L1 categories to include the new speech sounds.

Table 1

Non-native perceptual assimilation patterns and predicted discrimination accuracy, according to the PAM Model (Best, 1995)

Assimilation Patterns	Single-Category (SC)	Two-Category (TC)	Category Goodness (CG)	Uncategorized-Categorized (UC)	Uncategorizable (UU)	Nonassimilable (NA)
	2 non-native sounds = 1 L1 category	2 non-native sounds = 2 L1 categories	2 non-native sounds = 1 L1 category, but 1 non-native sound is a poorer sample of the L1 category	1 non-native sound = No native category. 1 non-native sound = 1 L1 category	2 non-native sounds = no L1 category	2 non-native sounds = not heard as speech
Predicted Discrimination Accuracy	Poor discrimination accuracy	Excellent discrimination accuracy	Moderate to very good discrimination accuracy	Very good discrimination accuracy	Poor to very good discrimination accuracy	Good to very good discrimination accuracy

In line with the PAM assumptions, Best and Tyler (2007) proposed PAM-L2 (Best & Tyler, 2007) as an extension of the original PAM to account for how L2 learners perceive non-native sounds. The authors posit that there are significant differences between naïve listeners, L2 learners, and bilinguals because the quantity and quality of second language experience and exposure differ among these groups. The model assumes that L2 listeners extract relevant information from the speech signal at three levels: (1) low-level gestures (articulatory information); (2) phonetic (invariant gestural relationships); and (3) phonological representations (mental organization of the language system). PAM-L2 posits that differences between L1 and L2 at these three levels pose diverse perceptual difficulty for L2 listeners. PAM-L2 argues that speech sounds are learned at the phonetic and phonological levels, being the phonetic categories established before the phonological ones. In addition, the model suggests that in L2 development, establishing new categories will depend on the dissimilarity between the L2 contrasts. PAM-L2 proposes four different cases in which L2 learners perceive L2 speech sound contrasts (see table 2):

1. In the first case, when only one of the L2 contrast member is assimilated to a L1 phonological category, being perceived as a good exemplar of that category, no learning is likely to occur for the particular member. However, if one of the L2 speech sounds is perceived as a deviant exemplar of an L1 phonetic category, it is possible that the listener establish a new phonetic category for the deviant L2 speech sound.
2. In the second case (corresponding to the category-goodness (CG) from the original PAM model), occurs when both members of the L2 contrast are assimilated to one L1 phonological category, but one of them is perceived as more deviant from the L1 category than the other. It

is likely that a new speech sound category is formed at the phonetic and phonological levels for the most deviant L2 member, but not the similar one.

3. The third case (single-category (SC) assimilation pattern), is when both members of the L2 contrast are assimilated to one L1 phonological category, but as equally good or poor exemplars of that category. Similar to PAM prediction, assimilation of both L2 speech sounds to one L1 may initially present a perceptual challenge for the listener to discriminate between the L2 members of the contrast. PAM-L2 posits that perceptual differentiation depends on whether both L2 speech sounds are perceived as poor or good exemplars of the L1 category.
4. In the last case, there is not L1-L2 assimilation (corresponding to the uncategorized-uncategorized (UU) pattern in PAM), when the two sounds of a given contrast are not clearly mapped onto a particular L1 category. PAM-L2 predicts that the listener could form new phonological categories for one or two L2 speech sounds.

Table 2

L2 learners perceptual assimilation patterns and predicted learning outcomes, according to the PAM-L2 Model (Best & Tyler, 2007)

	First case	Second case Category Goodness (CG)	Third case Single-Category (SC)	Fourth case Uncategorizable (UU)
Assimilation Patterns	1 L2 Speech sound = good exemplar of a L1 category. 1 L2 speech sound = deviant exemplar of a L1 category.	2 non-native sounds = 1 L1 category, but 1 non-native sound is a deviant exemplar of the L1 category	2 L2 speech sounds = 1 L1 category	2 non-native sounds = no L1 category
Predicted formation of L2 categories	Unlikely formation of new L2 categories for the L2 speech sounds that are good exemplars of the L1 category.	Likely formation of a new speech sound category for the more deviant exemplar.	Formation of a new L2 speech sound category depends on whether L2 phones are perceived as good or poor exemplars of the L1 category	Likely formation of new L2 categories for 1 or both L2 speech sounds

The PAM- L2 proposal has been implemented in a number of studies providing a framework to delineate predictions on initial perceptual difficulty for specific non-native contrasts, and to predict L2 learnability outcomes. For example, several studies have tested PAM and PAM-L2 predictions, showing that the assimilation patterns of non-native and L2 sound contrasts in different languages can predict discrimination outcomes in naïve listeners (Best, McRoberts, Sithole, 1988; Polka, 1991; Best, McRoberts & Goodell, 2001; Strange, Hisagi, Akahane-Yamada & Kubo, 2011), and in L2 learners (Guion, Flege, Akahane-Yamada & Pruitt, 2000).

One of the early studies that evaluated a PAM-based prediction was conducted by Best, McRoberts & Sithole (1988). The study investigated English-speaking infants' and adults' perceptions of apico-dental, palato-alveolar, and lateral-alveolar click consonants from Zulu. The discrimination of the selected contrasts was predicted to be good for both infants and adults, even without experience with the language, because English does not have click sounds in its inventory and clicks do not correlate to any speech sound in English. Therefore monolingual American English listeners should perceive the Zulu clicks as non-speech sounds that could be related to familiar sounds (such as the non-linguistic noises used for *giddy up*, *tick-tock* and *tsk-tsk* (as disapproval)). Best et al. (1988) found that both infants and adults discriminated the Zulu contrasts as non-assimilable (NA) patterns, meaning that what they heard was not similar to any speech sound in their native language.

Guion, Flege, Akahane-Yamada & Pruitt (2000) examined whether the PAM could be used to predict learning of English L2 sounds by three groups of Japanese L1 speakers that differed their English proficiency. Participants were asked to identify AE /b/, /v/, /w/, /θ/, /t/, /s/,

/ɹ/, /l/ and Japanese consonants /b/, /ɰ/, /t/, /d/, /s/, /ɾ/, /h/ followed by /a/, in Japanese- Japanese, Japanese-English, and English-English contrasts. The contrasts were presented in a carrier sentence ("This is ___"). In line with PAM's predictions, learners in their study did not discriminate an uncategorized-uncategorized (UU) contrast /ɹ /-/l/ and did not show improvement at higher levels of proficiency.

However, discrimination results of two uncategorized-categorized (UC) contrasts were different. Discrimination was good for the /ɹ/-/w/ contrast and improvement was also better as learners reached higher levels of proficiency. In comparison, discrimination of the /s/-/θ/ contrast was poor for all participants. The authors attributed differences in discrimination to how the L2 sounds were mapped to the L1. The /ɹ/ and /w/ were mapped to different sound categories in the L1, while the English /s/ was categorized as the Japanese /s/ and the English /θ/ fell between the two Japanese categories /s/ and /ɸ/. This result led the authors to suggest a revision to the PAM, to include a poor discrimination prediction of this contrast type (categorized-uncategorized) when the uncategorized sound is close in phonological space to the categorized sound.

Strange, Bohn, Nishi, & Trent (2005) conducted a study to examine the perceptual assimilation of North German (NG) vowels by American English (AE) listeners to determine whether acoustic and perceptual similarity varied as a function of the phonetic context. The stimuli consisted of 7 long NG vowels /i/, /e/, /ɑ/, /o/, /y/, /ø,/ and 7 short vowels /ɛ/, /ɪ/, /a/, /œ/, /ʏ/, /ʊ/ in the syllables bVp, bVt, dVt, gVk, gVt, presented in the carrier sentence "Ich habe CVC gesagt" ("I said CVC"). As a contrast condition, all AE vowels were embedded in similar CVC syllables (bVb, bVp, dVd, dVt, gVg, gVk) in the carrier sentence "I say the CVC on tape."

Forty-eight monolingual AE speakers with some background in phonetics and International Phonetic Alphabet (IPA) transcription participated in the study. Participants were asked to listen to the NG sentences and select an IPA symbol that most closely represented the vowel they had heard. They were presented with the same vowel a second time and were asked to rate the vowel on a 7-point scale ("very foreign sounding" versus "very English sounding"). Results showed that the AE listeners perceptually assimilated front rounded NG vowels to back rounded AE vowels, indicating that those NG vowels were good exemplars of AE categories and a single-category (SC) assimilation pattern. After a context-independent discriminant analysis, it was concluded that these perceptual assimilation patterns were predicted by spectral similarities (similar vocal tract and articulators' position and movement) among the vowels, and were not influenced by specific consonant context. The authors concluded that listeners develop a context-independent strategy (not influenced by the consonants surrounding the vowels) to evaluate cross-language vowel similarities. Observed PAM assimilation patterns allowed the authors to conclude that AE-speakers learners of German will have difficulty with some of the NG vowel contrasts that are also phonologically distinctive in AE such as [e/i] and [u/o]. In contrast, front and mid-low short vowels [ɪ/ɛ] or [ɔ/o] should not pose any difficulty in perceptual discrimination.

More recently, Strange et al. (2011) explored whether perceptual similarity predicts discrimination of AE vowels. Participants were naïve Japanese listeners. Stimuli consisted of eight AE vowels, presented in a disyllable context /hVbə/. AE vowels /i/, /ɪ/, /ɛ/, /æ/, /ɑ/, /ʌ/, /ʊ/, /u/ are mainly spectrally differentiated with duration as a secondary identification feature, while Japanese vowels /i/, /e/, /a/, /o/, /ɯ/ are primarily distinguished by both spectral and durational (one-mora two-mora) features. It was predicted that Japanese listeners would assimilate short /ɪ/, /ɛ/, /ʌ/, /ʊ/, intermediate /i:/, /u:/, and long /e$^{\text{ɪ}}$/, /æ:/, /ɑ:/, /ɔ:/, /ɔ:/, /o$^{\text{ʊ}}$/ AE vowels according to their duration categories. In the discrimination task, participants were presented with trials of three sounds (AXB); for each trial, participants were asked to decide if sound “X” was the same as “A” or “B”. In the categorization task, participants were required to indicate how similar AE vowels are to Japanese ones on a nine-point Likert scale (1= foreign-like, 9= Japanese-like).

Results showed that Japanese listeners were able to categorize AE vowels that were presented in contrasting pairs, based on the PAM predictions i:/ɪ, ɛ/æ:, ʌ/ɑ:, u:/ʊ, ɪ/ɛ, ʊ/ʌ, æ:/ɑ:, ɛ/ʌ, ɪ/ʊ. Four contrasts yielded accurate discrimination in a two-category pattern ʊ/ʌ, ɪ/ʊ, i:/ɪ, ɛ/ʌ, but performance on other contrasts was less accurate ɪ/ɛ, ʌ/ɑ:, ɛ/æ: with categorized-uncategorized and single-category patterns on spectral features (high/ front-back tongue position). The least accurately discriminated contrasts were æ:/ɑ:, u:/ʊ with single-category patterns and small category-goodness differences. These findings indicated that differences in duration ratios between AE vowels (1.5 to 1) and Japanese vowels (2 to 1) made it difficult for Japanese listeners to discriminate some AE vowel contrasts, especially those with intermediate and long duration characteristics. In addition, authors indicated that results from this study differ from previous ones (cf. Strange, Yamada, Kubo, Trent, Nishi & Jenkins, 1998; Bundgaard-Nielsen et al., 2011) and concluded that further research is necessary to account for dialectal variation, number of speakers (n = one for this study), and for the fact that perceptual assimilation patterns are different when the vowel contrasts are presented in citation form (vowel contrast embedded in a consonant context such as *bVb*) or in a sentence form (*I say bVb now*) (Strange et al., 1998; Strange, Yamada, Kubo, Trent & Nishi 2001).

These studies exemplify how the PAM model has been implemented to predict discrimination outcomes in a variety of experimental settings and language backgrounds. They also show that implementing the model requires control of a number of variables for the discrimination predictions to be observed. For instance, Guion et al. (2000) suggest the incorporation of an additional poor discrimination prediction to the categorized-uncategorized assimilation pattern that, according to PAM, should result in good discrimination. The findings of their study revealed that in a categorized-uncategorized assimilation pattern (English contrast /s/-/θ/ for Japanese listeners), the uncategorized sound /θ/ is close to the categorized one /s/, and therefore the former becomes hard to discriminate for Japanese listeners. According to Strange et al. (2011), another factor influencing the predicted discrimination of the PAM assimilation patterns is the form in which the contrasts are presented to the listeners, namely, citation or sentence forms.

Some questions remain unanswered with respect to the PAM model - for example, how listeners detect and extract articulatory/gestural invariant features in either L1 or L2, and where and how representations of those invariants are stored or become available for the experienced listener. Another question that PAM does not answer is how listeners cope with L1 and L2 invariants at the same time; neither does it address whether L2 influences the L1 phonological system during or after the learning process. Although PAM does posit that, in the process of learning an L2, learners can split their L1 categories and new categories can be established during L2 learning, it does not provide specific mechanisms for how the splitting of L1 categories occurs.

Another well-known model, the Speech Language Model (SLM) proposed by Flege (1995), shares some similarities with PAM and provides some insight into the relationship between speech perception and speech production in L2 adult learners. It also addresses the mutual influence of the L1 and L2 phonological systems.

2.1.2 The Speech Language Model (SLM). The Speech Language Model (SLM), formulated by Flege (1995), was primarily motivated by production errors observed in experienced adult L2 users. The goal of the model was to explain age-related constraints in native-like production of L2 vowels and consonants; the model focuses on bilingual adults' final speech perception and production performance, rather than on early learning stages, as PAM does. According to the SLM, L1 speech perception acquisition requires the adjustment of the perception system to the contrastive L1 phonetic elements, and the native language phonetic system is the basis for non-native sound learning. The model states that phonetic systems are flexible and can be reorganized over the lifespan to accommodate new L2 sounds. For this model, what the listener perceives is acoustic-phonetic information from the speech signal. This information is stored in long-term memory as phonetic categories.

Flege (1995) addresses the production and perception of non-native sounds within an SLM framework, suggesting that in some cases perception difficulties are reflected in the inaccurate production of L2 vowels and consonants. He argues that when learning the sounds of L2, adult learners start with established L1 phonetic categories that may interfere with the discrimination of phonetic differences, either between two sounds from the L2, or between one sound from L1 and one from L2. A crucial assumption of the SLM model is that the probability of learning to perceive the differences between non-contrastive L1 and L2 sounds decreases as the age of the learner increases (Baker, Trofimovich, Flege, Mack, & Halter, 2008). That is, early learners are likely to produce L2 vowels more accurately than late learners, but their productions are not always native-like (Flege, 1995).

The SLM assumes that L1 and L2 phonetic categories exist in a common phonological space, and that bilinguals try to separate that space in order to maintain a contrast between L1 and L2 categories. L1 and L2 categories interact in a bidirectional fashion, influencing the L1 categories as the L2 categories are formed. This interaction has been called the "interaction hypothesis" (or IH; Flege, 1992; Walley & Flege, 1999). This interaction depends on the age at which L2 is learned, and the development of the L1 phonetic system at the time L2 learning starts. It is likely that the L1 and L2 always influence each other to some degree in all learners (Baker & Trofimovich, 2005).

In classifying L2 sounds, the main SLM postulate is that it is easier for an L2 listener to perceive and produce speech sounds that are dissimilar from native phonemes (Flege, 1995). This model posits that adult bilinguals classify L2 sounds as identical, similar or new, compared to their phonetic categories. Each of these classifications has different implications in perception and production for adult bilinguals. According to Flege (1995), a new phonetic category for an L2 sound is created when bilinguals can auditorily differentiate the L2 sound from the closest L1 sound and from similar L2 sounds. It requires the listener (1) to detect the common phonetic features on multiple exemplars of a sound (such as AE vowel /u/ as produced by different AE speakers) while ignoring irrelevant acoustic differences between them, and (2) to differentiate multiple exemplars of a category from exemplars of other categories, ignoring common irrelevant features (Flege, 1995).

When the L2 sound is perceived as identical to the L1 sound, the SLM suggests a perceptual equivalence or process of assimilation (Flege, 2003). In this case, the formation of a new category is unlikely, and the production of that L2 sound is inaccurate. This classification is compatible with PAM's single-category assimilation pattern. SLM also suggests that when L1 and L2 sounds are perceived as identical they affect both languages' phonological systems, and production of both sounds becomes alike. When an L2 sound is classified as similar to a L1 sound, the formation of a new category depends on the perceived dissimilarity or dissimilation between the L1 and L2 sounds (see table 2). It is easier for the adult L2 learner to form a new category when the L2 sound is perceived as similar, but enough dissimilarity from L1 can be detected. This classification could be compared to the PAM's category-goodness assimilation pattern. In contrast, when an L2 sound is perceived as new, without a similar L1 counterpart, L2 adult learners create a new category. This could be seen as similar to the PAM uncategorized assimilation pattern.

Table 3.

SLM classification of L2 sounds (Flege, 1995)

	Identical L1-L2	**Similar L1-L2**	**New L2**
Perception/production L1 and L2	L2 sound Perception = Inaccurate L1/ L2 sound production = resemble one another	Depends on the degree of perceived phonetic dissimilarity between L1 and L2 sounds.	L2 sound = no L1 counterpart
New category formation	None	Depends on dissimilarities between L1 and L2 sounds	New category

A number of studies testing SLM principles have been implemented in a variety of language contexts, ages, and exposure environments. For example, Flege, MacKay & Meador (1999) examined the perception and production of English vowels by native L1-Italian/ L2-English bilinguals who differed in their age of arrival (AOA) to Canada (age 7-early, age 14-mid and age 19-late) and reported to use Italian at least 31% of the time. There was a fourth group of Italians (early-low) who arrived early, but used Italian less than 8% of the time. A control group of native English speakers was also included. The authors predicted that Italians who arrived early would have more accurate perceptions and more intelligible productions of the Canadian English vowels than those who arrived late. In the perception experiment, Italian listeners were required to listen to English and Italian vowels embedded in English words and Italian non-words that were arranged in 3 different contrasts (English-English, Italian-Italian, and English-Italian). The vowel contrasts were presented in triads, and participants were required to press one of three buttons to indicate if one of the vowels presented was different from the others. They also could press button 4 to indicate that they heard three instances of the same vowel. Results supported the predicted effect of AOA in native Italian speakers of English. The early group showed the most accurate perceptions, followed by the mid and then the late groups. There were not significant differences related to amount of L1 use between the early and the early-low groups, and their English vowel discrimination performance was not different from the native English control group. The authors suggested that unlike the late and mid groups, the early groups had formed long-term memory representations for the L2 vowels.

In the production experiment, Italian bilinguals were asked to read the English vowels /i/, /ɪ/, /e^{i}/, /ɛ/, /æ/, /o/, /ʌ/, /ɒ/, /ʊ/, /u/ embedded in words. Native Canadian English listeners (from

Ottawa, Canada) listened to the production to determine whether the target vowels were heard as intended. The late group (who arrived in Canada at 19 years of age) was rated as less accurate in terms of their vowel productions than the mid and early groups. Therefore, results of the production experiment supported the prediction that Italians who arrived in Canada at an early age (7 years) were able to produce English vowels in a native-like fashion. Results from this study provided evidence on the relationships between age and speech perception and production, showing that early learners may form L2 sound categories and produce L2 sounds with more accuracy.

Jia, Strange, Wo, Collado, & Guan (2006) conducted a study to examine how age-related differences, a crucial factor in SLM, interact with amount of exposure to the L2. They tested perception and production of AE vowels in three groups of native Mandarin speakers with different amounts of exposure to native English. The first group included Mandarin-speaking children, adolescents and young adults (7 - 20 years of age) who had lived in Beijing all their life and did not have exposure to native English input, but were involved in English instruction through school. Mandarin speakers who immigrated to New York City between 7 and 44 years of age formed the second group, which was divided in two sub groups according to length of residency: (1) past arrivals, who had lived in the U.S. for between 3 and 5 years, and (2) recent arrivals, who had been in the U.S. for two years or less.

The experimental stimuli included AE vowels /i/, /ɪ/, /eɪ/, /ɛ/, /æ/, /ʌ/, /ɑ/, /u/ that were selected based on their relationships to Chinese vowels. Only /i/ and /u/ have similar counterparts in Mandarin, while the AE vowels /ɪ/, /eɪ/, /ɑ/, are similar to contextual variants of Mandarin sounds, and AE vowels /ɛ/- /æ/ have no similar phonetic counterpart in Mandarin. Based on PAM and SLM, it was predicted that AE vowels that have a close phonetic distance /ɛ- æ/ and /ɑ- ʌ/ would be difficult to perceive and produce as they would be assimilated to a single Mandarin category. The AE contrasts /i-ɪ/, /i- eᴵ/, /æ-ɑ/ would be assimilated to Mandarin vowels, following a category-goodness type assimilation, with a relative difficulty depending on the perceived dissimilarities. The AE contrast /u/-/ɑ/ was predicted to be easy to perceive, as the phonetic distance between both vowels is large. The vowels were embedded in a disyllable word /dVpə/ and spoken by a female native English speaker. Participants were asked to listen to trials of three disyllable words and decide whether word 2 was the same as word 1 or word 3. In the production task, participants were instructed to listen and imitate each of the eight disyllables used in the perception task /dVpə/.

Results of the perception task showed that the three groups (China residents, recent U.S. immigrants, and longer-term U.S. immigrants) all performed above chance on the vowel discrimination task. However, the China residents' group obtained a lower discrimination score compared to the two groups in the U.S. These two groups (recent and past arrivals) did not differ from each other. The analysis also showed that for the group in China, older listeners with longer English-language instruction performed better in the discrimination task than younger participants who had received less English instruction. A partial correlation analysis indicated that in absence of the English instruction factor, there was a significant correlation between age and total discrimination accuracy. However, when removing the age factor, the relation between English instruction and discrimination performance was not significant. The opposite pattern was observed in the U.S. groups, showing that the recent arrival group performed significantly better on the task in general and on the most difficult contrasts /ɛ-æ/, /ɑ-ʌ/ than the past arrivals. Within the recent arrivals group, there were non-significant age differences, and only those who had spoken more English in social interactions tended to have a better overall performance. In the past arrival group, younger listeners performed significantly better than older ones. Moreover, those listeners whose mothers spoke English performed better than the rest of the group.

Results of the production task indicated that participants in China showed significantly lower production accuracy compared to both recent and past arrivals in the U.S., who did not differ in their overall production accuracy. There were significant differences between groups for specific vowels. For the three groups, the two easiest vowels were /u/ and /i/, while the most difficult ones were /æ/ and /ʌ/. The age factor influenced mostly the group of recent arrivals. The China residents' group and the longer-term immigrant group did not show an association between age and production accuracy.

Taken together, the results from perception and production tasks in this study provide more evidence on the relevance of the SLM model, indicating that as age increases accuracy in L2 perception and production decreases. It also suggests that L2 users who learn the language in an immersion setting perceive and produce L2 vowels more accurately than those who learn in a non-native classroom environment. The role of age was different in the three groups of listeners. In the Chinese non-naturalistic learners' group, older learners were at an advantage over young learners in perception and production. In contrast, age did not influence perception and production of recent arrivals to an English-dominant environment, while there was a younger-learner advantage in the past arrivals group.

A more recent study investigated how PAM and SLM principles could be applied to L2 learning in a non-naturalistic setting. Fabra & Romero (2012) tested perception and production of AE vowels of 3 groups of native Catalan speakers – mean age 22 – who had been exposed to the same amount of English through formal instruction (American English phonology class) in their own country. Catalan participants were assigned to one of three groups according to their proficiency, as measured by their scores on the pronunciation test in the phonology class (proficient, mid-proficient, and low-proficient). A perception test examined whether PAM assimilation patterns would predict perceptual difficulty for learners in non-naturalistic settings, and whether according to SLM postulates, native Catalan-speaking learners of English could establish new phonetic categories between L2 sounds and the closest L1 sounds.

In the two separate discrimination tasks, listeners at different levels of L2 English proficiency (high, mid and low) were presented with seven Catalan-English (C-E) contrasts /a-æ/, /a-ɑ/, /ɛ-ɛ/, /a-ʌ/, /i-i/, /i-ɪ/, /u-u/[1], and four AE vowel contrasts /i-ɪ/, /ɛ-æ/, /ɑ-ʌ/, /ʊ-u/. Both C-E and AE vowel contrast pairs were presented in a syllable (sVt). The authors predicted that AE vowels /ɑ/, /ʌ/, and /æ/ would be assimilated to Catalan vowel /a/, a single-category pattern that is predicted to be difficult to discriminate. The perceptual task required the participants to listen to three syllables spoken by three different speakers and decide whether vowel 1 or vowel 3 was the same as vowel 2. Results showed an overall difficulty to discriminate contrasts involving English /ɛ/, /æ/, /ɑ/, /ʌ/ or /ɑ/ and Catalan /ɛ/ and /a/, while English vowels /i/, /ɪ/ and /u/ contrasting Catalan /i/, and /u/ had more accurate discrimination scores. As predicted from PAM the AE /i-ɪ/, /u-ʊ/ contrasts were more accurately discriminated than /ɛ-æ/ and /ɑ- ʌ/. There

[1] Vowel symbols and letters are written down as they appear in the original paper. The contrasts

was a generally better discrimination performance in the high and mid-proficiency groups compared to the low-proficiency group. However, non-significant differences between groups were observed for the discrimination of the AE contrasts. High discrimination scores for the contrasts /i-ɪ/ and /u-ʊ/ indicated that Catalan listeners could have developed new categories for the English vowels /ɪ/ and /ʊ/, at least in the sVt context. In addition, results showing that Catalan listeners had difficulty perceiving the C-E contrasts /a-æ/, /a-ɑ/, /ɛ-ɛ/, /a-ʌ/, /i-i/, /i-ɪ/, /u-u/ suggested that Catalan listeners' representations for at least the AE vowels /ɑ/, /ʌ/, /æ/ share the same perceptual space as Catalan /a/, while English /i/, /ɛ/, /u/ belong to the same category as Catalan /i/, /ɛ/, /u/ respectively. The authors concluded that as suggested by the SLM, Catalan listeners created new categories at least for some English vowels, while L1 and L2 categories share a common phonological space. Non-significant differences in discrimination scores between the three groups suggested that formal instruction after adolescence in a non-native environment only produces small changes in the perceptual system.

The production experiment aimed to examine whether English proficiency influenced Catalan participants' productions of the English vowels. This time, a comparison group of native English speakers was included. Vowel production was elicited using monosyllabic words, in a varied consonantal context, containing the English vowels. Results showed that Catalan learners produced English vowels /ɪ/, /æ/, /ʊ/, and /u/ with higher accuracy than vowels /ʌ/, /ɛ/, /ɑ/ and /i/. There were not significant differences in production among the proficiency groups, but there were significant differences between the Catalan learners and the native English speakers. In this task results indicated that high proficiency levels in Catalan learners of English did not predict more accurate production of vowels.

As a general finding, the study indicated that even though L2 speech perception learning is possible, a non-naturalistic L2 learning environment produces only small changes in the speech perceptual systems of learners. It provides support for the SLM predictions concerning new category formation, showing that Catalan learners formed new categories for some of the English vowels tested. In addition, the study indicates that PAM-L2 can predict discrimination accuracy outcomes when L2 is learned in a non-naturalistic environment.

In sum, these studies show how the SLM provides a framework to better understand perceptual and production performance of adult L2 learners. Interestingly, researchers in these and other studies (Flege, Bohn & Jang, 1997; Guion, Flege, Akahane-Yamada, Pruitt, 2000; Morrison, 2002) have implemented the principles of both PAM and SLM to hypothesize about possible perceptual and production outcomes for L2 learners, showing that specific aspects of both models can be integrated in complementary ways. Results from the three studies above support several crucial tenets of the SLM. First, as age increases, accuracy in perception and production decreases. In Flege et al. (1999), early native Italian bilinguals showed significantly better perceptual and production performance than those who arrived in the L2 environment at 19 years of age. Interestingly, the exposure-age study by Jia et al. (2006) found that younger learners performed better than older ones in the target-language environment, and all ages performed better than those receiving instruction in a non-naturalistic environment. In non-naturalistic environments, older learners seem to have an advantage over the younger ones because they have had longer formal instruction. In this study, the learning environment interacted with the age factor, resulting in significantly different perceptual and production outcomes.

In Fabro & Romero (2012), it was shown that Catalan listeners in a non-naturalistic learning environment, with three proficiency levels (proficient, mid-proficient, low-proficient), did not differ in their perception and production of AE vowels. However, Catalan listeners significantly differed from native English speakers in production accuracy for AE vowels. This study did not control for the age factor, but indicated that young adult learners (mean age = 22 years) in a non-native environment only experience small perceptual changes affecting their perception and production of L2 speech sounds, regardless of their proficiency levels.

The second tenet of the SLM that is supported by the reviewed studies is the notion that phonetic categories are flexible and new L2 sounds can be incorporated over the life span. As stated in PAM and SLM, the formation of new categories depends on the ability to detect gestural-articulatory (PAM) or acoustic/phonetic (SLM) similarities and differences between L1 and L2 sounds (Best, 1995; Flege, 1995). The three studies discussed above show that L2 learners modify their phonetic categories to some extent after learning a non-native language. In sum, the influences of age and learning environment suggest that early learners establish new categories for new L2 sounds more easily than older learners (Flege et al., 1999; Jia et al., 2006; Fabro & Romero, 2012), and that immersion learning environments are more beneficial for the speech perception/production learning than a non-naturalistic classroom instruction (Jia et al., 2006; Fabro & Romero, 2012).

Since the SLM focuses on the production and perception performance of late bilinguals who are at or near the end of their L2 learning process, it does not provide an account for how phonetic discrimination occurs in the L2 adult learner, nor does it provide a developmental proposal that could account for the unfolding of L2 speech perception learning processes.

A more recent model, proposed by Escudero (2005), offers a linguistic approach to the non-native speech perception process, that accounts for development of L2 perceptual processing explaining the kinds of mechanisms, and tasks that are in place during the learning process.

2.1.3. Second Language Linguistic Perception (L2LP). The Second Language Linguistic Perception model (Escudero, 2005) has been developed on the basis of the Linguistic Perception (LP) model (Boersma, Escudero & Hayes, 2003). This account of L1 speech perception views the speech perception as a process of encoding acoustic-phonetic properties obtained from the speech signal. In addition, the model argues that speech perception is language-specific and implies the formation of phonological representations that depend on how the acoustic properties of the speech signal are encoded in perceptual mappings (phonetic categories). Crucial for this model is that phonetic and phonological processing occur at different level, but sequentially during the speech perception process.

The L2LP model proposes that L2 learners acquire the sounds of the second language with a separate perceptual system that is, initially, a copy of the L1 sound system, and changes with the L2 experience. According to the model, establishing the optimal L1 and L2 (location and shape of category boundaries, and relative use of auditory dimensions) serves as the initial point to determine the type of tasks the listener needs to accomplish to achieve optimal L2 perception. There are four major aspects in the process of L2 speech sound learning:

(1) Initial State. In this stage, a copy of the L1 perceptual system is the basis for learning L2 speech sounds. This system is gradually changes as the product of the L2 experience. Listeners will use their L1 perceptual parameters to map the acoustic values of L2 into speech sound representations. Escudero argues that there are two types of parameter that are in placed in the L2 initial state: Already categorized dimensions and non-previously categorized dimensions. For instance, in the initial state of Spanish learners of English, the dimension F1 is an example of the already categorized dimension, because Spanish speakers already mapped their L1 categories along this dimension. However, the dimension *duration* is an example of a non-previously categorized dimension, because duration is not used to categorize speech sounds in Spanish. Therefore, in the initial state listeners will map L2 speech sounds into their L1 perceptual categories. According to Escudero, in this case, Spanish listeners will perceive Southern British English (SBE) speech sounds /i/ and /ɪ/ as Spanish speech sound /i/, since they will use only their already-categorized dimension F1 to map those two L2 speech sounds, and won't use the duration dimension as it is a non-previously categorized dimension for Spanish listeners.

(2) Learning task. The goal of the learning task is to achieve an optimal perception of the L2 sound system. In order to determine the learning tasks, it is necessary to establish the differences between the leaner's initial state and the optimal L2 perception at two levels namely, perceptual and representational. This process in turn, yields two types of tasks (perceptual and representational). A Perceptual task may consist of changing and/or creating mappings, meaning that the listeners may need to modify existing parameters along an

already-categorized dimension (e.g. Spanish speakers learners of SBE need to shift their F1 parameters to categorize English /i/ and /ɪ/), and/or create new categories on a non-previously categorized dimension (e.g. Spanish-speaking learners of SBE would need to shift their F1 parameters to categorize English /i/ and /ɪ/), and/or create new categories on a non-previously categorized dimension (e.g. Spanish-speaking learners of SBE would need to create categories in a new duration dimension). The representational task, on the other hand, consists of changing the number of L2 categories on a specific perceptual dimension. Based on the initial state, Spanish learners will categorize English /i/ and /ɪ/ as their Spanish /i/, and as a result will experience confusion in the lexical storage of words that contain these vowel sounds. Therefore, Spanish learners will have to learn perceive the difference between the vowels and create a new category /ɪ/ in order to differentiate the words *sheep and ship*.

(3) Development. The creation of new phonetic or phonological representation in L2 involves the same learning devices (UG) mechanisms (distributional learning, Gradual Learning Algorithm -GLA) used in acquisition of the L1 sounds. Distributional learning mechanisms are used towards the creation of categories along a dimension that has not been previously categorized in the listener L1 language (e.g. durational dimension in Spanish learners), while the GLA allows the adjustment of the L2 categories.

(4) End state. For the L2LP model, learners may or may not achieve optimal perception of L2 speech sounds, depending on the amount and quality of the L2 input they receive. Therefore, the model does not predict specific learning outcomes for L2 learners, but states that the richness of the L2 input is more important than the amount of cognitive reorganization that

may occur in adulthood. The optimal end of state is characterized by the possibility of using both L1 and L2 language systems to the same extent, assuming that L1 is not influenced by the L2 speech sound representations.

In order to establish predictions about L2 perceptual learning, L2LP resorts to PAM perceptual assimilation patterns, proposing that listeners who assimilate two L2 speech sounds to one L1 category will find it more challenging task to create new categories. This contrasts with listeners who assimilate two L2 speech sounds to two L1 categories, because in this case, the task consists of shifting the perceptual boundaries in the already existent categories.

Escudero (2000) examined the perception of Scottish /i/, /ɪ/ by 30 L1-Spanish/L2-English listeners who learned their L2 after the age of 12, in comparison to native Scottish English-speaking participants. Stimuli consisted of several sets of continua with six steps each. The first set of the continua presented the prototypical values of the vowels, while three other sets contained vowels with maintained duration values, but manipulated spectral information. For these sets, good perceptual performance is expected only for listeners who relied on the spectral information. In the first part of the experiment, participants discriminated (same/different task) prototypical values of the vowels presented in pairs. The second and third parts consisted of a contrast identification test where participants pressed a "ship" or "sheep" image on a button to indicate which vowel (/ɪ/-/i/) was being presented.

Results indicated that 21 out of 30 L2 subjects demonstrated a native English-like vowel perception. However, 13 out of those 21 subjects used a cue weighting that differed from the one the English speakers had. One of the groups did not distinguish English /i/ and /ɪ/. A second group distinguished the vowel contrasts using only durational cues. A a third group used both spectral and durational cues, but duration had a higher weighting. And finally, the last group used both spectral and durational cues relying more on spectral cues just like English native speakers. The author suggested, therefore, that L2 learners could form phonetic categories by using L2 cues in a non-native way.

More recently, Escudero & Chládková (2010) compared the perception of two varieties of English, (American and Southern British) vowels by Peruvian Spanish listeners. They incorporated PAM postulates to the Second Language Linguistic Perception model (L2LP) proposed by Escudero in 2005. In the study, 40 monolinguals were asked to identify the 205 incoming stimuli (10 repetitions of 18 English vowels and 5 repetitions of the Spanish ones) with one of the responses (Spanish vowels) orthographically presented on a computer screen. The main finding in this study was that Spanish listeners perceived both varieties of English differently. In general, English vowels were assimilated to the Spanish vowel that had the closest F1 and F2 values. Therefore, AE and SSBE vowels with relatively high F2 values (/i/, /ɛ/) were assimilated to Spanish vowels with high F2 values (/i/, /e/). Similarly, non-native vowels with low F2 values (/ɔ/, /u/) were perceived as Spanish vowels (/o/, /u/). However, when analyzing F1 values in the identification responses, there was not a clear correspondence between high and low formant values for the SSBE vowels. Authors suggested that Spanish perceptual boundary between front and back vowels is different from that of SSBE, and as a consequence, many SSBE front vowels were classified as Spanish back vowels. In addition, various discrimination patterns were observed during the analysis: regarding the American English (AE) vowels, Spanish listeners labeled /ɪ, ɛ, æ/ as the Spanish /e/, reflecting a Single Category discrimination pattern. The AE contrast /ʊ/-/ɔ/ was perceived as the single Spanish vowel /o/. In terms of the SSBE, they identified the same three vowels /ɪ, ɛ /as Spanish /e/, and /æ/ was perceived as the Spanish /a/ (A multiple category assimilation, Boersma & Escudero, 2002) AE vowel /ɑ/ was assimilated to Spanish vowel /a/ above 70% of the time, while AE vowel /ʌ/ was assimilated to Spanish vowels /a/ and /o/ between 30% and 70% of the time. The SSBE /ɑ/ was partially assimilated to the Spanish /o/ and the SSBE pair /æ/-/ʌ/was assimilated to the Spanish /a/. Along the same lines, the SSBE vowel contrast /u/-/ʊ/ was identified as Spanish /u/.

Results from this study allowed the researchers to predict future L2 development outcomes for the Spanish listeners. Those predictions are based on specific English variety and specific performance on the assimilation of each contrast. For example, a Spanish learner of AE will need to split their Spanish /e/ category to incorporate English /ɪ, ɛ, æ/, while in the case of SSBE, the listener will find it more difficult to learn the /ɛ, æ, ɑ/ contrast.

Escudero and Boersma (2004), suggested that Spanish learners of two different varieties of English (Southern British and Scottish) go through different paths to perceive sounds from L2. In addition, other studies (Morrison, 2009 and Mayr & Escudero, 2010) have proposed that the perceptual learning is learner-dependent because each learner begins his/her perceptual learning with different perceptual backgrounds and experiences.

This model, in agreement with PAM & PAM-L2 and SLM postulates that listeners perceive non-native speech sounds based on their native language system, and that non-native perception predicts of future L2 development. The next section presents a direct comparison between the theoretical approaches of SLM, PAM, PAM-L2 and L2LP.

2.1.3. Comparing SLM and PAM & PAM-L2, and L2LP. There are similarities and differences in the assumptions underlying SLM and PAM& PAM-L2 models. The similarities between SLM and PAM are most evident in the shared recognition that perceptual learning processes continue to be in place through the lifespan and are actively used to perceive non-native speech sounds. Furthermore, both models consider that, native language differences, non-contrastive phonetic similarities, and dissimilarities between L1 and L2, determine how listeners perceive L2 speech sounds. In addition, the two models indicate that L2 learning is greatly influenced by the native language sound system. The models also agree on the existence of the L1 and L2 phonological categories in a common phonological space. The L2P2 incorporates relevant assumptions from SLM and PAM & PAM-L2, adding a more detailed description, explanation and prediction of perceptual processing underlying non-native speech sounds assimilation, and perceptual learning processing.

The main differences between the models reside in a number of key questions:

(1) What is the primary information that listeners extract from the speech signal? For PAM & PAM-L2, that primary information is articulatory gestures, while for SLM is the acoustic-phonetic information. This difference between SLM and PAM might be at least partially reconciled if we consider that acoustic phonetic cues (SLM) are embedded in the

articulatory gestures (PAM) that provide the listener with the necessary information to distinguish speech sounds.

(2) What level of abstraction is involved in the perceptual processing? PAM & PAM-L2 as well as L2LP assume that both phonological and phonetic levels interact in the acquisition of L2 sounds, while SLM posits that perceptual changes occur only at the phonetic level when new non-native categories are formed.

(3) What are the effects of listener age? While PAM & PAM-L2 does not directly address the effect of the age on the ability to assimilate non-native sounds, SLM makes the argument that perceptual discrimination of non-native speech contrasts is dependent on the age of acquisition of the foreign language, stating that older learners have more difficulties perceiving and producing the L2 sounds. L2LP agrees that age is a crucial factor that influence perception of non-native speech sounds, but argues that rich L2 input can overweight the effect of less plasticity in the adult brain.

(4) What is the target learning stage? PAM predicts the discriminability of non-native/ L2 sounds during early stages of L2 acquisition. Unlike PAM, PAM-L2 proposes assimilation patters for L2 learners, and SLM predicts discriminability of L2 sound in experienced learners. L2LP predicts perceptual learning outcomes from initial to advanced L2 learners.

Both PAM (and its extension PAM-L2) and SLM postulates are relevant for the proposed study. In terms of the population of interest, SLM focuses on adult bilinguals, (the population of interest of the study), and on the influence of experience and exposure in the perception of L2 contrasts. The latter points are also factors to consider in this study, since the participants are adults with different length of residence in the English speaking country (USA) and exposure to the L2 sounds. The predicted discrimination-identification outcomes of the specific AE vowel contrasts of interest, /ɑ-ʌ/, have been established based on the assimilation patterns proposed by PAM-L2.

However, it is clear from studies like those conducted by Strange et al. (e.g., 2005 and 2011) that various factors – apart from the acoustic information carried by the speech signal – influence the perception of non-native speech sound contrasts. In the next sections, research investigating the influence of various such factors is discussed.

2.2 Factors influencing the perception of foreign vowels

The large body of research on speech perception implementing models such as PAM/ PAM-L2 and SLM among others has demonstrated that there is a lot of individual variation in the perception of L2-sounds. Several factors determine listeners' perceptions of unfamiliar (speech sounds that they have not heard before) or L2 sounds (these are familiar sounds because they are part of the individual's second language). This section reviews how the concept of critical/optimal learning periods (Lenneberg, 1967) has been adapted to second language learning, and how research about this critical period and the age of the learner (Flege, 1991; Flege, 1995, Flege et al., 1999) has indicated that younger learners have an advantage over older ones in second language learning, and especially in acquiring the L2 phonological system (Bongaerts, Planken, & Schils, 1995).

It should be mentioned that it is not only the intrinsic traits of listeners that are decisive when learning a new speech sound system. The amount and quality of the listeners' linguistic experience (Kuhl & Iverson, 1995, Iverson, Kuhl, Akahane-Yamada, Diesch, & Tohkura, 2003), and the length of residency and exposure to the target language (Flege et al., 1997) interact with the age factor, so that there is a positive correlation between early exposure and naturalistic experience with the non-native language, and accuracy in perception and production of L2 language sounds (Bohn & Flege, 1990; Flege et al., 1997; Fox, Flege & Munro, 1995; Flege, Mackay, & Meador, 1999). In addition, perception of the relevant information from the speech signal is also influenced by the characteristics of the native and the second languages, and by the interactions between the different phonological systems (Flege, Schirru & Mackay, 2003).

The native language influences how listeners attend to the relevant cues in L2 sounds, affecting the perceptual discrimination and identification of vowels and consonants (Escudero & Boersma, 2004; Fox et al., 1995; and Jusczyk, 1992). Equally important, there are factors related to experimental design that have been found to affect listeners' performance in laboratory testing. For example, the context in which phonemes - vowels or consonants - are presented to the listener (Frenck-Mestre, Meunier, Espesser, Daffner, & Holcomb, 2005; Bent, Kewly-Port, & Ferguson, 2010), and the specific experimental tasks (Beddor & Gottfried, 1995) are known to have effects.

2.2.1 The Critical Period Hypothesis and age of the learner. Differences between child and adult learners of an L2 have motivated a lot of research in the language acquisition area. One of the most controversial issues in language learning has been whether or not there is a critical period during which it is possible to learn or acquire a language to a native-like ultimate attainment. The concept of the Critical Period Hypothesis (CPH) was adopted from biological sciences and applied to the study of language, originally by Lenneberg (1967), who argued that there are maturational constraints operating on the process of native language acquisition, with the first few years of life being the crucial time during which a first language can be acquired to native-speaker proficiency. According to Lenneberg, the critical period within which a language can be acquired in a native-like fashion ends by 12 years of age, after which attempts to acquire new languages will not result in native-like performance. The existence of the critical period for acquiring language has been supported by the study of cases such as Genie (Curtis, 1977), a child who was socially and linguistically deprived during her first years of life. Despite all efforts to teach Genie English, she never achieved typical adult language competence.

In terms of second language learning, the CPH has been extended to argue that after a critical period, it is not possible to learn a second language to native-like proficiency (Johnson & Newport, 1989; Long, 1990). Different authors have proposed a critical period for learning a second language with slightly different cutoffs. For instance, Krashen (1982) proposed that learning a second language as a native speaker happens only before 5 or 6 years of age. A more lenient range is the one proposed by Long (1990) and Pakwoski (1990), who extended the critical period for native-like language acquisition to the age of 15.

The claim of a maturational constraint on ultimate attainment in L2 acquisition is supported by a number of studies. Johnson and Newport (1989) examined L2 speakers' knowledge of grammatical structures of English using grammatical and ungrammatical pairs of 138 English sentences in a grammaticality judgment task. The participants were 46 native speakers of Chinese and Korean, who were students or professors at an American university and had lived in the U.S. for at least 3 years. The results showed a strong correlation between the percentage of correct responses and age at the time of arrival (AOA) to the host country, indicating that the later the arrival, the lower the accuracy of grammaticality judgment in the L2. In addition, the authors observed a significant correlation between accuracy and AOA for those listeners who had arrived in the U.S. by age 15. This correlation was not significant when the listeners had arrived after 17 years of age.

The effect of age of arrival is even more evident in the L2 phonology (Bongaerts, Planken, & Schils, 1995). Tsukada, Birdsong, Bialystok, Mack, Sung, & Flege (2005) compared discrimination and production of English vowel contrasts by North Korean (NK) adults and children who differed with respect to their length of residence in North America (3 vs. 5 years), and age-matched native English speakers (NE). The same participants were tested two times with an interval of 1 year between testing times (T1, T2). In the perceptual experiments, they tested discrimination of English vowel contrasts (/i-ɪ/, /e^{i}-ɛ/, /ɛ-æ/, /ɑ-ʌ/) that were considered difficult for Koreans, and one easy contrast (/i-ɑ/). There were no significant differences between T1 and T2 testing times. NK children were more accurate in discrimination of English vowels than NK adults, but less accurate than NE children. These findings do not suggest that adults cannot learn to perceive and produce the sounds of a second language since other factors may be in play, as proposed by Flege (2009). Flege suggests that the AOA factor must be considered in conjunction with other variables such as L1 proficiency, use of L1 and L2, neurological and cognitive functioning, and kind of L2 input received, such as formal vs. informal exposure or instruction.

In light of these findings, Knudsen (1999, 2004) uses the term “sensitive period” instead of “critical period”. He states that the brain circuits are significantly modified by experience during a sensitive period, causing certain patterns of connectivity to become highly stable and preferred. Therefore, plasticity that occurs beyond the end of a sensitive period, which is present in many circuits, changes connectivity patterns only within the structural and functional constraints established during the sensitive period. In general, the concept of critical period has been used to indicate that after a certain period of time there is lack of plasticity, whereas the concept of sensitive period has been used in reference to a reduced plasticity that continues through lifespan. This seems to be the case for most learning, especially in terms of second language. As an individuals age, the ability to perceive and produce the sounds of the second language entails an effortful adaptation of the established phonological system to incorporate the new phonemes (Flege, 1995). In most cases, experimental results have shown that adults attain different degrees of learning at specific linguistic levels, indicating that learning is possible. However, it may still be the case that the mechanisms used for such learning are different at different stages.

Although there is agreement on the effect of age on the acquisition/learning of L2 (Dekeyser, Alfi-Shabtay, & Ravid, 2010; Weber-Fox & Neville, 1999; Bongaerts et al., 1995; Moyer, 1999), studies focusing on the phonological level have provided evidence as to the capacity of adult learners to attain native-like levels of performance. For example, in a series of experiments conducted by Bongaerts et al. (1995), Dutch-speaking learners of English who were first exposed to English after 12 years of age were asked to read aloud sentences in English containing phones that are similar to and different from Dutch phonemes. Native speakers of Dutch rated the readings and a significant number of the participants' readings were judged as having native-like English pronunciation. Other studies have replicated the initial finding in Bongaerts et al. (1995) with similar results (e.g. Bongaerts, Summeren, Planken, & Shils, 1997; Bongaerts, Mennen, and Van der Slik, 2000).

Some studies have attempted to determine how other learner-related factors, such as motivation, interact with age of acquisition in the attainment of native-like proficiency for pronunciation. Moyer (1999) aimed to understand the role of motivation and instructional practices in L2 phonology learning. In her study, Moyer examined German pronunciation ratings of 24 Graduate students in German at the University of Texas at Austin, who were employed as teachers of the first four semesters in the German program. Native German speakers were included as a control group.

Participants had been exposed to native German for 2.7 years on average after age of 11. Four native German speakers who had recently arrived in the U.S. rated participants' productions and found that one participant received native-like ratings, and was even mistaken for a native German speaker. This participant had begun learning German at age 22 and a particular characteristic was his high motivation and interest in the German language and culture. Birdsong (2003) provides more evidence on the potential of adult learners to attain high levels of performance in the L2 under exceptional motivational levels. He examined 22 adult English-speaking learners of French living in France, who had been exposed to L2 after the age of 18. Two of these participants demonstrated native-like levels of phonetic/phonological achievement based on their pronunciation rating and acoustic analysis of their productions. Similar to the case in Moyer (1999), the participants with native-like ratings had high levels of motivation suggesting that age of first exposure alone cannot account for the variation in L2 attainment. It is clear that at least some adults, who are exposed to L2 after the age of 12, are able to attain native-like pronunciation.

In sum, the critical/sensitive period concept is still a focus of investigation and debate. A number of factors seem to interact with the age of learning in second language acquisition. The studies presented above show that age of learning influences native-like attainment at different L2 linguistic levels: morphology, syntax and phonology. Findings like those reported by Moyer (1999) and Bongaerts et al. (1995; 1997; 2003), suggest that age is not the only factor to determine how adults learn a second language. It seems that individual intrinsic factors such as motivation (Moyer, 1999) and aptitude (DeKeyser et al., 2010), interact with the age factor and affect second language learning in ways that have not been conclusively determined.

Given the background understanding that there is a critical/sensitive period within which exposure must occur in order for language learners to reach a native-like ultimate attainment in an L2, it is expected that the adult sequential Spanish-English bilingual participants in this proposed study will not perceive AE vowels in a native-like fashion, since they were first exposed to an English-dominant environment after the age of 15.

2.2.2 Native Language. There are also factors related to the properties of the native language that influence adult non-native speech perception. Research has shown that the perception of L2 speech sounds is largely determined by the native language system because non-native speech sounds are perceived according to their similarities to and dissimilarities from the native phonological system (Best, 1995; Best & Tyler, 2007; Flege, 1995). The L1 phonological system is thought to filter out the acoustic/phonetic parameters of new speech sounds that cannot be accommodated to the native sound system (Polivanov, 1932). The relative difficulty in perception of non-native contrasts depends on the relationship between L1 and L2 and the possibility of being able to establish a perceptual differentiation between phonetic contrasts (Abrahamson & Lisker, 1970).

As a result, the influence of the L1 phonological system is not the same for all non-native contrasts; rather, some L2 contrasts may be easier or more difficult for L2 learners (Polka, 1991). In a training study, Iverson and Evans (2007) investigated how native language vowel systems (Spanish, French, German, and Norwegian) influenced the cues that listeners use when they learn the English vowel system (e.g., formant transitions and duration). Experimental tasks consisted of: (1) identifying natural English vowels in quiet; (2) identifying English vowels in noise that had been signal processed to flatten formant movement or equalize duration; (3) perceiving best exemplars for L1 and L2 synthetic vowels in a five-dimensional vowel space that included formant movement and duration; and (4) rating how natural English vowels assimilated into their L1 vowel categories. The results showed that all listeners, regardless of their native language, learned new information from the English vowel system and assimilated the English vowels to their respective native language categories. They all used formant movement (spectral information related to the position of the articulators in the oral cavity), and durational cues to identify the English vowels. It was also observed that listeners with larger native language vowel systems, such as German and Norwegian, were more accurate at recognizing English vowels than were listeners with smaller native language vowel systems (Spanish and French). The authors concluded that despite the accuracy differences, all listeners perceived L2 vowels in a similar way, using the same cues.

Therefore, since the influence of the native language in the perception of non-native speech sounds is directly related to the L2 phonological system, it is possible to anticipate that the accuracy of identification of AE vowels by native Spanish speakers in this study may relate to their formation of new vowel categories corresponding to AE vowels, and their use of acoustic/phonetic cues in a similar fashion to native English listeners.

2.2.3 Experience with the non-native language. Adult monolingual (naïve) and L2 learners differ in how their experience with the non-native language influences their perception of the target-language speech sounds. The perceived relation between L1 and L2 sound systems influences L2 learners' speech sound perceptions. In addition, L2 learners are influenced by the degree and type of experience with the L2 sound system (Flege et al., 1997; Best and Strange, 1992). L2 learners' exposure to L2 speech sounds usually happens in two possible environments: (1) most common for adult learners is classroom instruction, usually in an L1 setting, where L1-accented teachers provided instruction that often includes modeling of incorrect phonetic details, and where the use of L2 is limited to the classroom instruction time; or (2) much less commonly for adults, immersion in the L2 environment. This is when listeners acquire L2 while living in the non-native language environment and are exposed to both formal instruction and a naturalistic environment outside the classroom (Best & Tyler, 2007). This type of exposure is considered more beneficial to the learner than classroom instruction in the L1 setting, since the naturalistic environment is rich in native linguistic input and opportunities to use the language.

Empirical studies indicate that learners' perceptions of non-native speech contrasts change in the course of learning, and learners with increased L2 experience in the L2 environment outperform less experienced learners from the same L1 background (Best &

Strange, 1992). Flege et al. (1997) studied the effects of English-language experience on non-native speakers' production and perception of English vowels along two continua: beat-bit /i-ɪ/ and bet-bat /ɛ-æ/. They compared native speakers of four diverse languages (German, Spanish, Mandarin and Korean) and assessed the relations between vowel production and perception accuracy. Within each language group, there were subgroups according to their English-language experience (measured by the length of residence in the U.S.). Flege and colleagues found significant differences between listeners from diverse language backgrounds. Germans perceived and produced the English contrast beat-bit /i-ɪ/ better than the other language groups regardless of L2 experience since this contrast is phonemic in German, but had difficulty perceiving and producing English vowel /æ/. Contrary to prediction, experienced Koreans perceived and produced the English contrast beat-bit /i-ɪ/ using spectral cues – as do native English speakers - but used durational cues to perceive and produce the vowel contrast bet-bat /ɛ-æ/. Flege et al. (1997) suggested that perceptual and production difficulties were observed for the English vowel contrast bet-bat /ɛ-æ/ because Korean has a phoneme with two variants that have similar qualities to /ɛ/ and æ/.

Native Spanish-speaking participants were found to rely on spectral cues more than durational cues, like native English speakers, when perceiving the vowel contrast /ɛ-æ/. However, this group did not seem to favor either spectral or temporal cues when perceiving and producing the contrast /i-ɪ/. There was a significant effect of language experience in the accuracy of perception and production of the vowel pairs in all background languages. Experienced listeners made greater use of spectral cues for both perception and production of the English vowels, again a pattern that is seen with native speakers of English. Flege et al. (1997) found that, despite the language background differences in perception of the English vowels, experienced speakers of German, Spanish, and Mandarin also relied more on spectral cues to perceive and produce English vowels compared to those with less experience. The effect of language experience (as measured by length of residency) has been found to vary according to other factors such as age related differences (Ito & Strange, 2009) and type of instructions (Jia et al., 2006).

2.2.4 Length of Residency (LOR). Living in the L2 speaking environment constitutes another factor that impacts the perceptual performance of adult L2 learners. Length of residency (LOR) in the L2 speaking environment is positively correlated with better perception and production performance (Flege et al., 1999, 2006; Guion et al., 2000). LOR has been found to interact with age of L2 acquisition and with extent of language use. Flege et al. (1999), as described in the discussion of the SLM speech model above, showed that young learners who are exposed early (by 7 years of age) to the L2 environment perform better in L2 perceptual and production tasks than late learners who are exposed to L2 in adulthood. Along the same lines, Jia et al. (2006) found that young learners who acquire their L2 in the target-language environment perform better than young learners receiving instruction in a non-naturalistic environment.

The effects of LOR are not limited to the ability to perceive and produce non-native vowels. Ito & Strange (2008) tested the perception of stop aspiration and glottal stop allophonic cues for word juncture in English (e.g. aspiration: *keeps talking vs. keep stalking,* glottal stop: *a nice man vs. an ice man*) by thirty Japanese late learners of English (between 15 and 45 years of age), and compared them to a control group of native English speakers. The length of residency

(LOR) of the Japanese listeners in the L2 environment ranged from 2 weeks to 2 years. Listeners were presented with two sentences visually (on a computer screen), and one of the two was also presented auditorily. They were asked to select which sentence they had heard by clicking on one of the onscreen choices. Results showed that, in general, Japanese perception of both cues was less accurate than the English-speaking control group, and within Japanese listeners, aspiration cues were more difficult to differentiate than glottal stops. Accuracy in perception of these cues positively correlated with LOR, indicating that the longer listeners had been in the English environment the more accurate was their perception of the cues. In addition, the findings suggest that different lengths of exposure are needed to develop accurate perception of different consonantal contrasts (learning to distinguish aspiration cues takes longer than learning to distinguish glottal stops, for Japanese listeners).

2.2.5 Attending to relevant vowel cues. Listeners can identify vowels and consonants by unconsciously attending to distinct phonetic-acoustic information in the speech signal. The speech signal carries information such as pitch, intonation, loudness, tones, duration, VOT, spectral features among others. The most relevant cues in this study are spectral and durational cues.

Native and second language listeners seem to attend to different dimensions of the consonants and vowels they hear. This has an impact on their ability to perceptually differentiate between L2 contrasts (Escudero & Boersma, 2004). Fox et al. (1995) and Jusczyk (1992) have proposed that when adults learn a new phonetic contrast, they need to re-allocate selective attention to the relevant acoustic cues and phonetic dimensions of L2 sounds (e.g. spectral and durational cues as described in chapter 1). In some cases, those dimensions are not present in

L1, posing a challenge for the listener who needs to identify the primary cues in order to correctly perceive the phonemes and form a new category. When L2 listeners cannot attend to specific acoustic differences that indicate a phonemic contrast in their L2, they tend to rely on secondary phonetic information for native speakers in order to make the distinction between the two phonemes (Flege et al. 1997; Iverson et al. 2003).

Languages such as Japanese use durational cues to contrast meaning, whereas some languages like English rely primarily on spectral cues and use durational cues as secondary information to distinguish between two or more phonemes (Bohn & Flege, 1990). Other languages like Spanish, Mandarin (Flege et al, 1997), Russian (Kondaurova & Francis, 2009), Portuguese (Rauber, Escudero, Bion & Baptista, 2005), Catalan (Cebrian, 2006) do not use durational cues to signal meaning contrasts at the level of phonemes. However, listeners of those languages do seem to rely on L2 durational information to identify the differences between specific contrasts. In order to account for this phenomenon, Bohn (1995) proposed the desensitization hypothesis, a perceptual principle that listeners apply independently of their native language background. This principle states that whenever spectral differences are insufficient to differentiate vowel contrasts, because listeners are not sensitized to the spectral differences, durational information will be used to differentiate the non-native vowel contrast (Bohn, 1995).

Hillebrand et al. (2000) examined the role of duration in perception of AE vowels that are spectrally similar and differ in duration /i-ɪ/, /u-ʊ/, /æ-ɛ/, /e-ɛ/, /ɑ-ʌ/, /ɔ-ɑ/. Stimuli consisted of synthesized versions of 300 utterances with different durations selected from a large, multi-talker

(45 men, 48 women, 46 children of 10-12-years of age) database of /hVd/ syllables including 12 vowels. The authors synthesized four versions of each utterance: one set of utterances had the original duration (vowel duration matched to the original utterance), a second set had a neutral duration (duration fixed at 272 ms, the grand mean across all vowels), the third set had a short duration (duration fixed at 144 ms, two standard deviations below the mean), and the fourth set had long duration (duration fixed at 400 ms, two standard deviations above the mean). Fifteen phonetically trained graduate students and 3 faculty members of the speech language pathology program at Western Michigan University participated in the study. Listeners were asked to identify (by selecting the corresponding phonetic symbol) each of 1200 syllables presented in random order.

Results suggested that 1) duration had a small overall effect on vowel identity since the great majority of signals were identified correctly at their original durations and at all three altered durations; 2) despite this, some vowels, especially /ɑ/, /ɔ-ʌ/ and /æ-ɛ/, were significantly affected by duration; 3) some vowel contrasts that differ systematically in duration, such as /i-ɪ/ and /u-ʊ/, were minimally affected by duration manipulations; 4) a simple pattern recognition model appears to be capable of accounting for several features of the listening test results, especially the greater influence of duration on some vowels than others; 5) given that a formant synthesizer does an imperfect job of representing the fine details of the original vowel spectrum, results using the formant-synthesized signals led to a slight overestimate of the role of duration in vowel recognition, especially for the shortened vowels. It should be noted that Hillebrand et al.'s study population was highly specialized: they were all monolingual English listeners with ample experience with phonetics. Therefore, although this approach yielded valuable information, the results cannot readily be generalized to other kinds of listeners, nor to speakers of other languages.

Cebrian (2006) studied the role of experience and the use of durational cues in the identification of the English vowel contrast /i-ɪ/ by Catalan listeners with different levels of experience with English. To establish whether naïve and L2 learners perceived similarity between L1 and L2 sounds, the author first tested Catalan listeners on identification and similarity rating of English vowels using a four-choice perceptual assimilation task. The response options were the Catalan vowels /i/, /ɛ/, /e/ and /eɪ/. In addition, Cebrian aimed to determine whether L2 spectral distinctions had a counterpart in L1 by measuring the phonetic distance between Catalan and English vowels. Results from this experiment showed that both Catalan listener groups had similar L2 to L1 assimilation patterns and goodness ratings. However, the groups differed in the assimilation rates for /ɛ/. Although it is not clear from the study what the differences between English and Catalan /ɛ/ are, one of the findings was that inexperienced Catalan listeners had greater percentage of assimilation from English /ɛ/ to the Catalan /ɛ/ (93%) compared to experienced listeners (78%). These results motivated the author to further study the identification of these AE contrasts in three groups of experienced Catalan listeners in Canada, experienced Catalan listeners in Barcelona, and inexperienced English listeners as a control group. All participants listened to the vowels /i/, /ɪ/, /ɛ/ randomly presented at four different durations (100, 150, 200, and 250 ms. Participants were asked to select the perceived vowel by selecting a corresponding word on a screen (beat, bit, bet). Findings revealed that all English speakers' responses across all duration conditions were based on vowel spectral information. In contrast, L2 listeners demonstrated stronger reliance on duration information.

Escudero (2000) has also investigated cue reliance of Spanish listeners when perceiving Scottish /i/ and /ɪ/ (see section 2 on Speech Perception Models). In this study, the cue reliance to signal the contrast was computed subtracting the score of the first stimulus from the seventh stimulus for each of the continua. The scores for cue reliance were calculated using the end points of all the possible continua for the number of elements taken for spectral and durational cues (7 continua for each cue). The cue weighting or phonetic attention paid to each cue was computed by dividing each reliance value by the sum of the reliance values for the two cues. For example, if the reliance for durational cues were 90 and the spectral one 30 and we wanted to calculated the durational cue weighting, then we would have to divide 90 by 120 and the result would be 0.75, which would mean that the durational weighting is 75% (Escudero, 2000). Findings from this study allowed Escudero to describe groups showing the same pattern of cue reliance.

A group that did not make distinctions between English /i/ and /ɪ/. Another group distinguished the vowel contrasts using only durational cues. And a third group that used both spectral and durational cues, with a higher weighting on duration. Lastly, there was a group used both spectral and durational cues relying more on spectral cues just like English native speakers. The author suggested that L2 learners could form phonetic categories using L2 cues in a non-native way.

In general, Escudero found that Spanish listeners make use of durational cues more than English listeners do (Escudero, 2000; 2005). These findings led to the proposal that when the duration dimension is not used in L1, this dimension can be regarded as a “blank slate” upon which listeners begin to classify the non-native contrasts (Escudero & Boersma, 2004). Escudero has described a developmental stage pattern to explain the acquisition of new contrast and how listeners change their acoustic reliance as they advance in their learning process of the English /i-ɪ/ contrast. According to Escudero, in stage 0, L1-Spanish speakers do not distinguish English /i/ and /ɪ/. During stage 1, the contrast is distinguished using only durational cues. In stage 2, listeners use both spectral and durational cues, but duration has a higher weighting. Finally, at stage 3, L2 listeners use both spectral and durational cues relying more on spectral cues just like English native speakers.

Taking all these factors together it can be concluded that age, native language, experience with L2, and length of residency (LOR) must be taken into account when analyzing perceptual performance of L2 learners. Results from the studies reviewed in this section support the notion of a younger-learner advantage in perception of L2 speech sounds. In addition, early exposure to the L2 environment from an early age results in a more accurate perception compared to L2 learners whose exposure occurred in adulthood (Flege et al., 1999; Jia et al., 2006). Moreover, the studies reviewed above have contributed significantly to the understanding of how L2 listeners use spectral and duration information, but several questions remain unanswered about the use of durational cues. For example: How long do L2 learners rely on durational information before using spectral cues to make distinctions between non-native phonemes? How do they learn to attend to spectral characteristics and ignore secondary acoustic information about the new phonemes? How do L2 listeners who have learned to use spectral cues to perceive L2 contrasts manage, when durational information is not available in the speech signal?

The characteristics of the proposed study participants (they have learned English in their home country with Spanish-speaking teachers of English, and they have been exposed to an L2 naturalistic environment after the age of 15) lead us to expect low accuracy in discriminating AE vowel contrasts /æ-ɛ/ and /ʌ-ɑ/, compared to native English speakers. However, in the case that these listeners are able to accurately distinguish between the AE vowels, it would be useful to know whether they are using the vowel cues in a native-like or non-native-like fashion.

2.2.6 Experimental Factors. In addition to all the factors identified above, experimental factors including the choice of type of stimuli (Strange, 2007), type of perceptual task (Beddor & Gottfried, 1995; Massaro & Cohen, 1983; Logan et al., 1991; Schouten, Gerrits, & Van Hessen, 2003), and the time of presentation of the stimuli (ISI, or inter-stimulus interval) (Strange, 1995; Werker & Logan, 1985) have all been shown to influence the perception of non-native speech sounds in adult listeners. Experimental decisions require the researchers to consider not only the individual, language, and experiential factors described above, but they also need to make careful decisions regarding experimental design, stimuli, and tasks. In this section, each of these factors will be addressed in turn.

2.2.6.1. type of stimuli. Decisions about which and how speech sounds are produced and presented to the listeners need to be made keeping in mind that there are multiple sources of acoustic variability in the speech signal, especially in vowels. Some sources of variability include natural or synthetic production, presentation (in citation form or carrier sentence), consonantal contexts, talkers, speaking rate, and prosody (Strange, 2007). Speech sounds that are synthetically produced may be useful to manipulate the acoustic properties of the sounds; however, they are not appropriate exemplars of typical communication. On the other hand, natural stimuli represent a more ecologic option, but they are not easy to manipulate and even small changes to the sounds would alter their fidelity. Therefore, using natural or synthetic stimuli depends on specific purposes of the research.

Gordon, Keyes, & Yung (2001) examined whether the perception of the AE contrast /ɹ/ - /l/ by twelve L1-Japanese (exposed to English after adolescence through immersion in the U.S.) and 12 monolingual English listeners differed in two conditions: naturally produced minimal pairs versus synthetic speech syllables. The natural speech stimuli consisted of four minimal pairs containing /ɹ/ and /l/ in four different positions within the word, and were produced by four speakers of American English (two males and two females). The synthetic stimuli, a /ɹa/-/la/ series, varied the values of formants F1, F2 and F3. After completing a language background questionnaire, all participants completed two identification tasks. The identification task for the natural stimuli consisted of circling the word (minimal pairs) on the response sheet that corresponded to the word they heard on the tape. Then, on the identification of synthetic stimuli, listeners had to circle "R" or "L" on the response sheet to indicate which syllable (/ɹa/-/la/) they thought they had perceived. In this study, American English listeners were able to accurately identify the natural stimuli, while the Japanese participants had difficulties perceiving the /ɹ/-/l/

sounds. This difference was significant for /ɹ/-/l/ produced in initial, cluster or medial word position, but non-significant when the sounds were in word final position.

For the synthetic stimuli, American listeners were able to accurately identify the syllables using cues from F1 and F3 formants. On the other hand, the accuracy of the Japanese listeners was related to highly variable use of F1 and F3 cues to identify the stimuli in both natural and synthetic speech. For example, listeners who had the best performance in the identification task with natural speech stimuli seemed to use both F1 and F3 cues to identify the synthetic syllables, as did American English native listeners. On the other hand, listeners with the poorest performance in the natural speech condition did not use either of the relevant identification cues. Among the Japanese listeners who seemed to use either F1 or F2 cues, there was one listener who seemed no to use F1 information. This listener had one of the best performances on the natural speech task. Conversely, some listeners who did not use F3 cue information had the worst performance in natural speech identification. By using both natural and synthetic speech stimuli to better understand the nature of the Japanese perception of AE contrasts, Gordon et al. were able to provide compelling evidence on cue weighting and identification processes in this population.

Another stimuli presentation factor to consider in a speech perception study design is the use of carrier sentences. Vowels contained in monosyllables or disyllables that are read by a speaker from a list (citation form) are spectrally and acoustically different from the same vowels when they are embedded in sentences that are more similar to continuous speech (Strange, 2007). Language is not processed as isolated sounds, so presenting the stimuli in a carrier sentence (e.g., *I say "bab" this time*) represents a more natural option. However, the challenge, especially of having vowels in a consonantal context, is the co-articulation effects since proceeding and following consonantal contexts affect the vowel sound and therefore influence its perception (Hillebrand, Clark & Nearey, 2001).

The selection of talkers to produce the stimuli represents another important factor. Individual talkers have different vocal tract dimensions, which results in variation in the production of speech sounds. Both native and non-native listeners adjust their perceptual parameters to find relevant common acoustic-phonetic features across talkers and ignore features that do not play a role in identifying specific speech sounds; however, this process seems to be more effortful for L2 listeners. For example, a study conducted by Bent, Kewley-Port, & Ferguson (2010) compared AE and Korean listeners' perception of AE vowels in noise /i/, /ɪ/, /e/, /ɛ/, /æ/, /ɑ/, /ʌ/, /o/, /ʊ/, /u/, produced by ten different talkers. The results showed that although there was variability in the accuracy scores in both language groups, AE listeners identified vowels more accurately than Korean listeners did. The patterns of perceptual errors in the Korean group were strongly influenced by across-talker variability, especially for two specific AE vowel pairs that have been documented as difficult for Korean listeners to perceive: /i-ɪ/ and /ɑ-ʌ/ (Nishi & Kewley Port, 2007; Tsukada, Birdsong, Bialystok, Mack, Sung & Flege, 2005). The authors concluded that cross-talker variability in the production of non-native speech sounds influences the perceptions of L2 listeners possibly due to their difficulty in forming native-like vowel categories. In addition, if sentences are used, it is important to note that sentence prosody and speaking rate also introduce acoustic variability in utterances (Fourakis, 1991). With so much variability in the acoustic characteristics of the vowels when they are in different contexts, Strange (2007) suggested that in order to understand the general L1/L2 phonetic relationships that influence L2 learners' perception of non-native speech sounds, it is necessary to establish comparisons of vowels produced in diverse phonetic, phonotactic and prosodic contexts.

2.2.6.2. type of task. Perceptual tasks can be grouped into discrimination and identification tasks. Discrimination tasks measure the ability of listeners to differentiate among stimuli without requiring labeling. That is, listeners only need to determine whether one of the sounds is the same or different from other sounds presented in the same trial. In identification tasks, one or more sounds are presented in the same trial, and listeners are asked to provide an explicit label to one or more stimuli. Speech sounds are presented individually with determined time between each stimulus (referred to as the inter-stimulus interval, or ISI). The two type of tasks are described more fully below.

2.2.6.2.1. discrimination tasks. Discrimination tasks are presented in various designs, a few of which are discussed here. In a Same-Different (AX) task, two stimuli are presented at each trial, and the listener is asked to decide whether the two sounds are the same or different. Reaction time (RT) is easy to measure in this task since it only depends on the presentation of the second "X" stimulus. AX discrimination tasks tap low level, sensory-based information about the speech signal. That is, listeners make their decisions based on acoustic/physical differences or similarities between A and X, instead of attending to more abstract phonemic, phonetic features, making it difficult to generalize listeners' responses to other conditions or other type of tasks (Logan et al., 1991). A speeded AX task requires listeners to respond as quickly as they can. Both stimuli are presented with very short ISI (usually between 100 and 500 ms). This task involves a trade off between accuracy and reaction time. Some listeners may not respond as accurately under the pressure of responding quickly. However, reaction times show less variability than in other tasks (McGuire, 2010).

The ABX discrimination task and its variants (AXB, XAB) consist of trials that include 3 different stimuli. Listeners compare which of the "A" and "B" stimuli is the same or most similar to "X". This type of task reflects the use of phonetic memory (Massaro & Cohen, 1983), since it requires that listeners attend to linguistic information contained in the speech sounds rather than their general acoustic properties. However, the ABX variant of this task is subject to a very strong bias toward the response "*B =X*", which could be due to the high memory load of the task (Schouten, Gerrits, and Van Hessen, 2003). By the time "X" is presented, memory for the A and B stimuli could be degraded. The AXB variant, in which the second stimulus is identical to either the first or the third one and is close in time to both, has yielded contradictory results. For instance, Van Hessen and Schouten (1999) reported that their subjects often ignored the third stimulus, thus annulling the expected advantage over ABX.

The oddball (category change) task (introduced in chapter 1), consists of the presentation of a stream of multiple and equal auditory stimuli that, at some points of the stream, are interrupted by "deviant" sounds. In behavioral experiments, or experiments designed to tap into conscious change detection, listeners are required to press a button, or otherwise indicate when they perceive the change in the stream of sounds. Passive listening to an oddball stimulus stream is used to detect neurophysiological changes associated with preconscious change detection. Oddball tasks are frequently used in neurophysiological studies, especially when the ERP components of interest are mismatch negativity (Näätänen, Paavilainen, Rinne, & Alho, 2007) and P300 (Frenck-Mestre, Meunier, Espesser, Daffner, & Holcomb, 2010). Discrimination tasks are easy to explain to listeners and do not require explicit knowledge of the nature of the similarities/differences between the speech sounds. When the goal is to obtain more explicit information from the listener, identification tasks are more appropriate. Some identification task paradigms are described below.

2.2.6.2.2. Identification tasks. Identification tasks require explicit labeling in response to one or more sounds are presented in the same trial. Identification tasks include:

(1) Yes-No identification, in which listeners are asked whether they heard "X" or not. This is a very easy task to create and has a low memory load – that is, listeners do not need to keep much information in their working memory while stimuli are being presented.

(2) Alternative forced-choice (AFC) tasks, in which, listeners are presented with an “X” stimulus, and given either two (2AFC) or four (4AFC) alternative- response options to decide which one corresponds to the stimulus they heard. These tasks make explicit reference to phoneme categories in the instructions. Listeners need to be familiar with the features of the response options and their specific labels in order to respond. In this case, listeners perceive the stimuli in a categorical way (Gerrits & Schouten, 2004). A problem that has been noted with tasks that require labels is that the type of label chosen in the design of identification tasks can be an important factor in experimental outcomes. For example, studies using orthographic labeling (giving the non-native phoneme an equivalent native name) may provide contradictory or unclear results because not all L2 sounds have an L1 counterpart, a situation that can lead the listener to classify phonemes into arbitrary categories that may not represent their true perception (Beddor & Gottfried, 1995; Strange 2007). As an alternative to orthographic labeling, some researchers have used International Phonetic Alphabet (IPA) symbols as response options in cross-language studies. After some training, listeners use IPA symbols to provide responses in identification tasks. Listeners are usually trained to memorize the relationship between specific IPA symbols and keywords containing the specific phonemes (e.g., Strange et al., 2005; Nishi & Kewley-Port, 2007).

2.2.6.2.3. Category of goodness tasks. Categories of goodness tasks require that listeners have some knowledge of the phonemes that are used as stimuli. The task requires listeners to listen to sounds, label them, and decide how good example of a specific category an “X” phoneme is. For example, in their study, Iverson & Kuhl (1996) asked adult Japanese listeners to identify the initial consonant in /ɹa/- /la/ syllables. Once listeners had identified the consonant, they were asked how well a token (sample) of /ɹ/ represented the /ɹ/ category in a scale from 1 (bad) to 7 (good). One possible limitation of this task is that rating depends on the intuitions of participants, rather than any objective criterion.

2.2.6.3. Interval of stimulus presentation (ISI). The inter stimulus interval (interval of time elapsed between the offset of one stimulus and the onset of the next) represents another aspect of the experimental manipulation that has been shown to affect the responses of L2 listeners. For example, Werker & Logan (1985) investigated how ISI interacts with auditory, phonetic, and phonemic processing in in non-native speech perception by testing the perception of Hindi dental and retroflex stop (CV) syllable contrasts by monolingual English listeners and a Hindi group for comparison. An AX discrimination task, in which listeners are required to say whether X is the same or different to A, was used. Crucially, the time interval between two stimuli (ISI) of each comparison varied (250 ms, 500 ms, and 1500 ms). The stimuli consisted of multiple natural tokens of a Hindi retroflex and dental contrast grouped in three types of pairs: (1) Physically identical (PI), corresponding to the A = X form; (2) Name identical (NI), including two different tokens of the same category (e.g. two dental stops or two retroflex stops); and (3) Different (DIFF), including one token of each category (one dental and one retroflex token). Werker & Logan predicted that, if English listeners were using auditory-sensory processing, they would discriminate physical differences in both NI and DIFF pairs. However, if they were using native language categorization (phonetic should this be phonological? level processing) they would be unable to discriminate either NI or DIFF pairs. On the other hand, Hindi listeners should easily discriminate DIFF pairs because the retroflex/dental stop contrast is phonological for their language.

Results showed that English listeners discriminated NI and DIFF in the shortest ISI condition (250 ms), reflecting acoustic-level processing. Discrimination was poor for DIFF pairs in the longest ISI condition (1500 ms), but improved with practice. In this case, monolingual

English listeners initially used phonetic level processing (native-language categorization), but reflected some phonemic processing (non-native categorization) as they went through the task. Discrimination remained poor for NI pairs in longer ISI conditions (500 and 1500 ms). These results suggest that English listeners did not have access to discriminate between dental or retroflex tokens differing in phonetic features that are not part of their language. Conversely, Hindi listeners discriminated DIFF, but not NI pairs during the longest ISI. The performance of Hindi listeners suggested that, with increased memory load due to the long ISI, this group used native-language categorization processing (phonemic level) to discriminate the pairs.

The findings from this study demonstrated that the timing of stimulus presentation can influence the processing level that listeners access while perceiving non-native speech segments. Phonemic processing (language-specific categorization) was reflected in the discrimination of DIFF pairs with the increased memory load in the long ISI conditions by native and non-native listeners, after some practice. Phonetic processing (language general categorization) was observed during less demanding task conditions (ISI 500), and auditory processing (discrimination based on physical differences) was evident in the minimally demanding task in the shortest ISI condition.

2.2.7. Implications for the experimental design of the proposed study. Since the purpose of the proposed study is to determine whether Spanish-English bilinguals perceive AE vowels using temporal and spectral parameters in a native-like fashion as compared to AE listeners, the use of completely naturally produced stimuli is not possible (Gottfried & Beddor, 1988), at least in part. For one of the conditions, the natural duration condition, vowels will be presented in naturally produced syllables in citation form. For the neutral duration condition, duration of the naturally produced vowels will be neutralized yielding edited natural stimuli. The experimental procedure will make use of the oddball paradigm in two different tasks (discrimination and identification). The oddball paradigm has been widely used in ERP studies looking at speech perception (Näätänen et al., 2007; Nieuwenhuis, Aston-Jones, & Cohen, 2005). In the proposed study, the *discrimination task* does not require conscious attention to the auditory stimuli. For this task, the ERP component of interest will be the mismatch negativity (MMN). In contrast, the *identification task* requires the listener to label each sound that is presented, allowing the elicitation of the second ERP component of interest, the P300. Based on the seminal work of Werker & Logan (1985), an inter-stimulus interval (ISI) of 800 ms has been selected for presentation of stimuli in the discrimination task, since this should provide optimal conditions for listeners to discriminate the vowels at the phonetic-phonemic level.

In sum, cross-language studies require the analysis, adaptation and modification of methodological variables that can be divided in two groups. The first group contains the variables that are related to the listener, such as intrinsic characteristics, native language, and experience with L2. The second kind of methodological variables relate to experimental design, and include type of stimuli, experimental tasks, and ISI. Trying to control each of these factors is a difficult process that requires the experimenter to make careful decisions about the most appropriate strategies that allow the study of language perceptual abilities in adults.

The following section will describe and compare the vowel inventory system of English and Spanish, to show how similarities and differences between the inventories are likely to influence the perception of AE vowels by adult Spanish learners of English.

2.3. Spanish and American English vowels

The vowel systems of Spanish and American English (AE) differ both in the size of their respective inventory and the number of relevant cues necessary for the vowel sounds to be perceived. These differences between the vowel inventories indicate that native listeners of the two respective languages rely more heavily on those phonemic cues that are relevant to identify their native phonemes. The vowel system in Spanish consists of five vowels, which is a small inventory compared to the eleven vowels of English /i/, /ɪ/, /e/, /ɛ/, /æ/, /ɑ/, /ʌ/, /o/, /ɔ/, /ʊ/, /u/, (Clopper, Pisoni, & Jong, 2005). The vowels in Spanish /i/, /e/, /a/, /o/, /u/, do not have a direct counterpart in the English vowel system (Hualde, 2005); they are described as high front and back /i/, /u/, which are shorter than the English /i/, /u/. Also, the tongue is lower for English /i/ and /u/ than for the same vowels in Spanish (Flege, 1989). Mid front and back Spanish vowels /e/, /o/, are pure monophthongs. Spanish has one low central vowel /a/, which is closer to the American English /ɑ/ than to American English /æ/. Significantly, the position of the tongue is also lower for English /ɑ/ than for Spanish /a/. In general, Spanish vowels are not called tense or lax, but they can be rounded /o/, /u/, and un-rounded /i/, /e/, /a/. Moreover, they only differ in spectral features, do not have as much formant movement as English vowels (see table 1) and may not be distinguished from one another by vowel duration differences (Harris, 1960). English vowels have been classified as tense or lax, a relative property that is determined in part by tongue root position. Tense vowels include /i/, /e/, /ɑ/, /o/, /ɔ/, /u/ while lax vowels are /ɪ/, /ɛ/, /æ/, /ʌ/, /ʊ/.

In order to perceive their native vowels, English speakers use two spectral dimensions: high-low (F1 formant), and front-back (F2 formant), derived from spectral cues and durational information. On the other hand, Spanish speakers use only the two dimensions of the spectral cues (high-low and front-back) for distinguishing their native vowel contrasts (Fox et al. 1995). Some authors have suggested that tense/lax vowel contrasts are distinguished by spectral and durational cues combined (Strange, 1989; Klatt, 1976; Geigerich, 1992).

Bradlow (1995) conducted a comparative acoustic analysis to determine the location of the English and Spanish vowels in the vowel space, and whether their location was related to language-specific features. One of the interesting findings was that the points of articulation of the 'common' vowels in English and Spanish /i/, /e/, /o/, /u/ were different. This difference was particularly evident in the F2 dimension, which varies relative to the position of the tongue in the oral cavity (back or front) to produce the vowel. The F2 values were higher for English front vowels /i/ (F2 = 2393 Hz) and /e/ (F2 = 2200) compared to the Spanish vowels /i/ (F2 = 2174 Hz) and /e/ (F2 = 1814 Hz). In general, English front vowels were located more peripherally in the vowel space than the Spanish front vowels (see figure 2).

With respect to English back vowels, F2 values were also higher for English /u/ (F2 = 1238 Hz) and /o/ (F2 = 1160 Hz) than for the Spanish vowels /u/ (F2 = 992 Hz) and /o/ (F2 = 1019 Hz) (see table 3 for English and Spanish vowels formant values). English vowels were positioned more centrally in the oral cavity than were Spanish vowels. These findings suggested that English vowels /i/, /e/, /o/, /u/ are articulated with a more fronted tongue position than the Spanish ones.

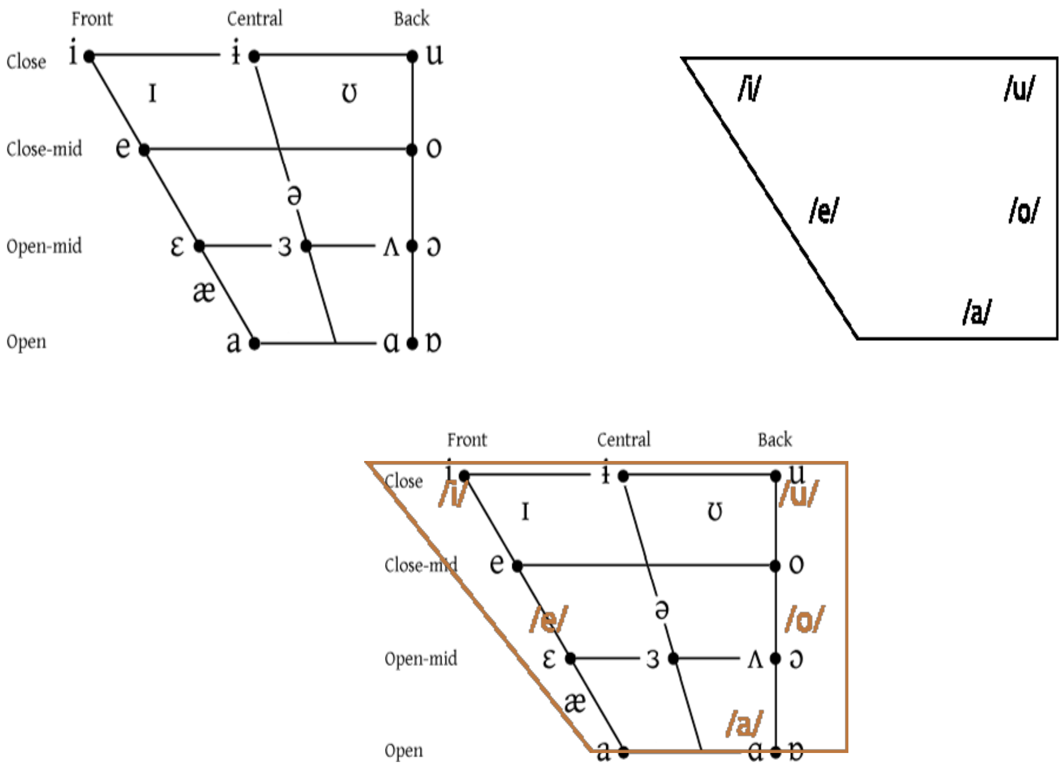

Figure 2. English and Spanish Vowels Quadrilateral

The quadrilaterals represent the arrangement of the vowels in the vowel space. Left quadrilateral represents the arrangement of the American English vowels in the vowel space. Right quadrilateral represents the Spanish vowels in the vowel space. Below, the Spanish vowel space superimposed on the American English vowels. Adapted from How Language Works by M. Gasser. Retrieved June 12, 2013 from http://www.indiana.edu/~hlw/PhonUnits/vowels.html. Copyright 2002 by Mike Gasser. Reprinted with permission.

Regarding L1 Spanish speakers who are learners of English, the most extensive work has been done on the perception of English vowel contrast /i/-/ɪ/, from different dialect variations such as American English, (Flege 1991; Flege, Munro, & MacKay, 1994; Fox et al., 1995; Flege et al., 1997), British English (Escudero 2000; Escudero & Boersma 2002; Escudero & Boersma 2004; Escudero & Chládková, 2010), and Canadian English (Morrison, 2006; 2008; 2009). Such studies have demonstrated that L1-Spanish listeners weight spectral and duration cues differently from native English speakers when perceiving and producing vowel contrasts – at least, during some learning stages (see Escudero, 2000).

A major aim of the proposed study is to determine how Spanish-English bilinguals perceive English phoneme contrasts that are not present in their language. Having a small-size vowel inventory, L1-Spanish listeners need to incorporate the new phonemes into their small repertoire. They might not need to learn vowels that are 'common' to both languages (/i/, /e/, /o/, /u/). However, there are language-specific differences in the articulation of these 'common' vowels that made them different between the languages. In addition to those vowels that are not common, L1-Spanish learners of English need to learn at least another seven monophthongs: /ɑ/, /ʌ/, /ɛ/, /æ/, /ɪ/, /ʊ/, /ɔ/. Spectral and durational features that are not employed contrastively in Spanish characterize these English vowels. Due the intrinsic differences and similarities between the two vowel inventories, it could be argued that these differences would represent some perceptual challenges for the L1-Spanish listener.

Iverson & Evans (2007) hypothesized that listeners from a language with a large vowel inventory would have more difficulty learning new vowels compared to listeners from a language with a small vowel inventory, because the larger vowel space would be already crowded, with no space for extra phonemes. To evaluate this hypothesis, native speakers of German (large vowel inventory) and Spanish (small inventory) were trained on non-native vowel discrimination over 5 sessions. German speakers demonstrated more improvement than the Spanish speakers. However, after other ten sessions of training, Spanish speakers attained the same performance level as the German speakers. Iverson concluded that German speakers had immediate access to more vowel categories than Spanish speakers, which they were able to use when they assimilated English vowels to their native language. This made the task of vowel learning easier for the German speakers than for the Spanish speakers who had less native categories to compare. Hence, these findings support the view that L1-Spanish listeners may encounter perceptual challenges when learning English vowels.

For the present proposed study, the American English (AE) vowel contrast /ɑ/-/ʌ/ (as in the words *dock and duck*) have been selected as targets, since research has shown that these contrasts represent difficulty for many Spanish speakers (Flege, 1991; Fox et al., 1995; Flege et al., 1997; Escudero & Chládková, 2010). In order to perceive this AE vowel contrasts /ʌ/-/ɑ/, L1-Spanish speakers need to attend to the all the relevant cues that signal distinctions between the members of the contrast pair (both spectral and durational cues). In the /ʌ/-/ɑ/ pair, vowel /ɑ/ could be described as low back, unrounded and tense, while the vowel /ʌ/ is mid central, unrounded and lax.

Table 4.

Duration and formant values of American English vowels adapted from Hillebrand et al., 1995

Vowel	**Duration**	**F1**	**F2**
i	306	437	2761
ɪ	237	483	2365
e	320	536	2530
ɛ	254	731	2058
æ	352	669	2349
ʌ	226	753	1426
ɑ	323	936	1551
ɔ	353	781	1136
o	326	555	1035
ʊ	249	519	1225
u	303	459	1105

Table 5

Spanish CVCV, English CVC, and English CVCV Mean Vowel Formants in Hertz. Taken from Bradlow, 1995.

Spanish	**CVCV**		**English**	**CVC**		**English**	**CVCV**
	F1 (SD)	**F2 (SD)**		**F1 (SD)**	**F2 (SD)**		**F1 (SD)**
i	286 (6)	2147(131)	i	268(20)	2393(239)	i	264(34)
e	458(42)	1814(131)	ɪ	463(34)	1995(199)	ɪ	429(20)
a	638(36)	1353(84)	e	430(45)	2200(168)	e	424(39)
o	460 (19)	1019(99)	ɛ	635(53)	1796(149)	ɛ	615(60)
u	322(20)	992(121)	æ	777(81)	1738(177)	æ	773(62)
			ʌ	640 (39)	1354(134)	ʌ	655(43)
			ɑ	780 (83)	1244(145)	ɑ	783(155)

Fox et al. (1995) conducted a study involving a multidimensional scaling analysis, which asked L1 Spanish learners of English (experienced and inexperienced) and native English listeners to rate the similarity between American English (AE) /i/, /ɪ/, /ei/, /ɛ/, /æ/, /ɑ/, /ʌ/ and Spanish (SP) /i/, /e/, /a/ vowels. Stimuli were produced by 3 native speakers of each language in a bVto context that was truncated to bV for presentation to the listeners. Participants heard 189 AE pairs, 189 SP pairs, and 189 AE-SP pairs. To rate the vowels English listeners seemed to use 3 spectral dimensions (high-low, front-back, and center-noncenter) as well as durational cues, while Spanish listeners relied primarily on the high-low dimension and did not use durational information to make decisions about the vowels (see table 5 for formant values of AE and Spanish vowels). There was a high similarity rate for AE /ʌ/-/ɑ/ assimilation to Spanish /a/. According to Best, this could be a Single Category assimilation pattern, which would be difficult to discriminate by Spanish listeners. Spectral analyses revealed that AE /i/ and Spanish /i/ are spectrally very similar; and that Spanish /a/ values fall between AE /æ/, /ɑ/, /ʌ/), and Spanish /e/ values fall between AE /ei/, /i/, /ɛ/. It was also found that the vowel space of more experienced L1-Spanish listeners was more similar to the vowel space of native English listeners, suggesting that the perceptual dimensions used by listeners to identify non-native vowels may change as result of improvement in L2 proficiency. The authors suggested that Spanish listeners were not sensitive to durational information when they needed to rate vowels' dissimilarity and hence, had difficulty in perceiving English vowel contrasts that differ with respect to both in duration and spectral quality. However, other authors have found that vowel duration (see table 4 for durational values for AE vowels) is the speech cue that might help Spanish speakers to differentiate among vowels with similar spectral values (Hillenbrand, Clark, & Houde, 2000; Escudero, 2000; Escudero & Boersma, 2004; Morrison, 2008).

2.4 Summary of proposed study experimental design.

This study aims to examine neurophysiological responses (ERPs) of adult sequential Spanish-English bilinguals (who started to learn the language in a non-native environment or after puberty) to AE vowel contrast /ɑ/-/ʌ/, compared to monolingual English-speaking listeners, in a task requiring perceptual discrimination under two listening conditions: (1) natural vowel duration and (2) neutral vowel duration. The neutral condition will allow this study to explore whether adult Spanish-English bilinguals rely more on duration or spectral cues to discriminate and identify the vowel contrasts compared to the monolingual English-speaking group. A number of studies have reported that this population has difficulty perceiving tense/lax vowel contrasts in English, most of the work has focused on the English contrast /i/-/ɪ/ (e.g. Fox et al., 1995; Bohn, 1995; Flege, 1997; Escudero, 2000; Morrison, 2006, 2008, 2009). These findings have motivated the exploration of the perception of other tense/lax contrasts at the neurophysiological level. A tense/lax AE vowel contrast /ɑ/-/ʌ/ contrast has been chosen for this experiment because these vowels have been reported to present perceptual challenges for adult Spanish-English bilinguals (Flege et al., 1994; Flege et al., 1997; Escudero & Chládkova, 2010) since listeners may not be sensitive to attend all relevant cues that signal distinctions between these vowel pairs.

The vowel contrasts are embedded in the consonantal context *bVb* to minimize coarticulation effects (Strange et al., 1998), and will be presented to the listeners using the oddball paradigm in two tasks while EEG is being recorded. For the first, a passive listening task in which attention is not required, the vowel stimuli (600) are presented to the listeners with an ISI of 800 ms to encourage the use of phonetic/phonemic level of representation of the vowels (Werker & Logan, 1985). For the second, an identification task that requires a conscious attentional response, each vowel stimulus (150) is presented upon listeners' response. At the neurophysiological level, the event related potentials of interest are the MMN (Mismatch Negativity), which is related to the first pre-attentional task; and the P300, related to the attentional task. Examining the perceptual processes occurring in L1-Spanish bilingual listeners at the millisecond level could provide insights into the nature of the phoneme information used by this language population to discriminate and identify non-native tense/lax contrasts.

3. Speech Perception In The Brain

This chapter will look at two levels of evidence regarding the perception of speech in the brain. The first section of this chapter presents an overview of the current neuroimaging research that has shed light on the brain structures that contribute to perception and production of speech in the monolingual and bilingual brain. The remainder of the chapter focuses on how one specific technique, namely the ERP method, has been implemented in several studies and with varied language populations to look at the millisecond neural processing that occurs during speech perception in L1 and L2 listeners.

3.1 Speech perception in the monolingual brain

Post-mortem case studies of people who suffered stroke or brain injury led to proposals by many researchers (notably Paul Broca (1861), Carl Wernicke (1874)) that production and comprehension of language can be localized to specific brain areas: the left superior frontal gyrus for production (Broca's area), and the left superior temporal gyrus for comprehension (Wernicke's area). Figure 3 shows some of the areas related to language and speech perception and production. This approach to the localization of language was later supported by experimental studies, such as those conducted by Penfield and Roberts (1959), who delivered electrical stimulation to the frontal lobes of awake patients who were undergoing surgery. Electrical stimulation of precentral cortical regions resulted in vocalizations, demonstrating the role of the frontal lobe in speech production.

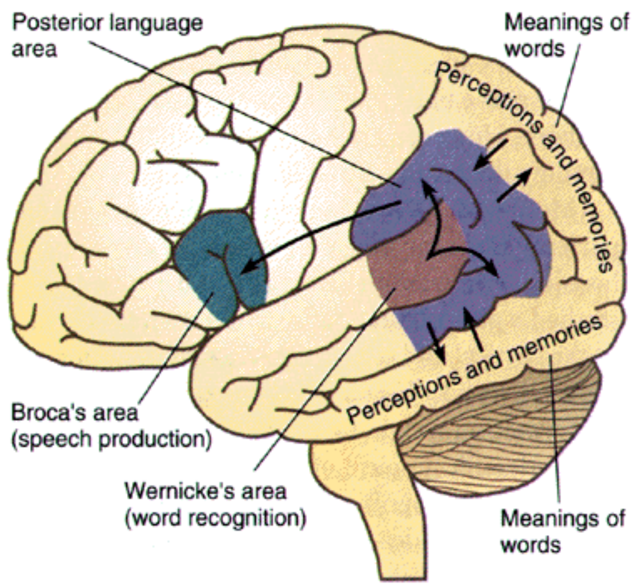

Figure 3. Brain areas related to perception and production of speech and language. Reprinted from http://everythingspeech.com/evaluation/aphasia/aphasia-2/Brain Areas associated with language. (2002). [Online image]. (2002). Retrieved June 12, 2013.

However, since Luria (1966) noted that patients with frontal lesions also showed comprehension deficits, Broca's area has been also associated with aspects of receptive language processing. Liberman and Mattingly (1985) proposed, in their Motor Theory of Speech Perception, that a motor component is crucial for the perception of speech. This proposal has been supported by studies that have used electrical stimulation (Schaffler, Luders, Dinner, Lesser & Chelune, 1993) and brain imaging (Bookheimer, 2002) showing that anterior language areas are involved in both perception and production of language (D'Ausilio, Craighero, & Fadiga, 2012). Hence, there is longstanding evidence that frontal brain regions are not only involved with speech production, but are also implicated in speech perception and language comprehension.

The developments of new technology used to study brain structures and functions have allowed researchers to better understand the interrelations of brain regions, cognitive processing and specific language abilities. Some of these techniques include:

(1) Functional Magnetic Resonance Imaging (fMRI), that measures brain activation as indexed by changes in the regional cerebral blood flow during specific cognitive tasks. This imaging technique provides very high (millimeter) spatial resolution, but its temporal resolution is limited by the speed of the hemodynamic response function. Event-related functional Magnetic Resonance Imaging (efMRI) is an extension of the fMRI method that permits more precise correlation between the timing of stimulus presentation and the hemodynamic response.

(2) Transcranial Magnetic Stimulation (TMS) uses an electromagnetic coil to induce depolarization or hyperpolarization in specific neuronal populations, resulting in activation changes. A variant of TMS, repetitive transcranial magnetic stimulation (rTMS), is under evaluation as a treatment tool for disorders including Parkinson's disease (González-García, Armony, Soto, Trejo, Alegría, & Drucker-Colín, 2011) dystonia (Borich, Arora, & Jacobson Kimberley, 2009), depression (Gross, Nakamura, Pascual-Leone & Fregni, 2007), and auditory hallucinations (Hoffman, Anderson,

Varanko, Gore, Hampson, 2008). The spatial accuracy of TMS relies on additional scanning measures (such as anatomical MRI), and the timing of effects is very variable.

(3) Positron Emission Tomography (PET) uses a radioactive tracer that is injected into the bloodstream. The tracer associates with glucose, which undergoes reuptake in active brain regions, resulting in differential rates of decay for the radioactive isotope. Hence, by tracking the rate of decay of the tracer, images of brain activation associated with stimulus processing can be derived. PET scans provide spatial accuracy in the range of several millimeters, but the temporal accuracy of this technique is extremely limited.

Current models of language processing in the brain do not support a view of distinct language regions like Broca's and Wernicke's areas working in isolation. Gow & Segawa (2009) and Mainy et al. (2008) have described a bidirectional circuit involving fronto-temporo-parietal areas. What is more, a number of brain imaging experiments have shown that areas in the left inferior frontal and premotor cortex (Broca's area) activate in conjunction with temporal areas (Wernicke's area) during processing of syllables and words (Siok, Jin, Fletcher & Tan, 2003; Wilson, Saygin, Sereno & Iacoboni, 2004; Zatorre, Evans, Meyer & Gjedde, 1992). It appears that activation of motor areas during speech perception is related to the sensory-motor nature of the perceptual representations (D'Ausilio, Craighero & Fadiga, 2012).

Specific areas of the temporofrontal neural circuitry respond differently to speech sounds depending on attentional demands. For example Hugdahl, Thomsen, Ersland, Mortem-Rimol and Niemi (2003) studied changes in neural activation as a result of directed attention to speech stimuli that differed in semantic content. Using event-related fMRI, the authors presented isolated Norwegian and Finnish vowels, pseudowords, and real words to adults in two conditions: (1) passive listening, and (2) directed attention. During the passive listening condition there was significant bilateral activation in the superior temporal gyri, suggesting a general involvement of these areas in response to all speech stimuli. On the other hand, during the directed attention condition, when attention was directed to pseudowords, bilateral activations were observed in middle temporal and some frontal regions. When attention was directed to vowel sounds, there was a significant increase in activation in the superior and medial temporal areas, especially left temporal areas. It was concluded that upper posterior sections of the temporal lobes and Heschl's gyrus (primary auditory cortex) are activated in response to speech sounds. However, more specific phonetic processing appears to occur more anteriorly and ventrally.

To examine the role of motor cortex in the perception of speech, Watkins and Paus (2004) applied TMS over left primary motor cortex to measure motor excitability during 4 conditions: (1) listening to speech (Speech condition); (2) viewing of speech-related lip movements (Lips condition); (3) viewing of eye and brow movements (Eyes condition); and (4) listening to and viewing noise (Control condition). In addition they used Positron Emission Tomography (PET) to determine whether changes in brain activation as indexed by cerebral blood flow were related to changes in evoked motor potentials from the lip muscles. The results of the experiment showed that the visual conditions (Lips and Eyes) did not produce any significant motor potentials, and did not differ from one another. In contrast, there were significant motor potentials in the Speech condition compared to the Eyes condition. These results support the previous findings indicating that listening to speech increases the excitability of the motor system that is related to speech production. Hence, there is a body of evidence supporting the view that speech perception processes are not confined to temporal areas, and that other frontal regions are also crucially involved.

Hickok and Poeppel (2000; 2004; and 2007) proposed a dual-stream model of speech processing that is bilaterally organized in the brain, and in which there are multiple routes to lexical access that act as parallel channels that permit processing of multiple and often redundant spectral and temporal cues in the speech signal. The authors support their proposal with evidence from patients with unilateral hemispheric damage, split-brain patients (patients who have no corpus callosum as a result of surgical sectioning), and individuals who have experienced Wada procedures (a procedure during which one hemisphere of the brain is anaesthetized to study lateralization of language and memory). The findings suggest that there is at least one pathway in each hemisphere to access the mental lexicon (the individual's mental repertoire of words).

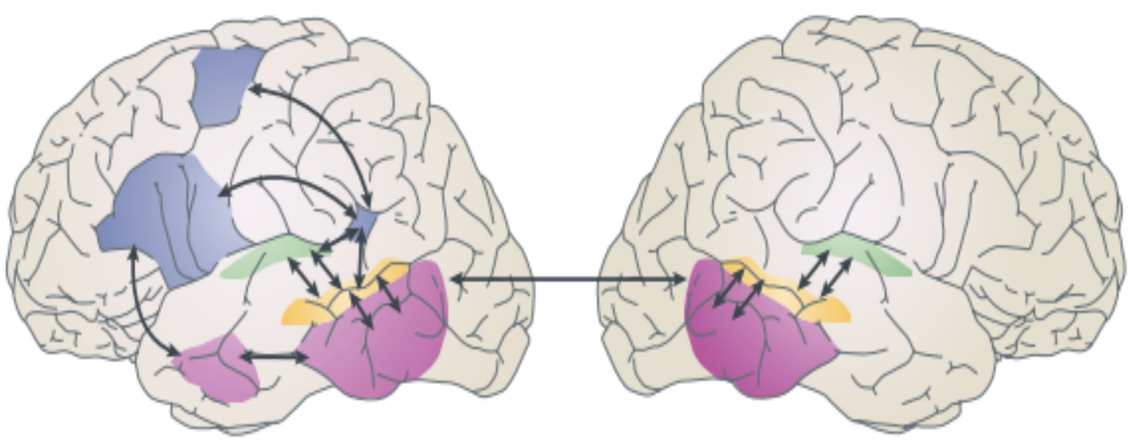

Figure 4. Anatomical locations of the dual-stream model components. Anatomical locations of the dual-stream model components. Adapted from Hickok & Poeppel (2007), p. 395. Copyright by Nature Publishing Group 2007. Green shades correspond to regions on the dorsal surface of the superior temporal gyrus (STG), involved in analysis of spectral and temporal cues. Yellow

shaded regions correspond to the posterior half of the superior temporal sulcus (STS), implicated in phonological-level processes. Pink areas represent the ventral stream in both hemispheres, associated with mapping between sound and meaning. Blue areas denote the dorsal stream in the left hemisphere, where sound is mapped onto articulatory representations.

Hickok and Poeppel's model of speech processing proposes that auditory processing is not specific to speech at the levels of the cochlear, brain stem, and thalamic nuclei. Early cortical stages at the Superior Temporal Gyrus (STG) are bilateral. Then, two streams are identified as processing different types of data: The *ventral stream* is related to speech comprehension, mapping sound into meaning. The *dorsal stream* is related to mapping sound into articulatory-based representations. Hickok & Poeppel (2004) suggest that the *dorsal stream* is left-dominant and the ventral stream is less left-lateralized with a bilateral component. According to this proposal, the *ventral stream* projects ventro-laterally to the Superior Temporal Sulcus (STS), the posterior, inferior temporal lobe (pITL) where the structures of the superior and middle portions of the temporal lobe act as an interface between sound-based representations in the STG and conceptual representations that are widely distributed (Damasio & Geshwind, 1984). On the other hand, *the dorsal* stream possibly includes frontal areas such as Broca's area, the frontal operculum and insula, the motor face area, and the dorsal premotor cortex. The dorsal stream projects dorso-posteriorly towards the parietal lobe, then to the frontal regions, and finally to the sylvian fissure at the boundary between parietal and temporal lobes.

The region at the Sylvian-Parietal-Temporal junction coordinates the mapping between auditory and motor representations, since articulatory gestures are planned in auditory space and then mapped onto motor representations. Hickok and Poeppel posit that both streams work bi-directionally. In *the ventral stream,* the pITL networks mediate the relations between sound-meaning for perception and production, while sectors of the STG are part of sub-lexical aspects of both perception and production. The temporal parietal system in the dorsal stream maps auditory speech representations onto motor representations, and there is simultaneous mapping from motor representations to speech representations. The authors also propose that speech perception tasks (those related to discrimination of speech segments – syllables, phonemes) rely mostly on the dorsal stream, whereas speech recognition tasks (those related to mapping sounds into meaning) rely more on the ventral stream and some of the left superior temporal gyrus (STG). The model proposed by Hickok and Poeppel (2000) not only brings together more evidence on the participation of specific neural structures on the perception and production processes of speech, but it also provides an account of the interactions between these structures through different projections, both bilateral (ventral stream) and left-dominant (dorsal stream). Hickok and Poeppel's proposal provides a framework for the development of functional hypotheses based on neuroanatomical-defined pathways. (As an example, a recent fMRI study that examines the ventral and dorsal pathways in processing L1 and L2 connected speech will be presented in the next section).

Taken together, research using neuroimaging tools has found that the traditional view of separate localizations for language production and comprehension provides (at best) an incomplete account of the brain structures involved in speech processing. The proposal by Hickok and Poeppel (2000, 2004 and 2007), is based on brain imaging evidence, and reveals a view of speech- and language-related brain structures as participating in complex and interactive bilateral neural circuitries. This is a more comprehensive account of the structures and connections that participate in perception and production of the speech signal. It also provides a way to interpret the role of motor representation in speech perception. The next section presents relevant studies showing that the representation of non-native speech sounds in the bilingual brain is both similar to, and different from, such representation in the monolingual brain.

3.2 Speech perception in the bilingual brain.

One of the major questions regarding how speech perception occurs in the brain of a bilingual individual concerns whether two languages recruit the same areas or distinct areas in the brain, and whether the neural mechanisms used by bilingual individuals to process L2 sounds are the same as those used by monolinguals. One research direction that shed light on this question examined bilingual adults with aphasia (a language disorder that results from a brain injury). Some of the observations of recovery of communication abilities in bilingual aphasia led to the conclusion that both languages must be localized in the same brain areas (see Fabbro, 1999, for a review). However, a number of patterns of recovery have been identified in bilingual patients with aphasia.

Paradis (1977) described five different patterns of language recovery in patients with aphasia:

(1) *Parallel recovery*, when both impaired languages improve in a similar way and extension. It accounts for about 65% of recovery cases in studies such as Frabbo (2001) and Paradis (2001).

(2) *Differential recovery,* when one of the languages improves to a greater extent than the other.

(3) *Selective recovery*, when only one of the languages recovers and is blended with inappropriate LM (language mixing).

(4) *Successive recovery,* when complete recovery of one of the languages precedes the recovery of the other.

(5) *Antagonistic recovery,* when recovery of L1 and L2 follow opposite patterns, and one of the languages improves as the other regresses.

Although these patterns of recovery are influenced by multiple factors, the variability in the observations led researchers to suggest that L1 and L2 must be supported by different brain networks (Junqué, Vendrell, & Vendrell, 1995, Kim, Lee, & Hirsch, 1997; Fabbro, 1999). An fMRI study conducted by Kim et al. (1997) on early and late bilinguals provided more evidence on the localization of L1 and L2 in the brain. Participants were asked to silently generate sentences, alternating between their L1 and L2, describing events that occurred during a specified period of the previous day. For late bilingual participants, fMRI revealed differential activations within Broca's area for their two languages. This was not seen for the early bilinguals, and neither group showed differences in Wernicke's area activations. Kim et al. proposed that these findings reflect different learning mechanisms in early vs. late bilingualism:

young children acquire grammar and phonological systems simultaneously and unconsciously because of the early and repeated exposure to both languages, whereas, when a second language is learned in adulthood, explicit learning is required. In early bilingualism, therefore, it seems that Broca's area does not need to undergo any modification subsequent to L2 exposure.

Recently, Ressel, Pallier, Ventura-Campus, Diaz, Avila and Sebastian-Galles (2012) conducted an MRI study comparing the brains of simultaneous or early sequential Spanish-Catalan bilinguals to a group of Spanish monolinguals matched for education, socio-economic status and musical experience. Using a manual volumetric measure of Heschl's gyrus (HG, primary auditory cortex) they confirmed previous findings indicating that bilinguals have larger HG than monolinguals (Golestani, Molko, Dehaene, LeBihan, & Pallier C, 2006; Wong, Warrier, Penhune, Roy, Sadehh, Parrish, & Zatorre, 2008). Results of this study also suggested that differences in the size of HG were related to experience with a second language and not to an innate feature.

Hesling, Dilharreguy, Bordessoules, and Allard (2012) used fMRI to examine differences in the ventral and dorsal pathways in two groups of adult late bilinguals with different L2 proficiency levels (high and moderate) while performing a comprehension task in both L1 (French) and L2 (English). Hesling et al. examined the ventral and dorsal pathways to determine degree of mastery of various speech components (prosody, phonology, semantics, and syntax) that are embedded in connected speech and that vary according to the degree of proficiency in native and foreign languages. The authors found that L1 and L2 connected prosodic speech stimuli share the same dorsal and ventral neural activation in highly proficient L2 subjects. Conversely, moderately-proficient L2 subjects only exhibited common L1 and L2 activations in the STS and the MTG (Medial Temporal Gyrus), and did not show significant activation in the ventral pathway while processing L2 and L1. Hesling et al. concluded that different processes of L2 are supported by differences in the integrated activity within distributed networks that included the left STSp (posterior Superior Temporal Sulcus), the left Spt (Sylvian Parietal Temporal region), and the left pars triangularis.

Archila-Suerte, Zevin, Ramos and Hernandez (2013), studied how neural mechanisms for speech perception change through childhood. In an fMRI study, the authors analyzed brain activation of two groups of Spanish-English bilingual children, younger (6 -8 years), and older (9-10 years), compared to groups of monolingual children, monolingual adults, and early sequential bilingual adults. Participants watched a silent movie while they were auditorily presented with English syllables /sæf/, /sɑf/, /sʌf/ (as in *hat*, *hot*, *hut*) recorded by a male English monolingual speaker. Results revealed that monolingual and bilingual listeners recruit different perceptual areas in childhood and adulthood for processing of non-native speech sounds. Monolingual children showed left-lateralized activation of the superior temporal gyrus (STG) in response to the native contrasts. There was stronger activation of the parahippocampal gyrus and right hemisphere areas in younger monolingual children compared to older ones. On the other hand, younger bilingual children showed a bilateral activation of the STG in response to non-native speech. Older bilingual children showed bilateral activation of the STG, the superior and inferior parietal lobules, and the inferior frontal gyrus.

The comparisons between monolingual children and monolingual adults revealed greater activations for the children in left thalamus, right precentral gyrus, and the right hippocampus. In contrast, monolingual adults showed more activation in bilateral hippocampus, the right temporal pole, and left medial temporal gyrus. The differences between bilingual children and bilingual adults were evident in older bilingual children (9 -10 years of age) who had more activity in inferior and superior parietal lobules, bilateral cingulate gyrus, and, bilateral precentral gyrus, suggesting more intense activation in sequential bilinguals in late childhood. Similar to previously reported findings, results of this study suggest that speech processing in childhood seems to gradually become left lateralized from an initial bilateral activation (Binder et al., 2000). Young bilinguals recruited similar areas as monolinguals (bilateral STG), while older bilinguals recruited areas such as bilateral STG, right middle frontal gyrus, bilateral superior and inferior parietal lobules, and bilateral inferior frontal gyrus (Binder et al., 2000; Joanisse, Zevin, & McCandliss, 2007).

These studies have reported structural differences in the size of the auditory cortex in monolinguals and bilinguals, indicating that learning a second language, at least in childhood, results in structural and functional brain changes. However, it is still unknown whether learning a second language in adulthood results in the same or similar changes in language related brain structures (Ressel et al., 2012). Studies using neuroimaging techniques such as MRI, fMRI, e-fMRI, TMS, and PET are contributing to a more detailed understanding of similarities and differences among monolingual and bilingual children, and adults, while controlling for factors such as age of acquisition (AOA) and experience with the language. The next section discusses the relevance of the ERP technique in the study of non-native speech perception. It reviews two ERP components, the Mismatch Negativity (MMN) and the P300, and presents research evidence on their suitability to shed light on the spectral and temporal cue processing of speech sounds in sequential bilinguals.

3.3 The event-related potential technique (ERP)

This section will briefly review some of the basic concepts (that were introduced in chapter 1) associated with the ERP components MMN and P300. Next, it presents how the components have been implemented in research focusing on non-native speech perception. Finally relevant ERP studies are described, discussing neurophysiological and behavioral evidence on perception of non-native phonemes in sequential adult bilinguals.

As mentioned in chapter 1, electroencephalography (EEG) has been utilized as an index of brain functions by recording the electrical activity generated by large populations of neurons as summed voltages at the scalp. It is a non-invasive procedure that can be applied repeatedly in individuals, and carries minimal risk. The high temporal resolution (millisecond precision) of this neurophysiological technique makes it very suitable for examining early brain responses to quickly processed auditory speech events. Although its spatial resolution is not as precise as that of functional magnetic resonance imaging (fMRI), it is possible to identify signal generators of specific components recorded using EEG. In particular, high-density EEG yields useful information about the spatial as well as temporal dimensions of brain activity.

Event-related potentials (ERPs) are derived from the continuously recorded EEG by averaging together the time-locked or synchronized brain responses to multiple instances of a cognitive event – such as a sound being heard, or a word being recognized. This averaging process enhances the signal to noise ratio of recordings and removes non-phase-locked activations so that (in principle) only activations related to the event of interest is represented in the averaged data – referred to as the event related potential. ERP components are characterized by simultaneous multi-dimensional online measures of polarity (negative or positive voltage deflections), amplitude, latency, and scalp distribution. Several ERP components have been identified as reflecting periods during which particular processes are occurring. The MMN and the P300 are two of the most common ERPs implemented to study second language processing.

3.3.1. The Mismatch Negativity (MMN) component of the ERP as a neurophysiological index of non-native speech perception. The mismatch negativity (MMN) is a neurophysiological brain response that is elicited by discriminable stimuli, usually through the oddball paradigm (Aaltonen, Niemi, Nyrke, & Tuhkanen, 1987; Alho et al., 1998; Bradlow et al., 1999; Cheour et al., 1998; Dehaene-Lambertz, Dupoux, & Gout 2000; Friederici, Friedrich, & Weber 2002). In the case of auditory stimuli, the MMN is an index of involuntary change detection in the stream of standard (same) stimuli, when a deviant (different) stimulus is presented (Escera et al., 2000). The elicitation of an MMN response requires that the central auditory system has formed a representation of redundant aspects of the auditory input before the occurrence of the deviant stimulus (Winkler, Cowan, Csépe, Czigler & Näätänen, 1996). The MMN response results from a mismatch between aspects of an incoming auditory stimulus and a short-term memory trace in the auditory cortex representing repetitive aspects of preceding auditory events, which usually lasts for a few seconds. Thus, there is no MMN elicitation to single sounds with no preceding sounds during the last few seconds (Korzyukov et al., 1999;

Näätänen & Picton, 1987). The MMN is usually presented as a peak in the *difference wave* obtained after subtracting the average time-locked response elicited by standard sounds, from that elicited by deviants.

Magnetoencephalography (MEG) is a technique that is related to EEG, but instead of measuring electrical potentials it measures the associated magnetic field fluctuations (Csépe, Pantev, Hoke, Hampson & Ross, 1992). MEG studies looking at the MMNm (the magnetic equivalent of the MMN) have shown signal maxima peaks at bilateral superior temporal cortices (Alho, 1995; Alho et al., 1998a and b; Csépe et al., 1992; Hari et al., 1984; Levänen et al., 1996). Research has indicated that the MMN component is generated by bilateral sources in the auditory and frontal cortices, and its fronto-central distribution has been attributed to the sum of the activity generated in superior temporal cortices (Jemel, Achenbach, Müller, Röpcke, & Oades, 2002). The maximum amplitude peak of the MMN response (voltage difference between a pre-stimulus baseline and the largest positive-going peak of the ERP) is usually observed at a latency (time from stimulus onset to the point of maximum positive amplitude within the time window) of 150 to 250 milliseconds (Näätänen, Paavilainen, Rinne & Alho, 2007).

The MMN component, which is a response to the deviant stimulus, is preceded by a complex of obligatory ERP responses to the standard stimulus (N1-P2) that also reflect central auditory processing of speech stimuli in the absence of conscious attention, indicating activation associated with a cognitive matching system that compares sensory inputs with stored memory (Tremblay, Piskosz, & Souza, 2003). The N1 component peaks around 100 ms after stimulus onset, while the P200 has a latency of 180 to 200 ms (Kreegipuu & Näätänen, 2011).

The amplitude and latency of the MMN are influenced by various factors, including:

(1) The magnitude of the deviance. The amplitude peak is greater and the latency is shorter in response to a more pronounced deviance between the standard and deviant stimuli in the stream of sounds (Escera, 2000). Large deviances can cause very short latencies for the MMN peaks that can result in MMN overlap with previous sensory ERP components (such as the auditory N100) (Campbell, Winkler, & Kujala, 2007).

(2) The probability of deviance occurrence. MMN amplitude decreases as the deviant-stimulus probability occurrence increases. For example, the MMN to the third deviant presented sequentially in the auditory stream, is smaller in amplitude to the first deviant presented after a number of standard stimuli (Näätänen, Paavilainen, Alho, Reinikainen, Sams, 1987; Haenschel, Vernon, Dwivedi, Gruzelier, Baldeweg, 2005). In order to maintain the low probability of the deviants, standard stimuli are usually presented between 85% - 80% of the trials, while deviant stimuli consist of the 15% to 20% of total trials.

(3) The length of the inter-stimulus interval (ISI). No MMN is elicited when the ISI is long (several seconds). When there is a shorter ISI between the standards, the MMN amplitude tends to get larger (Näätänen & Picton, 1987; Sabri & Campbell, 2001).

(4) Familiarity. MMN responses to familiar (native) speech contrasts are left lateralized and show higher amplitude and shorter latencies compared to MMN responses to unfamiliar (non-native or L2) contrasts (Zevin, Hia, Maurer, Rosania & McCandliss, 2010).

3.3.2. The P300 component as an index of attention allocation and cognitive effort to discriminate auditory stimuli. Discrimination of auditory stimuli generates a relatively large, positive-going waveform with a latency of about 300-800 ms (Toscano, McMurray, Dennhardt, & Luck, 2010) when elicited using auditorily-presented stimuli in adults (Polich & Kok, 1995). This is referred to as the P300, a positive deflection that, like the MMN, can be elicited through the oddball paradigm. However, unlike MMN, for P300 elicitation the participant must be actively engaged in the task of detecting deviant stimuli. Properties of the elicited P300 depend on listener performance during the task and on individual internal factors, rather than on the physical properties of the stimulus (Luck, 2005). The P300 component is considered an index of conscious cognitive processing because it is observed during conscious discrimination tasks.

Two subcomponents of the P300 have been described. One, the P3a, a frontal/central positivity that usually follows a MMN response, indexes orientation to a deviant stimulus in a passive task (Strange & Shafer, 2008), and is associated with novelty-related activations mostly in the inferior parietal and prefrontal regions (Linden, 2005). The other, referred to as the P3b, is generally observed as central/parietal positivity and is related to conscious attention to the stimuli (Katayama & Polich, 1996), with target-related responses in parietal and cingulate cortex. P300 effects have different generators based on modality, with visual stimuli eliciting responses from inferior temporal and superior parietal cortex, and auditory stimuli eliciting superior temporal responses (Linden, 2005). The P3b can be contrasted with the P3a, which has a more frontal distribution (with different neural generators) and is generally elicited by stimuli that are either novel or highly salient (Polich, 2003; 2004; 2007). For the purposes of this study, the P3b subcomponent of the P300 is of most interest since it will reflect listeners' perception of the AE vowel contrasts /ʌ/-/ɑ/ and /ɛ/-/æ/ during a conscious identification task in an oddball paradigm. The amplitude and latency of the P300 component is affected by:

(1) The improbability of the deviants: P300 increases in amplitude as the probability of a deviant stimulus occurrence decreases (Duncan-Johnson & Donchin, 1977; 1982; Johnson & Donchin, 1982).

(2) Attention and cognitive effort. Amplitude of the component is greater when the participant dedicates more effort to the task, but smaller when the participant is hesitant with respect to whether a given stimulus is a standard or a deviant (Polich, 2007; Polich & Kok, 1995).

(3) Location of the measured component. Latency is shorter over frontal areas and longer around parietal ones, and it also varies with the difficulty of discriminating the deviant from the standard stimuli (Polich, 2007). Amplitude of the P300 is also larger over parietal sites (Polich & Kok, 1995).

3.4 Neurophysiological studies of non-native phoneme perception

The increasing number of adult cross-language speech perception studies utilizing the ERP method is providing valuable data on covert language processes that occur at the millisecond level, long before any conscious behavioral response can be observed, through the use of perceptual discrimination or identification tasks. The MMN and P300 components have been well studied and their respective properties are understood to reflect subtle changes in speech segments (e.g. vowels, consonants) related to both acoustic and phonetic levels. While the MMN can be elicited without attention, it has been established that the P300 component indexes attention allocation and cognitive effort when the listener is focusing on detecting basic and high-order perceptual changes in speech (Polich, 2007).

Dehaene-Lambertz (1997) tested French adults in the perception of the retroflex/dental Hindi contrast /da/-/ɖa/ and the native French contrast /da/-/ba/. The stimuli were six syllables from a continuum ba-da-ɖa, each syllable with a length of 275 ms. The experiment consisted of two conditions (native and non-native). In each condition participants were presented with 3 different types of trials (Control, Within category, and Across category). Trials were composed of blocks of 4 syllables with an inter-trial interval of 4 seconds. The first three syllables in the block were identical while the last syllable marked one of the three trial types. Participants were asked to indicate whether the fourth syllable was the same or different from the other three.

Behavioral results showed that L1-French listeners perceived the phonemic boundary between the French /da/-/ba/, but not between the Hindi /da/-/ɖa/. In agreement with the behavioral results, the neurophysiological data yielded a large MMN elicited by native phonetic deviants while non-native or within-category deviants did not generate any significant mismatch negativity response. Dehaene-Lambertz concluded that participants’ ability to discriminate non-native phonetic contrast is related to a loss of auditory discrimination abilities. This conclusion is further supported by the fact that although infants are able to discriminate almost all phonetic boundaries used in any human language, this ability is “lost” through a reshaping of the perceptual mechanisms that occur during the first year of life (Dehaene-Lambertz, 1997). This loss of discriminative ability affects perception of some of the phonetic boundaries not used in the native language. Other researchers have suggested that phonetic discrimination abilities may be flexible enough to be modified by learning a second language in adulthood.

In an EEG and MEG study conducted by Näätänen et al. (1997), native Finnish listeners were presented with Finnish phoneme /e/ as the standard stimulus and the Finnish /ö/ or Estonian /õ/ as the deviant stimuli. The MMN responses of Finnish native speakers were significantly enhanced to the familiar Finnish /ö/ compared to the Estonian /õ/. In addition, generators of the MMNm associated with formation of phonemic traces for each language-specific sound were localized to left auditory cortex. The study provides evidence on the formation of memory traces for familiar speech sounds that correlates with the amplitude of the MMN response to the phonemic speech sounds (language-specific), compared to unfamiliar phonemes. The formation of language-specific memory traces for phonemes has been examined by other studies with other language populations such as English–Japanese (Phillips et al., 1995), and Finnish–Hungarian (Winkler et al., 1999).

Winkler et al. (1999) provided insight into the reorganization of brain mechanisms during second language learning. The authors tested the formation of new vowel representation after learning a non-native language. Participants included 10 Finnish speakers, 10 Hungarian speakers fluent in Finnish and 10 Hungarian speakers who had not previously been exposed to

Finnish. The stimuli included synthesized vowel contrast relevant to Finnish but not to Hungarian /æ/-/e/ and a vowel contrast evident in both languages /y/-/e/. Through the oddball paradigm, stimuli were presented in 2 blocks of 700 trials with stimulus onset asynchrony (SOA: the time between onset of one stimulus and onset of the next) of 1.2 sec. The vowel /e/ was presented as the standard stimulus for 82.5 % of trials, and vowels /æ/ and /y/ were deviants presented for 15 % and 2.5 % of the trials respectively. Results showed that the MMNs in Hungarian subjects who were fluent in Finnish were of similar amplitude to MMNs in Finnish participants. Analysis of the neurophysiological data revealed significant MMN responses from the native Finnish speakers and the fluent Hungarian speakers of Finnish to the deviant /æ/. However, the group of naïve listeners showed an absence of the MMN response to the Finnish deviant vowel.

The authors argued that the observed differences between the responses in the two Hungarian groups (fluent-naïve) were related to effects of learning and language experience, since the fluent group had developed representations for the Finnish phonemes during language training. The absence of MMN for the two Finnish /e/ sounds in the naïve group was argued to reflect their lack of detection of acoustic differences between the two sounds, which might have been perceived as variants of the Hungarian /ɛ/. Based on the pre-attentive processing reflected by the MMN responses of Finnish and Hungarian listeners, Winkler et al. (1999) suggested that acquiring fluent command of a foreign language in adulthood enhances the individual's ability to process the phonemes of that language pre-attentively, in a similar way to a native speaker.

Following Dehaene-Lambertz' work in 1997, Rivera-Gaxiola, Csibra, Johnson and Karmiloff-Smith (2000) explored the electrophysiological responses to different types of syllabic contrasts (English labial /ba/ vs. dental /da/; Hindi dental /da/ vs. retroflex /ɖa/) and a within category (two /ba/ tokens) by native English speakers. Participants were 50 native speakers of English. Stimuli were four syllables extracted from a synthesized continuum (/ba/-/da/-/ɖa/). The experiment consisted of 3 conditions (Hindi contrasts, Within category, and baseline). Using an oddball paradigm, with an ISI of 1.5 seconds, the stimuli were presented semi-randomly with standards representing 85% of all trials. MMNs were observed for the native (/da/ to /ba/), and the non-native (retroflex /ɖa/ to dental /da/) trial types. For the native (/da/ to /ba/), the distribution of the component was mostly prefrontal and right frontal. For the non-native contrast (retroflex /ɖa/ to dental/da/), the significant MMNs were observed over the left parietal region.

Results from this study contrast with those reported by Dehaene-Lambertz (1997). While Dehaene-Lambertz reported no significant MMN responses to non-native Hindi contrasts by L1-French naïve listeners, this study found that even though L1-English listeners did not perceive the differences between the Hindi contrasts in all behavioral conditions, there was neurophysiological evidence of perception at the pre-attentive level in all experimental conditions. The authors concluded that there is no loss of perceptual abilities in adulthood. More likely, they propose, there is a reshaping of the neural mechanisms for phoneme perception that occurs after learning a second language (Rivera-Gaxiola et al., 2000).

While behavioral studies demonstrate that L2 learners perceive non-native phonemes differently from native speakers, neurophysiological studies shed light on the possible neural mechanisms responsible for such differences. Behavioral and neurophysiological studies have also suggested that native and non-native listeners weigh spectral and durational cues differently.

L2-learners tend to weigh cues based on their native language cues, which may be inaccurate, especially if they do not have experience with the non-native cues they need to process. This observation has motivated a number of studies comparing perception of durational cues by listeners who are native speakers of languages that do and do not rely on durational cues to signal phonemic contrasts.

The temporal structure of speech sound patterns is contained in the auditory traces that contribute to the process of change detection (Näätänen, Paavilainen, Rinne & Alho, 2007). Therefore, the MMN is also elicited by changes in the temporal aspects of auditory stimulation such as sound duration (Deouell, Karns, Harrison & Knight, 2003; Grimm, Snik & Van Der Broek, 2004; Ylinen, Shestakova, Huotilainen, Alku & Näätänen, 2006). Cross-linguistic MMN differences in the processing of speech-sound duration were found by Minagawa-Kawai, Mori, Sato and Koizumi (2004), and by Ylinen, Huotilainen and Näätänen (2005; 2006).

Kirmse et al.'s (2008) study of MMN responses elicited by changes in vowel duration in adult L1 Finnish and German, showed that even without experience with durational cues in L1, German listeners were able to perceive the differences in non-native vowel duration at the pre-attentive level. These results support Bohn's desensitization hypothesis (1995), which holds that non-native listeners may rely on durational information to identify foreign vowels when the spectral cues are not easy to perceive. Even though German listeners perceived the duration-cued Finnish vowels, MMN amplitude and latency differences between both language groups suggest that the German speakers did not perceive the Finnish vowels in a native-like fashion at the neural level.

In a more recent study, Hisagi, Shafer, Strange and Sussman (2011) studied how selective attention to a non-native (Japanese) vowel duration contrast would yield improvement in discrimination as indexed by ERP responses for American English (AE) listeners. They predicted that Japanese listeners would pre-attentively discriminate the vowel duration contrast since it is phonemic in their language. On the other hand, American English listeners were expected to show less robust indices of pre-attentional perception of vowel duration because duration is not used as a primary phonemic cue in English. Participants included twenty-four native Japanese listeners who had been in the United States for less than 36 months, and 24 AE listeners who never studied Japanese and had no significant second language background.

Two types of stimuli were designed for this experiment. The auditory stimuli consisted of two nonsense words (NSW), *tado* vs. *taado*, naturally produced in a carrier sentence by a native Japanese speaker (Tokyo dialect). The NSW pair represented a vowel duration decrement (*tado*) and a vowel duration increment (*taado*). For the visual condition, stimuli consisted of four different sizes of pentagon and hexagon shapes. The study consisted of 2 different experiments (visual attend condition and auditory attend condition). Both tasks implemented a "categorical oddball" paradigm, presenting standard stimuli for 85% and deviant stimuli for 15% of 1,400 trials. An inter-stimulus interval (ISI) of 800 ms was used to present the NSW; the visual stimuli were presented with a different ISI (780 ms) to avoid presenting both visual and auditory stimuli at the same time. In the visual-attend task, participants were required to ignore auditorily presented NSW and silently count visual deviant shapes. On the other hand, the auditory-attend task required the participants to count auditorily presented speech sounds while watching, but not counting, visual stimuli.

A behavioral task followed the electrophysiological recording and consisted of identifying the same auditory stimuli presentation. Participants pressed a button when they heard a deviant stimulus. Only one block of 100 trials was used for each order or each behavioral task. In general, results from both conditions (auditory attend and visual attend) supported the initial hypotheses that non-native AE listeners' discrimination performances were poorer compared to those of the Japanese speakers. Similar MMN amplitudes were obtained from Japanese listeners during both experiments, whereas AE listeners showed larger amplitude MMNs in response to the deviant auditory stimuli during the attention (auditory) condition compared to the visual experiment. This finding supports the view that attention plays an important role in non-native perceptual performance. Based on these results, researchers concluded that L1 speech perception is automatic and attention is not necessary for differentiating between native phonemes, while focused attention is required to perceive non-native contrasts. This study shows that, when attention is directed to the relevant cues through instructions, naïve listeners perform more accurately in discriminating duration-based non-native contrasts than when the instruction is not directing their attention. Even though naïve listeners adjust their perception to extract new relevant cues from foreign phonemes, there are significant differences in performance between native and non-native listeners. The question now is how L2 listeners adjust their perceptual processes to extract the relevant non-native phoneme cues.

In addition to the studies of auditory perception, training studies are also providing valuable insights into how L2 learners modify their perceptual patterns after specific cue training in the laboratory. Ylinen, Uther, Latvala, Vepsäläinen, Iverson, Akahane-Yamada & Näätänen (2009) focused on training Finnish second-language learners of English to weigh duration and spectral cues. Twelve L1-Finnish and 13 L1-English speakers participated in the experiment. Researchers manipulated spectral and duration cues of English vowels /i/ and /ɪ/ to obtain two versions of minimal pairs containing the target contrast, one version with normal duration, and the other with modified vowel duration (ambiguous or equalized). A native male speaker of Southern England English produced the stimuli (45 minimal pairs) for the behavioral experiment. Nineteen pairs were used both in the pre-training, post-training and training sessions, while 26 pairs were used only during pre-training and post-training sessions.

Investigators expected English speakers to perceive the vowel contrasts without any difficulty since they would be listening to their native vowel prototypes and would have access to the spectral information needed to correctly identify the vowels, even when vowel duration was modified. In contrast, L2-English speakers of Finnish were expected to have difficulties discriminating between the English vowel contrasts because their native language experience would lead them to rely on durational cues rather than spectral information. Also, because the Finnish language has an /i/ vowel that is more similar to the English /i/ than to the English /ɪ/, which is not present in Finnish, researchers expected Finnish listeners to differ from English listeners in their responses to the stimuli. The behavioral task required the identification of minimal pairs containing the English vowel contrast in two versions (normal and modified duration). Participants were instructed to press a left or right button if they heard the word written on the left or right side of the computer screen. Stimuli in the MMN experiment included synthesized words *beat* and *bit*, presented with (a) normal duration and (b) equalized duration.

Using the oddball paradigm, stimuli were presented in 12 blocks, each including 434 stimuli. Each stimulus was presented as standard and deviant, in different blocks. ISI varied within blocks between 400 and 600 ms. In addition, one block of non-speech stimuli was included to assess the stability of the MMN across language groups and test sessions.

Pre training behavioral results showed that words used in the training were more difficult to identify than those used in the pre and post-training testing. L2-English listeners whose L1 was Finnish had a lower percentage of accuracy in identifying the modified-duration English vowels compared to the native English listeners, although they did not differ from native speakers in their ability to discriminate words with normal duration. Post-training evaluation results showed no differences between the groups, indicating that Finnish listeners improved their identification accuracy after training for duration modified vowel /i/, leading to performance resembling that of native English listeners. Finnish listeners had a larger MMN response to the deviant vowel at the left hemisphere and midline sites, compared to the pre-training MMN responses. However, there was not significant pre to post-training improvement in identification of modified-duration /ɪ/ in the Finnish listeners.

In sum, these findings suggest that second language learners change their weightings of specific cues when perceiving non-native vowels after training. They learn to use L2 relevant cues to identify vowels. L1-Finnish learners of English were effectively trained to weight spectral and duration cues in a similar way to English speakers, relying less on duration cues, and using relevant spectral information to discriminate the non-native vowels. The authors also concluded that speech-sound contrasts with multiple cues are the most difficult ones for foreign language learners.

The studies reviewed above show that listeners with diverse L1 backgrounds, with or without experience with a foreign language, and those undergoing specific auditory training, are able to perceive or learn to perceive non-native phonetic cues, improving foreign vowel perception. Naïve and L2-listeners modify, to certain extent, their perceptual parameters to adapt to the relevant information provided and perceive the non-native vowels. In some instances, it is a natural preference for duration information as stated by Bohn (1995); in others this ability is the result of training, or specific manipulation of cognitive demands (attention). However, there are significant perceptual differences affecting foreign vowel sound discrimination at the neural level, between naïve, native and non-native listeners. Native listeners show larger MMN amplitude and seem to be faster to detect the relevant cues that signal phoneme contrasts. Language experience has been a recurrent common characteristic in those non-native listeners, who show perception or improvement in perception of non-native phonemes compared to those who have not been exposed to the foreign language.

Neurophysiological studies examining L1-Spanish adult perception of non-native vowel contrasts are scarce. In a pilot study, Garcia and Froud (2007) described and compared adult sequential bilingual, Spanish-English brain responses (MMN and P300) to an English vowel contrast /i/-/ɪ/ during two conditions (pre-attentional and attentional) (see also Garcia & Froud, 2008; 2010). It was predicted that the MMN response would be greater in amplitude to the deviant stimulus /ɪ/ than to the standard /i/ only if subjects perceived the contrast at the pre-attentive level. Additionally it was predicted that the P300 amplitude would be influenced by how well subjects perceived the difference between the sounds at the attentive level. The target vowels were presented in the CVC (consonant-vowel-consonant) context (dit, dɪt), and were produced by two different female speakers. Participants included eight bilingual Spanish-English speakers (3 males), with a mean age of 29 years; and 5 native English speakers (4 males), whose mean age was 24 years, for a total of thirteen subjects who were all graduate students at various New York City educational institutions. All bilingual subjects learned English in their home countries after the age of 3, and none had lived in an English speaking country before coming to New York. Participants had lived in the English language environment for 16 months in average (range 6-30 months).

In two experiments run back-to-back, subjects (Spanish-English bilinguals and English monolinguals) were asked first to ignore 120 trials of binaurally presented English vowel sounds /i/-/ɪ/ and watch a silent movie while EEG was recorded. Then, they were asked to attend to 120 trials of the same vowels and decide, trial-by-trial, which of the two was being presented. Behavioral responses were recorded via button press. Results showed that English speakers were more accurate in identifying the contrast than the Spanish speakers. In addition, English speakers had a shorter reaction time for both sounds compared to Spanish speakers. The ERP results revealed that Spanish speakers did not perceive the difference between the two English vowels in either condition (pre-attentional or attentional). In contrast, native English speakers showed significantly greater negativity to deviants (MMN) in the pre-attentional condition, and greater positivity (P300) in the attentional condition. This pilot study provided preliminary neurophysiological data on the challenge faced by Spanish speakers who need to perceive specific AE vowel contrasts, in this case /i/-/ɪ/. The group of participants in the experiment learned English in their Spanish-speaking native countries after the age of 3, arrived in the U.S. during adulthood, and had been living in New York City for less than three years. None of them had lived in an English-speaking country before. Flege (1991) found that native Spanish-speakers adults who had lived in the United States since childhood produced English vowels more accurately than those who arrived in the country in adulthood, suggesting that short time of exposure to the L2 vowel sounds might have explained why these subjects were unable to identify the English contrast well enough to elicit the expected ERPs.

Most of the studies that have examined the perception of this contrast by Spanish speakers have examined aspects of behavior, rather than looking at brain responses that can shed light on the underlying representations and processes. However, Lipski, Escudero & Benders (2012) followed up on Escudero's (2009) study and looked at the neurophysiological (MMN) responses of L1-Spanish learners of Dutch to spectral and durational cues in Dutch vowels. Participants included 15 L1 Dutch listeners, and 15 L1 Spanish listeners who learned Dutch as adults. Stimuli were two sets of synthetic tokens corresponding to Dutch vowels /ɑ/ (durationally similar to Spanish /a/) and /aː/. Each varied along spectral (high and low) and durational (short and long) dimensions. Participants were instructed to watch a silent movie while the stimuli were presented. Each stimulus served as standard and deviant, in separate trial blocks. The ISI varied randomly between 850 and 960 milliseconds. Findings revealed consistency between behavioral and ERP results. The MMN at frontal sites was attenuated and latency was longer for Spanish than for Dutch listeners in spectral-cued contrasts, indicating that L1-Spanish listeners' responses to spectral cues are more effortful compared to the native Dutch listeners. This suggests that Spanish listeners have a preference for the use of durational cues as means of identifying the

Dutch vowel contrasts. In addition, results from this study showed that both groups had similar responses to durational cues.

These studies illustrate how research into adult cross-language speech perception has made use of ERP methods to better understand the underlying neural processes that characterize non-native phoneme perception in diverse language populations, in different diverse phonetic contexts, and under various experimental conditions. The MMN, a well-studied ERP component, has been successfully implemented to index early (100-250 milliseconds) pre-attentional detection of differences between native and non-native phonemes. Neurophysiological data supports what behavioral studies have claimed: there is no loss of perceptual abilities in adulthood, but there seems to be a reshaping of those perceptual processes needed to discriminate non-native phonemes (Best, 1995; Flege, 1995; Winkler, 1999; Best & Tyler, 2007; Kuhl, Stevens, Hayashi, Deguchi, Kiritani & Iverson, 2006). In addition, the perceptual reshaping of the highly specialized phonological system seems to happen as the result of learning a second language through regular classes, immersion, experience with the L2 and even laboratory training (Best & Tyler, 2007; Ylinen et al., 2009; Kaan, Wayland, Mingzhen & Barkley, 2007).

Neurophysiological studies that have examined cue weighting in non-native listeners suggest that phonetic cues are processed differently from native speakers, indicating that vowel duration is highly influenced by language experience (Kirmse et al., 2008). Moreover, such studies show that even when relying on L2 secondary cues (e.g. L1 Spanish listeners who rely on duration to perceive Dutch vowels, as shown in Lipski et al. 2012), L2 listeners accurately perceive some non-native contrasts (Lipski et al., 2012). Second language learners, therefore, can learn to weight primary cues in a native-like fashion. Results from studies such as those carried out by Escudero (2005) and Ylinen et al. (2009) show that adult speakers of a second language can learn to weight cues in a native-like fashion after some laboratory training.

Since there is extensive research on the perception of the AE vowel contrast (/i/-/ɪ/) and the perceptual challenge that it represents for the adult sequential Spanish-English bilingual, it is of interest to investigate whether other AE vowel contrasts present similar challenges to this listener group. The AE vowel contrast /ɑ/-/ʌ/ is of interest, because these two vowels have similar spectral values and duration (see values on table 1), and are therefore predicted to pose a challenge for the L1 Spanish listener. Based on the PAM & PAM-L2 model, L1 Spanish listeners are expected to assimilate both members of the AE vowel /ɑ/ to the Spanish /a/ and AE vowel /ʌ/ to Spanish /a/ and possibly to Spanish /o/ in a Two-Category assimilation pattern (Flege, 1991; Fox et al., 1995; Flege et al., 1997; Escudero & Chládková, 2010). The Single-Category assimilation pattern leads to the prediction that discrimination of both vowels would be poor for vowel /ɑ/, but discrimination could be better for vowel /ʌ/. Based on the theoretical framework presented in the previous chapters, chapter 4 states the research questions and hypotheses that guide the present study.

4. The proposed study

4.1 Research questions

Studies comparing perception of AE tense-lax vowel contrasts by Spanish speakers have mainly focused on the AE /i/-/ɪ/ vowel contrast (Escudero, 2000, 2005; Escudero & Boersma, 2002, 2004; Escudero & Chládková, 2010; Flege et al., 1994, 1997; Fox et al., 1995; Morrison, 2006, 2008, 2009) showing that Spanish listeners rely primarily on duration cues to discriminate non-native contrasts that differed in spectral and temporal characteristics. An explanation for this preference has been provided by Bohn (1995), who proposed that when spectral cues are not accessible, non-native listeners rely on durational information to perceive L2 contrastive vowels. This is referred to as the Desensitization Hypothesis (Bohn, 1995).

Nevertheless, in the case of Spanish speakers, this explanation may not apply because the primary cue to distinguish Spanish vowels is spectral information. This means that native speakers of Spanish will already be familiar with spectral differences between vowels. What may represent a perceptual challenge is that the spectral differences between English vowels (formant values) are smaller than the spectral differences between Spanish vowels. For instance, Spanish has one low central vowel /a/, which is closer to the American English /ɑ/ and /ʌ/ than to AE vowel /æ/. The F1 and F2 values of these vowels are very similar (see table 5 for the formant values of these specific vowels). English native listeners are familiar with the detection of these small spectral changes that signal differences between the vowels, but Spanish listeners do not have to make contrastive use of such small differences in their native language. Thus, it may be that when the small spectral differences between tense and lax vowel pairs are not available for

Spanish speakers to discriminate between two or more possible vowel sounds, they rely on other cues such as duration. Escudero & Boersma (2004) stated that when the duration dimension is not used in L1, this dimension can be regarded as a "blank slate" that listeners initially rely on to classify non-native contrasts (Escudero & Boersma, 2004). Escudero (2000) also described developmental stages related to use of durational and spectral cues by L1-Spanish listeners during the learning process, indicating that there is a stage where L1-Spanish listeners demonstrate native-like cue-weighting strategies when perceiving these vowels.

In this context, research questions have been derived that focus on L1-Spanish adult sequential bilinguals perception of AE vowels, and their perceptual reliance on specific vowel cues to discriminate and identify those vowel contrasts. Hence, the proposed study aims to address the following research questions.

Research Question 1

Do adult sequential Spanish-English bilingual listeners show indices of discrimination of AE vowel contrast /ɑ/- /ʌ/ at early stages of speech perception (at the pre-attentional and/or attentional levels), as reflected by behavioral (accuracy and reaction time) and neurophysiological measures (MMN and P300)?

Research Question 2

In the case that adult Spanish-English bilingual listeners show indices of discrimination of the AE vowel contrast /ɑ/- /ʌ/, do they rely more on durational differences to discriminate the vowels, or do they use spectral cues as the primary information to distinguish them, as reflected by behavioral (accuracy and reaction time) and the neurophysiological measures (MMN and P300)?

4.2 Hypotheses

Studies regarding the perception of these AE vowels by L1-Spanish listeners have demonstrated that Spanish listeners provide high similarity ratings for the AE vowels /ʌ/-/ɑ/ to Spanish /a/ and /o/. In Escudero and Chládkova, (2010), AE vowel /ɑ/ was assimilated to Spanish vowel /a/ above 70% of the time, while AE vowel /ʌ/ was assimilated to Spanish vowels /a/ and /o/ between 30% and 70% of the time. According to Best's PAM-L2 model, this finding constitutes a Single Category assimilation pattern for AE vowel /ɑ/, and possibly a Two Category pattern for vowel /ʌ/. Such a pattern would make it difficult for Spanish learners of English to differentiate this contrast, with an advantage on vowel /ʌ/ that could be easier to discriminate if it is assimilated in a Two-Category pattern. Additionally, Fox et al. (1995) concluded that the formant values for the Spanish /a/ falls between /ɑ/, /æ/, /ʌ/. Based on these findings, it is predicted that the AE vowel contrast /ɑ/-/ʌ/ should pose difficulty for the L1-Spanish listeners in the present experiment because each member of the contrast would be assimilated to the Spanish /a/.

Results from studies on the AE /i/-/ɪ/ vowel contrast have led researchers to predict that adult L1 Spanish listeners will rely more on durational cues when identifying the vowels. It is unclear whether L1 Spanish listeners will use the same strategy to discriminate and identify the AE contrast /ɑ/-/ʌ/. Despite such consistency in the findings regarding the use of durational cues by L1 Spanish listeners, Flege et al., (1997) concluded that native Spanish participants used spectral cues more than durational cues, similar to native speakers of English, when perceiving the vowel contrast /ɛ/, /æ/. Of interest is to determine whether adult sequential Spanish English bilinguals' perceptual discrimination of these vowel contrast (/ɑ/-/ʌ/) depends on changes signaled only by duration, or by both spectral and durational cues.

Several studies have found that experience with the language influences how non-native listeners weigh spectral and durational cues (Flege et al., 1997; Winkler et al., 1999; Ylinen et al., 2009). For example Escudero and Boersma (2004) suggested that Spanish learners of two different varieties of English (Southern British and Scottish) go through different paths to perceive sounds from L2. In addition, other studies (Morrison, 2009 and Mayr & Escudero, 2010) have proposed that the perceptual learning is learner-dependent because each learner begins perceptual learning with different perceptual backgrounds and experiences.

Therefore, it could be expected that listeners with more experience with the language, as measured by length of residency, and daily use of the language, rely on spectral and durational cues in a more native-like fashion compared to less experienced listeners. Since differences between AE vowels /ɑ/-/ʌ/ are signaled by both spectral and durational cues, it is expected that L1 Spanish listeners prefer durational to spectral cues for vowel discrimination / identification.

In order to determine whether adult sequential bilingual, Spanish-English listeners perceive the AE vowels and /ɑ/-/ʌ/, and demonstrate a stronger reliance on durational versus spectral cues, the perception of these AE vowels will be examined in two different testing conditions. In addition, the perceptual performance of adult sequential bilinguals, Spanish-English listeners will be compared to the perceptual performance of monolingual American English listeners.

In the first condition, natural vowel duration, listeners will be presented with AE /ɑ/-/ʌ/ as they were naturally produced by a native AE speaker (originally from Pennsylvania and living in New York). Listeners will hear the vowels with all the spectral and durational cues intact.

In the second condition, neutral vowel duration, the duration of each vowel /ɑ/-/ʌ/ will be neutralized – that is, duration values for each member of the vowel pair will be manipulated so that they do not significantly differ from one another. By neutralizing the vowel durations, listeners will be 'forced' to rely on only spectral cues to obtain information about the identity of the vowels. Listeners' performances in this condition will provide information on how indispensable are durational cues for L1 Spanish listeners to perceive the AE vowels /ɑ/-/ʌ/. The goal here is to examine whether adult sequential bilingual, Spanish-English listeners perceive these vowel contrasts based only on the distinctive spectral features of each vowel. Specific hypotheses for the behavioral and electrophysiological data are stated below.

4.2.1 Behavioral hypotheses

4.2.1.1 Natural vowel duration condition. In this condition participants are presented with the AE vowel pair /ɑ/-/ʌ/ with their natural duration values as produced by the native AE speaker, and they will carry out an identification task. In the case that adult sequential bilingual Spanish-English listeners have a non-native perception of the vowel contrasts (heavily influenced by their native language categories), accuracy is expected to be lower, and reaction time slower, compared to the native English group for AE vowel-pair contrast /ɑ/-/ʌ/. This will mean that L1 Spanish listeners will take longer than L1 English listeners to label these vowels due to their difficulty distinguishing between vowel pairs that differ primarily in spectral features. However, if participants have learned to perceive the very subtle differences between the AE vowel pair /ɑ/-/ʌ/, through their experience with the language, it is expected that their accuracy and reaction time on this task will resemble those of native English listeners.

4.2.1.2 Neutral vowel duration condition. Regardless of how well adult sequential bilingual Spanish-English listeners performed in the natural duration condition, it is expected that they will exhibit lower accuracy and slower reaction times when identifying both vowels, compared to monolingual English listeners. This pattern of performance would reflect a strategy whereby the bilingual listeners ignore informative spectral differences between the vowel pairs in favor of durational information. L1 English listeners are expected to rely primarily on spectral cues to identify the vowels, and their accuracy and reaction times should not be affected by the lack of duration differences since durational cues are secondary for them.

4.2.2 Neurophysiological hypotheses. Consistent with the behavioral predictions, ERP components will reflect whether L1-Spanish listeners have a native-like or non-native perception of the vowel contrasts. This will be evidenced by differences in the amplitude and latency of the ERP components MMN and P300 between groups.

4.2.2.1 Natural vowel duration condition. During the passive listening pre-attentional task, the Mismatch Negativity (MMN) is the ERP component of interest. It is expected that the allegedly challenging AE vowel contrast /ɑ/-/ʌ/ will elicit MMN responses of significantly smaller amplitude from the sequential bilingual Spanish-English listeners, compared to L1 English listeners. L1 English listeners are predicted to show significant MMN amplitude enhancement in response to the deviant vowels, reflecting pre-attentional discrimination of native speech sounds. The peak amplitude of the MMN component is expected to be observed around 150 -300 milliseconds after stimulus presentation.

During the attentional task, listeners are asked to identify (by button press) which sound they are hearing. In such a task, P300 is the target ERP component, since it indexes attentional allocation in a perceptual processing paradigm. The responses of the L1 Spanish listeners to the AE vowel contrast /ɑ/-/ʌ/ are predicted to elicit a smaller or non-significant P300 to the deviant vowel compared to the P300 response of L1-English listeners.

4.2.2.2. Neutral vowel duration condition. During both pre-attentional and attentional tasks, when durational cues are removed from the speech sounds, MMN and P300 responses to both vowel contrasts are predicted to be significant for L1 English listeners, but non-significant in L1 Spanish listeners when durational cues are removed from the speech sounds.

Table 6

Summary of the study hypotheses

Study Hypotheses		Natural Vowel Duration		Neutral Vowel Duration	
		English Group	Spanish Group	English Group	Spanish Group
Behavioral	**Accuracy**	Near 100% for both AE vowels /ɑ/-/ʌ/.	Lower than L1-English listeners for AE vowels /ɑ/-/ʌ/.	Near 100% for AE vowels /ɑ/-/ʌ/	Lower than in *natural vowel duration* condition, and lower compared to the L1-English listeners if they rely primarily on vowel duration information.
	Reaction Time	Spanish group will have longer RT than the English-speaking group in both experimental conditions			
Neurophysiological	**MMN**	Significant MMN amplitude enhancement to all deviant vowels, reflecting pre-attentional native speech sound discrimination.	Significant smaller MMN amplitude responses to AE /ɑ/-/ʌ/ compared to L1-English listeners.	Significant MMN amplitude enhancement to deviant vowels, showing that spectral cues are sufficient for to perceive the vowels.	MMN amplitude will reflect whether spectral cues are sufficient for L1-Spanish listeners to identify the vowels at the pre-attentional level.
	P300	Significant P300 amplitude enhancement to deviant vowels.	Significant smaller P300 amplitude responses to AE /ʌ/-/ɑ/ compared to L1-English listeners.	Significant P300 amplitude enhancement to deviant vowels.	P300 amplitude will reflect whether spectral cues are sufficient for L1-Spanish listeners to identify the vowels at the attentional level.

4.3. Research Design and Methods

This study followed a 2 x 2 x 2 x 2 mixed design. Four factors with 2 levels each included language group (Spanish vs. English as the L1), condition (natural vs. neutral vowel duration), vowels (/ɑ/-/ʌ/), vowel status (standard vs. deviant). The two different language groups (L1 Spanish listeners and monolingual American English listeners) constituted the between-subject factor, while the two different testing conditions (natural vowel duration and neutral vowel duration), the status (standard and deviant), and the vowels constituted within-subject factors.

4.3.1 Recruitment and consent form. This study conformed to all regulations established by the Teachers College Institutional Review Board. Participants were recruited from the student body at Teachers College, Columbia University, and from the local Spanish-speaking community. Participants were asked to participate in two experimental sessions; during the first, they heard vowels in the natural vowel duration condition; during the second, vowels were presented in the neutral vowel duration condition. On their first visit, participants read and signed the study consent form (see appendix 1), and were provided with ample opportunity to read and ask questions. The PI answered questions about the forms, procedures, data management and/or participant's rights as they arose, throughout the 2 sessions of the study. During the first visit participants also answered a language background questionnaire (see appendix 1) and during the second session they were screened for normal hearing.

4.3.2 Sample size and power calculations. Estimations of power and appropriate sample size for ERP are notoriously difficult (see, e.g., Picton et al., 2000 for an overview of some of the issues involved in statistical approaches to analyzing data from neuroimaging experiments). Power estimation requires knowledge of the expected percent signal change between two conditions (effect size), as well as estimates of the variability in signal change, and these are usually unknown in brain imaging studies. Signal-to-noise is typically low, due to repeated presentations of stimuli within each condition while subjects are recorded over a period of time.

The experiments described here took 45 (pre-attentive task) and 25 (attentional task) minutes of EEG recording time. The raw data consisted of continuous digital recordings (sampling 250 times per second) of voltage deflections at 128 different points on the participants' scalp. This means that, for this ERP experiment, a time series of approximately 675,000 (i.e., 250 samples per second x 60 seconds per minute x 45 minutes per experiment) data points for each of the 128 sensors for each experiment for each participant were captured. Within the time series data, there are two sources of variability of interest: within-subject time course variability (fluctuations from one time point to another) and within-subject experimental variability (variation in the effectiveness of the experimental conditions at producing a percentage signal change). Analyses of power and sample size for brain imaging data are therefore complex, and little work has been done on generation of power curves for ERP. Sample sizes and numbers of trials per condition have therefore been established with reference to available guidelines (e.g., Picton et al., 2000; Handy, 2005; Luck, 2005) and the previous experimental experiences of the sponsor. Additionally, experimental design parameters to reduce variability will be used where possible (e.g. within-subject variability can be minimized by ensuring trial-by-trial consistency: Handy, 2005; Luck, 2005).

This ERP study addressed two main components, the MMN and P300. These components have been shown to index speech perception of non-native vowels (Näätänen, 1995). Most studies of non-native vowel perception looking at MMN indices have included between 10 and 15 participants in the experimental group, and have presented 1200 trials approximately (Colin, Hoonhorst, Markessis, Radeau, De Tourtchaninoff, Foucher, et al. 2009; Ylinen et al., 2009, Hisagi, Miwako, Shaffer, Strange & Sussman, 2011). The proposed ERP experiment has been designed with 1200 trials in the pre-attentional task (designed to elicit the MMN component), and 300 trials in the attentional task (designed to elicit the P300 component). This design creates an acceptably tolerable length of recording time that meets recommendations for numbers of trials per condition (e.g., Luck, 2005; Handy, 2005). A review of the literature showed that, for some of the most comparable ERP studies (Colin et al., 2009; Hisagi et al., 2011, Kirmse et al., 2008, Lipski, Escudero, & Benders, 2012; Ylinen et al., 2009) on vowel perception using a similar participant group, a mean sample size of 12.4 was used (SD = 1.81). The recruitment of twenty participants for the study falls well within parameters utilized in the extant literature in this field.

4.3.3 Participants.

4.3.3.1 Talker (stimuli producer). A female native speaker of American English, a speech and language pathologist and doctoral student at Teachers College, who has training in IPA, recorded the stimuli to be presented during the EEG experiment tasks. The speaker originated from Pennsylvania and has been living in the New York City area for the past two years. Her accent was judged to be standard Northeastern American English by listeners who identified the vowels in the recordings (see table 8 below).

4.3.3.2 Listeners (participants). Ten adult sequential bilingual, Spanish-English listeners (Study group) and 10 monolingual American English speakers (Comparison group) were be recruited. The inclusion criteria for the experimental group required, first, that all bilingual subjects must have learned English in their home countries after the age of 3; and, second, that they came to live in the United States after puberty (> 15 years old). The comparison group consisted of 10 adult monolingual English-speakers, who reported not to be able to hold a conversation in any other language but English (see appendix 2 for demographic information).

4.3.4 Stimuli. The experimental stimuli consisted of naturally produced syllables containing the target vowels produced by a female AE speaker in citation form (bɑb -bʌb). The bilabial context /bVb/ has been chosen in order to minimize effects of consonant to vowel and vowel to consonant tongue co-articulation (Strange, 2007).

4.3.4.1 Stimuli Recording. Stimuli were recorded in a sound attenuated chamber at Teachers College using Sound Forge System 8.0. A Sound Level Meter microphone was placed 8 centimeters away from the talker's mouth. Talker read a written list of ten repetitions of the syllables, written in IPA, and was instructed to read them in citation form, at a normal speaking rate, enunciating each word clearly without exaggeration, stopping at the end of each page, and counting from 1 to 5 before starting on the next page to avoiding extra noise while turning pages. The talkers had breaks and water as needed to ensure her comfort during recording. All processing of the audio files was performed using Praat software (Boersma & Weenink, 2013).

All audio files were amplitude-normalized in Praat using the "Scale to Peak" function. The duration of the target vowels in the stimuli syllables /ɑ/-/ʌ/ (bʌb-bɑb) was modified in Praat, to have natural and neutral vowel duration for both testing conditions; however, all stimuli maintained the same consonant duration. Acoustical analysis in Praat 5.3 (Boersma & Wenink, 2013) was used to determine the vowel length for each speaker. Vowel duration was measured from the first positive peak in the periodic portion of each waveform to the constriction of the post-vocalic consonant (see figure 5). For the natural vowel condition, each vowel kept its own natural duration, while for the neutral-duration vowel condition each vowel was manipulated so that its duration matched that of its pair. To calculate the neutral values, first, the mean duration from each vowel in the pair was calculated to obtain the mean relative duration value that was imposed on the manipulated files to obtain the neutral vowel stimuli (see table 7). The duration of the stimuli with natural vowel duration was 411 ms, while the duration of the stimuli with the neutralized vowel duration were 377 ms.

Table 7

Mean vowel duration vowels of American English vowels /ɑ/-/ʌ/ as produced by a female American English talker in the bvb context. Means are calculated for the natural vowel duration and neutral vowel duration

Vowel	NATURAL DURATION			NEUTRAL DURATION		
	F1	F2	Duration	F1	F2	Duration
bɑb	903.964 Hz	1319.261 Hz	312 ms	905.976 Hz	1321.478 Hz	276 ms
bʌb	880.027 Hz	1545.569 Hz	240 ms	882.960 Hz	1549.062 Hz	

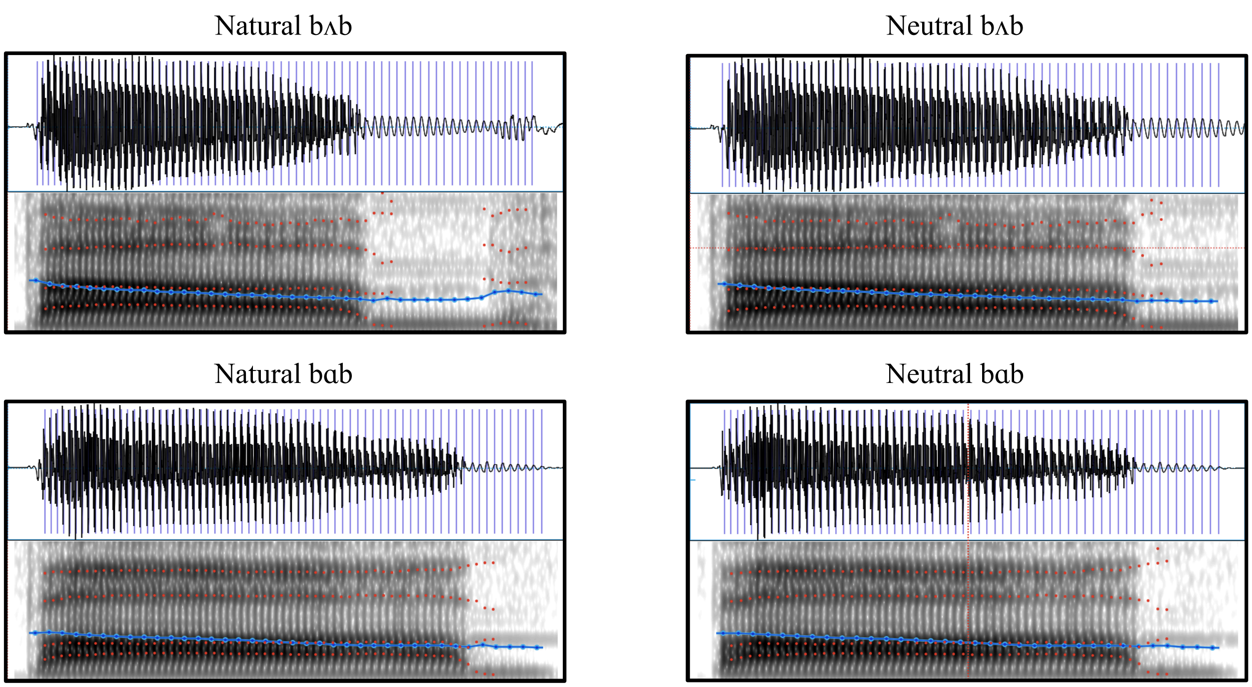

Figure 5. Spectrograms representing the four experimental syllables bʌb and bɑb in the natural (left) and neutral (right) vowel durations.

In order to ensure that the recorded vowel sounds were heard as the intended vowels by native American English listeners before being implemented as the experimental stimuli, 5 native American English listeners (from the New York area) were asked to listen to the 10 repetitions of each of the NSW. The instructions were identical to the ones used for the experimental participants: "Listen to the sounds coming through the speakers, they are made up words in containing American English vowels. If the word contains a vowel sound like in luck – gum (ʌ/*)*, please press button 1. If the word contains a vowel sound like in hot – mop (/ɑ/), please press button 2". Identification response percentages are presented in table 8 below.

Table 8.

Percent correct experimental stimuli identification accuracy by 5 native American English listeners

Syllable	Natural	Neutral
bɑb	97%	1%
bʌb	96%	97%

The high percent correct vowel identification accuracy obtained from five native American English listeners suggested that the recorded experimental stimuli contained the intended AE vowels.

4.3.5 Procedures. All participants read and signed a consent form. The principal investigator answered possible questions regarding procedures or participants' rights. Participants were encouraged to ask or comment their concern at any point of the study. All participants in the experiment were required to complete a language background questionnaire and passed a hearing screening after the experimental tasks.

Stimulus presentation and response collection were controlled using E-Prime software (Psychology Software Tools, Inc., Sharpsburg, PA) at 22.050 Hz through an external RME

Hammerfall DSP audio card connected to a Tannoy OCV 6 full bandwidth pendant speaker suspended 24 inches approximately directly above the participant's head. Timing offset of auditory stimuli was verified using a Cedrus StimTracker, and stimulus presentation volume was adjusted to the participant's comfort between 60 and 70 dB SPL.

4.3.5.1 Measurement of head size and vertex location. All EEG recordings took place in the Neurocognition of Language Lab, in the Department of Biobehavioral Studies at Teachers College, Columbia University. The EEG recording system in use in the lab is a 128-electrode high density HydroCel EEG system manufactured by Electrical Geodesics, Inc. The electrodes are held together in sensor "nets" by fine elastomer that keeps each electrode in a predictable geodesic position relative to all other electrodes in the array. The electrodes themselves are made of carbon embedded in small sponges that before use are soaked in an electrolyte solution to ensure optimal conductivity. This system permits the rapid and accurate application of large numbers of electrodes in high-density arrays with minimal time while maximizing participant comfort and safety. The circumference of each participant's head was measured to ensure the correct size sensor net was selected, and the vertex marked to ensure accurate placement of the net. Each participants was then fitted with the appropriate net. Once the sensor net was situated properly on the head the participant was seated in a chair in a sound attenuated room inside the laboratory that contains computer monitors to deliver the task instructions and stimuli. The sensor net was connected to an amplifier that was previously checked and calibrated.

Impedances (loss of signal between scalp and sensor) were measured by feeding a minute (400 microvolt) electrical field through each electrode, which is then 'read back' by the acquisition system so that the amount of signal loss can be calculated. The electrodes of the sensor net were adjusted as required so that impedances did not exceed 40kΩ. A response button box was provided for the participant to indicate the response to each trial presentation. Sound levels were measured, EEG recorded was initiated and stimulus presentation commenced.

4.3.5.2 Experimental Tasks and Instructions. Artifacts, such as eye blinks, can distort the data and so prior to all experimental sessions participants were instructed to avoid eye blinks and other body movements as much as possible.

4.3.5.2.2 First session of electrophysiological recordings (natural vowel duration condition). In the first recording session, the stimuli were presented with their natural vowel duration. Each session involved the two tasks, (pre-attentional and attentional). The instructions for each of the tasks are presented below. There were breaks built in to the experimental procedures to ensure participant comfort (every 300 trials in the pre-attentional task, and every 75 trials in the attentional tasks).

4.3.5.2.2.1 Pre-attentional Task

The first task was passive and required no behavioral response on the part of the participant. All participants were asked to ignore four blocks of 300 trials of binaurally presented AE vowels pair /ɑ/-/ʌ/ (bʌb-bɑb). To direct the participant's attention away from the auditory input, participants were required to watch a silent movie while EEG was recorded. The instructions to the participants were "Please watch the silent movie and ignore the sounds coming through the speaker". The stimuli were presented in an oddball paradigm, which consist of a series of equal (standard) stimuli that are replaced at some points by different (deviant) stimuli (see figure 6). A single token of each syllable was presented with an inter-stimulus interval (ISI) of 800 milliseconds. According to Werker and Logan (1985), an ISI of 800 milliseconds will tap into phonetic/phonemic levels of phoneme representation, rather than merely acoustic cues. Standard stimuli were present in 85% of the trials (256 out of 300) during each block, and the role of standards and deviants were reversed in the second and fourth blocks (e.g. if vowel /ɑ/ was the standard in the first block, it was deviant in the second block).

4.3.5.2.2.2 Attentional Task

In the attentional task, while EEG was recorded, participants were asked to decide, trial-by-trial, which vowel (bʌb or bɑb) was being presented in the oddball paradigm. The task included four blocks of 75 trails each. Participants were asked to select the response option corresponding to the vowel sound on each syllable they hear by clicking on the button that correspond to the number associated with the vowel sound. The presentation of the new stimulus was initiated when the participant responded to the current stimulus trial. The instructions for this task were: " Listen to the sounds coming through the speakers, they are made up words in containing American English vowels. If the word contains a vowel sound like in luck – gum, please press button 1. If the word contains a vowel sound like in hot – mop, please press button 2".

Participants were asked to read the instructions from the screen. The experimenter never read the keywords containing the target vowels in order to avoid priming the participants with his/her production of the vowel sounds. All instructions appeared in text form on the screen before the experimental sessions began and stayed in the screen during the session in order to avoid participants a working memory overload. Participants were provided with a stimulus response box labeled with numbers 1 and 2 corresponding to the keywords and vowels they were asked to identify. After reading the instructions, and asking possible questions, participants pressed a button to start the experiment.

4.3.5.2.3 Second session of electrophysiological recordings (neutral-vowel duration condition)

The second session (conducted on a different day from the first one) was identical to the first session, except the vowel stimuli had neutralized durations of 276 ms. After completing the second experimental session, all participants underwent a hearing screening.

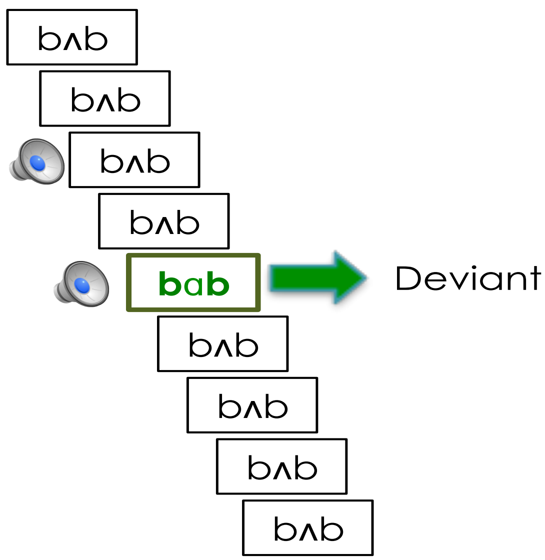

Figure 6. Oddball paradigm used as the experimental task in the study.

4.3.6. Data recording. All behavioral responses obtained during the attentional task (accuracy and reaction time) were recorded using the stimulus presentation software Eprime 2.0 (Psychology Software Tools, Pittsburg, PA). During the EEG data acquisition and preprocessing, scalp voltages were recorded with a high-density 128-channel hydrocel net connected to a high-input impedance amplifier (NetAmps200, Electric Geodesics Inc., Eugene, OR). Amplified analog voltages (0.1-100 Hz band pass) were digitized at 250 Hz. Individual sensors were adjusted until impedances were less than 40 kΩ, and all electrodes were referenced to the vertex (Cz) during recording. The net included channels above and below the eyes, and at the outer canthi, for identification of electro-oculographic artifacts (EOG, associated with eyeblinks and eye movements).

4.3.7. Data analysis.

4.3.7.1 behavioral data analysis. Behavioral responses (button press) in the attentional task were analyzed for each vowel /ɑ/-/ʌ/ in the two duration conditions (natural and neutral). Accuracy was measured by counting the correct responses out of the total number of trials presented, and error trials were omitted from the analysis. Percent correct responses were transformed in arcsine values to approximate a normalized distribution. Reaction time (RT) was considered as the time elapsed from the stimulus presentation to the button-press response. Similar to the accuracy scores, RT values were be log-transformed to approximate a normalized distribution, hence diminishing the likelihood of type I and type II errors (Cohen & Cohen, 1983). Accuracy and reaction time were investigated as dependent variables in three-factor mixed-designed ANOVA, to determine changes in vowel identification from condition 1 (natural-vowel duration) to condition 2 (neutral-vowel duration) between groups (experimental vs. control), and vowels (/ʌ/-/ɑ/). In addition, planned comparison (independent and paired-sample t-tests) will be carried out to determine statistically significant differences between and within groups.

4.3.7.2. Electrophysiological (EEG) data analysis

4.3.7.2.1. data pre-processing. A standard ERP analysis protocol was followed for the analysis of the EEG data (following principles described in detail in Picton et al., 2000; Luck, 2005; and Handy, 2005). The recorded raw EEG data was digitally filtered offline using a 0.1 High - 30 Hz Low bandpass filter, and then subjected to automatic and manual artifact rejection protocols for removal of movement and physiological artifacts (EKG, EMG, EOG). Noisy channels were marked as bad and removed from the analysis. Trials were discarded from analysis if they contain eye movements (EOG 70 μV), or if more than 20% of the channels are bad (average amplitude over 100 μV). In addition, error trials and timeout trials were also removed from the analysis process. Data were re-referenced to an average reference to eliminate the influence of an arbitrary recording reference channel (and to make it possible to incorporate data recorded from the vertex channel in analyses). Average referencing instead uses the average of all of the channels to better approximate the ideal zero reference values. Data were segmented into epochs of 800 milliseconds in length, 700 milliseconds following the onset of each stimulus, and a 100 millisecond pre-stimulus baseline period, to minimize the effects of long latency artifact (such as amplifier drift). Response-locked ERPs were computed within epochs starting at onset of the stimulus and lasting 700 milliseconds. Individuals' averaged data were grand-averaged within groups to enhance statistical power, and to reduce variance due to random noise.

4.3.7.2.2. EEG data analysis. The EEG data were analyzed according to a pre-determined region of interest for each ERP component, MMN (frontal-central region), and P300 (central-parietal region). The montage corresponding to the frontal-central area, where it is expected to generate the MMN response, included electrodes 7, 106, 13, 6, 112, 20, 12, 5, 118, 24, 19, 11, 4, and 124. The montage corresponding to the central-parietal area, where it is expected to generate the P300 response included electrodes 31, 55, 80, 37, 54, 79, 87, 42, 53, 61, 62, 78, 86, 93. Figure 6 below shows the two montages. Responses over sensor montages were examined during specific time windows post-stimulus onset (100-300 milliseconds, when MMN is expected, and a window of 250-500 milliseconds, when P300 can be observed).

Having an experimental designed that presented both vowels in standard and deviant roles, an identity MMN can be calculated to minimize the effect of the physical features of the stimulus in the MMN response. The identity MMN was obtained by subtracting the standard-stimulus responses from the deviant-stimulus responses separately for each vowel (Pulvermüller & Shtyrov, 2006).

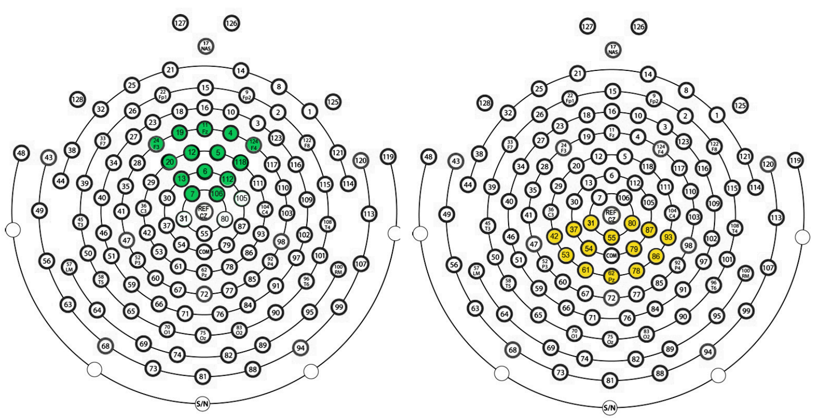

Figure 7. MMN and P300 montages. Left: MMN montage (frontal-central electrodes) in green. Right: P300 montage (central-parietal electrodes) in yellow.

The montaged data were exported in a format permitting further analyses using data analysis packages such as Excel, PASW and MATLAB. Repeated measures analyses of variance (ANOVAs) were conducted to test for main effects and interactions in the data (Dien & Santuzzi, 2005). The ANOVAs were followed by planned comparisons at each level of each significant variable in order to determine the sources of significant main effects and interactions. The dependent variable of interest here was mean amplitude (measured in microvolts) of the predetermined components of interest (MMN and P300). The data were submitted to a 4-factor repeated measures analysis of variance group (the study group of Spanish speakers vs. the comparison group of English speakers), condition (natural vs. neutral), vowel (/ɑ/-/ʌ/), and vowel status (standard vs. deviant). In the next chapter, I present the findings from the study.

RESULTS

This chapter presents the behavioral and neurophysiological results obtained from the Study (Spanish) and Comparison (English) groups during the experimental tasks. The report includes the behavioral and neurophysiological results separately. Chapter 6 (discussion) presents a synthesis and analysis of these findings.

5.1. Behavioral Results

The following results present behavioral data from ten native English speakers (8 female) in the Comparison group, and ten monolingual Spanish speakers (6 female) in the Study group. All participants gave informed consent and received compensation for participation in the study. The age of the English-speaking participants ranged from 20.4 to 30.9 (Mean= 24.01 years of age, SD = 3.48), while the age of the Spanish-speaking group ranged from 22.7 to 35.8 (Mean= 28.1 years of age, SD = 3.984). All participants reported being right-handed and passed a hearing screening at 20 dB HL at 1000 Hz, 2000 Hz, and 4000 Hz bilaterally.

Participants in the study group learned English in their home countries through formal English courses between the ages of 3 and 20 (Mean = 9.4, SD= 5.316) before coming to live in the New York area. The length of residence of the Spanish-speaking participants in USA before testing ranged from 6 months to 5 years (Mean = 1.6 years (SD = 1.424), and had an average age of arrival (AOA) of 26.8 years (SD = 4.666). Spanish-speaking participants reported using English from 25% to 100% percent of the day (Mean = 66.4%, SD = 28.28). English-speaking participants in the comparison group were all monolinguals and while all of them had taken second language classes during high school, all of them reported not being able to hold a conversation in any other language but English. (see demographic information in appendix 1).

All participants completed two perceptual tasks while EEG was recorded: first, a pre-attentional task, in which they ignored auditorily presented syllables containing the target vowels (/bʌb/ and /bɑb/) while watching a silent movie; second, an attentional task, in which participants pressed a button to indicate which vowel was being presented. The data analysis focused on reaction time (RT) and response accuracy (behavioral measures), and the amplitude of ERP components in specific time windows (neurophysiological measure). All procedures were carried out under IRB approval in the Neurocognition of Language Lab at Teachers College, Columbia University.

Accuracy and reaction times were submitted to arcsine and log transformations respectively to obtain data sets that are more normally distributed, hence reducing the probability of type I or II errors (Cohen & Cohen, 1983). Arcsine transformed accuracy and log-transformed reaction times for both vowels /ɑ/-/ʌ/ were analyzed through repeated measures ANOVAS with an alpha level of 0.05 for all statistical tests. An initial exploratory analysis of the data showed that accuracy scores from two Spanish speakers corresponding to vowel /ʌ/ were possible very low score outliers (32% and 20%). An independent samples t-test indicated that those scores were significantly different from the rest of the scores in both groups. Two different analyses of the accuracy data, including and excluding these outliers are presented below.

5.1.1. Accuracy. The Comparison (English) group identified the AE vowels (/ʌ/-/ɑ/) with a higher percent of accuracy compared to the Study (Spanish) group (see figure 1). The comparison group identified vowel /ʌ/ with 92 % accuracy during the *natural vowel duration condition,* and with 93% in the *neutral vowel duration condition.* The comparison group identified vowel /ɑ/ with a 98% accuracy in both *natural and neutral vowel duration conditions.* In contrast, the Study (Spanish) group identified vowel /ʌ/ with 60% accuracy in the *natural duration* and 81% of accuracy in the *neutral vowel duration condition*. Identification accuracy scores for vowel /ɑ/ were with 75% and 93 % in the *natural and neutral vowel duration conditions* respectively.

When the outliers were removed from the analysis, vowel identification accuracy percentage for vowel /ʌ/ changed from 60% to 83% in the *natural vowel duration condition*, and from 81% to 91% in the *neutral vowel duration condition.* Along the same lines, identification accuracy for vowel /ɑ/ increased from 75% to 82% in the natural vowel duration condition, and remained the same, 93%, in the *neutral vowel duration condition* (see table 9).

Table 9.

Average vowel Identification Accuracy for Comparison and Study groups in raw percentage, arcsine transformed, and arcsine back-transformed scores.

	Comparison Group				Study Group with outliers				Study Group without outliers			
	Raw	Arcsine Trans	Back Trans	SD	Raw	Arcsine Trans	Back Trans	SD	Raw	Arcsine Trans	Back Trans	SD
Nat/ʌ/	96.10	1.291	92.40	0.048	77.70	0.890	60.40	0.296	91.10	1.146	0.830	0.191
Nat/ɑ/	99.20	1.444	98.40	0.010	86.60	1.047	75.00	0.202	90.60	1.134	0.821	0.100
Neu/ʌ/	96.60	1.309	93.30	0.041	90.40	1.129	81.70	0.157	95.60	1.273	0.914	0.170
Neu/ɑ/	99.20	1.444	98.40	0.022	96.90	1.321	93.90	0.025	96.60	1.309	0.933	0.123

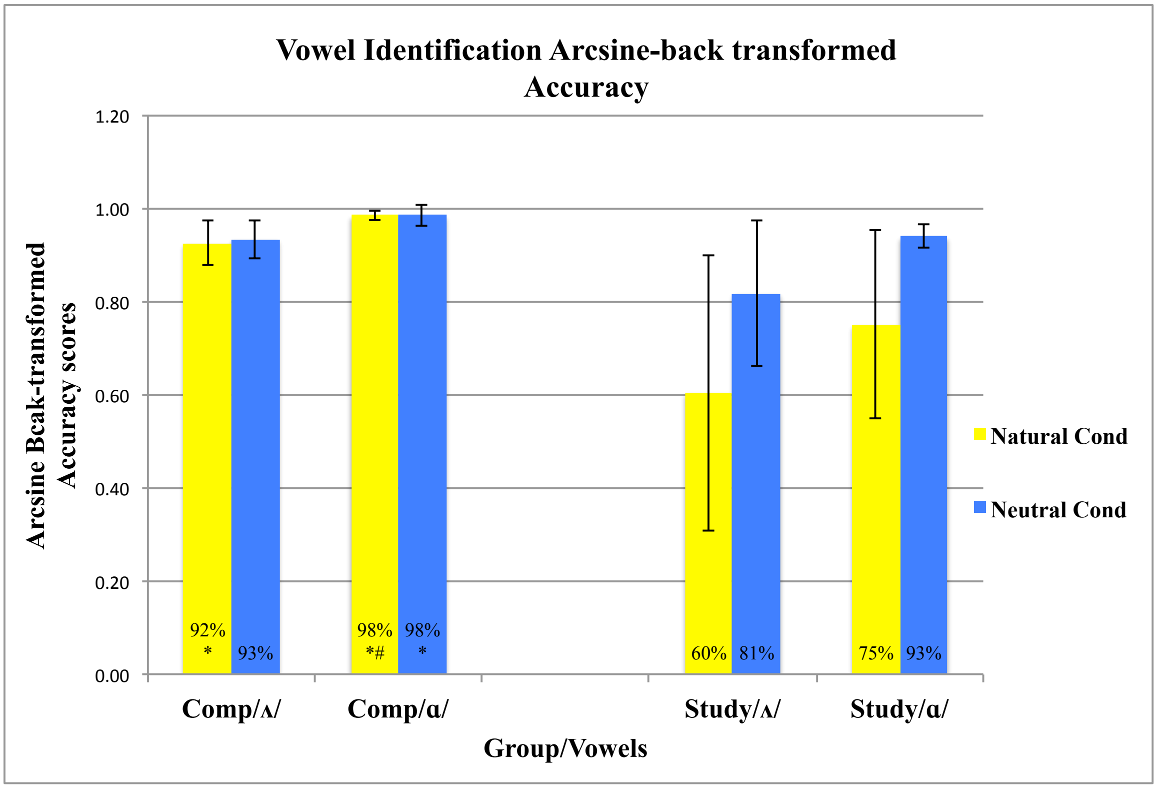

Figure 8. Accuracy. Mean vowel identification accuracy (Arcsine back-transformed) for the Comparison (N= 10) and Study (N= 10) groups during the natural and neutral vowel duration conditions. (* indicates between groups differences, while # sign indicates within group differences).

Accuracy scores (arcsine transformed) were analyzed with and without the outlier cases. All data were submitted to a mixed-designed 3-factor ANOVA comparing language group (Spanish vs. English), Condition (natural vs. neutral), and Vowel (/ʌ/-/ɑ/). Both analyses revealed a significant difference in vowel identification accuracy between Groups (with outliers: F (1, 18) = 1.067, $p = 0.010$, $\eta_p{}^2 = 0.317$; without outliers: F (1, 16) = 5.498, $p = 0.032$, $\eta_p{}^2 = 0.256$). Along the same lines, the main effect of Vowel was significant for both data sets (with outliers: F (1, 18) = 4.981, $p = 0.039$, $\eta_p{}^2 = 0.217$; without outliers: F (1, 16) = 4.684, $p = 0.046$,

$\eta_p{}^2 = 0.226$) suggesting that the Comparison group's accuracy was significantly better than the Study group's, and identification accuracy was different for each vowel in both groups. The main effect of Condition did not reach significance (with outliers: (Condition: F (1, 18) = 2.918, $p = 0.105$, $\eta_p{}^2 = 0.139$; without outliers: F (1, 16) = 1.151, $p = 0.299$, $\eta_p{}^2 = 0.067$), and there were no significant interaction effects (see Appendix 2).

Although interactions were not significant for the ANOVA, planned comparisons (independent samples t-tests) were conducted in order to explore vowel identification accuracy differences between the groups regarding specific vowels as hypothesized. Results indicated that the Comparison (English) group obtained significantly higher accuracy means compared to the Study (Spanish) group for both vowels (/ɑ/-/ʌ/) in the *natural vowel duration condition,* and only for vowel /ɑ/ in the *neutral vowel duration condition.* When the outliers are removed, the only significant difference between the groups was in vowel /ɑ/ in the *neutral vowel duration condition* (see table 10).

Table 10

Planned comparisons: Independent samples t-tests. Vowel identification accuracy differences between Comparison and Study groups

Vowel Identification Accuracy differences between Comparison and Study groups							
Vowel	Group	With outliers			Without outliers		
		Mean (SD)	*t* (18)	p	Mean (SD)	t (16)	p
Nat/ʌ/	Comparison	1.357 (0.191)	-2.252	0.050*	1.357 (0.325)	-1.305	0.210
	Study	1.023 (0.472)			1.219 (0.259)		
Nat/ɑ/	Comparison	1.483 (0.100)	-2.186	0.042*	1.483 (0.100)	-1.882	0.078
	Study	1.195 (0.404)			1.267 (0.347)		
Neu/ʌ/	Comparison	1.366 (0.174)	-1.401	0.225	1.366 (0.174)	-0.51	0.617
	Study	1.226 (0.307)			1.324 (0.181)		
Neu/ɑ/	Comparison	1.517 (0.124)	-2.745	0.013*	1.517 (0.124)	-2.983	0.009*
	Study	1.356 (0.138)			1.338 (0.130)		

*p<0.05

Paired-samples t-test showed significant vowel identification accuracy differences in the Comparison group, but not in the Study group. The comparison group was significantly more accurate to identify vowel /ɑ/ than vowel /ʌ/ in both *natural and neutral vowel duration conditions.* In contrast, there were no significant differences in vowel identification accuracy in the Study group with or without the outliers in the analysis (see table 11).

Table 11

Planned comparisons: Paired-samples t-tests. Vowel identification accuracy differences within Comparison and Study groups

Vowel Identification accuracy differences within groups									
Cond/ Vowel	Comparison Group			Study Group (with outliers)			Study Group (without outliers)		
	Mean (SD)	*t* (18)	p	Mean (SD)	t (18)	p	Mean (SD)	t (16)	p
Nat/ʌ/ Nat/ɑ/	1.357 (0.191) 1.483 (0.100)	-2.381	0.041*	1.023 (0.472) 1.195 (0.404)	1.267	0.237	1.219 (0.259) 1.267 (0.347)	-0.504	0.630
Neu/ʌ/ Neu/ɑ/	1.366 (0.174) 1.517 (0.124)	2.537	0.032*	1.126 (0.307) 1.356 (0.138)	-1.082	0.307	1.324 (0.181) 1.338 (0.130)	-0.202	0.846
Nat/ʌ/ Neu/ʌ/	1.357 (0.191) 1.366 (0.174)	-0.121	0.907	1.023 (0.472) 1.126 (0.307)	-2.079	0.067	1.219 (0.259) 1.324 (0.181)	-1.425	0.197
Nat/ɑ/ Neu/ɑ/	1.483 (0.100) 1.517 (0.124)	-0.636	0.541	1.195 (0.404) 1.356 (0.138)	-1.277	0.234	1.267 (0.347) 1.338 (0.130)	-0.556	0.595

*p<0.05

The inclusion of the two outliers (with respect to accuracy scores) in the Study group did not alter the within group statistical analyses. Therefore, the two outliers were included for all further analyses.

5.1.2 Reaction Time (RT). Figure 9 illustrates the reaction time (log back-transformed) scores from the Study and Comparison groups in the *natural and neutral vowel duration conditions.* The Comparison (English) group had shorter RTs to identify both vowels (/ɑ/ = 719.870 ms, and /ʌ/ = 755.969 ms) compared to the Study (Spanish) group (/ɑ/ = 1036.182 ms, and /ʌ/ = 1057.437 ms), during the *natural vowel duration condition*. This difference was maintained for the *neutral vowel duration condition*; in addition, both groups were faster to respond in the neutral than in the natural vowel duration condition (Comparison group (/ɑ/ = 572.698 ms, and /ʌ/ = 656.737 ms); Study group (/ɑ/ = 703.525 ms, and /ʌ/ = 718.916 ms) (See table 12).

A Mixed-design 3-factor ANOVA analysis comparing language group (Spanish-English), condition (natural vs. neutral), and vowel (/ɑ/-/ʌ/), showed that reaction time (RT) differences between the groups did not reach significance (F (1, 18) = 2.821, p = 0.110, $\eta_p{}^2$ = 0.135). However, significant main effects of Condition (F (1, 18) = 13.869, p = 0.002, $\eta_p{}^2$ = 0.435), and Vowel (F (1, 18) = 7.239, p = 0.015, $\eta_p{}^2$ = 0.287) suggested RT differences in vowel identification between groups related to the vowels and the experimental condition. There were no significant interaction effects. Table 12 presents average RT scores in milliseconds, in log-transformed, and log back-transformed values.

Table 12.

Average vowel identification reaction time for Comparison and Study groups in milliseconds, and log transformed scores

Condition/vowel	Comparison		Study	
	RT in milliseconds	Log Transformed	RT in milliseconds	Log Transformed
Nat/ʌ/	724.918	2.860	1057.437	3.024
Nat/ɑ/	719.870	2.857	1036.182	3.015
Neu/ʌ/	656.737	2.817	718.916	2.857
Neu/ɑ/	572.698	2.758	703.520	2.847

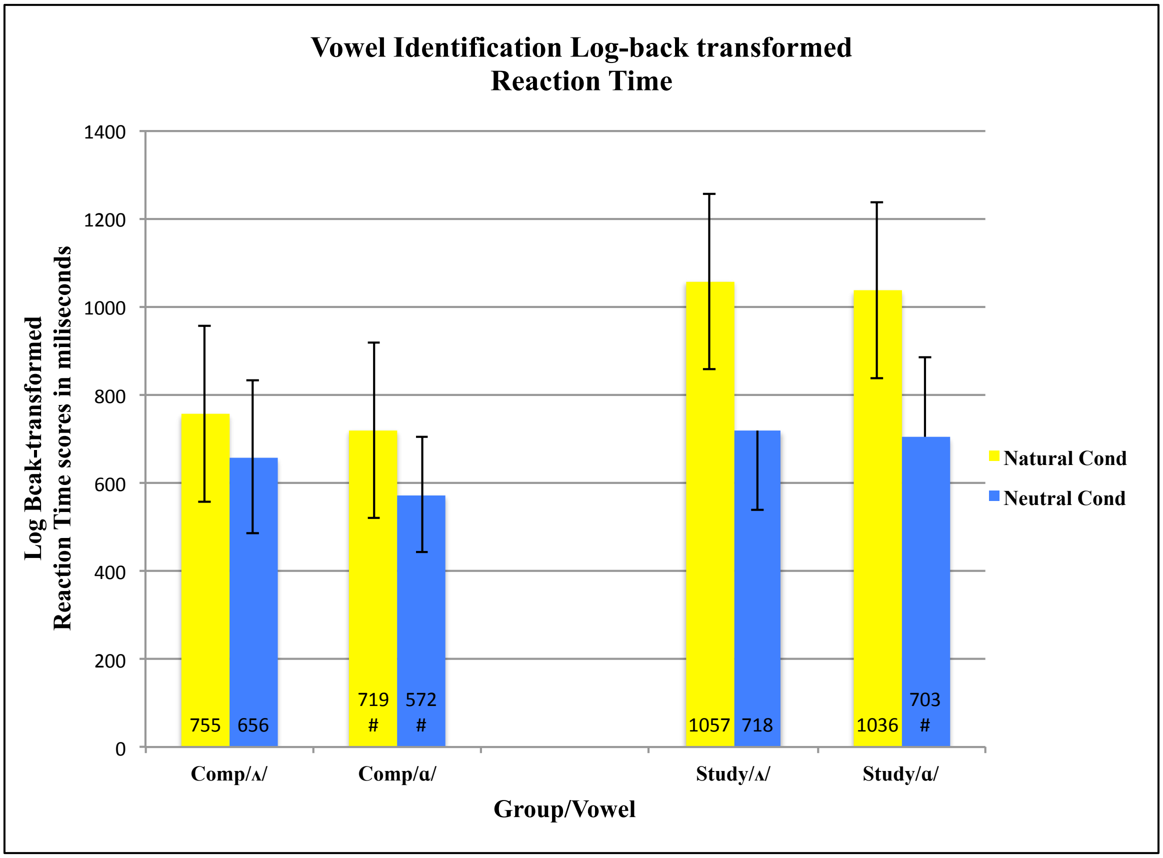

Figure 9. Reaction Time. Mean RT (Log back-transformed) for the study and comparison groups during the natural and neutral vowel duration conditions (* indicates between groups differences, while # sign indicates within group differences).

Planned comparisons (independent samples t-tests) revealed no significant group differences in RT for identification of vowels (/ɑ/-/ʌ/) in any conditions (see table 13).

Table 13

Planned comparisons: Independent samples t-tests. Vowel identification RT differences between Comparison and Study groups

Vowel Identification reaction time differences between Comparison and Study groups							
Vowel	Group	Natural			Neutral		
		Mean (SD)	*t* (18)	p	Mean (SD)	t (18)	p
/ʌ/	Comparison	2.863 (0.114)	1.641	0.118	2.804 (0.113)	0.662	0.516
	Study	2.982 (0.198)			2.839 (0.127)		
/ɑ/	Comparison	2.844 (0.105)	1.670	0.112	2.749 (0.089)	1.842	0.082
	Study	2.968 (0.209)			2.833 (0.114)		

* $p < 0.05$

Paired-samples t-tests revealed significant within group RT differences. The Comparison group was significantly faster to identify vowel /ɑ/ than vowel /ʌ/ in both *the natural* and *neutral vowel duration conditions.* The Study group was significantly faster to identify both vowels in the *neutral* than in the *natural vowel duration condition* (see table 14).

Table 14

Planned comparisons: Paired-samples t-tests. Vowel identification RT differences within Comparison and Study groups

Vowel Identification reaction time differences within groups						
Condition/Vowel	Comparison			Study		
	Mean (SD)	*t* (18)	p	Mean (SD)	t (18)	p
Nat/ʌ/- Nat/ɑ/	2.863 (0.114) / 2.844 (0.105)	1.154	0.287	2.982 (0.198) / 2.968 (0.209)	0.588	0.571
Neu/ʌ/- Neu/ɑ/	2.804 (0.113) / 2.749 (0.089)	2.755	0.022*	2.839 (0.127) / 2.833 (0.114)	0.433	0.675
Nat/ʌ/- Neu/ʌ/	2.863 (0.114) / 2.804 (0.113)	1.806	0.104	2.982 (0.198) / 2.839 (0.127)	2.535	0.032*
Nat/ɑ/- Neu/ɑ/	2.844 (0.105) / 2.749 (0.089)	3.583	0.006*	2.968 (0.209) / 2.833 (0.114)	2.66	0.026*

* $p < 0.05$

In summary, the behavioral results are in line with the behavioral hypothesis stating that compared to the Study group (L1 Spanish), the Comparison group (L1 English) would be significantly more accurate and faster to identify the AE vowels /ɑ/-/ʌ/ in both experimental conditions (*natural vs. neutral vowel duration*). The analysis of vowel identification accuracy data (with and without the outlier cases) revealed significant differences between the Study and Comparison groups, and those differences may be specific to the vowels, rather than the experimental condition. Including the outlier cases, the Comparison group showed a significantly higher vowel identification accuracy for both vowels /ɑ/-/ʌ/ in the *natural vowel duration* condition, and for vowel /ɑ/ in the *neutral vowel duration condition* compared to the Study group. However, without the outlier cases, the vowel identification accuracy differences between the groups remained only for vowel /ɑ/ in the *neutral vowel duration condition*, where the

Comparison group obtained higher mean identification accuracy scores. In addition, the Comparison group had a significantly higher vowel identification accuracy for vowel /ɑ/ compared to vowel /ʌ/ in both experimental conditions, while the Study group did not show any statistically significant difference in the vowel identification accuracy scores.

The analysis of vowel identification RT data showed that the Comparison and Study group did not differ in their RT to identify vowels /ɑ/-/ʌ/ in any experimental conditions (*natural vs. neutral vowel duration*). However, the comparison group was significantly faster to identify vowel /ɑ/ than vowel /ʌ/ in both conditions, and the Study group was significantly faster to identify both vowels in the *neutral vowel duration condition*.

5.2 Neurophysiological results

Analyses for the ERP components MMN and P300 are reported in this section for each vowel pair, with an alpha level of 0.05. The EEG data were pre-processed following standard ERP analysis protocols described in section 4.3.7.2.1 above (based on principles described in detail in Picton et al., 2000; Luck, 2005; and Handy, 2005). Processed data were then averaged over pre-determined regions of interest for each ERP component: fronto-central sensors (predicted location for the MMN component) and central-parietal sensors (predicted location for the P300). Mean amplitude was examined in specific time windows: MMN component was analyzed in a window between 100 – 300 ms after the stimulus onset, while the P300 component was analyzed between 250 – 500 ms. Individual mean amplitudes were averaged within groups for each of the components and submitted to factorial analysis of variance.

A 4-factor repeated measures analysis of variance was carried out, with factors group (study vs. comparison), condition (natural vs. neutral vowel duration), vowel pair /ɑ/-/ʌ/, and vowel status (standard vs. deviant). The ANOVA revealed two significant main effects: Condition: (F (1, 18) = 8.423, p = 0.009, $\eta_p{}^2$ = 0.319), and Status (F (1, 18) = 22.154, p < 0.001, $\eta_p{}^2$ = 0.552), indicating that the MMN mean amplitude response to the vowels differed in each condition (natural vowel duration vs. neutral vowel duration), and depending on the standard vs. deviant status. The main effect of Vowel Pair was not significant (F (1, 18) = 1.902, p = 0.185, $\eta_p{}^2$ = 0.096). A significant interaction between Status and Group (F (1, 18) = 10.330, p = 0.005, $\eta_p{}^2$ =0.365) suggests that the mean amplitude of the MMN component differed between the groups depending on the status of the vowels.

5.2.1 Pre-attentional task (targets the MMN component). Separate 3-factor repeated measures ANOVA (condition (natural vs. neutral), vowel (/ɑ/-/ʌ/), vowel status (standard vs. deviant)), and planned comparisons (paired-samples t-tests) were applied to further analyze the mean negativity amplitude differences observed within groups for the vowel contrast. ANOVA results from the Comparison group revealed a significant main effect of condition (F (1, 9) = 5.471, p = 0.044, $\eta_p{}^2$ = 0.378), and of vowel status (F (1, 9) = 25.162, p = 0.001, $\eta_p{}^2$ = 0.737) suggesting that L1-English listeners' responses to the vowel sounds differed in each condition (natural vowel duration vs. neutral vowel duration), depending on the vowel status (standard vs.

deviant). The main effect of vowel pair (F (1, 9) = 1.630, p = 0.234, $\eta_p{}^2$ = 0.153) was not significant for the Comparison group. No significant interaction effects were observed in this analysis.

Results from the 3-factor repeated measures ANOVA for the Study group (L1-Spanish) revealed no significant main effects (Condition: F (1, 9) = 3.676, p = 0.087, $\eta_p{}^2$ = 0.290; Vowel pair: F (1, 9) = 0.283, p = 0.607, $\eta_p{}^2$ = 0.031; Status: F (1, 9) = 1.479, p = 0.255, $\eta_p{}^2$ = 0.141). There were no significant interactions in this analysis.

5.2.1.1 Natural vowel duration condition. Figure 3 shows Study and Comparison groups' MMN responses to the vowels during the pre-attentional task in the natural vowel duration condition.

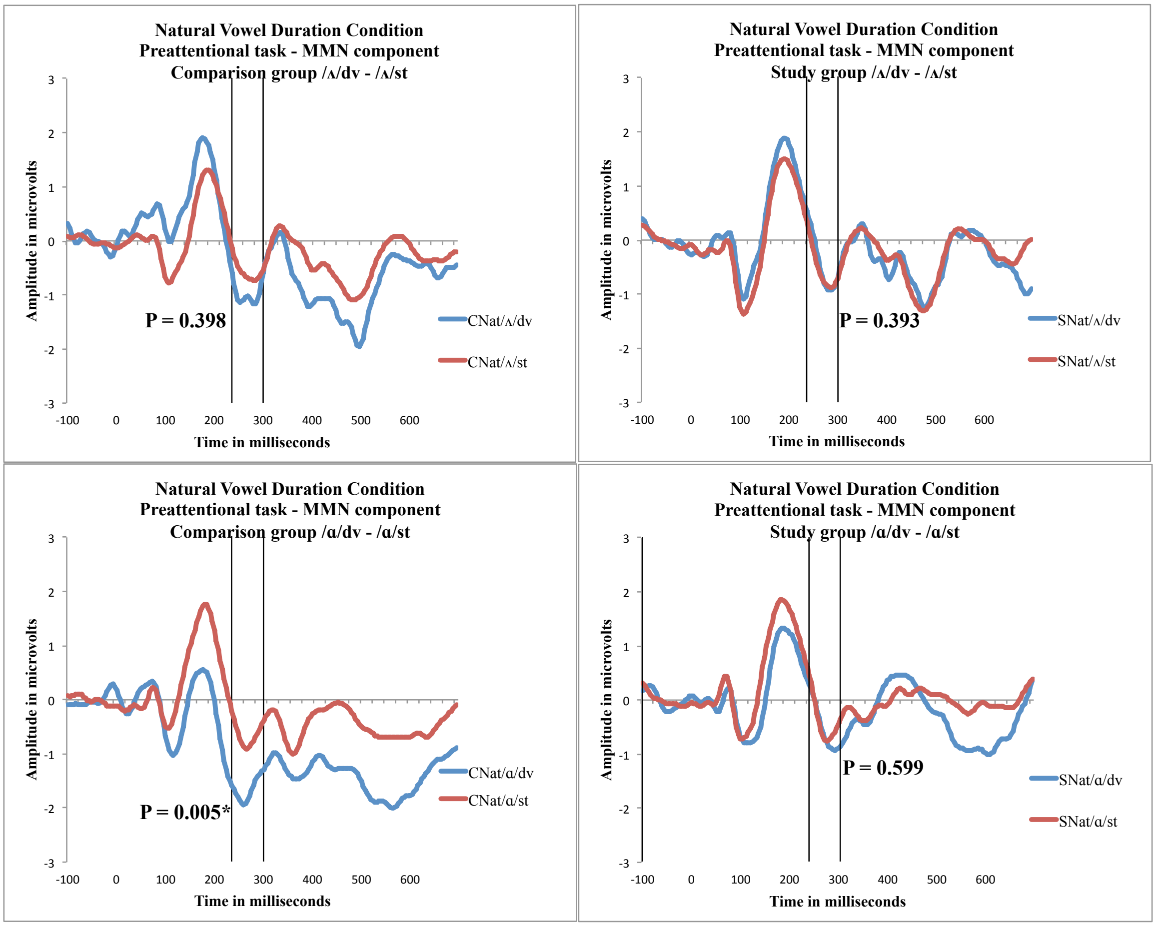

Figure 10. MMN natural vowel duration condition. MMN responses to vowels /ɑ/ and /ʌ/ during the pre-attentional task in the natural vowel duration condition. Left charts illustrate MMN responses from the Comparison (English) group, and right charts illustrate MMN responses from Study (Spanish) group.

Table 15 shows paired-samples t-tests for MMN mean amplitude for the Comparison and Study groups in the natural vowel duration condition. The results indicate that the Comparison

(English) group had a significantly greater negative amplitude response (MMN) to vowel /ɑ/ in the deviant status, but not to vowel /ʌ/.

Table 15

Pre-attentional MMN mean amplitude (µV) for Comparison (English) group in the natural vowel duration condition (Paired-samples t-tests)

Condition/Vowel	Mean amplitude in (µV)	SD	*t (9)*	Sig. (2-tailed)
Nat/ʌ/dv	-0.480	1.475	-0.888	0.398
Nat/ʌ/st	-0.175	1.041		
Nat/ɑ/dv	-1.366	1.844	-3.715	0.005*
Nat/ɑ/st	-0.227	1.098		

*$p < 0.05$

Planned comparisons revealed that the Study group had no statistically significant mean negative amplitude (MMN) response to the vowels /ɑ/ or /ʌ/ in their deviant status during the natural vowel duration condition (see table 16).

Table 16

Pre-attentional MMN mean amplitude (µV) for Study (Spanish) group in the natural vowel duration condition (Paired-samples t-tests)

Condition/Vowel	Mean amplitude in (µV)	SD	*t (9)*	Sig. (2-tailed)
Nat/ʌ/dv	0.110	0.774	0.897	0.393
Nat/ʌ/st	-0.035	0.639		
Nat/ɑ/dv	-0.046	1.048	-0.545	0.599
Nat/ɑ/st	0.100	0.649		

*$p < 0.05$

5.2.1.2. Neutral vowel duration condition. Results from the paired-samples t-tests obtained for the Comparison group, and presented in Table 17, showed that the mean amplitude of the MMN responses were significant to vowel /ɑ/ but not for vowel /ʌ/ in their deviant status during the *neutral vowel duration condition*.

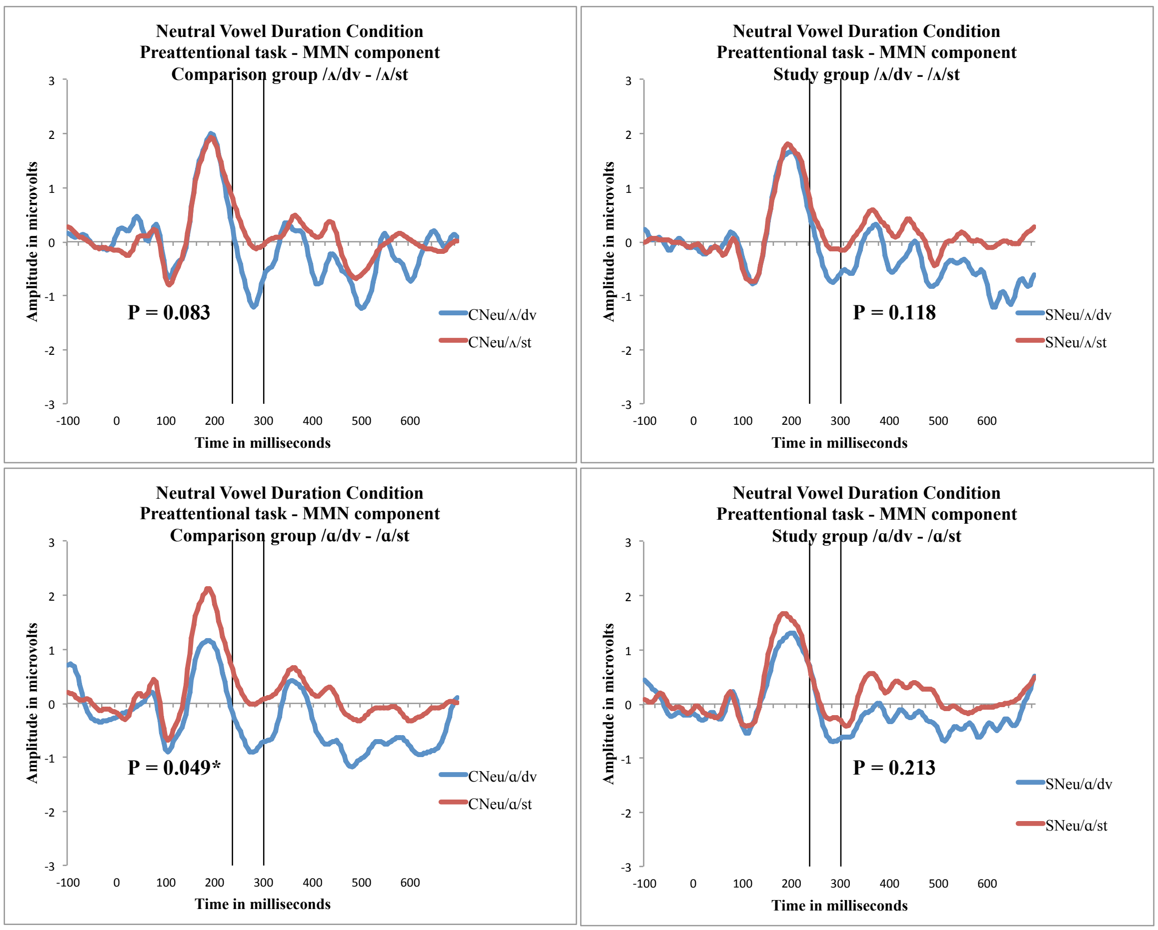

Figure 11. MMN neutral vowel duration condition. MMN responses to vowels /ɑ/ and /ʌ/ during the pre-attentional task in the neutral vowel duration condition. Left charts illustrate MMN responses from the Comparison (English) group, and right charts illustrate MMN responses from Study (Spanish) group.

Table 17

Pre-attentional MMN mean amplitude (µV) for Comparison group in the neutral vowel duration condition (Paired-samples t-tests)

Condition/VowelPair	Mean amplitude in (µV)	SD	*t (9)*	Sig. (2-tailed)
Neu/ʌ/dv	-0.064	1.386	-1.949	0.083
Neu/ʌ/st	0.538	0.637		
Neu/ɑ/dv	-0.282	1.750	-1.34	0.049*
Neu/ɑ/st	0.707	1.081		

*$p < 0.05$

Table 17 presents the Study group's MMN mean amplitude responses to vowels (/ɑ/-/ʌ/). The paired-samples t-test analysis indicated that there were not significant mean amplitude negativity responses to any vowel in the neutral vowel duration condition.

Table 18

Pre-attentional MMN mean amplitude (µV) for study group in the neutral vowel duration condition (Paired-samples t-tests)

Condition/VowelPair	Mean amplitude in (µV)	SD	*t (9)*	Sig. (2-tailed)
Neu/ʌ/dv	0.172	1.040	-1.728	0.118
Neu/ʌ/st	0.535	0.666		
Neu/ɑ/dv	0.155	0.880	-1.34	0.213
Neu/ɑ/st	0.363	0.576		

*$p < 0.05$

5.2.2 Attentional task (targets the P300 component). Figure 5 shows comparison and study groups' P300 responses to the vowel pairs during the attentional task in the natural vowel duration condition, in a time window between 250 and 500 ms after the stimulus onset, where the P300 response is expected.

A 4-factor mixed-design ANOVA group (Study vs. Comparison), condition (natural vs. neutral vowel duration), vowel /ɑ/-/ʌ/, and vowel status (standard vs. deviant), revealed no significant differences in P300 mean amplitude between the Study and Comparison groups during the attentional task (F (1, 18) = 2.212, p = 0.154, $\eta_p{}^2$ = 0.109). There was a significant main effect of Status (F (1, 18) = 29.265, p <0.001, $\eta_p{}^2$ = 0.619) indicating that P300 responses across the groups differed according to the vowel status (deviant vs. standard). No other significant main effects were observed in the analysis (Condition: F (1, 18) = 0.002, p = 0.961, $\eta_p{}^2$ = 0.000; Vowel: F (1, 18) = 0.520, p = 0.480, $\eta_p{}^2$ = 0.028).

The above finding was further confirmed by within-groups analyses. A repeated measures ANOVA for the Comparison group showed a significant main effect of Status (F (1, 9.000) = 38.202, p < 0.001, $\eta_p{}^2$ =0.809), but not Condition (F (1, 9) = 0.091, p = 0.770, $\eta_p{}^2$ = 0.010) or Vowel (F (1, 9) = 0.199, p = 0.666, $\eta_p{}^2$ = 0.022). Similarly, results from a repeated measures ANOVA analysis for the Study group revealed a significant main effect of Status (F (1, 9) = 7.528, p = 0.023, $\eta_p{}^2$ =0.455) only. The effect of other factors (Condition (F (1, 9) = 0.168, p = 0.692, $\eta_p{}^2$ = 0.018); Vowel (F (1, 9) = 0.341, p = 0.573, $\eta_p{}^2$ = 0.037) did not reach significance. These analyses confirm that the mean amplitude of the P300 response within each group depended on the status of the vowel (standard vs. deviant) regardless of the vowel duration condition. The next section presents results for each vowel duration condition separately.

5.2.2.1. Natural vowel duration condition. Figure 11 shows Comparison and Study groups' P300 responses to the vowel during the attentional task in the natural vowel duration condition.

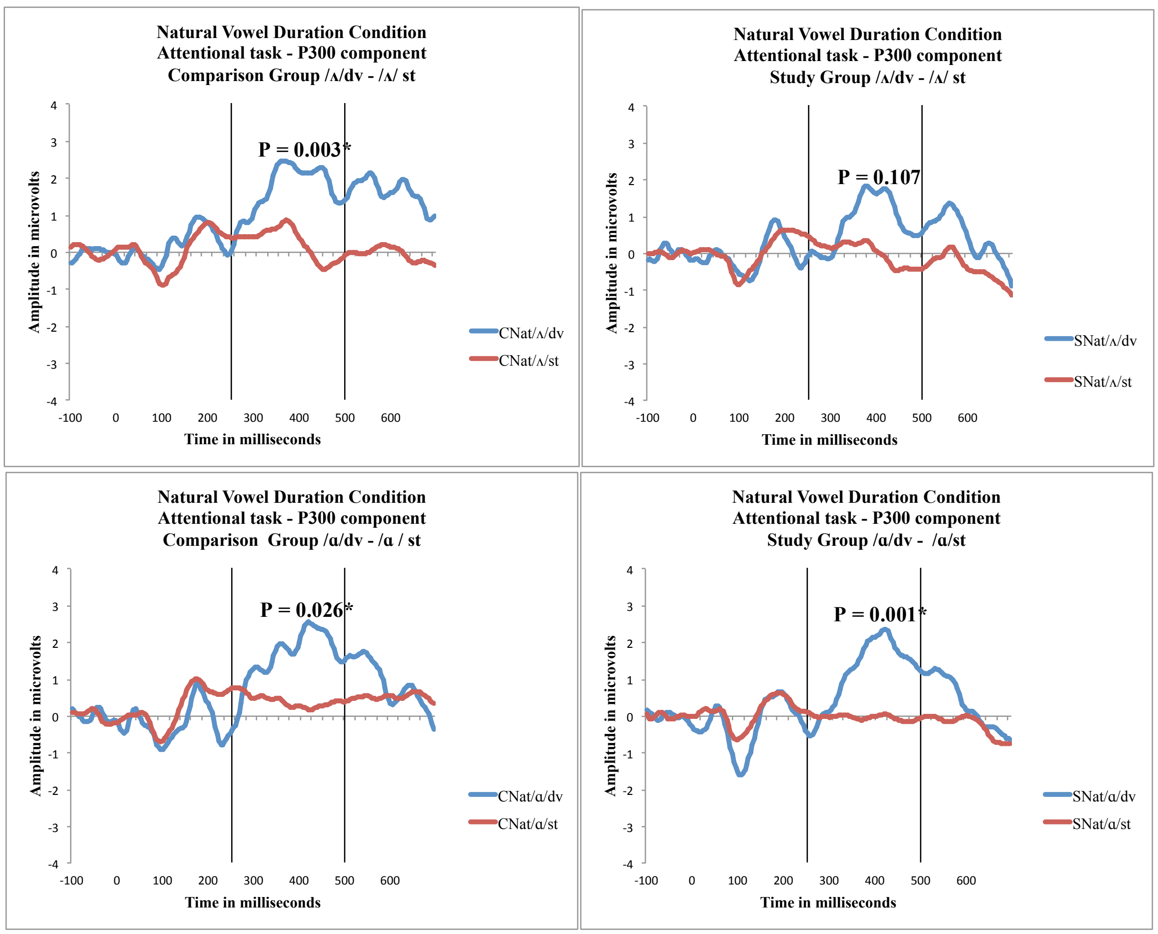

Figure 12. P300 natural vowel duration condition. P300 responses to vowels /ɑ/ and /ʌ/ during the pre-attentional task in the natural vowel duration condition. Left charts illustrate MMN responses from the Comparison (English) group, and right charts illustrate MMN responses from Study (Spanish) group.

Planned comparisons (paired samples t-tests) were applied to better understand the within-group mean amplitude response difference to the vowels. Table 19 presents the planned comparisons (paired-samples t-tests) indicating that the Comparison group (English) had significantly greater mean positivity amplitude responses to both vowels in their deviant status compared to their standard status during the *natural vowel duration condition*.

Table 19

Attentional P300 mean amplitude (μV) for Comparison group in the natural vowel duration condition (Paired-samples t-tests)

Condition/Vowel	Mean amplitude in (μV)	SD	*t (9)*	Sig. (2-tailed)
Nat/ʌ/dv	1.686	1.338	4.114	0.003*
Nat/ʌ/st	0.258	0.818		
Nat/ɑ/dv	1.595	1.067	5.124	0.001*
Nat/ɑ/st	0.420	0.748		

*$p < 0.05$

Paired-samples t-tests for the Study group (table 20) during the *natural vowel duration condition* showed a significantly greater positivity in the P300 time window in response to vowel /ɑ/, but not for vowel /ʌ/ in the deviant status.

Table 20

Attentional P300 mean amplitude (µV) for Study group in the natural vowel duration condition (Paired-samples t-tests)

Condition/Vowel	Mean amplitude in (µV)	SD	*t (9)*	Sig. (2-tailed)
Nat/ʌ/dv	0.848	2.399	1.791	0.107
Nat/ʌ/st	-0.179	0.745		
Nat/ɑ/dv	1.197	1.855	2.666	0.026*
Nat/ɑ/st	-0.285	0.671		

*$p < 0.05$

5.2.2.2. Neutral vowel duration condition. Figure 12 shows the P300 responses of the Comparison and Study groups during the attentional task in the neutral vowel duration condition.

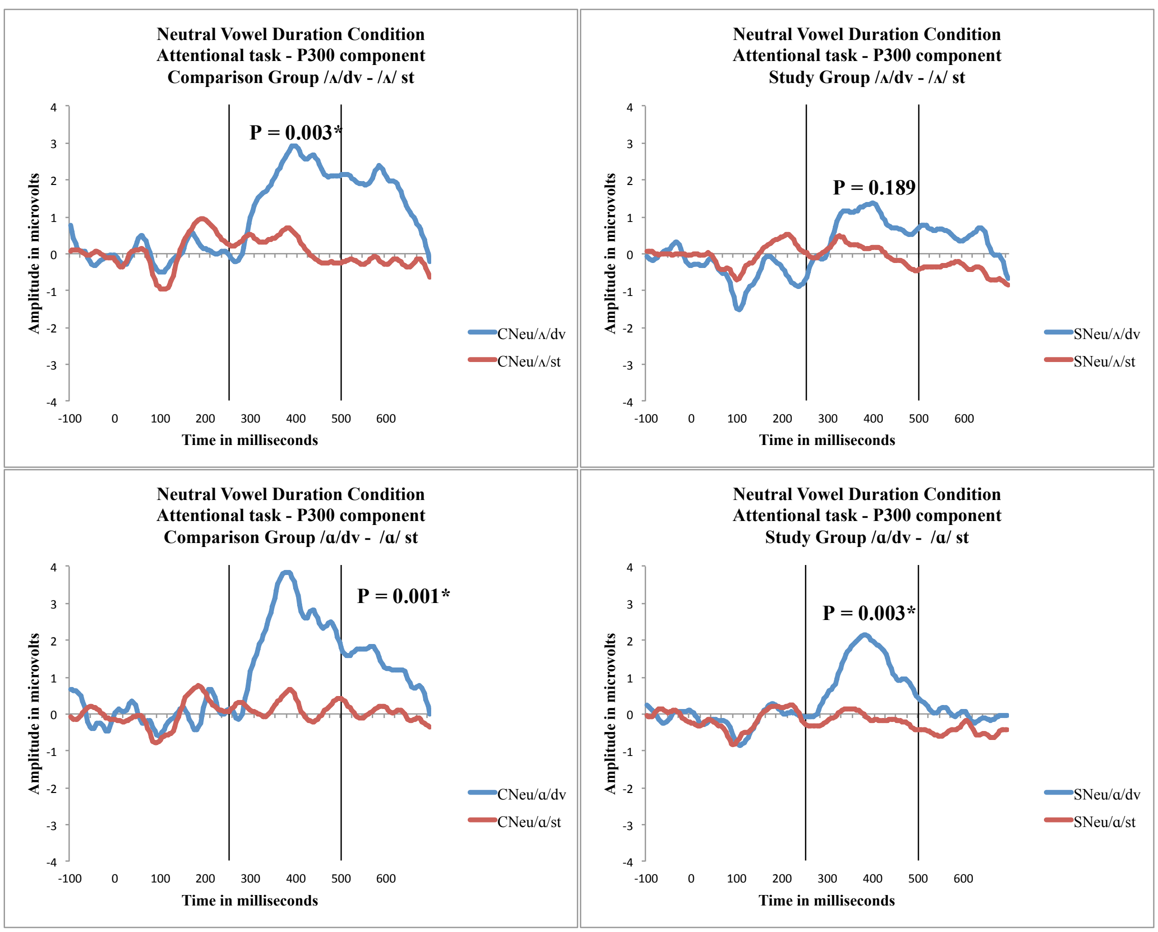

Figure 13. P300 neutral vowel duration condition. P300 responses to vowels /ɑ/ and /ʌ/ during the attentional task in the neutral vowel duration condition. Left charts illustrate P300 responses from the Comparison (English) group, and right charts illustrate P300 responses from Study (Spanish) group.

Planned comparisons (paired-samples t-tests) indicated that for the Comparison group (English), deviant status vowels elicited a significantly greater mean positivity amplitude response, corresponding to the P300 time window, to both vowels than those with standard status during the *neutral vowel duration condition* (see table 21).

Table 21

Attentional P300 mean amplitude (µV) for Comparison group in the neutral vowel duration condition (Paired-samples t-tests)

Condition/Vowel	Mean amplitude in (µV)	SD	*t (9)*	Sig. (2-tailed)
Neu/ʌ/dv	1.816	1.856	3.976	0.003*
Neu/ʌ/st	0.225	0.998		
Neu/ɑ/dv	2.202	1.817	5.242	0.001*
Neu/ɑ/st	0.177	1.133		

*$p < 0.05$

Planned comparisons (paired-samples t-tests) for the Study group, shown in table 14, revealed a significantly greater mean amplitude positivity response to vowel /ɑ/ but not for vowel /ʌ/ in deviant status during the *neutral vowel duration* condition.

Table 22

Attentional P300 mean amplitude (µV) for Study group in the neutral vowel duration condition (Paired-samples t-tests)

Condition/VowelPair	Mean amplitude in (µV)	SD	*t (9)*	Sig. (2-tailed)
Neu/ʌ/dv	0.695	2.515	1.422	0.189
Neu/ʌ/st	-0.204	0.736		
Neu/ɑ/dv	1.127	1.573	4.07	0.003*
Neu/ɑ/st	-0.408	0.560		

*$p < 0.05$

5.2.3. Summary of neurophysiological findings. The statistical analysis of the neurophysiological responses to the pre-attentional task revealed that the comparison group's negativity (MMN) responses for vowel /ɑ/ reached significance, but not for vowel /ʌ/ in their deviant status, in both the *natural and neutral vowel duration conditions*. Conversely, the Study group did not show significant MMN response to either of the vowels regardless of condition. In the attentional task, the Comparison group's positive (P300) responses to both vowels (/ɑ/-/ʌ/) reached significance in the expected time window (250 -500 ms) in both *natural and neutral vowel duration conditions*, while the Study group's responses were significant only to vowel /ɑ/ in both conditions.

5.3. General summary of findings

Taken together, behavioral and neurophysiological findings offer some responses to the study hypotheses, as follows:

(1) Vowel identification accuracy would be higher for the Comparison group (near 100%) to both vowels (/ɑ/-/ʌ/) in both conditions (*natural and neutral vowel duration*) compared to the Study group.

- In fact, compared to the Study group, accuracy was significantly higher for the Comparison group for both vowels in the *natural vowel duration condition*, and for vowel /ɑ/ in the *neutral vowel duration condition*. Unexpectedly, within-group analysis showed that the Comparison group was more accurate to identify vowel /ɑ/ than vowel /ʌ/ in both conditions, while the Study group did not show any vowel identification accuracy difference between the vowels.

(2) Vowel identification RT would be faster for the Comparison group in both conditions compared to the Study group.

- In this case, there were not significant differences between the groups in the vowel identification RT to both vowels in both conditions. However, within-group analysis showed that the Comparison group was significantly faster to identify vowel /ɑ/ in the *natural and neutral vowel duration conditions,* while the Study group was faster to identify both vowels in the neutral than the *natural vowel duration condition*.

(3) Comparison group would show significant MMN amplitude enhancement in response to the deviant vowels compared to MMN responses of significantly smaller amplitude from the Study group in both experimental conditions.

- As predicted, significant enhancement in MMN responses from the Comparison group were observed for vowel /ɑ/ only in both duration conditions. Although a significant MMN for both vowels were expected, this result may be in agreement with the significantly higher accuracy and faster RT responses to vowel /ɑ/. As expected, the negativity responses from the Study group did not reach significance, in agreement with the significantly lower accuracy scores. This finding could be explained by the fact vowel /ɑ/ has been described to be an anchor vowel in the vowel space (Polka & Bohn, 2010; Schwartz et al., 2005). Regarding the Study group, it may be inferred that this group had not formed a mental representation of the non-native vowel sounds corresponding to AE vowels /ɑ/-/ʌ/.

(4) Comparison group would show significant P300 amplitude enhancement in response to the deviant vowels compared to P300 responses of significantly smaller amplitude from the Study group in both experimental conditions.

- As predicted, significantly enhanced P300 responses from the Comparison group were observed for both vowels /ɑ/-/ʌ/ in both *natural and neutral vowel duration conditions* regardless of their deviant status. In contrast, during both conditions, P300 responses from the Study group were significantly greater only in the deviant status for vowel /ɑ/ but not for vowel /ʌ/. Significant P300 responses may be related to a later processing of the speech sounds and the access to attentional and cognitive resources to consciously identify each vowel being presented.

Chapter 6 will focus on a more comprehensive interpretation and discussion of the behavioral and neurophysiological results of the study.

5.4. Post – Hoc Analysis

Although the a priori hypotheses of the study focused on the specific time windows for the MMN and P300 responses (100-300 and 250-500 ms after stimulus presentation), a salient negativity peaking around 500 ms after stimulus onset can be observed in the grand-averaged waveforms for both the Comparison and the Study group during the pre-attentional task. Paired-samples t-tests were conducted to find out whether the mean amplitude of this negativity was significantly different for each vowel in its deviant status, compared to the standard status, for each group (See table 23).

Table 23

Post-hoc planned comparisons: Paired-samples t-tests. Pre-attentional late (~500 ms) negativity mean amplitude for Comparison and Study group

Condition/Vowel	Comparison			Study		
	Mean (SD)	*t* (9)	p	Mean (SD)	*t* (9)	p
Nat/ʌ/dv- Nat/ʌ/st	-0.766 (1.588) -0.337 (0.504)	-0.884	0.400	-0.386 (1.250) -0.311 (0.839)	-0.452	0.662
Nat/ɑ/dv- Nat/ɑ/st	1.398 (1.248) -0.467 (0.701)	-2.175	0.024*	-0.331 (1.390) -0.110 (0.674)	-0.687	0.510
Neu/ʌ/dv- Neu/ʌ/st	-0.309 (1.575) -0.030 (0.480)	-0.608	0.558	-0.498 (1.520) 0.148 (0.564)	-2.068	0.690
Neu/ɑ/dv- Neu/ɑ/st	-0.571 (1.188) -0.032 (0.732)	-1.349	0.210	-0.305 (0.750) 0.145 (0.853)	-1.709	0.122

The analysis of the mean amplitude differences within the groups regarding the late negativity observed during the pre-attentional task, showed that this negativity is only significant to vowel /ɑ/ in the deviant status during the *natural vowel duration condition* for the Comparison, and not for the Study group.

Similar negativity responses between 400 – 600 ms after stimulus presentation in MMN studies have been described as a response to pre-attentive discrimination between long and short sounds, when listeners are distracted from their task. It has been called the reorientation negativity (RON), and thought to reflect redirection of attention to the target task after involuntary attention shift (Escera, Alho, Schröger, & Winkler, 2000; Schröger & Wolff, 1998). In the case of the Comparison group, the late negativity response observed in the pre-attentive task may correspond to the RON since it is specific to vowel /ɑ/ in the natural vowel duration condition, when it is longer than vowel /ʌ/.

Since no other salient responses are apparent in the group waveforms for this condition, the post hoc analyses focused only on this late time window and no other components were subjected to analysis for the current study.

6. Discussion

This study aimed to determine whether adult sequential Spanish-English bilinguals perceive the AE vowel contrast /ɑ/-/ʌ/. Languages differ in the number and features of their speech sounds. Although infants are born with the ability to perceive speech sounds from different languages, within the first year of life the phonological system becomes specialized for preferential processing of the speech sounds that are salient for the native language. Behavioral studies have shown that, in order to accurately perceive non-native speech sounds, listeners must modify perceptual parameters along phonetic-acoustic dimensions in their phonological system. These modifications serve to either adjust existing categories from the L1, or create new categories that incorporate the new speech sounds (Best, 1995; Best & Tyler 2007; Boersma & Escudero, 2004; Escudero, 2000; 2005; Flege, 1997). This perceptual adjustment has been shown to be particularly challenging for sequential Spanish-speaking learners of English when perceiving some American English vowel sounds, because Spanish-speaking listeners may not be able to extract the relevant acoustic-phonetic cues that signal meaningful differences in the L2 (Escudero 2000; 2005; Boersma & Escudero, 2004). This is probably due to the fact that acoustic-phonetic differences among the English vowels are very subtle for Spanish-speaking adults to detect based on their native perceptual parameters.

Neurophysiological studies have shown that learning to perceive non-native speech sounds changes phonological representations at the level of the brain (e.g., Näätänen, 1997; Winkler, 1999). These changes are reflected by monitoring brain activations at very early stages of perceptual processing, when detection of language-specific parameters is automatic (Kirmse et al., 2008, Rivera-Gaxiola et al., 2000) as well as later in the processing stream, when (conscious) attention is involved in discrimination/identification processes (Hisagi et al., 2011, Lipski et al., 2012; Ylinen et al., 2009).

Acoustic differences between English vowels are signaled by spectral and durational cues, with spectral information being the primary channel for indicating phonemically-relevant changes (Escudero, 2000; 2005). In the Spanish phonological system, the opposite holds: vowels are differentiated by spectral features, while durational cues are not used. Interestingly, behavioral and neurophysiological studies have shown that native Spanish speakers rely more on durational cues to perceive some English vowels (/ɪ/-/i/: Boersma & Escudero, 2004; Escudero, 2000; 2005; Morrison, 2006; 2008; 2009), and some Dutch vowels (/aː/-/ɑ/: Lipski, 2012) at least at specific stages of L2 learning. While most of the research on perception of AE vowels has focused on the vowel contrast /ɪ/-/i/, little is known about other potential challenging vowel contrasts such as /ɑ/-/ʌ/. Therefore, the purpose of the study reported in this dissertation was to examine neurophysiological responses (ERPs) of adult sequential Spanish-English bilinguals to the AE vowel contrast /ɑ/-/ʌ/, compared to monolingual English-speaking listeners. Both groups of listeners underwent a passive listening (pre-attentive) and identification task (attentional) under two listening conditions: (1) natural vowel duration and (2) neutral vowel duration.

Based on behavioral and neurophysiological cross-language speech perception studies, it was predicted that monolingual English-speaking listeners would be significantly more accurate and faster than the Spanish-English bilinguals to identify the AE vowels /ɑ/-/ʌ/. Since these vowels are part of the native vowel inventory for the monolingual English-speaking listeners, it was predicted that their MMN and P300 responses to the listening and identification tasks would reflect pre-attentional and attentional discrimination of the vowels in both natural and neutral duration conditions. This would reflect a reliance primarily on spectral cues to discriminate between the two vowels. Conversely, since these vowels are not part of the native vowel inventory for the Spanish-English bilinguals, MMN and P300 responses provided outcome measures for evaluating perceptual processes. If the bilingual group perceived the vowel contrast at the pre-attentional (MMN) and attentional (P300) levels, relying on both spectral and durational cues in the natural vowel duration condition, or on durational cues alone in the neutral vowel duration condition, then the ERP responses were predicted to resemble those of the monolingual English-speaking listeners. On the other hand, non-significant neurophysiological responses (MMN and P300) to the vowel contrast would indicate that the Spanish-English bilingual group were not able to perceptually or attentionally discriminate the differences between the vowels /ɑ/-/ʌ/.

In this discussion, I will analyze study findings around each of the two research questions that were posed at the start of this dissertation.

Research Question 1. Do adult sequential Spanish-English bilingual listeners show indices of discrimination of AE vowel contrast /ɑ/- /ʌ/ at early stages of speech perception (at the pre-attentional and/or attentional levels), as indexed by behavioral (accuracy and reaction time) and neurophysiological measures (MMN and P300)?

Behavioral responses to the AE vowel contrast /ɑ/- /ʌ/ obtained during the attentional task revealed significant accuracy differences between the adult sequential Spanish-English bilingual listeners (the “Study group”), and the monolingual English group (the “Comparison group”), indicating, as expected, that the Comparison group was significantly more accurate to perceive the AE vowels than the Study group. On the other hand, there were no significant differences in reaction time between the groups. Within-group differences showed that the Comparison group was more accurate and faster to identify vowel /ɑ/ than vowel /ʌ/ in both natural and neutral vowel duration conditions, suggesting that vowel /ɑ/ was perceptually more salient than vowel /ʌ/.

Neurophysiological results indicated that the Study group did not show indices of discrimination of the AE vowel contrast /ɑ/- /ʌ/ at the pre-attentional level. This was indicated by non-significant MMN responses to the vowel contrast during the pre-attentional task in both duration conditions (natural and neutral vowel duration). However, the Study group’s neurophysiological data did reveal indices of attentional discrimination of the AE vowel /ɑ/, but not vowel /ʌ/, indicated by significant P300 responses to /ɑ/ in its deviant status during both natural and neutral vowel duration conditions. This finding suggest that the study group did not have pre-attentional access to mental representations corresponding to the non-native vowels /ɑ/- /ʌ/ between 100 ms and 300 ms after the stimulus onset at the pre-attentional level, but they had access to relevant acoustic-phonetic information about the vowels later in time, between 250 ms and 500 ms, when they were (consciously) attending to the stimuli.

Findings from the Study group suggest that there was no pre-attentional detection of differences between the non-native vowels, but attentional and cognitive resources that are recruited later in processing did facilitate the discrimination and identification of the challenging non-native vowel /ɑ/. These results support previous findings indicating that perception of some L2 phonemes is less pre-attentional, yielding MMN responses that are larger to native phonemes compared to non-native ones (Hisagi et al., 2011; Kirmse et al., 2008; Näätänen, 2007). The current findings are also in line with work showing that, when non-native listeners pay attention to L2 speech sound contrasts, detection of deviant features can improve (Hisagi et al., 2011).

The monolingual English group (Comparison group) showed expected indices of discrimination of the AE vowel contrast /ɑ/-/ʌ/ at the pre-attentional level as indicated by significant MMN responses to vowel /ɑ/ in its deviant status, during the pre-attentional task in both natural and neutral vowel duration conditions. Unexpectedly, this group did not show MMN responses as expected when /ʌ/ was the deviant sound, showing no indices of discrimination of the AE vowel /ʌ/ at the pre-attentional level in any condition. However, during a conscious decision making process, when attention was engaged in the task, the Comparison group did show a significant P300, indicating expected discrimination and identification of both vowels /ɑ/ and /ʌ/. These results suggest that for the Comparison group, attention also played a role in the discrimination and identification process for vowel /ʌ/.

The pattern of neurophysiological results was similar to that observed behaviorally, in that enhanced ERP responses (MMN and P300) were observed to the AE vowel /ɑ/ in both groups, further suggesting a perceptual vowel asymmetry favoring /ɑ/ over /ʌ/. Asymmetry in perception of native vowels by English monolingual listeners was not expected, since previous research (Polka & Bohn 1996; 2003; 2010) has shown that asymmetry effects in native vowel perception effect can be present in infancy, but are expected to disappear by adulthood. In general, however, asymmetries in vowel perception have been described as reflecting a preference for vowels that are located in the periphery of the vowel space due to their perceptual salience, as they provide an anchor for comparison (Polka & Bohn, 1996; 2003; 2010; Schwartz, Abry, Boë, Menard & Vallée, 2005). Polka and Bohn (1996) argued that one vowel in a contrast pair always plays the role of an anchor, regardless of its status in the listener's phonological system. Therefore, the observed perceptual preference for vowel /ɑ/ over vowel /ʌ/ in the Comparison group during pre-attentional and attentional tasks, and in the Study group in the attentional task, for both vowel duration conditions, could be interpreted as another exemplar of vowel perception phenomena. It appears that vowel /ɑ/ acted as the anchor vowel for both groups.

In line with the context of the Dispersion-Focalisation Theory (DFT) (Schwartz et al., 2005), vowels that have focal (or *close*) values for F1 and F2 offer a benefit for speech perception. In this case, vowel /ɑ/, for which F1 and F2 values are closer to each other (F1= 903.964, F2= 1319.61) is more perceptually salient than /ʌ/ (formant values: F1= 903.964, F2= 1319.261), and hence /ɑ/ becomes a reference for discrimination. Behavioral (Karypidis (2007); Nishi, Strange, Akahane-Yamada, Kubi, & Trent-Brown, 2008; Sebastián-Gallés, Echeverria, and Bosch (2005)) and neurophysiological studies (Vera Constán & Sebastian-Gallés, 2008) have shown vowel asymmetries that reveal consistently better discrimination for peripheral vowels in non-native listeners.

Although both groups showed a perceptual preference to AE vowel /ɑ/, there was a difference between the groups regarding the level at which this preference was apparent. The Comparison group showed a perceptual preference for AE vowel /ɑ/ at the pre-attentional level, where minimal cognitive resources are required. On the other hand, the Study group did not show indices of discrimination of any vowel in the contrast at the pre-attentional level; instead, the Study group showed discrimination of the AE vowel /ɑ/ only when attention, working memory load, and other cognitive resources were recruited to consciously identify the vowels. Along the same lines, attentional resources may have facilitated the Comparison group in the discrimination and identification of vowel /ʌ/ during the attentional task in both natural and neutral vowel duration conditions. However, due to its non-salient status, this vowel remained difficult for the Study group even when such resources were applied. Other studies have presented similar findings indicating that speech sound perception is facilitated and neurophysiological responses to specific speech sound contrasts are enhanced when conscious attention is directed to deviance in auditory speech stimuli (Hisagi et al., 2011; Sussman, Kujala, Halmetoja, Lyytinen, Alku, & Näätänen, 2004).

To summarize, the adult sequential Spanish-English bilingual listeners (Study group) showed indices of discrimination and identification of one vowel (namely /ɑ/) in the AE vowel contrast /ɑ/- /ʌ/, only during the attentional task, in both natural and neutral vowel duration conditions. In contrast, the Comparison group showed pre-attentional and attentional discrimination and identification of vowel /ɑ/, but only attentional discrimination and identification of vowel /ʌ/. Results from both groups can be explained by two factors: (1) Asymmetries in vowel perception that suggest a perceptual preference for vowels that occupy the periphery of the vowel space and are more perceptually salient, which is the case for vowel /ɑ/; (2) Attention as a facilitator in the perception of speech sounds at later stages (300 ms -500 ms) in perceptual processing.

Research Question 2. In the case that adult Spanish-English bilingual listeners perceive the AE vowel contrast /ɑ/- /ʌ/, do they rely more on durational differences to discriminate the vowels, or do they use spectral cues as the primary information to distinguish them?

Based on previous findings indicating that Spanish-speaking learners of English rely primarily on durational cues to identify English vowels /ɪ/-/i/ (Boersma & Escudero, 2004; Escudero, 2000; 2005; Morrison, 2006; 2008; 2009), and Dutch vowels /aː/-/ɑ/ (Lipski, 2012), it was expected that the Study group would show a reliance on vowel duration to perceive the AE vowel contrast /ɑ/-/ʌ/. However, the Study group showed no significant differences in discrimination between the two experimental conditions (natural vowel duration vs. neutral vowel duration). The significant P300 responses in the attentional task, during the *neutral vowel duration* condition, indicated that when attentional resources are recruited, sequential Spanish-English bilingual listeners discriminated AE vowel /ɑ/ but not /ʌ/. For this they must have been relying only on the available spectral differences, since durational cues were neutralized; therefore, durational information is not necessary for the Study group to detect the AE vowel /ɑ/ when it was presented as a deviant stimulus. Furthermore, as expected, the Comparison group did show indices of discrimination of AE vowel /ɑ/ at the pre-attentional level, and discrimination and identification of both vowels /ɑ/-/ʌ/ at the attentional level in both natural and neutral vowel duration conditions. These findings therefore indicate that durational information is a secondary acoustic-phonetic cue that is dispensable, if other cues are available, for both native and non-native discrimination of the AE vowels /ɑ/-/ʌ/.

The L2LP model (L2 Linguistic Perception; Escudero, 2005) proposes that in the initial learning stages Spanish learners will have two learning tasks in order to accurately perceive English vowels. One of the tasks is to create new categories along a new auditory dimension (length) that has not been previously implemented in the L1 because Spanish does not use duration as a contrastive cue for speech sound processing. The other learning task is to create extra categories along an already-used auditory dimension (height). To do so, learners need to redistribute or split their L1 perceptual parameters to incorporate the new non-native speech sounds. This model predicts that at the end of the learning process (end state), if there is enough rich L2 input, learners may reach a native-like perception of the L2 speech sounds. In the present study, it appears that the adult Spanish-English bilingual listeners in the Study group are at a stage in which they rely primarily on spectral information (similar to native English speakers) to perceive the vowel contrast /ɑ/-/ʌ/. However, the Study group must recruit attentional and cognitive resources in order to utilize spectral information for vowel discrimination.

To conclude, the adult sequential Spanish-English bilingual listeners (Study group) showed indices of discrimination and identification of AE vowel /ɑ/ but not /ʌ/ at the attentional level, when both spectral and durational information about the vowels was perceptually available in the natural vowel duration condition, but also when duration was neutralized leaving only spectral cues available to distinguish the vowels. These findings contrast with previous studies showing that Spanish-English bilingual adults rely more heavily on durational cues to identify some L2 vowels (in English: Boersma & Escudero, 2004; Escudero, 2000; 2005; Morrison, 2006; 2008; 2009; in Dutch: Lipski, 2012). The current findings show that Spanish-English bilinguals may use spectral and durational cues, like native English speakers, to perceive the English vowel contrast /ɑ/-/ʌ/. However, this cannot be described as an “end state” in the sense of Escudero (2005), since the neurophysiological evidence shows that these L2 learners are able to reach native-like discrimination only when they recruit attentional and cognitive resources to facilitate the perceptual process.

It remains to be seen in future research whether targeted learning strategies such as High Variability Phonetic Training (HVPT) (e.g. Bradlow et al., 1997; Lively et al., 1993), distributional learning (Maye & Gerken, 2000; Escudero, Benders & Wanrooij, 2011; Pająk, Bożena, & Roger Levy, 2011), could support the attainment of native-like L2 speech sound discrimination for adult sequential learners. This study has shown that behavioral discrimination does not necessarily constitute an “end state” for learning, since pre-attentional processing is needed for the rapid, accurate, online recognition of speech sounds in the L2 (Näätänen et al., 2007).

6.1. Study limitations

Learning to perceive the sounds of a second language has been shown to be influenced by multiple factors that include learner characteristics such as age of L2 learning (Lenneberg, 1969; Long, 1990; Pakwoski, 1990), the learning environment and learning/teaching method (Best & Tyler, 2007; Jia et al., 2006), Age of Arrival to the L2 environment (Bongaerts et al., 1997; Flege et al., 1997; Tsukada et al., 2005), motivation and daily use of L2 (Piske, MacKay, & Flege, J., 2001).

Language variables can also influence perception of non-native speech sounds. Such variables include: native language (Abrahamson & Lisker, 1970; Best, 1995; Best & Tyler, 2007; Flege, 1995), interaction of the L1 with L2 (Escudero & Boersma, 2004, Iverson & Evans, 2007; Polka, 1991), and experience with L2 (Ito & Strange, 2009; Kuhl & Iverson, 1995; Iverson et al., 2003). One of the limitations of this study is that participants in the Study group came from diverse language backgrounds, and there was a wide range of age of acquisition of the language (from 3 to 21 years; mean = 28.01, SD = 3.48). The heterogeneity of the Study group participants may have masked effects, especially related to age of acquisition. Participants who learned at 3 years of age may have a different perceptual processing of the experimental vowels compared to those who learned after puberty. An analysis grouping the participants in early, mid and late learners could have helped to elucidate perceptual differences in non-native speech vowels by age of L2 learning. However, the number of participants in the Study group (N= 10) did not permit creation of subgroup comparisons without loss of statistical power.

The Comparison group participants also came from diverse dialect backgrounds, even though all were native speakers of AE. Cross-dialectal spectral variation has been described in different dialects of AE vowels (Fox & Jacewicz, 2009), indicating that the perceptual parameters to discriminate the experimental vowels may be slightly different between listeners from different AE dialects. These factors may have introduced variability in behavioral and neurophysiological responses to the vowel stimuli in this experiment, which were all produced by one native speaker of AE.

In addition, experimental design decisions in laboratory settings affect the way listeners perceive non-native segments (Beddor & Gottfried, 1995). Researchers typically present the target speech segments either in a carrier sentence (similar to running speech) or in a citation form (isolated production of speech sounds, words or syllables). The choice of presentation format influences the spectral characteristics of the target speech sounds (Hillebrand et al., 2001). Also, presenting speech sounds produced by one or multiple talkers, or presenting multiple tokens from the same speaker, provides a more nuanced representation of realistic communication (Bent et al., 2010). Presentation of highly variable stimuli (multiple talkers, multiple tokens and multiple phonetic, phonotactic and prosodic contexts) allows the analysis of the influence of L1 and L2 in cross-language studies (Strange, 2007).

The experimental stimuli themselves were naturally spoken speech sounds produced by a female AE speaker of the Philadelphian dialect living in New York City. Apart from the fact that only one dialectal variant of AE was represented in the stimulus stream, there are additional limitations associated with the stimuli. For example, including only one token of each of the vowels for the experimental manipulation means that this experiment did not reflect the token variability that is found in real life speech. Extending the number of trials was deemed to be too onerous for the participants, and pilot results showed that including variable stimuli attenuated the amplitude of neurophysiological responses that index acoustic change-detection (MMN). Nevertheless, an experimental design that incorporates at least some aspects of the within- and between- talker variability that characterizes speech perception in real life might enhance the ecological validity of the paradigm. Also, including a vowel contrast that was predicted to be easy for both language groups (e.g. /i/-/u/) could have provided a baseline comparison for the behavioral and neurophysiological responses observed in the perceptually challenging vowel pair (/ɑ/-/ʌ/).

The experimental conditions (natural and neutral vowel duration) were not counterbalanced, which might have influence the results observed in terms of reaction time showing no differences between the groups, but within groups. The expected differences between groups showing a faster reaction time in the comparison group over the study group, could have been observed by counterbalancing the conditions.

The native speakers were at ceiling in their accuracy responses, which suggests that the introduction of some noise into the speech signal might have helped to make the task harder for the native speakers, so that the ceiling effects could have been removed. Another factor that could have potentially contributed to ceiling effects in the Comparison group was the low uncertainty paradigm that was implemented in the experiment. Having a three-way contrast instead o two could have also helped to elucidate more specific accuracy differences between the groups.

Despite its limitations, the present study sheds light on some of the mechanisms underlying the perception of American English vowels /ɑ/-/ʌ/ perception by monolingual English and adult sequential Spanish-English bilingual listeners. In the next section, I describe some of the additional research questions that are emerging from this first step.

6.2. Future directions

This study focused on examining neurophysiological responses to the perception of the vowel contrast /ɑ/-/ʌ/, and very informative behavioral data (accuracy and reaction time) were collected during the attentional task. However, there is no behavioral characterization of the perception of these vowels in these adult sequential Spanish-English bilinguals. As a first step, defining the perceptual boundaries for these vowels for both monolingual English and Spanish-English bilingual groups would permit the development of more specific predictions about possible assimilation patterns for Spanish-English bilinguals, based on speech perception models such as the PAM-L2.

The findings of this study raised questions that constitute next steps in the investigation of AE vowel perception by adult sequential Spanish-English bilingual and English monolingual listeners.

(1) Are there within group neurophysiological differences in discrimination of the AE vowel contrast /ɑ/-/ʌ/ between Spanish-English bilinguals who acquired the language early in life (before the age of 10) and those who learned it later in life (after 10 years of age)? Since the range age of L2 acquisition in the Study group was large (3 to 21 years of age), it would be interesting to know whether there are within group pre-attentional differences that are determined by the age of acquisition of English. In this case, increasing the number of participants in the Study group would permit comparison between early, mid and late learners of English. It could be expected that earlier non-native learners show more native-like MMN

responses to the vowel contrast compared to late learners. It would be valuable to add a group of monolingual Spanish listeners so as to establish an initial state in the perception of this vowel contrast. In addition, it would be interesting to examine how participant factors, such as Age of L2 learning (early vs. late), L2 exposure (long vs. short), or Length of Residency in an AE-dominant environment, correlate with the pre-attentional neurophysiological responses to this vowel pair.

(2) Are there neurophysiological differences related to the latency of the ERP components that reflect discrimination of the AE vowel contrast /ɑ/-/ʌ/ between Spanish-English bilinguals and English monolinguals? This study focused on amplitude of the neurophysiological responses to the AE vowel contrast /ɑ/-/ʌ/, but did not address possible differences in the latency of the responses, which could be an important indicator specific language group difference in the perception of the target vowels. It would be beneficial for this study to conduct a latency analysis that complements the voltage mean amplitude differences observed between the groups.

(3) How do the behavioral and neurophysiological responses in discrimination and identification of the AE vowel contrast /ɑ/-/ʌ/ relate to production of the same vowels in AE and Spanish-English bilinguals? Since the goal of these perceptual processes is to communicate effectively with other people, it is of interest to establish a relationship between how listeners perceive native and non-native speech sound contrast and how they actually produce them.

(4) How did the lexical status of the stimuli shaped the neurophysiological responses in both language groups? The syllables used in this study bɑb (real word) and bʌb (nonsense word) constituted might have influenced the behavioral and neurophysiological perceptual responses in

both Spanish-English bilinguals and English monolingual groups. A future study looking presenting stimuli that constitute a real word vs. real word contrast, and nonsense word vs. nonsense word contrasts would provide more information on how whether the allegedly perceptual salience of vowel /ɑ/ is indeed related to its vowel properties as indicated by hypotheses such as Natural Vowel Referent (Polka & Bohn, 1996; 2002; 2010) or Dispersion-Focalisation Theory (Schwartz et al., 2005), or instead, it could be explained by the fact that the "preferred" vowel /ɑ/ was presented in a real word (bɑb), while the other vowel /ʌ/, was presented in the context of a nonsense word (bʌb).

(5) In this study durational cues were removed from the vowels to see whether Spanish-English bilingual adults relied more on durational cues (as expected according to Bohn's desensitization hypothesis), or if spectral cues were sufficient to perceive the AE vowel contrast /ɑ/-/ʌ/. Although the results showed that spectral cues were sufficient, it would be interesting to investigate how native and non-native listeners would perceive the vowel pair /ɑ/-/ʌ/, when duration is not removed but lengthened.

(6) How could attentional and other cognitive resources serve as a the basis for an auditory training program that focuses on teaching second language learners to attend relevant non-native vowel cues during the L2 learning process? One significant finding in this study was that attention facilitated discrimination and identification of AE vowel /ʌ/ in the Comparison group, and vowel /ɑ/ in the Study group. Recent studies have shown that discrimination of non-native contrasts can improve after training using distributional learning strategies (Escudero et al., 2011; Wanrooij, Escudero & Reijmaker, 2013). Other studies have found that perceptual training of non-native speech sounds modifies neurophysiological responses at the pre-attentional level, indicating that perception of L2-relevant cues can be learned (Tremblay et al.,

2001; Ylinen,Uther, Latvala, Vepsäläinen, Iverson, Akahane-Yamada et al., 2009). The design of acoustic-phonetic training that can be implemented not only in a lab setting, but also in the language learning classroom, could potentially help learners to form new non-native speech sound categories. The final goal would be to evaluate whether auditory training generates changes at pre-attentional levels indicating the formation of non-speech sounds categories resembling native listeners. The effectiveness of such a program in the formation of non-native speech sound categories, could be evaluated by implementing the ERP method, focusing on pre-attentional measures of speech perception such as the MMN (Nataanen, 1997; Winkler, 1999; Tremblay et al., 2001; Ylinen et al., 2009). In these ways, neuroscientific research can contribute to the development of effective and ecologically valid classroom interventions, as well as providing an objective means to evaluate the efficacy of various pedagogical strategies for adult L2 learning.

6.3. Conclusions

The objective of the current study was to examine neurophysiological responses (ERPs) of adult sequential Spanish-English bilinguals to American English vowel contrasts /ɑ/-/ʌ/ compared to monolingual English-speaking listeners, in two tasks requiring perceptual discrimination and identification under two listening conditions (natural and neutral vowel duration).

Study findings indicate that adult sequential Spanish-English bilinguals are less accurate than English monolinguals in discriminating the AE vowel contrastive pair /ɑ/-/ʌ/. In addition, native English listeners are more accurate in identification of vowel /ɑ/ than vowel /ʌ/. These behavioral findings are in line with ERP results, suggesting that Spanish-English bilinguals do not show neurophysiological indices of perception of the AE vowel contrast /ɑ/-/ʌ/ the pre-attentional level. However, when attentional and other cognitive resources are recruited, discrimination and identification improve, at least towards the most perceptually salient vowel in the pair, vowel /ɑ/. The Comparison group showed pre-attentional neurophysiological indices of discrimination towards AE vowel /ɑ/, but not to vowel /ʌ/. When attentional resources were recruited, the Comparison group showed neurophysiological indices of discrimination of both vowels (/ɑ/-/ʌ/) in the contrast. The apparent perceptual preference for AE vowel /ɑ/ was common to both groups at the behavioral and neurophysiological levels for the Comparison group, and at the neurophysiological level for the Study group. Non-significant differences between natural vs. neutral vowel duration conditions suggested that Spanish-English bilinguals do not use durational information as the most important cue to discriminate and identify AE vowel contrast (/ɑ/-/ʌ/) at the attentional level. Instead, they seem to rely on spectral cues primarily, as native English listeners do.

The present study sheds light on some of the mechanisms underlying the early perceptual processing of American English vowels /ɑ/-/ʌ/ by monolingual English and adult sequential Spanish-English bilingual listeners. In particular, it reveals new information about perceptual discrimination (enhanced perceptual salience for anchor vowels), processes in cross-language speech perception (the unexpected availability of spectral cues for vowel discrimination), and dissociations between behavioral discrimination and neurophysiological evidence for attentional recruitment (hence a lack of pre-attentional access to mental representations even when native-like perceptual mechanisms can be accessed). Findings from this study therefore add to the extant behavioral and neurophysiological evidence about cross-linguistic speech perception, specifically vowel discrimination, and differences between monolingual and sequential bilingual listeners with respect to attentional and pre-attentional processing (cf. Escudero & Chlädkova, 2010; Lipski et al., 2011).

Learning to perceive the speech sounds of a language constitutes a fundamental requirement for effective communication. Recognition and comprehension of challenging non-native vowels is usually aid by contextual cues that provide semantic and syntactic information. However there are always instances in which context does not help, and fine-grained detection of acoustic-phonetic cues is necessary to convey a message. By revealing dissociations between superficially competent but underlyingly effortful processing (as in the case of adults who are able to carry out discrimination tasks only when attentional and cognitive resources are recruited), studies like this one can provide greater understanding of the processes that can interfere with, or support, the acquisition of native-like speech sound discrimination in adult second language learning. Findings from neuroscientific investigations of cross-linguistic speech perception therefore offer a foundation for translational work that could allow curriculum designers and teachers help adult second language learners develop their full learning potential from the very start of the L2 learning process.

REFERENCES

Aaltonen, O., Niemi, P., & Nyrke, T. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology*, *24*, 197–207. Retrieved from http://www.sciencedirect.com/science/article/pii/0301051187900020

Abdi, H. (2010). The greenhouse-geisser correction. *Encyclopedia of Research Design. Sage Publications*, 1–10. Retrieved from http://www.utd.edu/~herve/abdi-GreenhouseGeisser2010-pretty.pdf

Abramson, A.S, & Lisker, L. (1970). Discriminability along the voicing continuum: Cross language tests. *Proceedings of the 6th International Congress of Phonetic Sciences.* (pp. 569-573) Prague: Academia, Czechoslovak Academy of Sciences.

Alho, K., Connolly, JF., Cheour, M., Lehtokoski, A., Huotilainen, M., Virtanen, J.,… Ilmoniemi, RJ (1998a). Hemispheric lateralization in pre-attentive processing of speech sounds. *Neuroscience Letters*, 258:9–12.

Alho, K., Winkler, I., Escera, C., Huotilainen, M., Virtanen, J., Jäskelääinen, IP., & Pekkonen, E. (1989b). Processing of novel sounds and frequency changes in the human auditory cortex: magnetoencephalographic recordings. *Psychophysiology*, 35:211–24.

Aoyama, K. (2003). Perception of syllable-initial and syllable-final nasals in English by Korean and Japanese speakers. *Second Language Research*, *19*(3), 251–265. doi:10.1191/0267658303sr222oa

Archila-Suerte, P., Zevin, J., Ramos, A. I., & Hernandez, A. E. (2013). The neural basis of non-native speech perception in bilingual children. *NeuroImage*, *67*, 51–63. doi:10.1016/j.neuroimage.2012.10.023

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child--adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech*. 54 (4), 317-342. DOI:10.1177/0023830908099068

Beddor, P. S., & Gottfried, T. L. (1995). Methodological issues in cross-language speech perception research with adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 207–232). Baltimore, MD: York Press.

Bent, T., Kewley-Port, D., & Ferguson, S. H. (2010). Across-talker effects on non-native listeners' vowel perception in noise. *The Journal of the Acoustical Society of America*, *128*(5), 3142–51. doi:10.1121/1.3493428

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

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775. doi:10.1121/1.1332378

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345. Retrieved from http://psycnet.apa.org/journals/xhp/14/3/345/

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*(3), 305–330. Retrieved from http://www.haskins.yale.edu/sr/SR109/SR109_07.pdf

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language speech learning: In Honor of James Emil Flege*. (pp.13–34). Amsterdam: Benjamin.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., & Possing, E.T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex,* 10 (5), 512–528.

Birdsong, D. (2003). Authenticité de prononciation en français L2 chez des apprenants tardifs anglophones: Analyses segmentales et globales. *Acquisition et Interaction en Langue Étrangère*, 18, 17-36.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis*, 20, 3–49.

Boersma, P., P. Escudero & R. Hayes (2003). Learning abstract phonological from auditory phonetic categories: an integrated model for the acquisition of language-specific sound cate- gories. In M. J. Sole, D. Recansens & J. Romero (eds.), *Proceedings of the 15th International Con- gress of Phonetic Sciences,* 1013-1016. Barcelona: Causal Productions. Boersma.

Boersma, P., & Weenink, D. (2013). *Praat: doing phonetics by computer* [computer program]. Version 5.3.53, retrieved 9 July 2013 from http:// www.praat.org/.

Bohn, O.-S. (1995). Cross language speech perception in adults: first language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross language research* (pp. 279–304). Timonium, MD: York Press.

Bohn, O.-S. & Flege, J. E. (1990) Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, 11, 303-328.

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience,* 25, 151–188.

Bongaerts, T., Planken, B., & Schils, E. (1995). Can late starters attain a native accent in a foreign language? A test of the critical period hypothesis. In D. Singleton & Z. Lengyel (Eds.), (pp. 30–50). *The age factor in second language acquisition.* Clevedon, England: Multilingual Matters Limited.

Bongaerts, T., Mennen S. & Van derSlik, F. (2000). Authenticity of pronunciation in naturalistic second language acquisition: the case of very advanced late learners of Dutch as a second language, *Studia Linguistica*, 54, 298-308.

Bongaerts, T., Summeren, Ch., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition,* 19, 4: 447-465. doi: 10.1017/S0272263197004026

Bongaerts, G. (2003). Eingefleischte Sozialität. Zur Phänomenologie sozialer Praxis. *Sociologia Internationalis*, 36, 246 – 260.

Borich M., Arora, S., Jacobson Kimberley, T. (2009). Lasting effects of repeated rTMS application in focal hand dystonia. *Restorative Neurology and Neuroscience,* 27, 55–65. doi: 10.3233/RNN-2009-0461

Bosch, L., & Sebastian-Galles, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, *46*(2-3), 217–243. doi:10.1177/00238309030460020801

Bradlow, A, R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America, 97*(3), 1916 – 1924. doi.org/10.1121/1.412064

Bradlow, AR., Kraus, N., Nicol, TG., McGee, TJ., Cunningham, J., Zecker, SG., & Carrel, TD. (1999). Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children. *Journal of the Acoustical Society of America,* 106, 2086–96.

Brain areas associated with language. (2002). [Online image]. (2002). Retrieved June 12, 2013 from http://everythingspeech.com/evaluation/aphasia/aphasia-2/

Brattico, E., Winkler, I., Näätänen, R., Paavilainen, P., & Tervaniemi, M. (2002). Simultaneous storage of two complex temporal sound patterns in auditory sensory memory. *NeuroReport* 13: 1747 – 1751.

Broca, P. (1861). Remarques sur le siége de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bulletin de la Société Anatomique de Paris,* 6, pp. 330–357.

Browman, C. P. & L. Goldstein (1989). Articulatory gestures as phonological units. *Phonology*,6, 201-251.

Brown, D. (2000). *Principles of language learning and teaching.* 4th ed. New York: Pearson.

Butler, Y. & Hakuta, K. (2004). Bilingualism and second language acquisition. In T. Bhatia &

W. Ritchie (eds.), *Handbook of bilingualism* (pp. 114-144). Malden, MA: Blackwell

Campbell, T.A., Winkler, I., & Kujala, T. (2007). N1 and the mismatch negativity are spatiotemporally distinct ERP components: Disruption of immediate memory by auditory distraction can be related to N1. *Psychophysiology*, 44, 530–540. doi:10.1111/j.1469-8986.2007.00529.x

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*(3), 372–387. doi:10.1016/j.wocn.2005.08.003

Ceponiene, R., Choeur, M., & Näätänen, R. (1998). Interstimulus interval and auditory event-related potential in children. Evidence for multiple generators. *Electroencephalography and Clinical Neurophysiology, 108,* 345 – 354.

Chomsky, N. (1957). *Syntactic structures*. Mouton, The Hague.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience,* 1, 351–3.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.

Colin, C., Hoonhorst, I., Markessis, E., Radeau, M., De Tourtchaninoff, M., Foucher, A., & Collet, G., (2009). Mismatch negativity (MMN) evoked by sound duration contrasts: an unexpected major effect of deviance direction on amplitudes. *Clinical neurophysiology*, *120*(1), 51–9. doi:10.1016/j.clinph.2008.10.002

Coopler, C., Pisoni, D., & Jong, K. (2005) Acoustic Characteristics of the vowel system of six regional varieties of American English. *Journal of the Acoustical Society of America, 118*(3), 1661 – 1676.

Cowan, N., & Morse, P. (1986). The use of auditory and phonetic memory in vowel discrimination. *The Journal of the Acoustical Society of America*, *79*(February). Retrieved from http://link.aip.org/link/?JASMAN/79/500/1

Csépe, V., Pantev, C., Hoke, M., Hampson, S., Ross, B. (1992). Evoked magnetic responses of the human auditory cortex to minor pitch changes: localization of the mismatch field. *Electroencephalography Clinical Neurophysiology*, 84 (6), 538–48.

Curtiss, J. (1977). *Genie: a psycholinguistic study of a modern-day "wild child".* New York: Academic Press.

D'ausilio, A., Craighero, L., & Fadiga, L. (2012). The contribution of the frontal lobe to the perception of speech. *Journal of Neurolinguistics,* 25, 238 – 235 doi:10.1016/j.jneuroling.2010.02.003.

Damasio, A., & Geschwind, N, (1984). The neural basis of language. *Annual Review of Neuroscience, 7*: 127–147.

Dattalo, P. (2008). *Determining sample size: Balancing power, precision, and practicality.* New York: Oxford University Press.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, *8*(4), 919–924. Retrieved from http://journals.lww.com/neuroreport/Abstract/1997/03030/Electrophysiological_correlates_of_categorical.21.aspx

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *Journal of Cognitive Neuroscience*, *12*(4), 635–647. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/089892900562390

Dehaene-Lambertz, G., & Peña, (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *NeuroReport*, 12: 3155-3158.

DeKeyser, R., Alfi-Shabtay I., & Ravid, D. (2010). Crosslinguistic evidence for the nature of age effects in second language acquisition. *Applied Linguistics,* 31, 413 – 438 doi:10.1017/S0142716410000056

Deouell, LY., Karns, CM., Harrison, TB., & Knight, RT. (2003). Spatial asymmetries of auditory event-synthesis in humans. *Neuroscience Letters*, 335, 171–174.

Dien, J., & Santuzzi, A. M. (2005). Application of repeated measures ANOVA to high-density ERP datasets: A review and tutorial. In T. C. Handy (Ed.) *Event-related potentials: A methods handbook*. (pp. 57-82). Cambridge: MIT Press.

Eimas, P. D. (1978). Developmental aspects of speech perception. In R. Held, H. W. Leibowitz, & H. L. Teuber (Eds.). *Handbook of sensory physiology.* (Vol. 8). Berlin: Springer-Verlag.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Escera, C., Alho, K., Schröger, E. & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neurootology*, 5, 151-166.

Escudero, P. (2000). *Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels in Spanish speakers*. (Unpublished M.Sc. thesis). University of Edinburgh, United Kingdom.

Escudero, P. (2005). *Linguistic perception and second language acquisition*. LOT Landelijke Onderzoekschool Taalwetenschap. Retrieved from http://www.lotpublications.nl/publish/articles/001488/bookpart.pdf

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(04). doi:10.1017/S0272263104040021

Escudero, P., & Chládková, K. (2010). Spanish listeners' perception of American and Southern British English vowels. *The Journal of the Acoustical Society of America*, *128*(5), 254–259. doi:10.1121/1.3488794

Escudero, P., Benders, T.,& Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *Journal of the Acoustical Society of America*, 130, 206–212.

Evans, B. G., Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America* 121(6), 3814-3826

Fabbro, F. (1999). Aphasia in multilinguals. In: F. Fabbro (Ed.), *Concise encyclopedia of language pathology* (pp. 335–41). Oxford: Pergamon Press

Fabbro, F. (2001). The bilingual brain: Bilingual aphasia. *Brain and Language* 79, 201–210. doi:10.1006/brln.2001.2480

Flege, J. (1991). The interlingual identification of Spanish and English vowels: Orthographic evidence. *Quarterly Journal of Experimental Psychology, 43A*(3), 701 – 731

Flege, J. (1995). Second language speech learning theory findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

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

Flege, J. E. (2003). Methods for assessing the perception of vowels in a second language: A categorial discrimination test. In E. Fava, & A. Mioni (Eds.), *Issues in clinical linguistics* (pp. 19–44). Padova: UniPress.

Flege, J. E. (2009). *Give input a chance! In T. Piske and Young-Scholten, M. (Eds) Input Matters in SLA. Bristol: Multilingual Matters, Pp. 175-190.* Retrieved from http://jimflege.com/files/Flege_in_Piske_2009.pdf

Flege, J., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics*, *34*(2), 153–175. doi:10.1016/j.wocn.2005.05.001

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

Flege, J., & Fletcher, K. (1992). Talker and listener effects on degree of perceived foreign accent. *The Journal of the Acoustical Society of America*, *91*(January), 370–389. Retrieved from http://link.aip.org/link/?JASMAN/91/370/1

Flege, J., MacKay, I., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America, 106*, 2973 - 2987.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5 Pt 1), 3125–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7759652

Flege, JE., Schirru, C., & Mackay, I.R.A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication,* 40, 467-491.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second language acquisition. *Journal of Memory and Language, 41(1),* 78–104.

Fogerty, D., & Kewley-Port, D. (2009). Perceptual contributions of the consonant-vowel boundary to sentence intelligibility. *The Journal of the Acoustical Society of America*, *126*(2), 847–57. doi:10.1121/1.3159302

Fowler, C. A. (1986). An even approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics,* 14, 1-38.

Fourakis, M. (1991). Tempo, stress, and vowel reduction in American English. *Journal of the Acoustical Society of America*, 90, 1816-1827.

Fox, R., Flege, J. E., & Munro, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: a multidimensional scaling analysis. *The Journal of the Acoustical Society of America*, *97*(4), 2540–51. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7714272

Fox, R., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America*, 126(5), 2603 – 2618.

doi:10.1121/1.3212921

Friederici, AD., Friedrich, M., & Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *NeuroReport*, 13, 1251–4.

Frenck-Mestre, Ch., Meunier, C., Espesser, R., Daffner, K., & Holcomb, P. (2010). Perceiving nonnative vowels: The effect of context on perception as evidenced by event-related brain potentials. *Journal of Speech, Language and Hearing Research,* 48, 1-15.

Fox, R., & Jacewicz E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustic Society of America*, 126(5), 2603-2618. doi: 10.1121/1.3212921

Garcia, P; & Froud, K. (2007, May). *N100 & P300 responses to native and nonnative vowels in late Spanish-English bilinguals.* Poster session presented at the annual Cognitive Neuroscience Society meeting. New York. NY.

Garcia, P; Froud, K. (2008) *Speech perception in late bilinguals: An EEG study on vowel perception.* [PowerPoint Slides]. Paper presented at the 2008 conference of the American Association of Applied Linguistics. Washington, DC.

Garcia, P; & Froud, K. (2010, April). *Longitudinal change in MMN responses to English vowels by a Spanish-speaking learner of English.* Poster session to be presented at the annual Cognitive Neuroscience Society meeting. Montreal, Canada.

Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum.

Gasser, M. (2002). English and Spanish vowel quadrilateral [Online image]. Retrieved June 12, 2013 from http://www.indiana.edu/~hlw/PhonUnits/vowels.html.

Giegerich, H. (1992). *English Phonology: An Introduction*, Cambridge: Cambridge University Press.

Genesee, F., Paradis, J. & Crago, M. (2004). *Dual language development and disorders: A* *handbook on bilingualism and second language learning*, Baltimore, MD: Brookes.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & psychophysics*, *66*(3), 363–76. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15283062

Gibson, J.J. (1966). The senses considered as perceptual systems. Boston: Houghton Mifflin.

Gibson, J.J. (1979). The Ecological Approach to Visual Perception. Boston: Houghton Mifflin.

Golestani N., Molko N., Dehaene S., & LeBihan D, Pallier C. (2006). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex,* 17, 575– 582. doi:10.1093/cercor/bhk001

González-García, N., Armony, J., Soto, J., Trejo, D., Alegría, M., & Drucker-Colín, R. (2011). Effects of rTMS on Parkinson's disease: a longitudinal fMRI study. *Journal of Neurology*. 258, 1268–1280. DOI 10.1007/s00415-011-5923-2

Gordon, P., Keyes, L., Yung, Y-F. (2001). Ability to perceive non-native contrasts. Performance on natural and synthetic speech stimuli. *Perception and Pyscophysics,* 63(4), 746-758.

Gottfried, T. L. & Beddor, P. S. (1988). Perception of spectral and duration information in French vowels. *Language and Speech,* 31, 57-75.

Gow, D. W. Jr., & Segawa, J. A. (2009). Articulatory mediation of speech perception: a causal analysis of multi-modal imaging data. *Cognition*, 110, 222–236.

Grimm, S., Snik, A., & Van Der Broek, P. (2004). Differential processing of duration changes within short and long sounds in humans. *Neuroscience Letters*, 356, 83 -86.M.

Gross M., Nakamura L., Pascual-Leone A., Fregni F. (2007). Has repetitive transcranial magnetic stimulation (rTMS) treatment for depression improved? A systematic review and meta-analysis comparing the recent vs. the earlier rTMS studies. *Acta Psychiatrica Scandinavica*. 116: 165–173. doi: 10.1111/j.1600-0447.2007.01049.

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*, 2711–2724. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10830393

Haenschel, C., Vernon. DJ., Dwivedi, P., Gruzelier, JH., & Baldeweg, T. (2005). Event- related brain potential correlates of human auditory sensory memory- trace formation. *Journal of Neuroscience,* 25 104 94–501.

Handy, T. C. (2005). *Event-Related potentials: A Methods handbook.* Cambridge: Bradford/MIT Press.

Hari R., Hälämäinen M., Ilmoniemi R., Kaukoranta E., Reinikainen K., Salminen J., …Sams, M. (1984). Responses of the primary auditory cortex to pitch changes in a sequence of tone pips: neuromagnetic recordings in man. *Neuroscience Letters*, 50, 127–32.

Harnsberger, J. D. (2000). A cross-language study of the identification of non-native nasal consonants varying in place of articulation. *Journal of the Acoustical Society of America* 108, 764-783.

Harris, J. (1969). *Spanish phonology*. Cambridge, Mass: MIT Press.

Hesling, I., Dilharreguy, B., Bordessoules, M., & Allard, M. (2012). The neural processing of second language comprehension modulated by the degree of proficiency: a listening connected speech FMRI study. *The open neuroimaging journal*, *6*(1), 44–54. doi:10.2174/1874440001206010044

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in cognitive sciences*, *4*(4), 131–138. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10740277

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1-2), 67–99. doi:10.1016/j.cognition.2003.10.011

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature*, *8*(May), 393–402.

Higawara, R. (2009). Wide band spectrograms of the vowels of American English in a /b__d/ context [Online image]. Retreived June 12, 2013 from http://home.cc.umanitoba.ca/~robh/howto.html#formants

Hillebrand, J. M., Clark, M. J., & Houde, R. a. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, *108*(6), 3013–22. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11144593

Hillenbrand, J., Clark, M., & Nearey, T. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America*, 109, 748-763.

Hisagi, Miwako; Shafer L. Varlerie; Strange, W; Sussman, E. (2011). Perception of a Japanese Vowel Length Contrast by Japanese and American English Listeners: Behavioral and Electrophysiological measures. *Brain*, 89–105. doi:10.1016/j.brainres.2010.08.092.Perception

Hoffman, R.E., Anderson, A., Varanko, M., Gore, J., Hampson, M. The time course of regional brain activation associated with onset of auditory/verbal hallucinations. British Journal of Psychiatry 2008; 193: 424-425

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, *119*(5), 3059. doi:10.1121/1.2188377

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

Hugdahl, K., Thomsen, T., Ersland, L., Morten Rimol, L., & Niemi, J. (2003). The effects of attention on speech perception: An fMRI study. *Brain and Language*, *85*(1), 37–48. doi:10.1016/S0093-934X(02)00500-X

Ito, K., & Strange, W. (2009). Perception of allophonic cues to English word boundaries by Japanese second language learners of English. *The Journal of the Acoustical Society of America*, *125*(4), 2348–60. doi:10.1121/1.3082103

Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *The Journal of the Acoustical Society of America*, *99*(2), 1130–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8609297

Iverson, P., Kuhl, P. K., Akahane-yamada, R., & Diesch, E. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, 47–57. doi:10.1016/S0

Iverson, P., Evans, B. G. (2007). Auditory training of English vowels for first-language speakers of Spanish and German. *16th International Congress of Phonetic Sciences*. (pp.1625-1628).

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866–77. doi:10.1121/1.3148196

Jakoby, H., Goldstein, A., & Faust, M. (2011). Electrophysiological correlates of speech perception mechanisms and individual differences in second language attainment. *Psychophysiology,* 48, 1516 – 1530. doi 10.1111/j.1469-8986.2011.01227.x

Jemel, B., Achenbach, C., Müller, BW., Röpcke, B., & Oades, R. (2002). Mismatch negativity results from bilateral asymmetric dipole sources in the frontal and temporal lobes. Brain Topography, 15, 13–27.

Jia, G., Strange, W., Wu, Y., Collado, J., Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *Journal of the Acoustic Society of America,* 119(2), 1118-1130. doi: 10.1121/1.2151806

Joanisse, M.F., Zevin, J.D., & McCandliss, B.D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using FMRI and a short- interval habituation trial paradigm. *Cerebral Cortex,* 17(9), 2084–2093. doi:10.1093/cercor/bhl124

Johnson, J., & Newport, E. (1989). Critical period effects in sound language learning: The influence of maturation state on the acquisition of English as a second language. CognitivePsychologist, 21, 60–99

Junqué, C., Vendrell, P., & Vendrell, J. (1995). Differential impairments and specific phenomena in 50 Catalan-Spanish bilingual aphasic patients. In M. Paradis (Ed.), Aspects of bilingual aphasia (pp. 177–210). Oxford, UK: Pergamon Press.

Jusczyk, P. (1992): "Developing phonological categories from the speech signal." *Phonological development: models, research, implications*, edited by Charles A. Ferguson, Lise Menn & Carol Stoel-Gammon. Timonium, Md.: York Press.

Jusczyk, P.W., & Hohne, E.A. (1997). Infants' memory for spoken words. *Science*, 277, 1984 – 1986.

Karypidis, C. (2007). Order effects and vowel decay in short-term memory: The neutralization hypothesis. *In Proceedings of the 16th international congress of phonetic sciences*, 2007, pp. 657–660

Katayama, J., & Polich J. (1996). P300 from one-, two-, and three-stimulus auditory paradigms. *International Journal of Psychophysiology,* 23, 33–40.

Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *122*(4), 2365–75. doi:10.1121/1.2773986

Kim, H. S., Relkin, N. R., Lee, K. M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, 388, 171–174.

Kinoshita, K., Behne, D. M., & Arai, T., "Duration and F0 as perceptual cues to Japanese vowel quantity", *Proceedings of the 7th International Conference on Spoken Language Processing, 2002.* Online:http://www.splab.ee.sophia.ac.jp/papers/2002/2002_03.pdf , accessed on 2 Sep, 2009.

Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schröger, E., & Jacobsen, T. (2008). Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, *67*(2), 131–43. doi:10.1016/j.ijpsycho.2007.10.012

Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustic Society of America*, 59(5), 1208 – 21.

Knudsen, E. I. (1999). Early experience and critical periods. In M. J. Zigmond (Ed.), Fundamental Neuroscience (pp. 637–654). San Diego, CA: Academic Press.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of cognitive neuroscience*, *16*(8), 1412–25. doi:10.1162/0898929042304796

Kockhar, R. (2005). The occupational status and mobility of Hispanics. Pew Hispanic Center. Retrieved from: http://www.pewhispanic.org/files/reports/59.pdf

Kohnert, K., & Bates, E. (2002). Balancing bilinguals II: Lexical comprehension and cognitive processing in children learning Spanish and English. *Journal of Speech, Language, and Hearing Research, 45*, 347-359.

Kondaurova, M. V, & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America*, *124*(6), 3959. doi:10.1121/1.2999341

Korpilahti, P., Krause, C. M., Holopinen, I., & Jang, A. H. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language,* 76, 332 – 339.

Korzyukov, O., Alho, K., Kujala, A., Gumenyuk, V., Ilmoniemi, RJ., Virtanen, J., … Näätänen, R. (1999). Electromagnetic responses of the human auditory cortex generated by sensory-memory based processing of tone-frequency changes. *Neuroscience Letters*, 276, 169–72.

Krashen, S. (1982). *Principles and practice in second language acquisition.* New York: Pergamon Press.

Krashen, S. (1988). *Second Language Acquisition and Second Language Learning*. Prentice-Hall International.

Kreegipuu, K., & Näätänen, R. (2011). The Mismatch Negativity. In Kappenman, E., & Luck, S. (Eds.) *The Oxford Handbook of Event-Related Potential Components*. (pp 145-156). New York: Oxford University Press.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Science*, 97, 11850 -11857

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M. & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society* B, 363, 979-1000

Kuhl, P., & Iverson, P. (1995). Linguistic experience and the perceptual magnet effect. Retrieved from http://discovery.ucl.ac.uk/23733/

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9* (F13-F21).

Kuhl, P. K. (2009). Early language acquisition: Neural substrates and theoretical models. In M. S. Gazzaniga (Ed.), The Cognitive Neurosciences, 4th Edition (pp. 837-854). Cambridge, MA: MIT Press.

Lengeris, A. (2009). Individual differences in second-language vowel learning. Retrieved from http://discovery.ucl.ac.uk/19029/

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

Levänen S., Ahonen A., Hari R., McEvoy L., & Sams, M. (1996). Deviant auditory stimuli activate human left and right auditory cortex differently. *Cerebral Cortex,* 6:288–96.

Liberman, a M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4075760

Linden, D. (2005). The P300: Where in the brain is it produced and what does it tell us? *The Neuroscientist*, 11, 563. doi: 10.1177/1073858405280524

Lipski, S. C., Escudero, P., & Benders, T. (2012). Language experience modulates weighting of acoustic cues for vowel perception: An event-related potential study. *Psychophysiology*. doi:10.1111/j.1469-8986.2011.01347.x

Lisker, L. & Abramson, A. (1970). The voicing dimension: some experiments in comparative phonetics. *Proceedings 6th International Congress of Phonetic*

Sciences: 563-567.

Lively, S., Logan, J., & Pisoni, D. (1993). Training Japanese listeners to identify English /r/ and /l/. II The role fo phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*(3), 1242 - 1255.

Lively, S. E., Pisoni, D. B., Yamada, R. a, Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, *96*(4), 2076–87. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7963022

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

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12,* 251-285.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of research on language acquisition: Vol. 2. Second language acquisition* (pp. 413-468). New York: Academic Press.

Luck, S. J. (2005). *An introduction to the event-related potential technique.* Cambridge: MIT Press.

Luria, A. (1966). The higher cortical function in man. New York: Basic Books.

Major, R. (2001). Foreign accent: The ontogeny and phylogeny of second language phonology. Mahwah, NJ: Lawrence Erlbaum.

Mainy, N., Jung, J., Baciu, M., Kahane, P., Schoendorff, B., Minotti, L.,…Lachaux, J. (2008). Cortical dynamics of word recognition. *Human Brain Mapping*, 29, 1215–1230.

Massaro D. W. and Cohen, M. M. (1983). Phonological constraints in speech perception. *Perception and Psychophysics, 34*, 338-348.

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In C. Howell, S. Fish, & T. Keith-Lucas (Eds), *Proceedings of the 24th Annual Boston University Conference* on Language Development, 522–533. Somerville, MA: Cascadilla Press.

McGuire, G. (2010). *A brief primer on experimental designs for speech perception research*. Retrieved from http://people.ucsc.edu/~gmcguir1/experiment_designs.pdf

Meisel, J. (1989). Early differentiation of languages in bilingual children. In *Bilingualism across the lifespan: Aspects of acquisition, maturity and loss*. In K. Hyltenstam. & L. Obler, (eds), 13-40. Cambridge, UK: Cambridge University Press.

Meisel, J. (2001). The simultaneous acquisition of two first languages: Early differentiation and subsequent development of grammars. In J. Cenoz & F. Genesee (Eds). *Trends in Bilingual Acquisition*. (pp.11-41). Amsterdam: Benjamins.

Minagawa-Kawai, Y., Mori, K., Sato, Y., Koizumi, T. (2004). Differential cortical responses in second language learners to different vowel contrasts. *NeuroReport,* 15, 899–903.

Morrison, G.S. 2002. *Effects of L1 duration experience on Japanese and Spanish listeners' perception of English high front vowels*. Unpublished master's thesis, Simon Fraser University, Burnaby, British Columbia, Canada.

Morrison, G.S. (2006). Methodological issues in L2 perception research, and vowel spectral cues in Spanish listeners' perception of word-final /t/ and /d/ in Spanish. In M. Diaz-Campos (Ed.), *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology* (pp. 35–47). Somerville, MA: Cascadilla Proceedings Project.

Morrison, G. S. (2008). L1-Spanish Speakers' Acquisition of the English /i /-/I/ Contrast: Duration-based perception is not the initial developmental stage. *Language and Speech*, *51*(4), 285–315. doi:10.1177/0023830908099067

Morrison, G. S. (2009). L1-Spanish speakers' acquisition of the English /i/-/I/ Contrast II: Perception of vowel inherent spectral change. *Language and Speech*, *52*(4), 437–462. doi:10.1177/0023830909336583

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in second language acquisition*, (March 1996), 81–108. Retrieved from http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=36620

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear & Hearing, 16*(1), 6-18.

Näätänen, R. (1992). *Attention and Brain Function.* Hillsdale. NJ: Lawrence Earlbaum.

Näätänen, R., & Alho, K. (1997). Mismatch negativity – The measure for central sound representation accuracy. *Audiology and Neuro-Otology,* 2, 341 – 353.

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective attention effect on evoked potential reinterpreted. *Acta Psychologica,* 42, 313 – 329.

Näätänen, R., Lehtokoski, A., Lennest, M., Luuki, A., Alliki, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Advances*, *161*, 4–6. Retrieved from http://audition.ens.fr/P2web/Maria/Naatanen97.pdf

Näätänen, R., Lehtokoski, A., Lennest, M., Cheour, M., Huotilainen, M., livonent, A., ... & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Letters to Nature*, 3851, 432 - 434.

Näätänen, R., Pakarinen, S., Rinne, T., Takegata, R. (2004). The mismatch negativity (MMN)—towards the optimal paradigm. Clinical Neurophysiology 115, 140–144.

Näätänen, R., & Picton T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of component structure. *Psychophysiology*, 24, 375–425.

Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., Sams, M. (1987b). Interstimulus interval and the mismatch negativity. In: Evoked Potentials III, The Third International Evoked Potentials Symposium (Eds. C. Barber, & T. Blum), pp. 392–397. Butter- worths, Boston.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, *118*(12), 2544–90. doi:10.1016/j.clinph.2007.04.026

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: influence of training sets. *Journal of speech, language, and hearing research : JSLHR*, *50*(6), 1496–509. doi:10.1044/1092-4388(2007/103)

Nishi, K., Strange, W., Akahane-Yamada, R., Kubi, R., & Trent-Brown, S. A. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. Journal of the Acoustical Society of America, 124, 576–588.

Nittrouer, S. (2002). Learning to apprehend phonetic structure from the speech signal: The hows and whys. *In Investigations in Clinical Phonetics and Linguistics*. Edited by F. Windsor, M. L. Kelly, and N. Hewlett. London, England: Lawrence Erlbaum.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. (2005). Decision making, the P3, and the Locus-coeruleus norepinephrine system. *Psychological Bulletin,* 131(4), 510 -532. doi: 10.1037/0033-2909.131.4.510

Pająk, Bożena, & Roger Levy. 2011. Phonological generalization from distributional evidence. In L. Carlson, C. Holscher, & T. Shipley (Eds.). *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* 2673–2678. Austin, TX: Cognitive Science Society.

Pakarinen, S., Takegata, R., Rinne, T., Huotilainen, M., & Näätänen, R. (2007). Measurement of extensive auditory discrimination profiles using mismatch negativity (MMN) of the auditory event-related potential (ERP). *Clinical Neurophysiology*, 118, 177-185.

Pakarinen, S., Lovio, R., Huotilainen, M., Alku, P., Näätänen, R., & Kujala, T. (2009). Fast multi-feature paradigm for recording several mismatch negativities (MMNs) to phonetic and acoustic changes in speech sounds. *Biological psychology*, *82*(3), 219–26. doi:10.1016/j.biopsycho.2009.07.008

Paradis, M. (1977). Bilingualism and aphasia. In H. Whitaker & H. A. Whitaker (Eds.), *Studies in neurolinguistics* (Vol. 3, pp. 65–121). New York: Academic Press.

Paradis, M. (2001). Bilingual and polyglot aphasia. In R. S. Berndt (Ed.), *Handbook of neuropsychology* (2nd ed.) (pp. 69–91). Oxford, UK: Elsevier Science.

Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics, 11,* 73–89.

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

Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S., et al. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human Brain Mapping*, 19, 170–182.

Pew Research Hispanic Center (2011). Statistical portrait of foreign-born population in the United States. Retrieved from: http://www.pewhispanic.org/files/2013/01/PHC-2011-FB-Stat-Profiles.pdf

Phillips, C., Marantz, A., McGinnis, M., Pesetsky, D., Wexler, K., Yellin, E., Poeppel, D., Roberts, T., Rowley, H. (1995). Brain Mechanisms of Speech Perception: A Preliminary Report. In C. Schütze, K. Broihier & J. Ganger (Eds.), *Papers on Language Processing and Acquisition* (pp. 153-191). Cambridge, MA: MITWPL

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., … Roberts, T. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *Journal of Cognitive Neuroscience*, 12, 1038–1055.

Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning, 44*, 493-527.

Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology & Neuro-Otology,* 5, 111 -139.

Piske, T.,MacKay, I. R.A., and Flege, J. E. (2001). Factors affecting the degree of foreign accent in an L2: A review. *Journal of Phonetic*s, 29, 191–215.

Polich J. (2003). Overview of P3a and P3b. In: Polich J (Ed.), *Detection of change: event-related potential and fMRI findings.* Boston, MA: Kluwer.

Polich J. (2004). Clinical application of the P300 event-related potential. *Physical Medicine and*

*Rehabilitation Clinics of North America***,** 15, 133–61.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, *118*(10), 2128–48. doi:10.1016/j.clinph.2007.04.019

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological Psychology, 41*(103-146).

Polivanov, E. (1932). La perception des sons d'une langue e´trange`re. Travaux du Cercle Linguistique de Prague, 4, 79-96.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *The Journal of the Acoustical Society of America*, *89*(June), 2961–2977. Retrieved from http://link.aip.org/link/?JASMAN/89/2961/1

Polka, L., Bohn, O.-S., 1996. A cross-language comparison of vowel perception in English-learning and German-learninginfants. *The Journal of the Acoustical Society of America*, 100, 577–592.

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221–231. doi:10.1016/S0167-6393(02)00105-X

Polka, L., & Bohn, O.-S. (2010). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467–478. doi:10.1016/j.wocn.2010.08.007

Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Progress in neurobiology*, *79*(1), 49–71. doi:10.1016/j.pneurobio.2006.04.004

Pons, F., Albareda-Castellot, B., & Sebastián-Gallés, N. (2012). The interplay between input and initial biases: asymmetries in vowel perception during the first year of life. *Child development*, *83*(3), 965–76. doi:10.1111/j.1467-8624.2012.01740.x

Raphael, L., Borden, G., & Harris, K. (2011). *Speech Science Primer: Physiology, Acoustics and Perception of Speech* (6th ed). Baltimore, MD. Lippincott Williams & Wilkins

Rallo Fabra, L. (2005). Predicting ease of acquisition of L2 speech sounds: A perceived dissimilarity test. Vigo International Journal of Applied Linguistics, 2, 75–92.

Rallo Fabra, L., & Romero, J. (2012). Native Catalan learners perception and production of English vowels. Journal of Phonetics (40) 491-508. doi:10.1016/j.wocn.2012.01.001

Rauber, A. S., Escudero, P., Bion, R., & Baptista, B. O. (2005). The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese. *Proceedings of Interspeech,* 2005, 2913–2916.

Ressel, V., Pallier, C., Ventura-Campos, N., Díaz, B., Roessler, A., Ávila, C., & Sebastián-Gallés, N. (2012). An effect of bilingualism on the auditory cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *32*(47), 16597–601. doi:10.1523/JNEUROSCI.1996-12.2012.

Rinne, T., Alho, K., Ilmoniemi, R.J., Virtanen, J., Näätänen, R. (2000). Separate time behaviors of the temporal and frontal mismatch negativity sources. *Neuroimage,*12: 14–9.

Rivera-Gaxiola, M., Csibra, G., Johnson, MH., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural and Brain Research Journal,* 111(1-2), 13 – 23.

Sabri, M., & Campbell, KB. (2001). Effects of sequential and temporal probability of deviant occurrence on mismatch negativity. *Cognitive Brain Research*, 12, 171–80.

Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In R. Siegler & & D. Kuhn (Eds.). *Handbook of child development* (Vol. 6, pp. 58–108). New York: Wiley.

Sams, M., & Nätäänen, R. (1991) Neuromagnetic responses of the human auditory cortex to short frequency glides. *Neuroscience Letters*, 121, 43–46.

Schaffler, L., Luders, H., Dinner, D., Lesser, R., & Chelune, G. (1993). Comprehension deficits elicited by electrical stimulation of Broca's area. *Brain*, 116, 695–715.

Sheldon, A. and Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3, 243–61.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002) E-Prime Reference Guide. Pittsburgh: Psychology Software Tools Inc.

Schouten, B., Gerrits, E., & Van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication,* 41, 71 – 80.

Schwartz, J.-L., Abry, C., Boë, L.-J., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion–Focalisation Theory. *Speech Communication*, *45*(4), 425–434. doi:10.1016/j.specom.2004.12.001

Schwartz, B. & Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12: 40-72.

Seliger, H., Krashen, S., & Ladefoged, P. (1982). Maturational constraints in the acquisition of second languages. In S. Krashen, R. Scarcella, & M. Long (Eds.), *Child-adult differences in second language acquisition* (pp. 13–19). Rowley, MA: Newbury House.

Siok, W. T., Jin, Z., Fletcher, P., & Tan, L. H. (2003). Distinct brain regions associated with syllable and phoneme. *Human Brain Mapping*, 18, 201–207.

Skinner, B.F. (1957). *Verbal Behavior*. NY: Appleton. [Spanish translation: Conducta Verbal. Mexico, Trillas (1957).

Strange, W. (1992). Learning non-native phoneme contrasts: Interactions among subjects, stimulus, and task variables. In E. Tohkura, Vatikiotis-Bateson. & Y. Sagisaka (Eds.), *Speech Perception, Production and Linguistic structure* (pp. 197–219). Tokyo: Ohmsha.

Strange, W. (1995). *Speech perception and linguistic experience: Issues in cross-language research*. Timonium: York press. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Speech+Perception+and+Linguistic+Experience+Issues+in+Cross-Language+Research#0

Strange, W. (2007). Cross-language phonetic similarity of vowels: Theoretical and methodological issues. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language speech learning: In Honor of James Emil Flege*. (pp. 35–55). Amsterdam: Benjamin.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., and Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *The Journal of the Acoustical Society of America,* 109, 1691–1704.

Strange, W., Bohn, O.-S., Nishi, K., & Trent, S. a. (2005). Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America*, *118*(3), 1751. doi:10.1121/1.1992688

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36*(2), 136 - 145.

Strange, W., Hisagi, M., Akahane-Yamada, R., & Kubo, R. (2011). Cross-language perceptual similarity predicts categorial discrimination of American vowels by naïve Japanese listeners. *The Journal of the Acoustical Society of America*, *130*(4), EL226. doi:10.1121/1.3630221

Strange, W. & Shafer, V.L. (2008). Speech perception in second language learners: the re-education of selective perception. Zampini, M., & Hansen, J. (Eds). *Phonology and Second Language Acquisition*. Cambridge University Press.

Strange, W., Yamada, R. A., Kubo, R., Trent, S. A., Nishi, K., and Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of . Phonetics,* 26, 311–344.

Sussman, E., Kujala, T., Halmetoja, J., Lyytinen, H., Alku, P., & Näätänen, R. (2004). Automatic and controlled processing of acoustic and phonetic contrasts. *Hearing Research*, *190*(1-2), 128–40. doi:10.1016/S0378-5955(04)00016-4

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.) *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125-144). Oxford, England: Oxford University Press.

Schwartz, J., Abry, C., Boë, L., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion-Focalisation Theory. *Speech Communication,* 45, 425 -434.

Tabors, P.O. (1997). *One Child, two languages: A guide for preschool educators*. Baltimore, MD: Paul H. Brookes Publishing Co.

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological science*, *21*(10), 1532–40. doi:10.1177/0956797610384142

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear and hearing*, *22*(2), 79–90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11324846

Tremblay, K., Piskosz, M., & Souza, P. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clinical Neurophysiology*, *114*(7), 1332–1343. doi:10.1016/S1388-2457(03)00114-7

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, *33*(3), 263–290. doi:10.1016/j.wocn.2004.10.002

U.S. Department of Labor Employment and Training Administration Report (2005). English proficiency: What employers need for their Spanish-speaking workforce. Online: www.doleta.gov/reports/pdf/English_Proficiency_Report.pdf, accessed on October, 2012

Vance, T.J. (1987). An introduction to Japanese phonology. State University of New York Press, Albany, NY.

Vygotsky, Lev S. (1978). *Mind in society: Development of higher psychological processes*. In Cole, M., John-Steiner, V., Scribner, S., Souberman, E. (Eds). Cambridge, MA: Harvard University Press.

Walley, A., & Flege, J. (1999). Effects of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics*, 27, 307-332

Wanrooij, K., Escudero, P., and Raijmakers, M. (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal phonetics*. In press.

Watkins, K., & Paus, T. (2004). Modulation of motor excitability during speech perception: the role of Broca's area. *Journal of cognitive neuroscience*, *16*(6), 978–87. doi:10.1162/0898929041502616

Weber-Fox, C. and Neville, H.J. (1999). Functional neural subsystems are differentially affected by delays in second-language immersion: ERP and behavioral evidence in bilingual speakers. In D. Birdsong (Ed), *New perspectives on the critical period for second language acquisition*, (pp. 23-38). Hillsdale NJ: Lawrence Erlbaum.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Attention, Perception, & Psychophysics*, *37*(1), 35–44. Retrieved from http://www.springerlink.com/index/B12T035531515841.pdf

Werker, J. F. & Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development,* 7, 49– 63.

Wernicke, C. (1874). Der Aphasische Symptomencomplex: Eine Psychologische Studie Auf Anatomischer Basis, Whitefish, MT. Kessinger Publishing.

Winkler, I., Cowan, N., Csépe, V., Czigler, I., & Näätänen R. (1996). Interactions between transient and long-term auditory memory as reflected by the mismatch negativity. *Journal of Cognitive Neuroscience*, 8, 403–15.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, a, Czigler, I., et al. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*(5), 638–42. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10442032

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception and Psychophysics, 26(2),* 95–104.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7, 701–702.

Wong, P., Warrier, C., Penhune, B., Roy,K., Sadehh, A., Parrish, T., Zatorre R. (2008) Volume of left Heschl's gyrus and linguistic pitch learning. *Cerebral Cortex* (18) 828–836. Medline.

Vera-Constán, F., & Sebastián-Gallés, N. (2008). Asymmetries in speech perception: Perceptual salience at the lexical level. Poster presented at the 29th international congress of psychology, Berlin.

Ylinen, S., Huotilainen, M., & Näätänen, R. (2005). Phoneme quality and quantity are processed independently in the human brain. *Neuroreport*, *16*(16), 1857–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16237342

Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., & Näätänen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: a cross-linguistic study. *Brain research*, *1072*(1), 175–85. doi:10.1016/j.brainres.2005.12.004

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256, 846–849.

Zevin, J. D., Datta, H., Maurer, U., Rosania, K. a, & McCandliss, B. D. (2010). Native language experience influences the topography of the mismatch negativity to speech. *Frontiers in human neuroscience*, *4*(November), 212. doi:10.3389/fnhum.2010

APPENDICES

1. Consent Forms

INFORMED CONSENT FORM FOR NATIVE SPANISH SPEAKERS

Title of the project: Perception of American English vowels by adult sequential Spanish – English bilinguals: An EEG study.

DESCRIPTION OF THE RESEARCH: You are invited to participate in a research study on how adult Spanish-English bilinguals perceive the vowels of English. It is well known that when a person learns a second language in adulthood, there are some differences in the way the second language (L2) is learned, perceived and produced. One problem for adult Spanish speakers learning English is the different vowel sounds used in English. We want to find out whether Spanish – English bilinguals and English monolingual speakers brains attend to the same or different auditory information when perceiving some English vowels. We will do this by asking you to participate in some tasks that involve identifying different vowels, and by monitoring your brain activity as you listen to some vowel sounds.

You have been asked to participate in this study as an EXPERIMENTAL PARTICIPANT. As an experimental participant you will provide us with information about how the brains of adult Spanish-English bilinguals respond to American English vowel sounds the same stimuli that we will be presenting to native English speakers.

PROCEDURES: In this research project, you will be asked to come to the Neurocognition of Language laboratory and complete 2 sessions of testing and brain data collection.

The first time you come to the lab, we will ask you to fill out a language background questionnaire. Then, we will ask you read a list of words in English while your voice is recorded. Recording will take around five minutes. After that, we will get ready to record brain data. The recording of brain data, or electroencephalography (EEG), involves the following steps. Your head size will be measured and you will have a net placed on your head that contains sensors within small sponges that sit directly on the scalp. The sponges are first soaked in a weak salt solution (potassium chloride), which helps pick up small electrical signals. The minute signals generated by brain activity are recorded through the sensors. In the first listening task, while we record EEG, you will listen to sounds through earphones while watching a silent movie. This allows us to gather information about how your brain is processing speech sounds. Following on from this, the investigator will teach you some symbols used to transcribe speech sounds from different languages (symbols from the International Phonetic Alphabet, or IPA). You will listen to some sounds from English, and practice remembering which sound and symbol go together. Finally, we will once again ask you to carry out a listening task while we record your brain data; however, in the second listening task, you will be asked to push one of two buttons to identify the sounds you hear, while EEG is recorded.

The second time you come to the lab, you will carry out the same tasks as before, but we will change the vowel sounds slightly.

RISKS AND BENEFITS:

Participation in research always involves some risk. In this study we will use physiological recording techniques to monitor your brain activity. As with all physiological recording, there is a minimal risk of electric shock. The amount of electric shock risk could be compared to the risk of using a toaster or a hair dryer. This is minimized by using a special isolated amplifier, and by ensuring that you are never connected to earth ground.

There is a small risk of skin irritation, associated with application of the sensor net to your scalp. We minimize this risk by careful choice of electrolyte, which is a simple salt solution. There is also a small risk of skin infection, minimized by careful and complete disinfection of electrodes. The sensor net will be wet when applied, and this may be slightly uncomfortable at first. However, towels are provided so as to minimize discomfort and to protect your clothing.

The experimental and training tasks can be repetitive, and you may find them somewhat boring and/or difficult to complete. However, you can take breaks during the experiment and training and continue only when you feel ready.

There is no direct benefit to you for participating in the study. We hope that your participation will help us understand more about how bilingual adults perceive non-native vowels, and how the brain responds to training programs in auditory perception.

If you feel uncomfortable or concerned with the net application or the procedures used, feel absolutely free to discuss them with the experimenter. You may stop participating at any time with no penalty whatsoever.

REIMBURSEMENT

We will make small cash payments to thank you for your time and participation, at the end of the study. Payments are $15 for each of the two occasions when we record your brain activity. The total amount of reimbursement for your participation, if you complete all the study requirements, will be $ 30.

CONFIDENTIALITY:

Your privacy is VERY important to us, and we are extremely careful to protect your identity.

Computer files will be stored on password-protected computers, which can be accessed only by members of the research team. Data files are identified by numbers, which are assigned separately to each person. The only place where your name and your identifying number will be

stored together is on this consent form. Digital files will be kept indefinitely and may be used for future analyses.
We ask you to provide contact information on this form also, so that we can keep in touch with you during the study in order to make an appointment for your second visit; however your contact information will NEVER be disclosed to anyone. You will be given a copy of this form to keep, and the only other copy will be stored in a locked filing cabinet in the laboratory.

When we report results from our studies (e.g. at meetings to discuss research, or in professional journals), we usually report results from many people together, as averages. We NEVER use names when reporting or discussing data.

TIME INVOLVEMENT: Your participation will take approximately one to two hours on the first day, and another one to two hours on the second day that you take part in the study, plus travel time.

HOW WILL RESULTS BE USED: The results of the study will be used in the dissertation of the principal investigator, in professional reports for publication in journals, and for presentation at professional and academic conferences.

CONSENT:

I agree that I ____[Name] am willing to take part in the study entitled Perception of American English vowels by adult Spanish-English bilinguals: An EEG study.

I have had an opportunity to ask questions about the study, and I understand what is involved.

Signed: ____

Date (mm/dd/yyyy): _______//_______//_________

Please also sign the Participant's Rights form (attached).

Teachers College, Columbia University

PARTICIPANT'S RIGHTS

Principal Investigator: ____

Research Title: ____

- I have read and discussed the Research Description with the researcher. I have had the opportunity to ask questions about the purposes and procedures regarding this study.
- My participation in research is voluntary. I may refuse to participate or withdraw from participation at any time without jeopardy to future medical care, employment, student status or other entitlements.
- The researcher may withdraw me from the research at his/her professional discretion.
- If, during the course of the study, significant new information that has been developed becomes available which may relate to my willingness to continue to participate, the investigator will provide this information to me.
- Any information derived from the research project that personally identifies me will not be voluntarily released or disclosed without my separate consent, except as specifically required by law.
- If at any time I have any questions regarding the research or my participation, I can contact the investigator, who will answer my questions. The investigator's phone number is (347)207-8517.
- If at any time I have comments, or concerns regarding the conduct of the research or questions about my rights as a research subject, I should contact the Teachers College, Columbia University Institutional Review Board /IRB. The phone number for the IRB is (212) 678-4105. Or, I can write to the IRB at Teachers College, Columbia University, 525 W. 120th Street, New York, NY, 10027, Box 151.
- I should receive a copy of the Research Description and this Participant's Rights document.
- If video and/or audio taping is part of this research, I ( ) consent to be audio/video taped. I ( ) do NOT consent to being video/audio taped. The written video and/or audio taped materials will be viewed only by the principal investigator, and members of the research team.
- Written, video and/or audio taped materials ( ) may be viewed in an educational setting outside the research

  ( ) may NOT be viewed in an educational setting outside the research.
- My signature means that I agree to participate in this study.

Participant's signature: ________________________________ Date:____/____/____

Name: ________________________________

Investigator's Verification of Explanation

I certify that I have carefully explained the purpose and nature of this research to __________________________________ (participant's name) in age-appropriate language. He/She has had the opportunity to discuss it with me in detail. I have answered all his/her questions and he/she provided the affirmative agreement (i.e. assent) to participate in this research.

Investigator's Signature: ____

Date: ______________________

Participant Contact Information

Please provide us with an email address and phone number where we can contact you for the duration of the study. We will never share this information with anyone.

Name: ___________________________________

Email: ____________________________________

Telephone: ________________________________

Participation Record

EEG session 1

EEG session 2

INFORMED CONSENT FORM FOR NATIVE ENGLISH SPEAKERS

Title of the project: Perception of American English vowels by adult sequential Spanish – English bilinguals: An EEG study.

DESCRIPTION OF THE RESEARCH: You are invited to participate in a research study on how adult Spanish-English bilinguals perceive the vowels of English. It is well-known that when a person learns a second language in adulthood, there are some differences in the way the second language (L2) is learned, perceived and produced. One problem for adult Spanish speakers learning English, is the different vowel sounds used in English. We want to find out whether Spanish – English bilinguals and English monolingual speakers attend to the same or different auditory information when perceiving some English vowels. We will do this by asking you to participate in some tasks that involve identifying different vowels, and by monitoring your brain activity as you listen to some vowel sounds.

You have been asked to participate in this study as a CONTROL PARTICIPANT. As a control participant you will provide us with a comparison of how the brains of native speakers of English respond to the same American English vowel sounds that we will be presenting to Spanish speakers who speak English as a second language.

PROCEDURES: In this research project, you will be asked to come to the Neurocognition of Language laboratory and complete 2 sessions of testing and brain data collection.

The first time you come to the lab, we will ask you to fill out a language background questionnaire. Then, we will ask you read a list of words in English while your voice is recorded. Recording will take around five minutes. After that, we will get ready to record brain data. The recording of brain data, or electroencephalography (EEG), involves the following steps. Your head size will be measured and you will have a net placed on your head that contains sensors within small sponges that sit directly on the scalp. The sponges are first soaked in a weak salt solution (potassium chloride), which helps pick up small electrical signals. The minute signals generated by brain activity are recorded through the sensors. In the first listening task, while we record EEG, you will listen to sounds through earphones while watching a silent movie. This allows us to gather information about how your brain is processing speech sounds. Following on from this, the investigator will teach you some symbols used to transcribe speech sounds from different languages (symbols from the International Phonetic Alphabet, or IPA). You will listen to some sounds from English, and practice remembering which sound and symbol go together. Finally, we will once again ask you to carry out a listening task while we record your brain data; however, in the second listening task, you will be asked to push one of two buttons to identify the sounds you hear, while EEG is recorded.

The second time you come to the lab, you will carry out the same tasks as before, but we will change the vowel sounds slightly.

RISKS AND BENEFITS:

Participation in research always involves some risk. In this study we will use physiological recording techniques to monitor your brain activity. As with all physiological recording, there is a minimal risk of electric shock. The amount of electric shock risk could be compared to the risk of using a toaster or a hair dryer. This is minimized by using a special isolated amplifier, and by ensuring that you are never connected to earth ground.

There is a small risk of skin irritation, associated with application of the sensor net to your scalp. We minimize this risk by careful choice of electrolyte, which is a simple salt solution. There is also a small risk of skin infection, minimized by careful and complete disinfection of electrodes. The sensor net will be wet when applied, and this may be slightly uncomfortable at first. However, towels are provided so as to minimize discomfort and to protect your clothing.

The experimental and training tasks can be repetitive, and you may find them somewhat boring and/or difficult to complete. However, you can take breaks during the experiment and training and continue only when you feel ready.

There is no direct benefit to you for participating in the study. We hope that your participation will help us understand more about how bilingual adults perceive non-native vowels, and how the brain responds to training programs in auditory perception.

If you feel uncomfortable or concerned with the net application or the procedures used, feel absolutely free to discuss them with the experimenter. You may stop participating at any time with no penalty whatsoever.

REIMBURSEMENT

We will make small cash payments to thank you for your time and participation, at the end of each of your two visits to the lab. Payments are $15 for each of the two occasions when we record your brain activity. The total amount of reimbursement for your participation, if you complete all the study requirements, will be $30.

CONFIDENTIALITY:

Your privacy is VERY important to us, and we are extremely careful to protect your identity.

Computer files will be stored on password-protected computers, which can be accessed only by members of the research team. Data files are identified by numbers, which are assigned separately to each person. The only place where your name and your identifying number will be stored together is on this consent form. Digital files will be kept indefinitely and may be used for future analyses.
We ask you to provide contact information on this form also, so that we can keep in touch with you during the study in order to make an appointment for your second visit; however your contact information will NEVER be disclosed to anyone. You will be given a copy of this form to keep, and the only other copy will be stored in a locked filing cabinet in the laboratory.

When we report results from our studies (e.g. at meetings to discuss research, or in professional journals), we usually report results from many people together, as averages. We NEVER use names when reporting or discussing data.

TIME INVOLVEMENT: Your participation will take approximately one to two hours on the first day, and another one to two hours on the second day that you take part in the study, plus travel time.

HOW WILL RESULTS BE USED: The results of the study will be used in the dissertation of the principal investigator, in professional reports for publication in journals, and for presentation at professional and academic conferences.

CONSENT:

I agree that I ____[Name] am willing to take part in the study entitled Perception of American English vowels by adult Spanish-English bilinguals: An EEG study

I have had an opportunity to ask questions about the study, and I understand what is involved.

Signed: ____

Date (mm/dd/yyyy): _______//_______//_________

Please also sign the Participant’s Rights form (attached)

Teachers College, Columbia University

PARTICIPANT'S RIGHTS

Principal Investigator: ____

Research Title: ____

- I have read and discussed the Research Description with the researcher. I have had the opportunity to ask questions about the purposes and procedures regarding this study.
- My participation in research is voluntary. I may refuse to participate or withdraw from participation at any time without jeopardy to future medical care, employment, student status or other entitlements.
- The researcher may withdraw me from the research at his/her professional discretion.
- If, during the course of the study, significant new information that has been developed becomes available which may relate to my willingness to continue to participate, the investigator will provide this information to me.
- Any information derived from the research project that personally identifies me will not be voluntarily released or disclosed without my separate consent, except as specifically required by law.
- If at any time I have any questions regarding the research or my participation, I can contact the investigator, who will answer my questions. The investigator's phone number is (347)207-8517.
- If at any time I have comments, or concerns regarding the conduct of the research or questions about my rights as a research subject, I should contact the Teachers College, Columbia University Institutional Review Board /IRB. The phone number for the IRB is (212) 678-4105. Or, I can write to the IRB at Teachers College, Columbia University, 525 W. 120th Street, New York, NY, 10027, Box 151.
- I should receive a copy of the Research Description and this Participant's Rights document.
- If video and/or audio taping is part of this research, I ( ) consent to be audio/video taped. I ( ) do <u>NOT</u> consent to being video/audio taped. The written, video and/or audio taped materials will be viewed only by the principal investigator and members of the research team.
- Written, video and/or audio taped materials ( ) may be viewed in an educational setting outside the research

  ( ) may <u>NOT</u> be viewed in an educational setting outside the research.
- My signature means that I agree to participate in this study.

Participant's signature: ________________________________ Date:____/____/____

Name: ____________________________

Investigator's Verification of Explanation

I certify that I have carefully explained the purpose and nature of this research to ________________________________ (participant's name) in age-appropriate language. He/She has had the opportunity to discuss it with me in detail. I have answered all his/her questions and he/she provided the affirmative agreement (i.e. assent) to participate in this research.

Investigator's Signature: _______________________________________

Date: _____________________

Participant Contact Information

Please provide us with an email address and phone number where we can contact you for the duration of the study. We will never share this information with anyone.

Name: _________________________________

Email: __________________________________

Telephone: ______________________________

Participation Record

EEG session 1

EEG session 2

2. Language Background Questionnaire

The questionnaire used for this study has been adapted with permission from a questionnaire designed by Dr. Erika Levy.

Language Background Questionnaire

Please complete this questionnaire to the best of your knowledge and add any information you feel might be relevant (use the back of the paper if needed).

Name: ____________________________________Participant number: ______________
Date: ___________________________e-mail address: _____________________________
Address: ____
Telephone Numbers: (Home) __________________ (Work) ______________________
Date of Birth: ________________ Gender: _________
Birthplace: ____________________________ _________________________________
Town/City State/Country
Occupation: ____________________Number of years of education after high school ___
How did you find out about this study? ____

Places in which you have lived for more than 1 year:
City/State/Country Years
________________________________ from age _______ to age _______
________________________________ from age _______ to age _______
________________________________ from age _______ to age _______
________________________________ from age _______ to age _______
If you have lived in more places please check here _____ and continue on the back.

Parent 1's Birthplace: ____
Languages parent 1 spoke fluently: __________________________________
Parent 2's Birthplace: ____
Languages parent 2 spoke fluently: __________________________________
Parent 3's Birthplace: ____
Languages parent 3 spoke fluently: __________________________________

What languages were spoken in your home when you were growing up? (for example, by parents, guardians, grandparents, or relatives) ___________________________________

What languages are spoken in your home now? _________________________________

What languages do you speak fluently and understand without effort?
1. _____________________ 2. _______________________ 3. ___________________

What language(s) did you speak/understand as a child (before going to school)?
1. _______________________2. _______________________ 3. ___________________

What language(s) were used in your classrooms in elementary school?
1. _______________________2. _______________________ 3. ___________________

If you have studied English, please answer the following questions as accurately as you can. If your native language is English, please continue to the asterisk (*) on Page 4:

How old were you when you started learning English? __________

School

How many years did you take English in high school: ______, college: ______, graduate school: _ other: _?

Did you have any native speakers of the language as teachers or tutors? **No____ Yes____ Don't know** ________
If **yes**, please specify the number of semesters or months with native speakers as teachers or tutors: ________________________________

How many years of English (in total) have you studied? _________

Overall in your English classes, what percent of the time was the focus on pronunciation (on average)? ____

If focus on pronunciation differed in initial years versus in later years, please specify:
____

Overall in your English classes, what percent of the time was devoted to informal activities (i.e., natural conversation in real-life situations)? _____________ Please explain: _________________
____
____

How long ago did you last take an English class? _______________

Experience in English-speaking country

How many years and/or months have you been living in the United states? ___________________

____

How many years and/or months have you received academic instruction in English? ___________
____
____

What percent of the time do you spend speaking English with English-speaking friends/colleagues? (Please explain) ____
____

What is the primary language that you speak at home?______________________

What percent of the time of the day do you speak English and in what capacity? ______________
____
____

How important is it to for your professional success to know English? (Please explain) _______
____

Rating

On a scale from 1-5 (with 1 being "like a native English person" and 5 being "very foreign") how would you rate your English in terms of:

	Like native **English**				Very **Foreign**
a. General proficiency:	**1**	**2**	**3**	**4**	**5**
b. Pronunciation:	**1**	**2**	**3**	**4**	**5**
c. Grammar:	**1**	**2**	**3**	**4**	**5**
d. Comprehension:	**1**	**2**	**3**	**4**	**5**
e. Reading:	**1**	**2**	**3**	**4**	**5**

Comments if you have any:____
____

On a scale from 1-5 (with 1 being "It is very important to me" and 5 being "It is not important at all to me") how would you rate your desire to be as proficient as possible in the following areas of English?

Very important to me **Not at all important to me**

f. General proficiency:	**1**	**2**	**3**	**4**	**5**
g. Pronunciation:	**1**	**2**	**3**	**4**	**5**
h. Grammar:	**1**	**2**	**3**	**4**	**5**
i. Comprehension:	**1**	**2**	**3**	**4**	**5**
j. Reading:	**1**	**2**	**3**	**4**	**5**

Comments if you have any:____
____

***What (other) language(s) did you study as a foreign language in school?**

A.________________________ **B.**________________________ **C.**______________________

Please answer the following questions about the languages you listed above:

Regarding Language A: How old were you when you started learning it? ____________

How many years total (approximately) did you study it? ____________

On a scale from 1-5 (with 1 being "like a native speaker of the language" and 5 being "very foreign") how would you rate your skills today in the foreign language in terms of:

	Like native **speaker**			Very **foreign**	
a. General proficiency:	**1**	**2**	**3**	**4**	**5**
b. Pronunciation:	**1**	**2**	**3**	**4**	**5**
c. Grammar:	**1**	**2**	**3**	**4**	**5**
d. Comprehension:	**1**	**2**	**3**	**4**	**5**
e. Reading:	**1**	**2**	**3**	**4**	**5**

Comments if you have any:____

____

Regarding Language B:

On a scale from 1-5 (with 1 being "like a native speaker of the language" and 5 being "very foreign") how would you rate your skills today in the foreign language in terms of:

	Like native speaker				Very foreign
a. General proficiency:	**1**	**2**	**3**	**4**	**5**
b. Pronunciation:	**1**	**2**	**3**	**4**	**5**
c. Grammar:	**1**	**2**	**3**	**4**	**5**
d. Comprehension:	**1**	**2**	**3**	**4**	**5**
e. Reading:	**1**	**2**	**3**	**4**	**5**

Comments if you have any:____

____

Regarding Language C:

On a scale from 1-5 (with 1 being "like a native speaker of the language" and 5 being "very foreign") how would you rate your skills today in the foreign language in terms of:

Like native Very

speaker **foreign**

a. General proficiency: **1** **2** **3** **4** **5**

b. Pronunciation: **1** **2** **3** **4** **5**

c. Grammar: **1** **2** **3** **4** **5**

d. Comprehension: **1** **2** **3** **4** **5**

e. Reading: **1** **2** **3** **4** **5**

Comments if you have any:____

____

Talent

On a scale from 1-5 (with 1 being "very talented" and 5 being "not talented at all") how would you rate your talent for the following skills?:

	Very talented				**Not talented at all**
Ability to imitate sounds in foreign languages	**1**	**2**	**3**	**4**	**5**
Talent in learning languages	**1**	**2**	**3**	**4**	**5**
Musical talent	**1**	**2**	**3**	**4**	**5**

Comments: ____

____

Have you ever studied Phonetics (the scientific study of speech sounds)? **YES / NO**

If YES, have you ever done phonetic transcription? **YES / NO**

If YES, how much? ________________________________

Do you have normal hearing (as far as you know)? **YES / NO**

Did you at any time have therapy for a speech, reading or other language problem?______ Please specify: ____

Which hand do you write with? (Circle one):	Right	Left	Either
Which hand do you throw a ball with?	Right	Left	Either
Which hand do you wave good-bye with?	Right	Left	Either
Which hand do you hold a spoon in?	Right	Left	Either
Which of these do you consider yourself?	Right-handed	Left-handed	Ambidextrous

Comments if any ____

What do you consider your racial/ethnic background to be? Check all that apply.

(**Optional: You need not answer)**

Caucasian ____ Native American ___

African American ____ Pacific Islander ______

Hispanic ___ Asian American _____

Other-please specify __________________________

3. Demographic information

Number	Group	Gender	Handedness	Age	Place of Birth	Native Language	AoA	English Instruc	LOR	English proficiency	Daily use of English
930-933	Study	F	Right	30	Santiago, Chile	Spanish	10	8.5	3	1	70%
931-937	Study	M	Right	26	Monterey, Mexico	Spanish	7	5	5	2	69%
938-942	Study	F	Right	27	Viña del Mar, Chile	Spanish	8	10	1.5	2	40%
941-944	Study	F	Right	24	Santo Domingo, DR	Spanish	4	14	0.6	2	80%
946-948	Study	M	Right	30	Mexico City, Mexico	Spanish	3	15	1	2	90%
949-950	Study	F	Right	22	Manizales, Colombia	Spanish	20	2	1.5	3	70%
959-962	Study	M	Right	33	Limache, Chile	Spanish	14	8	2	2	25%
960-969	Study	F	Right	28	Iquique, Chile	Spanish	14	15	0.6	3	40%
984-995	Study	F	Right	35	Caracas, Venezuela	Spanish	12	6	0.6	3	100%
987-992	Study	M	Right	26	Puebla, Mexico	Spanish	12	8	1	2	80%
935-945	Comparison	M	Right	22	Washington	English					
961-968	Comparison	F	Right	20	New Hyde Park, NY	English					
973-981	Comparison	F	Right	23	Des Moines, IA	English					
975-989	Comparison	F	Right	23	Danbury, CT	English					

985-988	Comparison	F	Right	23	Englewood, NJ	English
986-1000	Comparison	M	Right	30	Brooklyn, NY	English
993-1003	Comparison	F	Right	25	Artesia, NM	English
997-1006	Comparison	F	Right	23	Staten Island, NY	English
1004-1007	Comparison	F	Right	30	Chicago, Il	English
1011-1013	Comparison	M	Right	22	Syosset, NY	English