

# Perception of American English vowels by sequential Spanish–English bilinguals\*

PAULA B. GARCÍA

Boys Town National Research Hospital

KAREN FROUD

Department of Biobehavioral Sciences, Teachers College - Columbia University.

(Received: May 04, 2015; final revision received: June 30, 2016; accepted: July 04, 2016; first published online 13 September 2016)

*Research on American-English (AE) vowel perception by Spanish–English bilinguals has focused on the vowels /i/-/ɪ/ (e.g., in sheep/ship). Other AE vowel contrasts may present perceptual challenges for this population, especially those requiring both spectral and durational discrimination. We used Event-Related Potentials (ERPs), MMN (Mismatch Negativity) and P300, to index discrimination of AE vowels /a/-/ʌ/ by sequential adult Spanish–English bilingual listeners compared to AE monolinguals. Listening tasks were non-attended and attended, and vowels were presented with natural and neutralized durations. Regardless of vowel duration, bilingual listeners showed no MMN to unattended sounds, and P300 responses were elicited to /a/ but not /ʌ/ in the attended condition. Monolingual listeners showed pre-attentive discrimination (MMN) for /a/ only; while both vowels elicited P300 responses when attended. Findings suggest that Spanish–English bilinguals recruit attentional and cognitive resources enabling native-like use of both spectral and durational cues to discriminate between AE vowels /a/ and /ʌ/.*

Keywords: ERPs, MMN, P300, vowel perception, second-language learners, bilingualism, attention, durational and spectral cues

## 1. Introduction

One of the greatest challenges that second-language (L2) learners face is the perception and production of non-native speech sounds. Languages differ as to which acoustic-phonetic cues (e.g., spectral, durational, voicing cues) signal distinctions between speech sounds. Therefore, native language experience with specific cues may influence the perception of speech sounds in a second language (Iverson, Kuhl, Akahane-yamada & Diesch, 2003).

Several models have been proposed to account for the process of L2 speech sound acquisition (Best, 1995; Best & Tyler, 2007; Flege, 1995; Escudero, 2005). These share the perspective that learning speech sounds of a foreign language occurs under the influence of the already-established native language system. The specific nature of perceptual/acoustic similarities between L1 and L2 separates these models. The Speech Learning Model

(SLM: Flege, 1995) focuses on speech perception and production of non-native speech sounds, and posits that when the learner is familiar with the L2 sound system, dissimilarities between L2 speech sounds and the closest L1 congeners facilitate phonetic learning. The Perceptual Assimilation Model (PAM: Best, 1995) provides an account for non-native speech perception in naïve listeners by predicting the discriminability of non-native speech sound contrasts, depending on how specific contrasts are assimilated to L1 speech categories. The PAM-L2 (Best & Tyler, 2007), an extension of PAM, has focused on L2 learners. This model proposes four scenarios in which L2 phonological categories are assimilated to pre-existing L1 speech categories, and predicts the likelihood of L2 new category formation under each set of conditions. In the first of these four scenarios, only one member of an L2 phonological contrast is assimilated to an existing L1 category. In this case, the other member of an L2 contrast could be perceived as EITHER not being part of any L1 category, OR assimilated to two L1 categories. Therefore, the discrimination and formation of a new L2 category is predicted for this member of the contrast. In a second scenario, both L2 phonological categories are assimilated to a single L1 category. However, one member of the contrast is perceived as being more deviant than the other, rendering the formation of a new L2 category contingent on the extent to which this member of the contrast is distinct from the L1 category. A third scenario describes a more difficult process, where both L2 categories are perceived as equally good exemplars of only one L1

\* The research reported here was partially funded by a grant to the first author (Vice President's Student Research in Diversity Grant from Teachers College, Columbia University), and by P20 GM109023 (Boys Town National Research Hospital). It formed part of the first author's doctoral dissertation work. The authors thank the following people for assistance during the study: Dayna Moya, Guannan Shen, Lisa Levinson, Heather Green, Felicidad García and Trey Avery in the Neurocognition of Language Lab, Teachers College, Columbia University. We are deeply thankful to Kanae Nishi, Michael Gorga, Rachel Scheperle, Ben Kirby, Erika Levy and Laura Sánchez for their comments and suggestions on the manuscript and on earlier versions of this work.

Address for correspondence:

Paula García, Boys Town National Research Hospital, 555 North, 30<sup>th</sup> Street, Omaha, NE 68131

[paula.garcia@boystown.org](mailto:paula.garcia@boystown.org)

category, and the formation of a new L2 category is predicted to be unlikely. In the fourth scenario, there is no L1-L2 phonological assimilation, and therefore either sound (or both) could be easy to categorize, depending on how the specific L2 sound compares to other L1 categories (Best & Tyler, 2007; Flege, 1995). Similar to PAM-L2, the L2 Linguistic Perception model (L2LP; Escudero, 2005, van Leussen & Escudero, 2015) states that new L2 speech sounds are influenced by the production of those sounds in L1 environments. In a so-called 'new scenario', similar to the second scenario in PAM-L2, the learner has difficulty learning non-native contrasts, because the L2 contrast is acoustically close to a specific L1 sound. In a 'similar scenario', two L2 sounds are acoustically close to two different L1 sounds; this makes learning of new speech sound contrasts easier, because there is no need to create a new category, and the learner simply shifts existent perceptual categories to accommodate the L2 contrast. Finally, the 'subset scenario', in which one non-native sound is perceived as belonging to multiple L1 categories, represents less of a challenge for the learner. Based on these models the perceptual challenges encountered by Spanish-speaking learners of AE are predicted for certain AE vowel contrasts – namely, those in which L2 contrasts are acoustically similar to L1 speech sounds.

The vowel systems of Spanish and AE differ both in the number of contrasting categories and in the number of cues necessary for the vowel sounds to be accurately distinguished. The vowel system in Spanish consists of five vowels (/i/, /e/, /a/, /o/, /u/), which is a small inventory compared to the 11 monophthong vowels of American English (/i/, /ɪ/, /e/, /ɛ/, /ɜ/, /ʌ/, /æ/, /ɑ/, /ɔ/, /ʊ/, /u/; Bradlow, 1995; Clopper, Pisoni & De Jong, 2005). Unlike vowels in AE, Spanish vowels are differentiated only spectrally; durational cues are not used to signal lexical differences (Cebrian, 2006; Hammond, 2001; Harris, 1969). Despite the overlapping use of phonetic symbols for Spanish and some AE vowels, Spanish vowels do not have direct counterparts in the English vowel system, because of these representational differences.

Although studies have shown that consonants are relevant for lexical access (Nespor, Peña & Mehler, 2003), vowels play a decisive role in intelligibility (Bent, Bradlow & Smith, 2007). For example, in tasks that require listeners to change a nonsense word into a real word, listeners tend to change vowels rather than consonants (Cutler, Sebastián-Gallés, Soler-Vilageliu & Van Ooijen, 2000; Van Ooijen, 1996), suggesting that vowel intelligibility is a heavily-weighted cue for speech perception (Fogerty & Kewley-Port, 2009; Kewley-Port, Burkle & Lee, 2007). In addition, previous studies have shown that L2 learners have more difficulties learning L2 vowels than consonants (Munro & Derwing, 2008).

Distinctions between some English vowel pairs (e.g., /i/-ɪ/) constitute common perceptual confusions for

Spanish speakers. Extensive work has been done on perception by Spanish speakers of the English vowel contrast /i/-ɪ/, from different dialect variations (AE: Bohn, 1995; Flege, 1991; Flege & Munro, 1994; Fox, Flege & Munro, 1995; Flege, Bohn & Jang, 1997; British English: Escudero, 2005; Escudero & Boersma, 2004; Escudero & Chládková, 2010; Canadian English: Morrison, 2006; 2008; 2009). These studies demonstrated that L1-Spanish listeners rely mostly or exclusively on durational cues to distinguish between these vowels, at least at some stages in the learning process (Escudero, 2005), while native AE speakers use spectral cues as primary and duration as a secondary cue. Interestingly, the reliance of Spanish-speaking listeners on durational cues has also been demonstrated for the perception of Dutch vowels (long /a:/ and short /a/, spectrally similar to Spanish /a/; Lipski, Escudero & Benders, 2012), suggesting that Spanish listeners may rely on durational cues specifically for perceiving non-native vowel contrasts that are similar to native Spanish vowels. A similar reliance on durational cues is observed when other listeners who employ no durational cues in their L1 phonology process English vowel contrasts (e.g., Mandarin-English: Bohn, 1995; Russian-English: Kondaurova & Francis, 2008).

To explain this preference for durational cues despite the lack of specific phonological relevance in certain languages, Bohn (1995) proposed that when spectral cues are not accessible (e.g., when spectral differences between speech sounds are small, as for AE vowels), non-native listeners rely on psychoacoustically more salient durational information to perceive L2 vowel contrasts. Conversely, Escudero (2005) argues that L1 Spanish listeners rely more on durational cues because this dimension is a 'blank slate' (since it is not a phonetic cue in Spanish), and it is therefore easier for learners to create new speech-sound categories along this dimension.

With respect to similarity/proximity between Spanish and AE vowels, contrasts other than the much studied /i/-ɪ/ pair, such as /ɑ/-ʌ/, may be equally or even more challenging. For the /ɑ/-ʌ/ contrast this could be due to the similarity of both non-native vowels to the Spanish vowel /a/, and to the existence of other mid to low AE vowels (/ɑ/, /ʌ/, /æ/) that could cause perceptual confusion. Specifically, Fox et al. (1995) compared acoustic characteristics of AE and Spanish vowels and found that formant values for the Spanish low central vowel /a/ fall between those for English /ʌ/ and /ɑ/. Acoustic comparison of AE vowels and Spanish (Madrid) vowels (Bradlow, 1995) revealed that mean formant values for Spanish /a/ (F1 = 683 Hz, F2 = 1353 Hz) are very close to formant values for AE /ʌ/ (F1 = 640 Hz, F2 = 1354 Hz) in CVC contexts. AE vowel /ɑ/ had a higher F1 (780 Hz), while the F2 (1244 Hz) was lower compared to AE /ʌ/ and Spanish /a/. Spanish-speaking learners of English may find it difficult to distinguish among these

non-native vowels because they are not part of their native vowel inventory, they are so close to each other in the vowel space, and they are also close to the native Spanish vowel /a/. Escudero and Chládková (2010) reported that Peruvian Spanish speakers classified AE vowels /ɑ/-/ʌ/ as Spanish /a/ 99% and 53% of the time respectively. Based on these findings, it is predicted that the AE vowel contrast /ɑ/-/ʌ/ should pose difficulties for L1-Spanish listeners because both vowels would be perceived as similar to Spanish /a/. However, as reported for the vowel pair /ɪ/-/i/ (Bohn, 1995; Flege 1991; Flege & Munro, 1994; Fox et al., 1995; Flege et al., 1997), it is possible that at least at some point in their learning process, L1-Spanish listeners rely primarily on durational cues to differentiate these vowels.

Perceiving difficult non-native speech sound contrasts may also be affected by attentional factors. Gordon, Eberhardt and Rueckl (1993) found that native listeners weighted speech cues differently when distracted from or attending to speech sounds, indicating that speech sound classification is influenced by attention. Similarly, Hisagi, Strange, Shafer and Sussman (2010) found differences in neurophysiological responses to Japanese vowel duration contrasts between Japanese and AE listeners when they were distracted from the vowel sounds.

Speech perception processes like the ones described so far unfold over milliseconds, and by the time a behavioral response is recorded, perceptual processing has already taken place in the brain (Hickok & Poeppel, 2007). To further understand how acoustic-phonetic cues are weighted when perceiving non-native contrasts, it is helpful to investigate early brain responses indexing sensory perception of speech sounds. One way of examining brain processes associated with speech sound distinctions is through electrophysiological methodologies, such as event related potentials (ERP).

Two specific ERP components have been identified as having specific importance in speech sound processing, and attentional resource allocation: the Mismatch Negativity (MMN) and the P300. Both the MMN and the P300 have been well studied and their respective properties are understood to reflect subtle changes in speech segments (e.g., vowels, consonants) at both acoustic and phonetic levels.

The MMN is a negative voltage deflection that peaks at 150 to 250 milliseconds (ms) after the onset of an unexpected or 'deviant' stimulus in a series of expected or 'standard' stimuli. It has been shown to reflect attention-independent processing of change detection in auditory stimuli, which makes it appropriate for use as an index of central auditory stimulus representation (Näätänen, 1995; Näätänen, Paavilainen, Rinne & Alho, 2007). The MMN also indexes language-specific representation of speech sounds (Näätänen, Lehtokoski, Lennes, Cheour, Huotilainen, Iivonen Vainio, Alku, Ilmoniemi, Luuk,

Allik, Sinkkonen & Alho, 1997), and it has been shown that native speech sounds elicit larger MMN amplitudes and shorter latencies than non-native speech sounds (Näätänen & Alho, 1997; Ylinen, Shestakova, Huotilainen, Alku & Näätänen, 2006; Kirmse, Ylinen, Tervaniemi, Vainio, Schröger & Jacobsen, 2008; Lipski & Mathiak, 2008; Hisagi et al., 2010). Furthermore, the MMN is also elicited by changes in temporal aspects of auditory stimulation, such as sound duration (Deouell, Karns, Harrison & Knight, 2003; Grimm, Snik & Van Der Broek, 2004; Ylinen et al., 2006). Speakers of languages that use durational differences to contrast meaning show enhanced MMN responses to changes in vowel duration, relative to speakers of languages that do not use durational cues (Minagawa-Kawai, Mori, Sato & Koizumi, 2004; Ylinen, Huotilainen & Näätänen, 2005; Ylinen et al., 2006; Tervaniemi, Jacobsen, Rottger, Kujala, Widmann & Vainio, 2006; Kirmse et al., 2008; Hisagi et al., 2010; Nenonen, Shestakova, Huotilainen & Näätänen, 2005). Moreover, amplitude and latency of the MMN change as a result of perceptual training (Ylinen, Uther, Latvala, Vepsäläinen, Iverson, Akahane-Yamada & Näätänen, 2010), suggesting that second language learners can change their weightings of specific cues for non-native vowel discrimination. The MMN response is usually preceded by an earlier (around 100ms after stimulus onset) negative component, and a positive component (200ms after stimulus onset) elicited by both the standard and the deviant sounds, namely the N1 and P2 respectively (Tremblay, Kraus, McGee, Ponton & Otis, 2001). These responses have been associated with the detection of differences in physical properties of the stimuli, and they often overlap with the MMN (Campbell, Winkler & Kujala, 2007).

MMN responses to speech sounds have been observed in both left and right hemispheres of the brain. For example, Csépe (1995) reported that vowels elicited larger MMN responses from right hemisphere generators, and MMN responses to plosive consonants showed larger amplitudes from the left hemisphere. It has also been reported that MMN responses to the syllable /da/ showed similar amplitude over both hemispheres when the syllable signaled a pitch change, but the response amplitude was larger from the left hemisphere when a phonetic change was signaled (Sharma, Kraus, McGee & Nicol, 1997). In addition, symmetric MMN responses to non-native speech sounds can become left-lateralized after perceptual training (Tremblay, Kraus, Carell & McGee, 1997). According to Näätänen (2001), both hemispheres contribute to analysis of acoustic characteristics of speech stimuli, while speech-specific analysis occurs mainly in the left hemisphere.

While the MMN can be elicited without attention, it has been established that the P300 component indexes attention allocation and cognitive effort when listeners

focus on detecting basic and high-order perceptual changes in speech (e.g., Polich, 2007). The P300 is a positive voltage deflection that peaks around 300–800ms after stimulus onset when elicited by auditory stimuli in adults (Polich & Kok, 1995; Toscano, McMurray, Dennhardt & Luck, 2010). This component is considered an index of updating working memory representations, attention allocation and cognitive effort (Näätänen, 1990; Donchin & Coles, 1988) since it can be elicited during conscious auditory discrimination tasks (Polich, 2007). A subcomponent of the P300, the P3a, indexes attention orientation to a distractor stimulus that is not task-relevant (Spencer & Polich, 1999), and its peak amplitude is observed over central scalp locations that are associated with anterior cingulate cortical generators (Dien, Spencer & Donchin, 2003). In contrast, the P3b subcomponent has a maximum peak observed at parietal scalp electrodes, and has been associated with context-updating processes and memory storage related to temporo-parietal association cortex (Polich, 2007). Hisagi et al. (2010) studied how selective attention to a non-native vowel duration contrast could yield improvement in discrimination as indexed by ERP responses to Japanese vowels in AE listeners. They found that, during a visual-attend condition where listeners were attending to shapes on a screen and not to auditory input, AE listeners had attenuated MMN responses to the Japanese vowel durational contrast compared to native speakers of Japanese. Conversely, in the auditory-attend condition, which required listeners to count deviant stimuli, AE listeners showed enhanced MMN responses. P300 responses to the durational contrast were similar between groups, again suggesting that attention modulates perception of non-native speech sound contrasts.

Against this background of behavioral and neurophysiological experimentation in cross-linguistic speech perception, two questions motivate this study: 1. Do adult sequential Spanish–English bilingual listeners show discrimination and/or identification of the AE vowel contrast /a/-/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); and 2. In the case that adult Spanish–English bilingual listeners perceive the AE vowel contrast /a/-/ʌ/, do they rely more on durational differences to discriminate the vowels, or do they use spectral cues?

The study consisted of two perceptual listening tasks, carried out under two different testing conditions, by both Spanish–English bilingual and AE monolingual listeners. In the first condition, *NATURAL VOWEL DURATION*, listeners were presented with AE /a/-/ʌ/ tokens spoken by a native AE speaker, with all spectral and durational cues intact. In the case that bilingual listeners had non-native perception of the vowel contrasts (heavily influenced by

their native language categories), we expected them to be less accurate and slower to identify the vowels compared to the AE monolingual group, and that these behavioral differences would be accompanied by attenuated MMN (pre-attentional discrimination of the vowel sounds) and P300 (attentional identification) responses compared to monolingual listeners. However, if bilingual listeners had learned to perceive the subtle differences between the AE vowel pair /a/-/ʌ/, through their experience with the language, their behavioral and neurophysiological measures were expected to resemble those of monolingual listeners.

In the second condition, *NEUTRAL VOWEL DURATION*, the duration of each vowel was neutralized – that is, duration for each member of the vowel pair was manipulated so that they did not differ from one another. By neutralizing vowel duration, listeners could only use spectral cues to obtain information about the identity of the vowels. It was expected that, under these listening conditions, the bilingual group would exhibit lower accuracy and slower reaction times when consciously identifying both vowels, and attenuated MMN and P300 responses compared to monolingual AE listeners. Such responses would indicate that bilingual listeners ignored informative spectral differences between vowels in favor of durational information, which is absent. Alternatively, if spectral information is sufficient for the bilingual group to discriminate and/or identify the AE vowels, the behavioral and neurophysiological responses would be similar to those elicited from the monolingual group. Monolingual AE listeners were expected to rely primarily on spectral cues to identify the vowels, and to be unaffected by the lack of durational differences, since durational cues are secondary for them.

## 2. Methods

### 2.1. Participants

Eleven adult sequential bilingual Spanish–English listeners (bilingual group) and 14 monolingual AE listeners (monolingual group) consented and received compensation to participate in the study. All participants were right-handed and passed a hearing screening at 20 dBHL (1000, 2000, and 4000 Hz bilaterally). No participants reported any history of neurological, hearing or language-related disorders. All procedures were carried out under IRB approval in the Neurocognition of Language Laboratory at Teachers College, Columbia University.

Bilingual participants were from various countries in Latin America (6 female, mean age 28.01 years, SD 3.98), learned English in their home countries through formal English courses between the ages of 3 and 20 (mean age 9.4 years, SD 5.31), their formal instruction in English ranged from 2 to 15 years (mean 9.1 years,

SD 4.39), and they came to live in the United States (New York area) between 19 and 34 years of age (mean arrival age 26.8 years, SD 4.66). The length of residence of the Spanish-speaking participants in the USA before testing ranged from 6 months to 5 years (mean 1.6 years, SD 1.42). Bilingual participants reported using English from 25% to 100% percent of the day (mean 66.4%, SD 22.59). This bilingual group is a typical representation of adult immigrants who have learned their L2 from a combination of classroom instruction with non-native teachers, followed by naturalistic exposure when immigrating to the host country after puberty. It may be that initial classroom instruction in the home countries is not sufficient for these listeners to acquire L2 vowel categories resembling those of monolingual listeners (Peltola, Kujala, Tuomainen, Ek, Aaltonen & Näätänen, 2003). However, exposure to native L2 speech sounds after immigration to the U.S. may have positively influenced their perception of the non-native vowel sounds (Winkler, Kujala, Tiitinen, Sivonen, Alku, Lehtokoski, Czigler, Csépe, Ilmoniemi & Näätänen, 1999). Although there is a wide range in the age of initial exposure to non-native sounds (19–35 years of age), no participants were exposed prior to adulthood. Best and Tyler (2007) suggested that, while 6–12 months of experience with the non-native sounds is enough to show significant L2 perceptual learning in second language learners, there is little perceptual benefit from additional experience.

The 14 adult monolingual English-listeners (8 female, mean age 24.01 years, SD 3.48) were from various regions in the United States. All reported no foreign language instruction beyond high school level; all were unable to hold a conversation in any language other than English.

Five participants (1 bilingual, 4 monolingual) were excluded from ERP analysis due to large numbers of artifacts (more than 20% noisy channels) in the EEG data; this left ten participants in each group for analysis.

## 2.2. Stimuli

The experimental stimuli consisted of four naturally produced monosyllables containing the target vowels produced by a female AE talker in citation form (/bab/-/bAb/). Recently, an examination of the effect of consonant context on vowel sensorimotor adaptation revealed that contexts with mainly inter-articulator co-articulation and dynamic articulatory patterns, such as bilabials and stops, facilitate vowel sensorimotor adaptation, compared to contexts with larger intra-articulator co-articulation and static articulatory postures such as alveolars and fricatives (Berry, Jaeger, Wiedenhoeft, Bernal & Johnson, 2014). Hence, the bilabial context /bVb/ was chosen, to minimize effects of consonant-to-vowel and vowel-to-consonant tongue co-articulation (Strange, Weber, Levy, Shafiro, Hisagi & Nishi, 2007).

### 2.2.1. Stimuli recording

Stimuli were recorded in a sound-attenuated chamber using Sound Forge System 8.0. A Shure SM58 Dynamic Microphone was placed 8 centimeters away from the talker's mouth. A female talker was instructed to read a list of ten repetitions of the syllables, written in IPA, at a normal speaking rate, enunciating each word clearly without exaggeration.

All audio files were amplitude-normalized in Praat 5.3 (Boersma & Wenink, 2013) using the 'Scale to Peak' function. The duration of target vowels in the stimulus syllables /bAb/-/bab/ was modified in Praat, to provide stimuli with natural and neutralized vowel durations for both testing conditions. Vowel duration was measured from the first positive peak in the periodic portion of each waveform to the constriction of the post-vocalic consonant. For the natural-duration vowel condition, each vowel kept its own natural duration. For the neutral-duration vowel condition, duration of both vowels was manipulated to a neutral value derived from the combined mean durations for each vowel in the pair. This mean relative duration was imposed on the manipulated files to obtain the neutral vowel stimuli (see Table 1). Figure 1 shows the spectrograms of the stimuli in both natural and neutral duration conditions.

In order to ensure that recorded vowel sounds were heard as the intended vowels by native AE listeners before being implemented as experimental stimuli, five native AE listeners (from the New York area) were asked to listen to 10 repetitions of each syllable. The instructions were identical to those later provided to the experimental participants: "Listen to the sounds coming through the speakers; they are made up words 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." Correct identification response percentages for vowel /ɑ/ were 97% and 100%, and for vowel /ʌ/ were 96% and 97%, in the natural and neutral conditions respectively. The high vowel identification accuracy obtained from five native AE listeners suggested that the recorded experimental stimuli contained the intended AE vowels.

## 2.3. Experimental Tasks

Participants were seated in a sound-attenuated chamber in front of a computer monitor that displayed a movie without sound. They were instructed to watch the movie and ignore the sounds coming through the speaker. Auditory stimuli were presented at 65 dB SPL through an external RME Hammerfall DSP audio card connected to a Tannoy OCV 6 full bandwidth pendant speaker suspended 27 cm directly above the participant. Timing offset of auditory stimuli was verified using a Cedrus StimTracker. The experimental tasks were divided into

Table 1. *Vowel duration and formant frequencies for AE vowels /a/-/ʌ/ as produced by a female AE talker in the CVC contexts /bab/ - /bʌb/. Vowel duration means are calculated for the neutral vowel duration vowels.*

Vowel	NATURAL DURATION			NEUTRAL DURATION		
	F1 Hz	F2 Hz	Duration ms	F1 Hz	F2 Hz	Mean Duration (SD)
/a/	903.964	1319.261	312	905.976	1321.478	273 ms (.050)
/ʌ/	880.027	1545.569	240	882.960	1549.062	273 ms (.050)

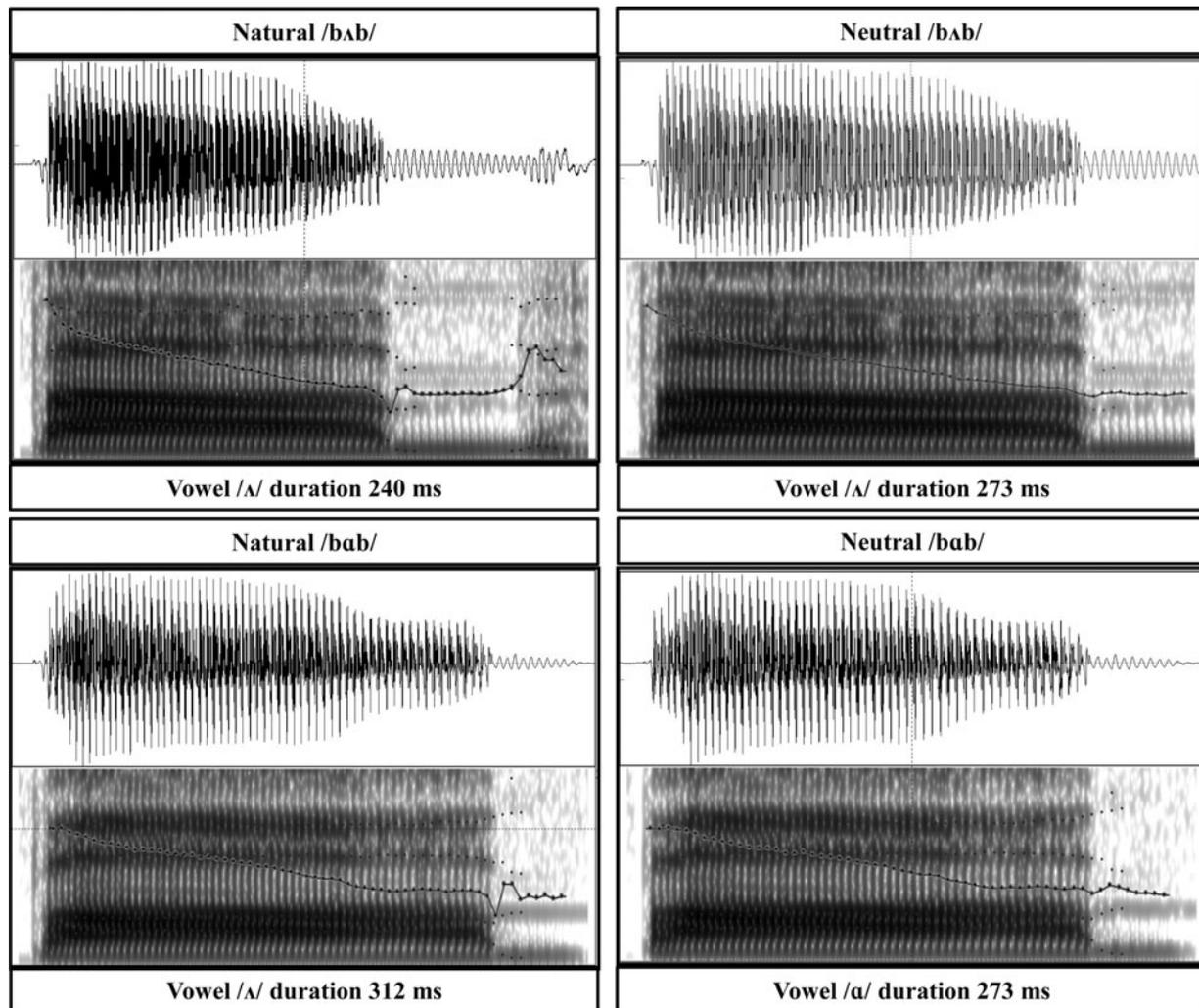


Figure 1. Paired time waveforms and spectra representing the four experimental syllables bʌb and bab in the natural (left) and neutral (right) vowel duration conditions.

two different sessions. In each session, listeners performed a passive non-attended and an attended task. The instructions for each of the tasks are presented below. Breaks were built into the experimental procedures to ensure participant comfort (every 300 trials in the pre-attentive task, and every 75 trials in the attentional task).

**2.3.1. Electrophysiological recordings**

**a. Non-attended task**

In a passive task requiring no behavioral response from the listeners, all participants were asked to ignore auditorily presented AE vowel stimuli (the syllables /bab/-/bʌb/). Participants watched a silent movie while EEG was recorded. The instructions to participants were: “Please

watch the silent movie and ignore the sounds coming through the speaker". Stimuli were presented in an oddball paradigm with an inter-stimulus interval (ISI) of 800ms. According to Werker and Logan (1985), an ISI of 800ms will tap into phonetic/phonemic levels of phoneme representation, not only acoustic cues. Standard stimuli were presented in 85% of the trials (256 out of 300) during each block, and the roles of standards and deviants (44 deviant trials per block, for a total of 88 deviants per condition) were reversed in the second and fourth blocks (e.g., if vowel /a/ was standard in the first block, it was deviant in the second).

### **b. Attended task**

In the attended task, while EEG was recorded, participants decided via button press, trial-by-trial, which vowel (/Λ/ or /a/) was being presented in the oddball paradigm. The task included four blocks of 75 trials each. The presentation of the next stimulus was initiated when the participant responded. The instructions for this task were: "Listen to the sounds coming through the speakers; they are made up words 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." As in the Non-attended task, standard stimuli were 85% of the trials during each block, and the roles of standards and deviants were reversed in the second and fourth blocks. All instructions appeared in text form on the computer screen prior to the experimental sessions, and stayed onscreen to minimize additional working memory load.

### **2.3.2. Second session of electrophysiological recordings**

The second session (conducted on a different day from the first) was identical to the first session, except for the experimental condition. Participants always completed the Non-attended task (MMN) before the Attended task to avoid possible learning effects. However, the order of presentation for the natural/neutral vowel duration conditions was counterbalanced across participants; those who heard natural duration stimuli on the first day of recording heard neutral duration stimuli on the second day, and vice versa.

All behavioral responses obtained during the experimental tasks were recorded through E-Prime 2.0 (Psychology Software Tools, Pittsburg, PA).

### **2.4. EEG Recording**

During EEG data acquisition, scalp voltages were recorded with a high-density 128-channel hydrocel net connected to a high-input impedance amplifier (NetAmps300, 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 below 40 kΩ (Ferree, Luu, Russell & Tucker, 2001), and all electrodes were referenced to the vertex (Cz) during recording. Electrodes above and below the eyes, and at the outer canthi, allowed for identification of electro-oculographic artifacts (EOG, associated with eye blinks and eye movements).

### **2.5 Data analysis**

Behavioral responses in the attended task were analyzed for each vowel /a/-/Λ/ in the two duration conditions (natural and neutral). Accuracy was measured by counting correct responses out of all trials presented, and error trials were omitted from analysis. Ninety percent (90%) of trials per participant were included in statistical analysis. Percent correct responses were transformed to Rationalized Arcsine Units (RAU – Studebaker, 1985) to approximate a normalized distribution. Reaction time (RT) was recorded as the time elapsed from onset of stimulus presentation to execution of a button-press response. RT values were log-transformed to approximate a normal 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 a three-factor mixed-designed ANOVA, to determine the significance of changes in behavioral indices of vowel identification between condition 1 (natural vowel duration) to condition 2 (neutral vowel duration) between groups (bilingual vs. monolingual) and between vowels (/a/-/Λ/). In addition, planned comparisons (independent and paired-samples t-tests) were conducted to determine statistically significant differences within and between groups.

#### **2.5.1. EEG data pre-processing and analysis**

Recorded raw EEG data were digitally filtered offline using a 0.1-30 Hz bandpass filter, and subjected to automatic and manual artifact rejection protocols for removal of movement and physiological artifacts. Noisy channels were interpolated using spherical spline modeling, based on recorded data from surrounding sensors. Data were re-referenced to the average to eliminate the influence of an arbitrary recording reference channel.

Recorded data were segmented into epochs of 800ms: 700ms following the onset of each stimulus, and a 100ms pre-stimulus baseline period, to minimize the effects of long latency artifacts (such as amplifier drift). Trials were discarded from analysis if they contained eye movements (variance greater than  $>70 \mu\text{V}$  in one epoch), or if more than 20% of the channels were noisy (average amplitude over  $100 \mu\text{V}$  in one epoch). Only recordings with more than 75% of uncontaminated trials were included in analysis. The average numbers of accepted trials across

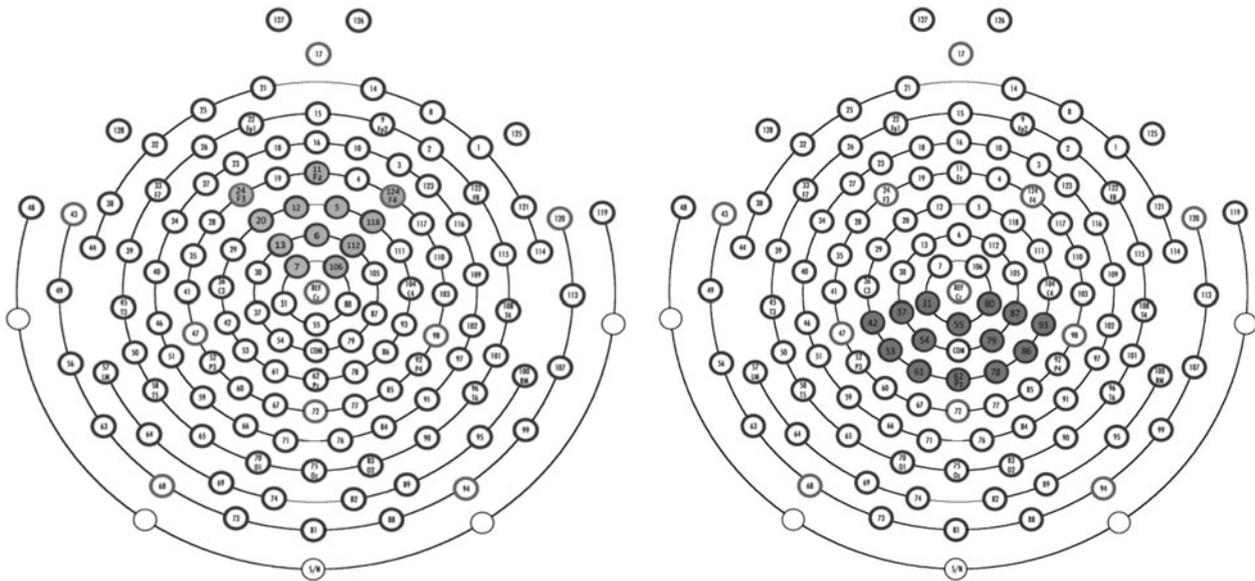


Figure 2. The figure on the left side represents the electrodes analyzed for the MMN response. The left-most line of electrodes correspond to the left-hemisphere montage (electrodes 7, 13, 20 and 24), electrodes in the center correspond to the central montage (electrodes 11, 6, 12, 5), and the right most line of electrodes correspond to the right-hemisphere montage (electrodes 106, 112, 118 and 124). The figure on the right side represents the electrodes analyzed for the P300 response. The electrodes correspond to the central-parietal area (31, 55, 80, 37, 54, 79, 87, 42, 53, 61, 62, 78, 86, and 93).

groups were 482 standards ( $SD = 41.47$ ) and 87 deviants ( $SD = 0.95$ ) in the Natural condition and 504 standards ( $SD = 6.50$ ) and 87 deviants ( $SD = 2.07$ ) in the Neutral condition. Response-locked ERPs were computed within epochs, starting at stimulus onset. Individuals' averaged data were grand-averaged within groups to enhance statistical power and reduce variance due to random noise.

### 2.5.2. Extraction of the components

EEG data were analyzed according to a pre-determined region of interest for each ERP component: for MMN this was the frontal-central region, and for P300 the central-parietal region. To investigate possible hemispheric differences in MMN responses to speech stimuli for each group, the amplitudes of MMN responses were analyzed in three different electrode montages: Left (electrodes 7, 13, 20 and 24); Central (electrodes 11, 6, 12, 5); and Right (electrodes 106, 112, 118 and 124). The P300 montage corresponding to the central-parietal area included electrodes 31, 55, 80, 37, 54, 79, 87, 42, 53, 61, 62, 78, 86, and 93 (see figure 2). Responses over sensor montages were examined during specific time windows post-stimulus onset (100–300 ms, when MMN is expected, and 250–500 ms, when P300 is typically observed).

MMN is usually presented as a difference wave, obtained by subtracting the ERP to standard stimuli from the ERP response to deviants. Since the subtraction involves brain responses to two physically different speech sounds, it is possible that physical differences in the stimuli might elicit different early components

associated with physical features of the sounds, and this could influence the MMN. To ensure that the potential contributions of other components were removed, the identity MMN technique (Pulvermüller & Shtyrov, 2006) was implemented. This requires subtracting the ERP elicited in response to standard stimuli from the ERP elicited to the same stimuli presented as deviants. This is possible in this experiment, since both vowel sounds were presented in standard and deviant status in separate blocks in the experiment. Therefore, the average responses to standard stimuli for each participant in each condition were subtracted from responses to physically identical deviants (e.g., deviant /a/ - standard /a/). Difference waves were computed for each participant, and individual difference waves were grand-averaged by condition (natural and neutral duration), vowel (/a/-/Λ), and group (bilingual and monolingual). The grand averaged negative (MMN) peaks for the two vowels in each condition were derived by examining grand averaged waveforms for the right, central and left montages during the time window 100–300ms post stimulus onset. MMN mean amplitude was calculated as the mean voltage during a 60ms interval centered around the most negative peak observed in each condition and each hemisphere (see Table 4 for the MMN peak latencies). Similarly, peak latency of the P300 component was defined as the largest positive peak in the 250–500ms time window, identified by examining the grand averaged difference waveforms for the central-parietal montage. P300 mean amplitude was calculated as the mean voltage during a

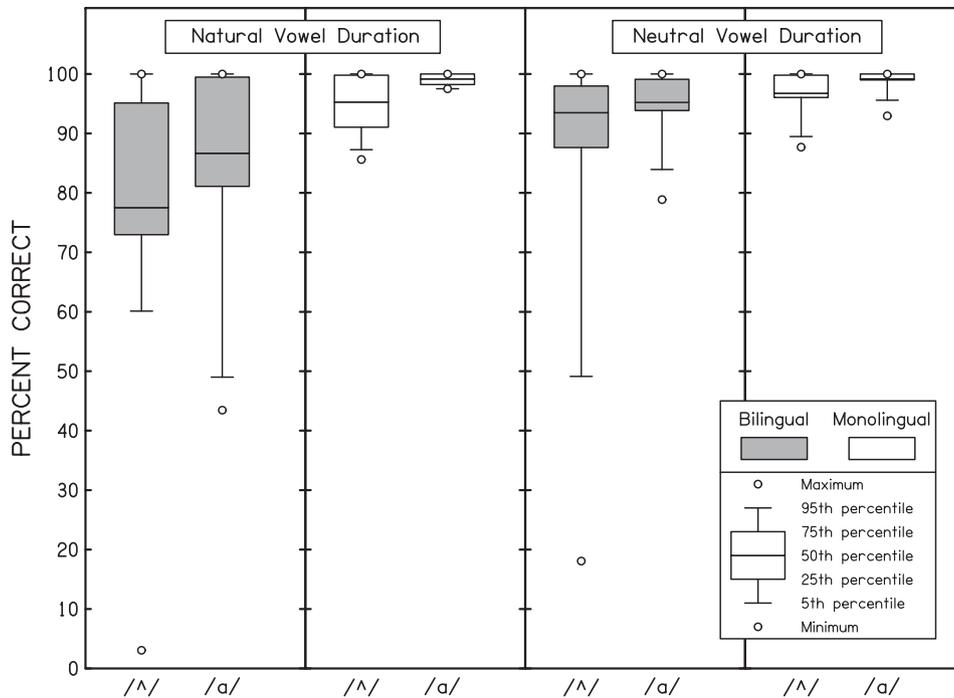


Figure 3. Vowel identification accuracy (percent correct). Bilingual (N =10) and Monolingual (N=10) groups during the natural and neutral vowel duration conditions.

60ms interval period centered on the most positive peak latency. Individual grand-averaged difference waves were subjected to independent-samples t-tests to determine whether they were significantly different from zero. Individual mean amplitude and latency measures were submitted to repeated measures analyses of variance (ANOVA) with factors group (bilingual vs. monolingual), montage (Left vs. Central vs. Right), condition (natural vs. neutral), and vowel (/a/-/ʌ/). The Greenhouse-Geisser correction was reported to adjust the associated degrees-of-freedom when the sphericity assumption was violated. All analyses were conducted with an alpha level of 0.05.

### 3. Results

#### 3.1. Behavioral Results

##### 3.1.1. Accuracy

Figure 3 shows the median, interquartile ranges, 95<sup>th</sup> and 5<sup>th</sup> percentiles and outliers for accuracy (percent correct) in both groups. The bilingual group identified /ʌ/ with 77.53% and 87.63% median accuracy in the natural and neutral vowel duration conditions, respectively. Median identification accuracy scores for /a/ were 86.67% and 95.22% in the natural and neutral vowel duration conditions, respectively. In contrast, the monolingual group identified /ʌ/ with 95.28% and with 96.77%

median accuracy in the natural and neutral vowel duration conditions, and /a/ with 99.18% median accuracy in both natural and neutral vowel duration conditions.

The distribution of these accuracy scores was examined for skewness and kurtosis. Accuracy scores included outliers on the lower end of the scale corresponding to vowel /ʌ/ in the natural (3.0%) and neutral (18%) vowel duration conditions. Based on standardized values for skewness (-2.917) and kurtosis (9.449) the distribution of scores was negatively skewed and peaked. Accuracy was re-examined after cases identified as outliers were removed. No additional outliers were identified and the distribution appeared to be approximately normal (skewness = -1.456 and kurtosis = 1.784).

To determine whether outlier scores influenced accuracy differences between the bilingual and monolingual groups, the ANOVA was conducted twice, including and excluding outliers. Both analyses (outliers included:  $F(1, 18) = 7.880, p = .012, \eta_p^2 = .304$ ; outliers excluded:  $F(1, 17) = 7.005, p = .017, \eta_p^2 = .292$ ) revealed a significant difference in accuracy between groups.

Independent samples t-tests were conducted to explore vowel identification accuracy differences between groups. Results indicated that the bilingual group obtained significantly lower accuracy compared to the monolingual group for only the vowel /a/, in both vowel duration conditions (natural and neutral) (see Table 2).

Table 2. Vowel identification accuracy for bilingual and monolingual groups. RAU (Rationalized Arcsine Units) were used in the statistical analysis. Independent samples *t*-tests.

Vowel	Group	Accuracy			
		Mean Raw percentage (SD)	Mean RAU(SD)	<i>t</i> (18)	<i>p</i>
Natural /Λ/	Bilingual	77.529 (30.159)	83.131 (38.553)	-1.854	.080
	Monolingual	95.279 (5.263)	107.132 (13.766)		
Natural /α/	Bilingual	86.667 (20.226)	97.429 (28.225)	-2.167	.044*
	Monolingual	99.176 (1.088)	117.296 (6.596)		
Neutral /Λ/	Bilingual	87.626 (24.748)	97.045 (29.892)	-1.285	.215
	Monolingual	96.772 (4.100)	110.066 (11.516)		
Neutral /α/	Bilingual	95.218 (6.292)	106.481 (13.287)	-2.636	.017*
	Monolingual	99.177 (2.216)	119.488 (8.184)		

\**p* ≤ .05

3.1.2. Reaction time

Figure 4 illustrates reaction time scores for both groups in natural and neutral vowel duration conditions.

The reaction time (RT) scores (Log-transformed) were submitted to a mixed-design 3-factor ANOVA comparing language group (Bilingual vs. Monolingual), condition (natural vs. neutral), and vowel (/α/ vs. /Λ/). Results showed that RT differences between groups were not significant ( $F(1, 18) = 3.283, p = 0.087, \eta_p^2 = .154$ ). However, a significant main effect of Condition ( $F(1, 18) = 10.323, p = .005, \eta_p^2 = .364$ ) showed that both groups were slower to identify vowels in the natural condition compared to the neutral condition. These unexpected results indicate that durational cues in the vowels were not indispensable for any group to make decisions about vowel identity. There were no significant interactions. Planned comparisons revealed no significant group differences in RT for identification of vowels (/α-/Λ/) in any conditions (see Table 3).

In summary, compared to the monolingual group, Spanish-English bilingual listeners were significantly less accurate at identifying vowel /α/ in both experimental conditions (natural vs. neutral vowel duration). The groups did not differ in accuracy for identifying AE vowel /Λ/. This finding could be due to the larger variability in the bilingual group (SD = 38). Similarly, although the

bilingual group was slower than the monolingual group to identify the AE vowels, the difference was not statistically significant.

3.2 Neurophysiological results

3.2.1 Non-attended task – MMN component

A typical characteristic of MMN topography is a frontal-central maximum negativity while a positive deflection (polarity inversion) is observed at mastoid electrodes (Näätänen, 1990; Sussman, Wrinkler, Kreuzer, Saher, Näätänen & Ritter, 2002). An analysis comparing amplitudes of the neurophysiological response at mastoid electrodes and the frontal (Fz) electrode in each group revealed no amplitude difference for the bilingual group ( $F(1, 9.000) = 1.007, p = .342, \eta_p^2 = .101$ ). However, for the monolingual group, this difference was significant ( $F(1, 9.000) = 6.502, p = 0.031, \eta_p^2 = .101$ ). This amplitude difference, and the presence of the expected polarity inversion, suggests that MMN responses were elicited in the monolingual, but not in the bilingual group.

One-sample *t*-tests indicated that for the bilingual group, difference waves were different from zero only over right montage electrodes in the neutral condition for both vowels /Λ/ and /α/ ( $p = .049$ ), and over central electrodes for the neutral vowel /α/ ( $p = .019$ ). For the monolingual

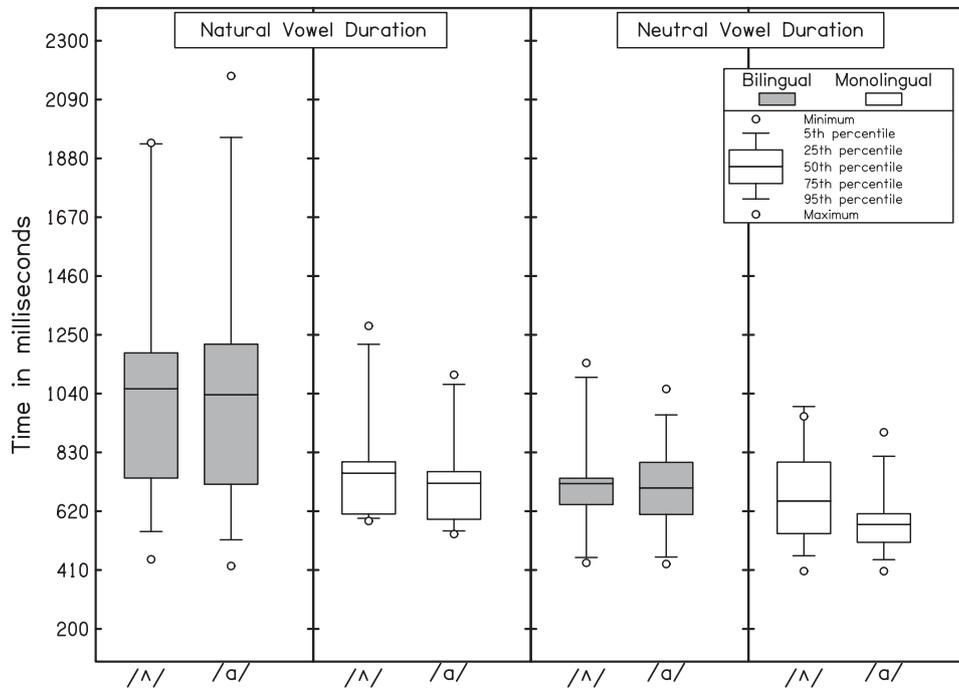


Figure 4. Reaction time for vowel identification (ms). Bilingual (N = 10) and Monolingual (N = 10) groups during the natural and neutral vowel duration conditions.

Table 3. Vowel identification RT differences between monolingual and bilingual groups. Independent samples *t*-tests.

Vowel	Group	Reaction Time			
		Mean Raw milliseconds (SD)	Mean LogTran (SD)	<i>t</i> (18)	<i>p</i>
Natural/ʌ/	Bilingual	1057.437 (507.193)	2.974 (.192)	1.918	.710
	Monolingual	755.976 (231.556)	2.811 (.188)		
Natural/a/	Bilingual	1036.182 (535.753)	2.955 (.203)	1.686	.109
	Monolingual	719.870 (192.049)	2.835 (.097)		
Neutral/ʌ/	Bilingual	718.916 (217.939)	2.837 (.128)	.703	.491
	Monolingual	656.737 (173.425)	2.800 (.111)		
Neutral/a/	Bilingual	1036.182 (535.753)	2.828 (.116)	1.754	.096
	Monolingual	572.698 (131.814)	2.747 (.090)		

\* *p* < .05

Table 4. Non-attended task. MMN difference wave mean peak latencies (milliseconds) and amplitude ( $\mu V$ ) for Bilingual (Spanish–English) and Monolingual (English) groups in the natural and neutral vowel duration conditions (independent samples *t*-tests for amplitude differences are shown).

	Group	Peak latencies(ms)	Mean( $\mu V$ )	SD	t(9)	Sig(2-tailed)
<b>Left_Natural /<math>\Lambda</math>/</b>	Bilingual	240	.313	.623	1.587	.147
	Monolingual		−.365	.882	−1.311	.222
<b>Left_Natural /<math>\alpha</math>/</b>	Bilingual	152	−.527	.987	−1.689	.126
	Monolingual		−.779	1.081	−2.28	.049*
<b>Left_Neutral /<math>\Lambda</math>/</b>	Bilingual	240	−.301	.65	−1.464	.177
	Monolingual		−.683	1.443	−1.498	.168
<b>Left_Neutral /<math>\alpha</math>/</b>	Bilingual	228	.156	.967	.512	.621
	Monolingual		−1.023	1.669	−1.939	.084
<b>Right_Natural /<math>\Lambda</math>/</b>	Bilingual	240	−.038	.726	−.168	.870
	Monolingual		−.546	1.535	−1.125	.290
<b>Right_Natural /<math>\alpha</math>/</b>	Bilingual	216	−.022	.627	−.115	.911
	Monolingual		−2.122	1.022	−6.562	.001*
<b>Right_Neutral /<math>\Lambda</math>/</b>	Bilingual	268	−.555	.773	−2.274	.049*
	Monolingual		−.817	.821	−3.148	.012*
<b>Right_Neutral /<math>\alpha</math>/</b>	Bilingual	176	−.433	.601	−2.28	.049*
	Monolingual		−1.335	0.991	−4.26	.002*
<b>Central_Natural /<math>\Lambda</math>/</b>	Bilingual	248	.24	0.636	1.192	.264
	Monolingual		−.558	1.739	−1.016	.336
<b>Central_Natural /<math>\alpha</math>/</b>	Bilingual	220	−.117	.711	−.523	.614
	Monolingual		−1.285	1.265	−3.213	.011*
<b>Central_Neutral /<math>\Lambda</math>/</b>	Bilingual	280	−.62	.929	−2.11	.064
	Monolingual		−1.062	1.495	−2.248	.051
<b>Central_Neutral /<math>\alpha</math>/</b>	Bilingual	176	−.691	.746	−2.859	.019*
	Monolingual		−1.274	1.387	−2.904	.017*

\**p* < .05

group, difference waves were significantly different from zero over electrodes in left, central and right montages in the natural condition only to vowel / $\alpha$ / ( $p = .049$ ,  $p < .001$ , and  $p = .011$ , respectively). However, in the neutral condition, difference waves were different from zero for both vowels over central (/ $\alpha$ /:  $p = .017$ ; / $\Lambda$ /:  $p = .051$ ) and right electrodes (/ $\alpha$ /:  $p = .002$ ; / $\Lambda$ /:  $p = .012$ ), but not over left montage electrodes (/ $\alpha$ /:  $p = .168$ ; / $\Lambda$ /:  $p = .084$ ). Table 4 presents means and SD of the difference waves for each group by vowel, condition and montage.

Figures 5, 6, 7 show MMN difference waves for bilingual and monolingual groups in response to the non-attended task, in natural and neutral vowel duration conditions, per electrode montage (left, central, right respectively). ANOVA revealed that the groups differed in the amplitude of MMN responses to the vowels ( $F(1.000, 18.000) = 17.384$ ,  $p = 0.001$ ,  $\eta_p^2 = .491$ ). A significant main effect of montage ( $F(1.000, 18.000) = 4.513$ ,  $p = .025$ ,  $\eta_p^2 = .200$ ), and a significant interaction between

montage, condition, vowel and group ( $F(1.000, 18.000) = 6.769$ ,  $p = .018$ ,  $\eta_p^2 = .273$ ) suggest that mean MMN amplitudes differed between the groups at each montage site, for each vowel and condition.

ANOVA conducted on MMN peak latencies (in ms) revealed a significant main effect of vowel ( $F(1.00, 18.000) = 24.785$ ,  $p < 0.001$ ,  $\eta_p^2 = .579$ ), and a significant interaction between montage, condition and vowel ( $F(1.907, 34.328) = 6.769$ ,  $p = .031$ ,  $\eta_p^2 = .178$ ). Independent samples *t*-tests showed that, compared to the bilingual group, the monolingual listeners' MMN response was faster (shorter latency) only to / $\Lambda$ /, in the natural condition over right hemisphere electrodes ( $p = .035$ ) and in the neutral condition over left hemisphere electrodes ( $p = .013$ ). Table 5 shows mean MMN peak latency in milliseconds for each group by vowel, condition and montage.

In summary, the bilingual group did not show significant MMN responses to any of the vowels in the

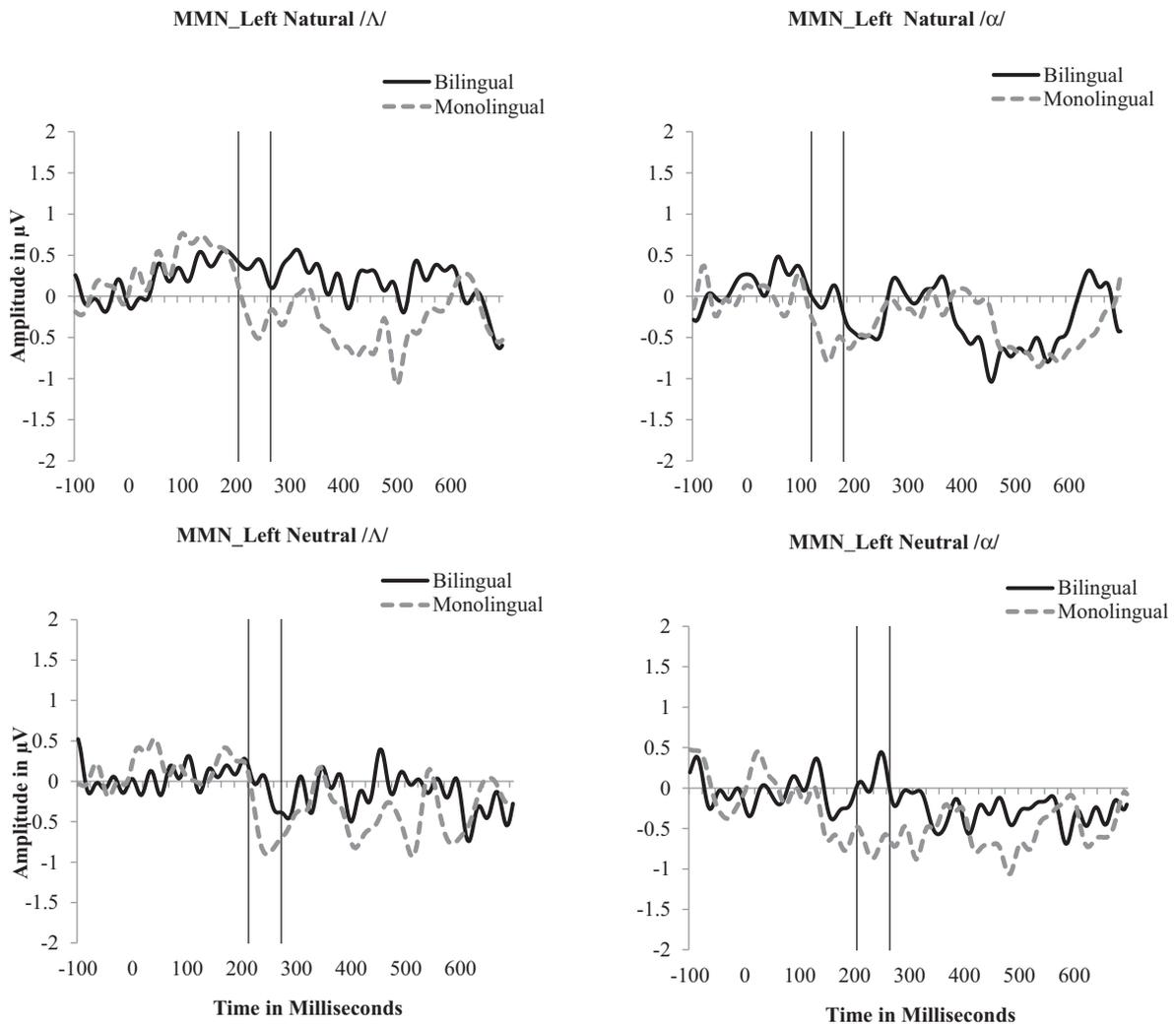


Figure 5. MMN difference waves in response to vowels /ʌ/ (left) and /ɑ/ (right) during the non-attended task in the natural and neutral vowel duration conditions in bilingual and monolingual listeners at the left hemisphere (electrodes 7, 13, 20 and 24). The vertical lines indicate the 60 ms around the most negative peak used for analysis.

deviant status in the natural condition, but showed a right-lateralized MMN response to both vowels in the neutral vowel duration condition, suggesting that when duration was neutralized the bilingual group processed these vowels based on spectral properties. Unexpectedly, the monolingual group showed a significant MMN response to /ɑ/ in the natural duration condition, over right, left and central electrode montages, but not to /ʌ/. Since these are native vowels for the monolingual group, we expected no differences in MMN responses to either vowel. Like the bilingual group, monolingual listeners had significant right-lateralized MMN responses to both vowels in the neutral vowel duration condition. These responses from the monolingual listeners suggest that processing /ɑ/ involved acoustic and phonetic analysis, as evidenced by the presence of the MMN response in all three montages (right, central and left) and both natural

and neutral conditions. Conversely, the responses of the bilingual listeners resembled those of the monolingual group only in the neutral condition, and reflected right-lateralized acoustic processing of the speech stimuli.

### 3.2.2 Attended task (targeting the P300 component)

Independent samples t-tests showed that difference wave amplitudes between 250 and 500ms were significantly different from zero in the bilingual group in response to vowel /ɑ/ in both conditions (natural /ɑ/,  $p = .048$ ; neutral /ɑ/,  $p = .010$ ). In the monolingual group, difference waves in response to both vowels in both conditions were significantly different from zero ( $p < .005$ ). See Table 6 for summary statistics and t-tests.

ANOVAs revealed a significant main effect of vowel ( $F(1.000, 18.000) = 9.851, p < .006, \eta_p^2 = .354$ ), indicating that P300 responses between groups differed according to

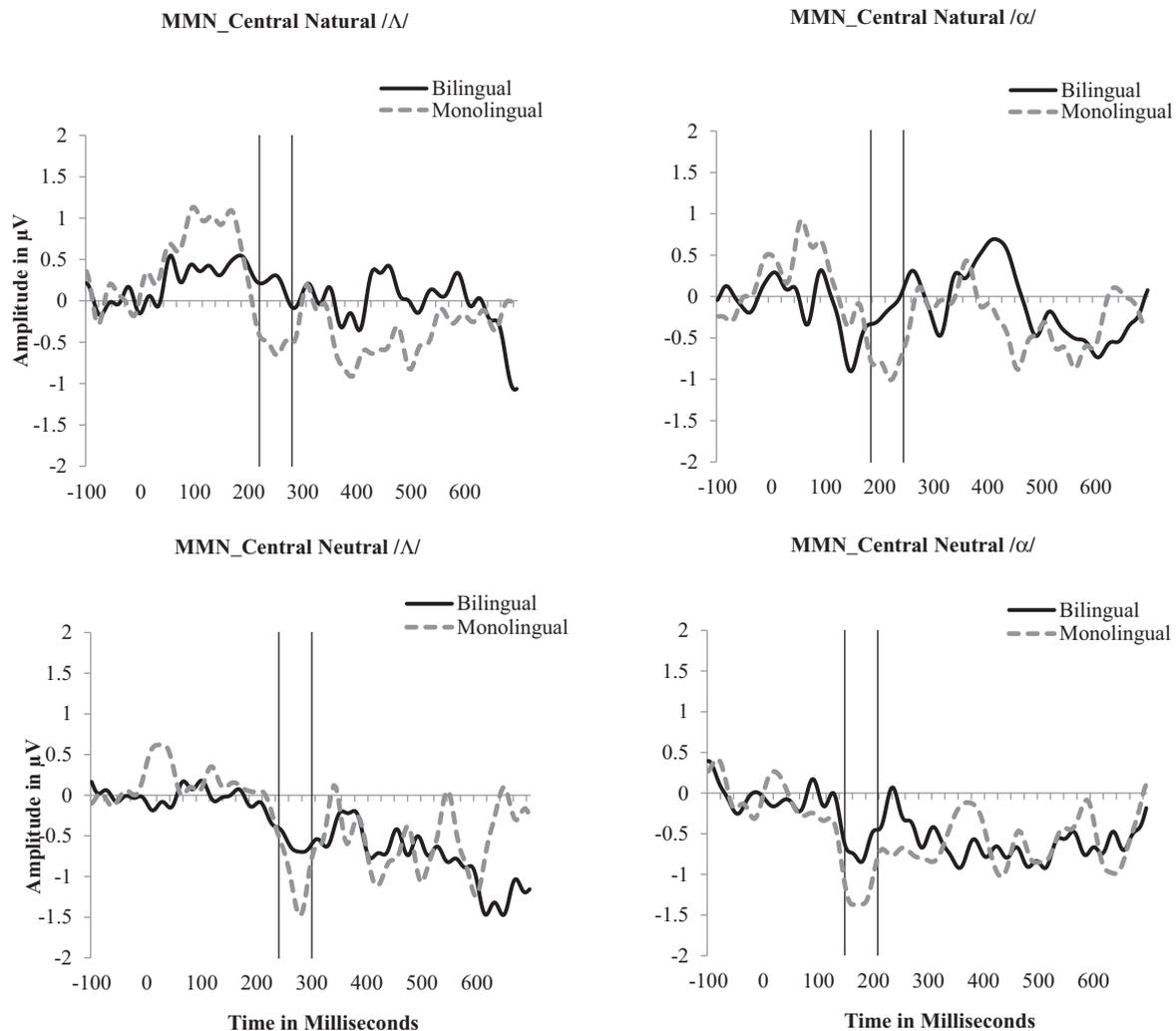


Figure 6. MMN difference waves in response to vowels /ʌ/ (left) and /ɑ/ (right) during the non-attended task in the natural and neutral vowel duration conditions in bilingual and monolingual listeners at the central electrodes (11, 6, 12, 5). The vertical lines indicate the 60 ms around the most negative peak used for analysis.

vowel. No other significant main effects or interactions were observed in the analysis. Figure 8 below shows bilingual and monolingual listeners' P300 responses to the vowels during the attended task in natural and neutral vowel duration conditions.

There were no significant differences in mean peak latency between groups for the P300 component ( $F(1, 18) = 4.443, p = .049, \eta_p^2 = .407$ ). See Table 7.

In summary, regardless of vowel duration, the bilingual group did not show significant MMN responses in the non-attended task. However, during the attended task, they had significant P300 responses to vowel /ɑ/, but not /ʌ/. In contrast, the monolingual group showed significant MMN responses to vowel /ɑ/, but not /ʌ/ in the non-attended condition, while having significant P300 responses to both vowels in the attended condition.

#### 4. Discussion

The purpose of this study was to examine behavioral and 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 carried out passive listening (non-attended) and identification (attended) tasks under two listening conditions: (1) natural vowel duration and (2) neutral vowel duration. Neurophysiological studies have shown that learning to perceive non-native speech sounds changes brain responses known to index aspects of phonological representation (e.g., Näätänen & Alho, 1997; Winkler et al., 1999). Such changes should be reflected by brain activations at very early stages of perceptual processing, when detection of

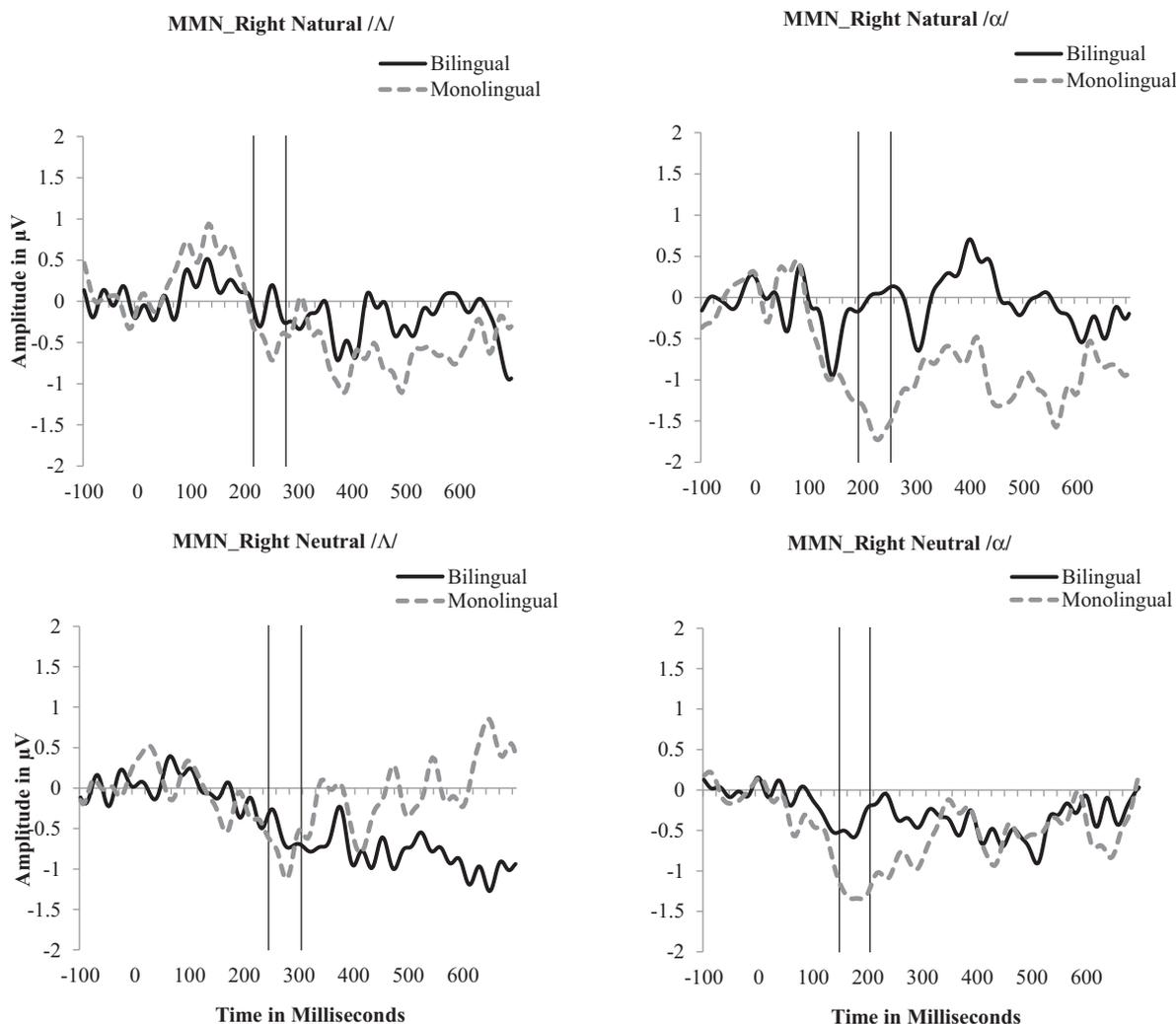


Figure 7. MMN difference waves in response to vowels /ʌ/ (left) and /ɑ/ (right) during the non-attended task in the natural and neutral vowel duration conditions in bilingual and monolingual listeners at the right hemisphere (electrodes 106, 112, 118 and 124). The vertical lines indicate the 60 ms around the most negative peak used for analysis.

language-specific parameters is automatic (Kirmse et al., 2008, Rivera-Gaxiola, Csibra, Johnson & Karmiloff-Smith, 2000; Ylinen et al., 2010; Tamminen, Peltola, Toivonen, Kujala & Näätänen, 2013) as well as later in the processing stream, when attention is involved in discrimination/identification processes (Hisagi et al., 2010; Lipski, Escudero & Benders, 2012). The present study capitalized on the availability of ERPs indexing each of these stages of processing, and aimed to identify specifically the differences associated with bilingual adults’ discrimination of a vowel contrast specific to AE: /ʌ/ versus /ɑ/.

Behaviorally, it was predicted that the bilingual group would be significantly less accurate and slower than the monolingual listeners to identify the AE vowels /ɑ/-/ʌ/, because these vowels are not part of their native vowel inventory. Neurophysiologically, it was predicted

that if the bilingual group perceived the vowel contrast in the non-attended (MMN) and attended (P300) tasks, relying on both spectral and durational cues in the natural vowel duration condition, or on spectral cues alone in the neutral vowel duration condition, then the ERP responses would resemble those of monolingual English-speaking listeners. On the other hand, non-significant neurophysiological responses (MMN and P300) to the vowel contrast would have indicated that the Spanish–English bilingual group was not able to pre-attentively discriminate between, or consciously identify, the AE vowels /ɑ/-/ʌ/.

Conversely, since these vowels are part of the native vowel inventory for monolingual English-speaking listeners, it was predicted that their MMN and P300 responses to the listening and identification tasks would reflect discrimination (non-attended task) and

Table 5. Non-attended task. MMN mean latency (milliseconds) for Bilingual (Spanish–English) and Monolingual (English) groups in the natural and neutral vowel duration conditions (independent samples *t*-tests for latency differences are shown).

	Group	Mean Latency (ms)	SD	t(18)	Sig (2-tailed)
<b>Left_Natural /Λ/</b>	Bilingual	225	29.996	1.713	.104
	Monolingual	201	32.608		
<b>Left_Natural /a/</b>	Bilingual	257	28.413	1.449	.164
	Monolingual	239	27.116		
<b>Left_Neutral /Λ/</b>	Bilingual	243	47.406	2.759	.013*
	Monolingual	193	32.21		
<b>Left_Neutral /a/</b>	Bilingual	258	26.141	-.229	.822
	Monolingual	261	28.538		
<b>Right_Natural /Λ/</b>	Bilingual	252	53.78	2.287	.035*
	Monolingual	208	26.022		
<b>Right_Natural /a/</b>	Bilingual	270	37.393	1.762	.095
	Monolingual	244	26.697		
<b>Right_Neutral /Λ/</b>	Bilingual	244	41.183	0	1
	Monolingual	244	35.377		
<b>Right_Neutral /a/</b>	Bilingual	239	42.742	-.697	.495
	Monolingual	250	27.468		
<b>Central_Natural /Λ/</b>	Bilingual	232	44.838	.093	.927
	Monolingual	230	51.058		
<b>Central_Natural /a/</b>	Bilingual	235	37.797	-.957	.351
	Monolingual	253	44.173		
<b>Central_Neutral /Λ/</b>	Bilingual	222	53.487	.658	.519
	Monolingual	209	32.21		
<b>Central_Neutral /a/</b>	Bilingual	252	34.544	.101	.921
	Monolingual	251	27.263		

\**p* < .05

identification (attended task) of the vowels in both natural and neutral duration conditions. This would reflect a reliance primarily on spectral (rather than durational) cues to discriminate between the two vowels.

Behavioral responses to the AE vowel contrast /a/- /Λ/ obtained during the attentional task revealed significant differences in identification accuracy between groups, indicating, as expected, that the bilingual group was significantly less accurate in identifying the AE vowel /a/ than the monolingual group. On the other hand, there were no significant differences in identification accuracy for /Λ/, possibly due to the large variability in accuracy scores for the bilingual group. Similarly, differences in reaction time between groups did not reach significance. It is possible that the labeling task was too easy, so that the lack of cognitive demands reduced any possible group difference.

Neurophysiological results indicated that, when not attending to the vowel sounds, bilingual listeners did not show indices of discrimination of the AE vowel contrast

/a/- /Λ/ in the natural vowel duration condition. This was indicated by non-significant MMN responses to the vowel contrast over left hemisphere sensors during the non-attended task. However, significant MMN responses to the vowels in the neutral condition over the right hemisphere suggest that the bilingual group did engage in acoustic processing of the duration-neutralized vowels (Näätänen, 2001).

The bilingual group's neurophysiological data revealed indices of attentional identification of the AE vowel /a/, but not /Λ/, indicated by significant P300 responses to /a/ in its deviant status during both natural and neutral vowel duration conditions. This finding suggests that the bilingual group may not have had pre-attentional access to mental representations corresponding to the non-native vowels /a/- /Λ/ between 100 and 300ms after stimulus onset, but they do appear to be able to access relevant acoustic-phonetic information about the vowels later in time, between 250 and 500ms, when they were (actively) attending to the stimuli. In addition, these findings indicate

Table 6. Attended task. P300 difference wave mean peak latency (milliseconds) and amplitude ( $\mu V$ ) for Bilingual (Spanish–English) and Monolingual (English) groups in the natural and neutral vowel duration conditions (independent samples *t*-tests for amplitude differences are shown).

	Group	Peak latencies (ms)	Mean ( $\mu V$ )	Std Dev	t(9)	Sig (2-tailed)
<b>Natural /<math>\Delta</math>/</b>	Bilingual	444	1.401	2.656	1.668	.130
	Monolingual		2.481	2.128	3.686	.005*
<b>Natural /<math>\alpha</math>/</b>	Bilingual	428	2.275	3.307	2.176	.048*
	Monolingual		2.979	2.228	4.17	.003*
<b>Neutral /<math>\Delta</math>/</b>	Bilingual	444	.987	2.702	1.156	.278
	Monolingual		2.852	1.605	5.618	.001*
<b>Neutral /<math>\alpha</math>/</b>	Bilingual	368	3.069	2.983	3.253	.010*
	Monolingual		4.071	2.528	5.092	.001*

\**p* < .05

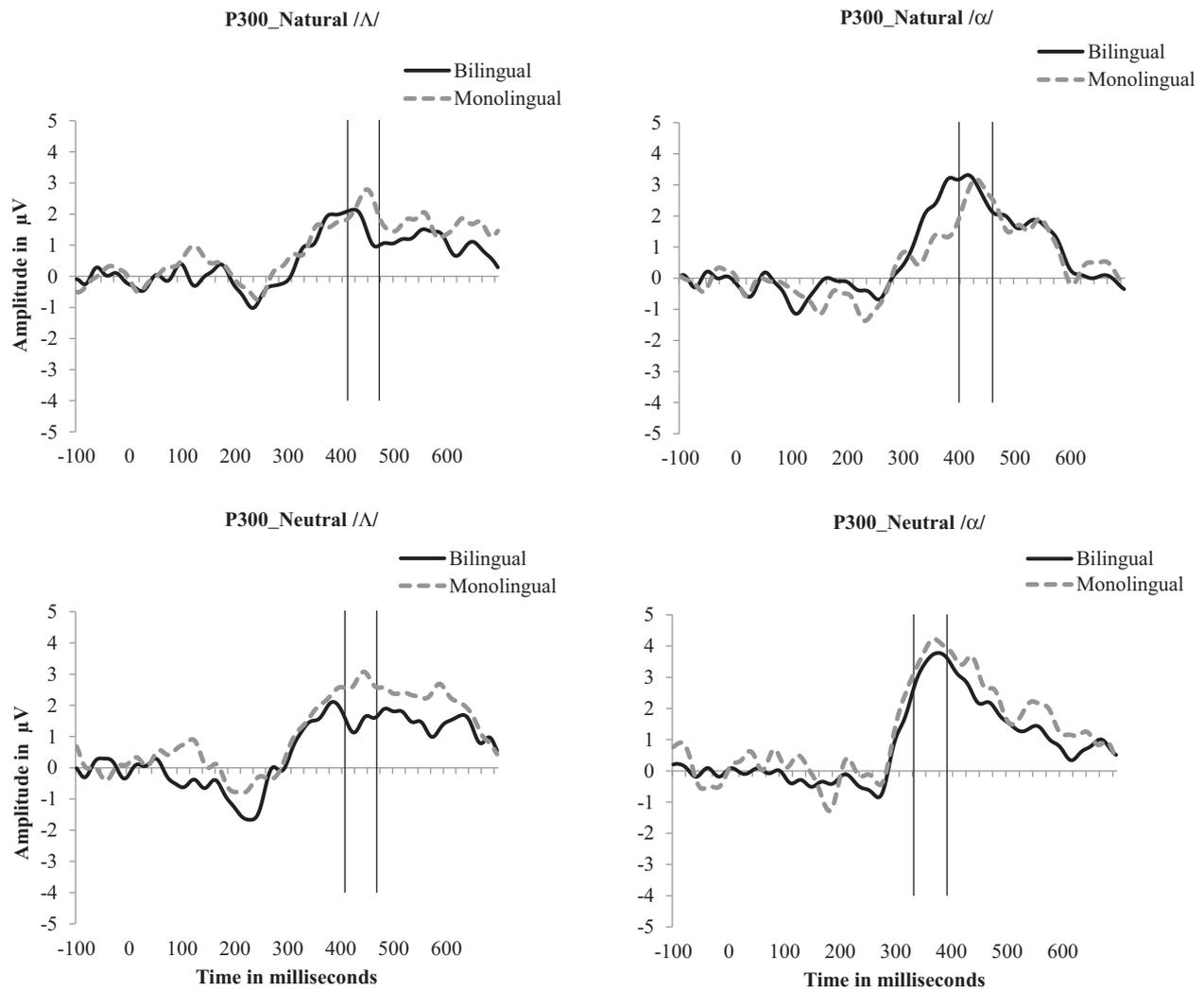


Figure 8. P300 responses to vowels / $\Delta$ / (left) and / $\alpha$ / (right) during the attended task in the natural and neutral vowel duration condition in bilingual and monolingual listeners at central-parietal electrodes (31, 55, 80, 37, 54, 79, 87, 42, 53, 61, 62, 78, 86, and 93). The vertical lines indicate the 60 ms around the most positive peak used for analysis.

Table 7. *Attended task. P300 mean latency (milliseconds) for Bilingual (Spanish–English) and Monolingual (English) groups in the natural and neutral vowel duration conditions (independent samples t-tests for latency differences are shown).*

	Group	Mean latency ms	Std Dev	t(18)	Sig (2-tailed)
<b>Natural /ʌ/</b>	Bilingual	388	29.300	.446	.661
	Monolingual	381	43.586		
<b>Natural /ɑ/</b>	Bilingual	436	42.957	1.999	.061
	Monolingual	402	31.281		
<b>Neutral /ʌ/</b>	Bilingual	380	47.608	−.851	.406
	Monolingual	396	35.650		
<b>Neutral /ɑ/</b>	Bilingual	407	42.575	1.343	.196
	Monolingual	386	26.077		

\* $p < .05$

no automatic detection of differences between the non-native vowels, but attentional and cognitive resources that are recruited later in processing did facilitate the identification of the challenging non-native vowel /ɑ/. Our results support previous findings indicating that perception of some L2 phonemes may not occur pre-attentively, and that MMN responses are larger to native phonemic contrasts compared to non-native ones (Kirmse et al., 2008; Näätänen, Paaivlainen, Rinne & Alho, 2007). The current findings are also in line with work showing that, when non-native listeners attend to L2 speech sound contrasts, detection of deviant features can improve (Hisagi et al., 2010; Ong, Burnham & Escudero, 2015). This suggests that attention may facilitate the perception of non-native speech sounds.

This study's results are influenced by factors such as age of acquisition of the language (early vs. late), length of exposure to sounds of the L2, use of the second language, and type of L2 learning (e.g., classroom vs. immersion). Peltola, Tamminen, Toivonen, Kujala and Näätänen, (2012) found that the nature of an L2 learning experience can influence pre-attentive responses to non-native vowels. They observed that balanced Finnish–Swedish bilinguals (who learned both languages from birth and used them daily) had longer MMN latencies to vowels compared to dominant Finnish–Swedish bilinguals (who learned both languages sequentially, and had a high L2 proficiency level). They proposed that dominant bilinguals, who were hypothesized to have separate representational systems for each language, can inhibit native language categories and access non-native categories faster than balanced bilinguals, hypothesized to have just one language system and therefore a larger pool of categories to search. More recently, Tamminen et al. (2013) observed longer MMN latencies and overall smaller amplitudes in response to non-native Swedish vowels in balanced Finnish–Swedish bilinguals compared

to monolingual Finnish listeners, suggesting that even skilled bilinguals show slower and smaller effects of pre-attentive processing than native speakers.

Along similar lines, sequential bilingual listeners in this study showed smaller MMN amplitudes in response to non-native vowel contrasts than the monolingual group. It is possible that, similar to findings in Peltola, Kujala, Tuomainen, Aaltonen & Näätänen (2003), the extensive classroom instruction experience (14–15 years) and the various lengths of exposure to AE speech sounds (6 months to 5 years) were not sufficient to allow the bilingual listeners to develop and/or access native-like categories for classifying the non-native vowels, at least pre-attentively (at MMN latencies).

The desensitization hypothesis proposed by Bohn (1995) suggests that, when spectral differences are not available for listeners to differentiate vowel contrasts, they will rely on durational differences. In addition, the L2LP model (Escudero, 2005; van Leussen & Escudero, 2015) posits that L2 learners whose first language does not use duration contrastively may begin to use this new dimension, hence relying on durational cues at some stages of the learning process, and eventually learning to weight primary and secondary acoustic-phonetic cues in a native-like fashion. The behavioral and neurophysiological findings from this study suggest that the bilingual participants may have access to spectral information in the AE vowels, and may have reached the stage at which, when attentional resources are engaged, they rely primarily on spectral cues to distinguish /ʌ/ from /ɑ/. This discriminative ability did not, however, appear to assist them in the opposite contrast, distinguishing /ʌ/ from /ɑ/. This may be counter-intuitive, but the present study also provides evidence that the vowel /ʌ/ may be perceptually less salient than /ɑ/, even for monolingual native speakers of AE; this point is addressed further below.

The monolingual group from the present study showed expected indices of discrimination of the AE vowel contrast /a/-/ʌ/ during the non-attended task, as indicated by significant MMN responses to /a/ in its deviant status, during both natural and neutral vowel duration conditions. Unexpectedly, this group did not show statistically significant MMN responses when /ʌ/ was the deviant sound in the natural duration condition. They did, however, show indices of automatic discrimination over the right hemisphere during the neutral duration condition, reflecting the detection of acoustic differences between the speech sounds (Näätänen, 2001). During a conscious decision-making process, when attention was engaged in the task, the monolingual group showed a significant P300, indicating expected identification of both /a/ and /ʌ/ in natural and neutral conditions. These results suggest that, unexpectedly, attention also played a role for the monolingual group in identifying /ʌ/. It is possible that the representation of the mid-central AE vowel /ʌ/ is not as defined as the low-back vowel, which is also peripheral in relation to the central one, rendering /ʌ/ less salient than /a/, even for monolingual AE speakers.

The pattern of neurophysiological results was similar to that observed behaviorally, in that enhanced ERP responses (MMN and P300) were observed in response to /a/ in both groups, further suggesting a perceptual vowel asymmetry favoring /a/ over /ʌ/. Asymmetry in vowel perception has been described as a phenomenon in which a vowel presented against the background of another vowel is more easily discriminated when the standard vowel is more peripheral in the vowel space, acting as a perceptual ‘anchor’, than if the presentation is in the reverse order (Polka & Bohn, 2003; 2010). In general, vowel perception asymmetries have been described as reflecting perceptual preferences for vowels located in the periphery of the vowel space, as they provide an anchor for comparison (Polka & Bohn, 1996; 2003; 2010; Schwartz, Abry, Boë, Ménard & 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 /a/ over /ʌ/ in the monolingual group during the non-attended, and in the bilingual group in the attended, task, for both vowel duration conditions, could be interpreted as an instance of asymmetric vowel perception phenomena. It appears that /a/ acted as the anchor vowel for both groups.

The Natural Vowel Referent framework (Polka & Bohn, 2010) suggests that asymmetry effects in native vowel perception can be present in infancy, but such effects are expected to reduce for native vowel contrasts and enhance for non-native contrasts by adulthood (Polka & Bohn, 2010). However, previous studies have shown asymmetries in English-speaking adults discriminating native vowel contrasts (Cowan & Morse, 1986; Repp &

Crowder, 1990). The current findings are more in line with the Dispersion-Focalization Theory (DFT) (Schwartz et al., 2005). Under this view, vowels that have focal (or close) values for F1 and F2 offer a benefit for speech perception. In the current study, /a/ has close values for F1 and F2 (F1=903.964, F2=1319.61), and appears to be more perceptually salient than /ʌ/ (formant values: F1=880.027, F2=1545.569). Hence, /a/ becomes a reference for discrimination. Behavioral studies (Karypidis, 2007; Nishi, Strange, Akahane-Yamada, Kubi & Trent-Brown, 2008; Sebastián-Gallés, Echeverría, and Bosch, 2005) have shown vowel asymmetries that reveal consistently better discrimination for peripheral vowels in non-native listeners; and neurophysiological studies (Aaltonen, Niemi, Nyrke & Tuhkanen, 1987; Sharma & Dorman, 2000) have found larger MMN responses to deviant vowels with low F1 and high F2 values. An alternative explanation of the apparent perceptual salience of the AE vowel /a/ in this study could be that the targets /bʌb/ and /bab/ were hyper-articulated (Johnson, Flemming & Wright, 1993), despite our efforts to ensure that they were not over-enunciated during recording; this could have rendered the formant values for /a/ more extreme.

Although both groups showed perceptual preference for AE vowel /a/, there was a difference between groups regarding the level at which this preference was apparent. The monolingual group showed a perceptual preference for the AE vowel /a/ in the passive listening task, where minimal cognitive resources are required. In contrast, the bilingual group did not show discrimination of either target vowel when attention was not required; instead, they showed discrimination of /a/ only when attention, working memory load, and other cognitive resources were recruited to consciously identify the vowels. Similarly, attentional resources may have facilitated identification of /ʌ/ during the attentional task in both natural and neutral vowel duration conditions for the bilingual group. However, due to its non-salient status, this vowel remained difficult for bilingual listeners even when such resources were recruited. Shafer, Schwartz and Kurtzberg (2004) have suggested that attention is important for discriminating difficult speech sound contrasts, even for native listeners. Other studies have presented similar findings indicating facilitation of speech sound perception and enhanced neurophysiological responses to specific speech sound contrasts when conscious attention is directed to auditory speech stimuli (Hisagi et al., 2010; Sussman, Kujala, Halmetoja, Lyytinen, Alku & Näätänen, 2004).

An additional source of potential confound in this study, though difficult to control, is the dialectal heterogeneity of the monolingual participants, even though all were native AE speakers. Cross-dialectal spectral variation has been described in different dialects of AE vowels (Fox & Jacewicz, 2009), indicating that perceptual parameters for discriminating vowels may

be slightly different among listeners from different AE dialects. However, it is clear that all AE listeners use spectral information as a primary cue for vowel discrimination. These factors may have introduced variability in behavioral and neurophysiological responses to the vowel stimuli in this experiment.

The lexical frequency of the syllables used in this study (/bab/ and /bAb/) might have influenced behavioral and neurophysiological responses in both Spanish–English bilinguals and English monolingual groups. Both syllables are real words in American English, but while /bab/ is high frequency (91.4902 instances per million), /bAb/ is much lower frequency (1.6272 per million) (Marian, Bartolotti, Chabal & Shook, 2012). This suggests that /bab/ may have been more familiar to listeners in both groups, and raises the question whether observed asymmetries could be partially explained by the fact that /a/ was presented in the context of a more frequent real word (/bab/), while /A/ was presented in the context of a less frequent word (/bAb/). The view that the apparent perceptual salience of /a/ is related to its representational properties is supported by the Natural Vowel Referent hypothesis (Polka & Bohn, 1996; 2003; 2010) and Dispersion-Focalization Theory (Schwartz et al., 2005), as discussed; however, a future study specifically directed at examining the lexical properties of carrier stimuli (e.g., contrasting real and pseudo words) would provide valuable insights into other possible causes for the vowel asymmetry affecting this particular contrast.

Although the perceptual salience of /a/ may explain the MMN asymmetry observed in monolingual listeners, the fact that neutralization of the durational cues supported discrimination of both vowels at the acoustic level raises questions about how the unavailability of secondary cues would make the acoustical distinction between these vowels apparently easier for monolingual listeners at the pre-attentive level. Based on previous findings indicating that Spanish-speaking learners of English rely primarily on durational cues to identify English vowels /i/-/i/ (Escudero & Boersma, 2004; Escudero 2005; Morrison, 2006; 2008; 2009), and Dutch vowels /a/-/a/ (Lipski et al., 2012), it was expected that the bilingual group would also show reliance on vowel duration to discriminate the AE vowel contrast /a/-/A/. However, the bilingual group showed no differences in discrimination between the two experimental conditions (natural vowel duration vs. neutral vowel duration). In a training study comparing learning lexical tones in unattended vs. attended conditions, Ong et al. (2015) found that when attending to the stimuli, without explicitly directing attention to relevant speech cues, listeners learned the tones. The significant P300 responses in the attended task in this study indicated that, when attentional resources were recruited, the bilingual group discriminated /a/ but not /A/. For this to occur, they must have been

relying only on available spectral differences, since durational cues were neutralized; therefore, durational information was not necessary for the bilingual group to detect the AE vowel /a/ when it was presented as a deviant stimulus. Furthermore, as expected, the monolingual group did show indices of discrimination for /a/ when not consciously attending to the vowel sounds, and identification of both vowels /a/-/A/ when attention was required, 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 /a/-/A/.

The L2 Linguistic Perception model (L2LP; Escudero, 2005) proposes that, in initial learning stages, Spanish learners of English will create new categories along a new auditory dimension (length) that has not been previously implemented in their L1 because Spanish does not use duration as a contrastive cue for speech sound processing. Such listeners will also have to create extra categories along an already-used auditory dimension (height). This model predicts that at the end of the learning process, if there is enough L2 input, learners may achieve native-like perception of L2 speech sounds. In the present study, it appears that the adult Spanish–English bilingual participants are at a stage in which they rely primarily on spectral information (similar to native English speakers) to perceive at least one of the AE vowels, namely /a/, in the vowel contrast /a/-/A/. However, they must recruit attentional and cognitive resources in order to utilize spectral information for vowel discrimination. Given the present findings, it seems that even after mainly classroom instruction, limited exposure to the native English sounds, and adult age of arrival to an English-speaking environment, sequential bilingual listeners can recruit cognitive resources that permit access and identification of non-native speech sounds, even when these contrasts are not available pre-attentively.

Although we expected the bilingual group to show response patterns similar to those reported for the English vowel contrast (/i/-/ɪ/), this was not the case. It is possible that different L2 vowel contrasts require different perceptual adjustments in the use and weight of specific acoustic-phonetic cues. In addition, the unexpected finding that native AE listeners showed no significant MMN responses to /A/ generates new questions regarding the status of that vowel in the phonetic representations of native listeners; therefore, generalization of findings at this stage may not be possible.

To conclude, the objective of this study was to examine neurophysiological responses (MMN and P300 event-related potentials) of adult sequential Spanish–English bilinguals to AE vowel contrasts /a/-/A/ compared to monolingual English-speaking listeners, in two tasks

requiring perceptual discrimination and identification under two listening conditions: natural and neutral vowel duration. In general, study findings indicate that adult sequential Spanish–English bilinguals are less accurate than English monolinguals in discriminating the AE vowel contrastive pair /a/-/ʌ/. Bilingual listeners did not show neurophysiological indices of perception of the AE vowel contrast /a/-/ʌ/ at the pre-attentional level. However, when attentional and other cognitive resources were recruited, identification improved, at least towards the most perceptually salient vowel in the pair, /a/. The monolingual group showed pre-attentional neurophysiological indices of discrimination towards AE vowel /a/, but not to /ʌ/. When attentional resources were recruited, the monolingual group showed neurophysiological indices of identification of both vowels in the contrast. The apparent perceptual preference for /a/ was observed at the behavioral and neurophysiological levels for the monolingual group, and at the neurophysiological level for the bilingual group. Non-significant ERP differences between natural and neutral vowel duration conditions suggested that the bilingual group did not use durational information as the most important cue to discriminate and identify AE vowel contrast /a/-/ʌ/ at the attentional level. Instead, they seemed to rely on spectral cues primarily, as do native English listeners.

These findings strongly suggest that Spanish–English bilingual listeners are able to recruit attentional resources for some non-native speech sound contrasts that are signaled by spectral and durational cues, and this approach has implications for perceptual learning in second language acquisition.

## References

- Aaltonen, O., Niemi, P., Nyrke, T., & Tuhkanen, M. (1987). Event related brain potentials and the perception of a phonetic continuum. *Biological Psychology*, *24*, 197–207.
- Best, C. T. (1995). A direct realistic 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., & 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.
- Berry, J., Jaeger, J., Wiedenhoef, M., Bernal, B., & Johnson, M. (2014). *Consonant context effects on vowel sensorimotor adaptation*. Paper presented at 15<sup>th</sup> Annual Conference of the International Speech Communication Association, INTERSPEECH, Singapore
- 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.
- Bent, T., Bradlow, A., & Smith, B. (2007). Segmental errors in different word positions and their effects on intelligibility of non-native speech: All's well that begins well. In O.-S. Bohn & M. J. Munro (eds.), *Language experience in second language speech learning : In honor to James Emil Flege*, pp 331–347.
- Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America*, *97*, 1916–1924. doi.org/10.1121/1.412064
- Campbell, T., 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
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Clopper, C. G., Pisoni, D. B., & De Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America*. *118*, 1661–1676. <http://dx.doi.org/10.1121/1.2000774>
- Cowan, N., & Morse, P. (1986). The use of auditory and phonetic memory in vowel discrimination. *The Journal of the Acoustical Society of America*, *79*, 500–507.
- Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & Van Ooijen, B. (2000). Constrains of vowel and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition*, *28*, 746–755.
- Csépe, V. (1995) On the origin and development of the mismatch negativity. *Ear and Hearing*, *16*, 91–104.
- Deouell, L. Y., Karns, C. M., Harrison, T. B., & Knight, R. T. (2003). Spatial asymmetries of auditory event-synthesis in humans. *Neuroscience Letters*, *335*, 171–174.
- Dien, J., Spencer, K. M., & Donchin, E. (2003). Localization of the event-related potential novelty response as defined by principal components analysis. *Cognitive Brain Research*, *17*, 637–650.
- Donchin, E., & Coles, M.G.H. (1988). Is the P300 component a manifestation of cognitive updating? *Behavioral Brain Science*, *11*, 357–427.
- Escudero, P. (2005). Linguistic perception and second language acquisition. Explaining the attainment of optimal phonological categorization. LOT Dissertation Series, 113, Utrecht University.
- Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*. 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, 254–259. doi:10.1121/1.3488794
- Ferree, T. C., Luu, P., Russell, G. S., & Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clinical Neurophysiology*, 112, 536–544.
- Flege, J. (1991). The interlingual identification of Spanish and English vowels: Orthographic evidence. *Quarterly Journal of Experimental Psychology*, 43, 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., 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. E., & Munro, M. J. (1994). Auditory and categorical effects on cross-language vowel perception. *The Journal of the Acoustical Society of America*, 95, 3623–3641.
- 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, 847–857. doi:10.1121/1.3159302.
- 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, 2540–2551.
- Fox, R., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America*, 126, 2603–2618. doi:10.1121/1.3212921.
- Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, 25, 1–42. doi:10.1006/cogp.1993.1001.
- 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.
- Hammond, R. M. (2001). *The Sounds of Spanish: Analysis and Application (With Special Reference to American English)*. Cascadilla, Somerville, MA.
- Harris, J. (1969). *Spanish phonology*. Cambridge, Mass: MIT Press.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature*, 8, 393–402.
- Hisagi, M., Shafer, V., Strange, W., & Sussman, E. (2010). Perception of a Japanese Vowel Length Contrast by Japanese and American English Listeners: Behavioral and Electrophysiological measures. *Brain Research*, 1360, 89–105. doi:10.1016/j.brainres.2010.08.092
- 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
- Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, 3, 505–528.
- 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*, pp. 657–660.
- 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, 2365–75.
- 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*, 67, 131–143.
- 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, 3959–3971. doi:10.1121/1.2999341
- Lipski, S. C., Escudero, P., & Benders, T. (2012). Language experience modulates weighting of acoustic cues for vowel perception: An event-related potential study. *Psychophysiology*, 49, 638–650. doi:10.1111/j.1469-8986.2011.01347.x
- Lipski, S. C., & Mathiak, K. (2008). Auditory mismatch negativity for speech sound contrasts is modulated by language context. *NeuroReport*, 19, 1079–83. doi:10.1097/WNR.0b013e3283056378
- Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE* 7(8): e43230. doi:10.1371/journal.pone.004323
- 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. (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 (eds.), *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology* pp. 35–47. Somerville, MA: Cascadilla
- 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, 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, 437–462. doi:10.1177/0023830909336583
- Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel perception. *Language Learning*, 58, 479–502.
- Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive functions. *The Behavioral and Brain Sciences*, 13, 201–288.

- Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear & Hearing, 16*(1), 6–18.
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*, 1–21.
- Näätänen, R., & Alho, K. (1997). Mismatch negativity – The measure for central sound representation accuracy. *Audiology and Neuro-Otology, 2*, 341–353.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*: 432–4.
- 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
- Neonen, S., Shestakova, A., Huotilainen, M., Näätänen, R. (2005). Speech-sound duration processing in a second language is specific to phonetic categories. *Brain and Language, 92*:26–32.
- Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio, 2*: 221–247.
- 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.
- Ong, J. H., Burnham, D., & Escudero, P. (2015). Distributional Learning of Lexical Tones: A Comparison of Attended vs. Unattended Listening. *PLoS ONE 10*(7): e0133446. doi:10.1371/journal.pone.0133446.
- Peltola, M., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., & Näätänen, R. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters, 352*, 25–28. doi:10.1016/j.neulet.2003.08.013
- Peltola, M., Tamminen, H., Toivonen, H., Kujala, T., & Näätänen, R. (2012). Different kinds of bilinguals – Different kinds of brains: The neural organization of two languages in one brain. *Brain and Language, 121*, 261–166.
- 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.
- Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *The Journal of the Acoustical Society of America, 89*, 2961–2977.
- Polka, L., & Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America, 100*, 577–592.
- Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication, 41*, 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*, 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*, 49–71. doi:10.1016/j.pneurobio.2006.04.004
- Repp, B. H., & Crowder, R. G. (1990). Stimulus order effects in vowel discrimination. *Journal of the Acoustical Society of America, 88*, 2080–2090.
- Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural and Brain Research Journal, 111*, 13–23.
- 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*, 425–434. doi:10.1016/j.specom.2004.12.001
- Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*, 240–255.
- Sharma, A., & Dorman, M.F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America, 107*, 2697–2703.
- Sharma, A., Kraus, N., McGee, T.J., & Nicol, T.G. (1997). Developmental changes in P1 and N1 central auditory responses elicited by consonant-vowel syllables. *Electroencephalography and Clinical Neurophysiology, 104*, 540–545.
- Shafer, V.L., Schwartz, R. G., & Kurtzberg, D. (2004). Language-specific memory traces of consonants in the brain. *Cognitive Brain Research, 18*, 242–254.
- Spencer, K., & Polich, J. (1999). Post-stimulus EEG spectral analysis and P300: attention, task, and probability. *Psychophysiology, 36*, 220–32.
- Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., and Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects, *Journal of the Acoustical Society of America, 122*, 1111–1129. <http://dx.doi.org/10.1121/1.2749716>.
- Studebaker, G. (1985). A “rationalized” arcsine transform. *Journal of Speech and Hearing Research, 28*, 455–462.
- 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*, 128–40. doi:10.1016/S0378-5955(04)00016-4
- Sussman, E., Winkler, I., Kreuzer, J., Saher, M., Näätänen, R., & Ritter, W. (2002). Temporal integration: Intentional sound discrimination does not modify stimulus-driven processes in auditory event synthesis. *Clinical Neurophysiology, 113*, 909–920.
- Tamminen, H., Peltola, M., Toivonen, H., Kujala, T., & Näätänen, R. (2013). Phonological processing

- differences in bilinguals and monolinguals. *International Journal of Psychophysiology*, 87, 8–12. doi:[10.1016/j.ijpsycho.2012.10.003](https://doi.org/10.1016/j.ijpsycho.2012.10.003)
- Tervaniemi, M., Jacobsen, T., Rottger, S., Kujala, T., Widmann, A., & Vainio, M. (2006). Selective tuning of cortical sound-feature processing by language experience. *European Journal of Neuroscience*, 23, 2538–2541.
- 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, 1532–1540. doi:[10.1177/0956797610384142](https://doi.org/10.1177/0956797610384142)
- Tremblay, K., Kraus, N., Carell, T., & McGee, T. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. *Journal of the Acoustical Society of America*, 102, 3762–3773
- 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, 79–90.
- van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology*, 6, 1000.
- Van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*, 24, 573–583.
- Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Attention, Perception, & Psychophysics*, 37, 35–44.
- Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Ilmoniemi, R., & Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, 36, 638–42.
- Ylinen, S., Huottilainen, M., & Näätänen, R. (2005). Phoneme quality and quantity are processed independently in the human brain. *NeuroReport*, 16, 1857–1860.
- Ylinen, S., Shestakova, A., Huottilainen, M., Alku, P., & Näätänen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: a cross-linguistic study. *Brain research*, 1072, 175–185. doi:[10.1016/j.brainres.2005.12.004](https://doi.org/10.1016/j.brainres.2005.12.004)
- Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. (2010). Training the brain to weight speech cues differently: A study of Finnish second-language users of English. *Journal of Cognitive Neuroscience*, 22, 1319–1332. doi:[10.1162/jocn.2009.21272](https://doi.org/10.1162/jocn.2009.21272)