Assessing the double phonemic representation in bilingual speakers of Spanish and English: An electrophysiological study

Adrián García-Sierra a,⁎, Nairán Ramírez-Esparza a, Juan Silva-Pereyra b, Jennifer Siard c, Craig A. Champlin c

a Institute for Learning & Brain Sciences, University of Washington, United States
b Universidad Nacional Autónoma de México – Iztacala, Mexico
c University of Texas at Austin, United States

⁎ Corresponding author. Address: Institute for Learning & Brain Sciences, University of Washington, Box 357988, Seattle, WA 98195, United States.
E-mail address: gsa@uw.edu (A. García-Sierra).

Available online 23 April 2012

Event Related Potentials (ERPs) were recorded from Spanish–English bilinguals (N = 10) to test pre-attentive speech discrimination in two language contexts. ERPs were recorded while participants silently read magazines in English or Spanish. Two speech contrast conditions were recorded in each language context. In the phonemic in English condition, the speech sounds represented two different phonemic categories in English, but represented the same phonemic category in Spanish. In the phonemic in Spanish condition, the speech sounds represented two different phonemic categories in Spanish, but represented the same phonemic categories in English. Results showed pre-attentive discrimination when the acoustics/phonetics of the speech sounds match the language context (e.g., phonemic in English condition during the English language context). The results suggest that language contexts can affect pre-attentive auditory change detection. Specifically, bilinguals’ mental processing of stop consonants relies on contextual linguistic information.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

It is well known that speech perception differs in bilinguals and monolinguals (Flege & Eefting, 1986; Flege, MacKay, & Meador, 1999; Mackain, Best, & Strange, 1981; Sundara & Polka, 2008). However, little is known about the way speech perception in bilinguals is affected by the language they are using at the moment. The present investigation explores bilinguals’ ability to discriminate speech sounds that have similar acoustic characteristics in English and Spanish, but have different phonemic-meaning in each of these languages. For example, the English /ga/ is acoustically similar to the Spanish /ka/. Therefore, Spanish–English bilinguals may perceive an acoustic–phonetic speech sound based on the language they are using at the moment – bilinguals may develop two phonemic representations associated with a single acoustic–phonetic speech sound.1 This phenomenon has been called bilinguals’ double phonemic boundary or bilinguals’ double phonemic representation (Elman, Diehl, & Buchwald, 1977; García-Sierra, Diehl, & Champlin, 2009). The term “double” indicates that bilinguals may perceive an acoustic event as two different speech sounds depending on the language context.

The double phonemic boundary in bilinguals has been assessed primarily in the perception of stop consonants. Voicing is one of the primary features that differentiates the stop consonants of a language. Stop consonants can take one of two voicing values: voiced or voiceless. Voiced and voiceless stops vary in the timing of the release of lip closure, or other constriction of the vocal tract associated with the consonant, and onset of the voicing of the vowel. The release of the consonant is accompanied by a strong burst of air, or aspiration, which persists until the onset of voicing. The time period between the release of the consonant and the onset of the voicing of the vowel is defined as voice onset time or VOT (Abramson & Lisker, 1970, 1972; Lisker & Abramson, 1970). Therefore, VOT is a temporal property that robustly specifies voicing in stop consonants. Accordingly, by varying the amount of VOT, it is possible to generate a continuum ranging from voiced/short-aspirated sounds (e.g., a clear ‘ga’) to voiceless/long-aspirated sounds (e.g., clear ‘ka’).

In the English language, voiceless stop consonants have large time periods of aspiration (large VOT duration) and voiced stop consonants have short aspiration (short VOT duration). In contrast, Spanish language voiceless stop consonants have short aspirated periods and voiced stops are not aspirated, with voice onset before the release of the consonant. Voiced stops in the Spanish language are also described as pre-voiced or negative VOT because voicing...
occurs before the release of the consonant articulation. It is important to note that voiceless sounds in Spanish (/p, t, k/) have VOT/aspiration durations that are similar to voiced sounds in English (/b, d, g/). In fact, when monolingual speakers are asked to identify the speech sounds from a voiced-to-voiceless continuum, the placement of the boundary dividing voiced from voiceless sounds will depend on their native language. For example; a velar stop with +15 ms of VOT denotes a clear /ga/ for English monolingual speakers but, it denotes a clear /ka/ for Spanish monolingual speakers.

The fact that the placement of the phonemic boundary depends on the native language raises important questions regarding second language acquisition. For example, would phonemic categories in bilinguals resemble monolinguals’ phonemic categories? Or will bilinguals develop independent phonemic categories for speech sounds that share acoustic–phonetic information but have different phonemic – meaning in each of their languages? If the latter is true, then it would be indicative that bilinguals’ mental representations of stop consonants depend on previous linguistic information (i.e., lexical, syntactic, semantic information of the language in use). Researchers have hypothesized that bilinguals can develop double phonemic representations; that is, perception can change depending on the language used at the moment. However, studies on double phonemic representation in bilinguals have yielded contradictory findings. Although most studies have tested double phonemic representation using a speech continuum, there have been methodological differences, especially in the way that language contexts are established and in the use of monolingual control groups.

Caramazza, Yeni-Komshian, Zurif, and Carbone (1973) asked French–English bilinguals to categorize 3 VOT continua differing in place of articulation (/b-p/, /d-t/, and /g-k/). Language contexts were established by brief conversations in English or French before the perceptual task. The results showed no shift in bilinguals’ phonemic boundaries as a function of language contexts. However, bilinguals’ phonemic boundaries were not aligned with those of monolingual speakers of English or French. These results suggested that the phonemic boundaries of L1 (first language learned) and L2 (second language learned) are merged into a single category that incorporates the phonetic inventories of both languages. The authors concluded that bilinguals do not shift between phonetic rules; rather they adopt a single rule that fits the needs of both languages. The same finding was reported by Williams (1977), who replicated Caramazza et al. (1973) with bilingual speakers of Spanish and English. Although some bilingual participants did show phonemic-boundary-shifts in accordance with language contexts, the language effect was not present in the group average.

In a follow-up study, Elman and colleagues (1977) proposed that language contexts should be established by exposing bilinguals to the language of interest throughout the perceptual task. Indeed, the investigators presented precursor sentences in the language of interest, without excessive exposure to the acoustics of the precursor sentences. The results showed that both groups were affected by the acoustic history of the sentences, but bilinguals showed stronger boundary-shifts than monolinguals. Furthermore, the magnitude of the phonemic-boundary-shift correlated with L1/L2 proficiency (confidence in using Spanish and English) in bilinguals, but not in monolinguals. In other words, large boundary shifts were correlated with high confidence in using English and Spanish. However, since both groups showed language effects, it is clear that the acoustics of the precursor sentences influenced the phonemic boundary shift.

In the present investigation we assessed bilinguals’ double phonemic representation by means of Event Related Potentials (ERPs). In contrast to behavioral paradigms, ERPs provide a continuous measure of the processes occurring between a stimulus and response, making it possible to determine which stage or stages of processing is affected by a specific experimental manipulation (Luck, 2005; see Stryrov & Pulvermüller, 2007 for a review). We assessed speech discrimination in the form of the Mismatch Negativity or MMN (Näätänen, 1992). The MMN reflects electric brain activity associated with early auditory processing in change detection (Näätänen, 1992; Näätänen, Gaillard, & Mantysalo, 1978; Näätänen & Michie, 1979). The MMN is elicited by presenting a repetitive sound that establishes an auditory memory trace for a given sound. Then, a new sound that differs from the memory trace in frequency (or localization, or duration, or intensity, etc.) is presented. The degree of deviance between the memory trace and the new sound is reflected by the ERP amplitude, so that the MMN response increases as the acoustic differences between standard (memory trace) and deviant increase (Tiitinen, May, Reinikainen, & Näätänen, 1994). It has been shown that this ERP component can also be recorded in response to speech sounds (syllables or words) that differ in phonetic properties (Aaltonen, Niemi, Nyrke, & Tuhkanen, 1987; Diesch & Luce, 1997a, 1997b; Näätänen, 2001; Winkler et al., 1999a). Therefore, it can be used to test speech discrimination of consonants and vowels. The MMN has been used in cross-linguistic studies to assess discrimination of speech continua varying in frequency and time domains (Rivera-Gaxiola, Csibra, Johnson, & Karmiloff-Smith, 2000; Rivera-Gaxiola, Johnson, Csibra, & Karmiloff-Smith, 2000; Shafer, Schwartz, & Kurtzberg, 2004; Sharma & Dorman, 2000).

The MMN is an excellent tool to explore bilingual’s double phonemic boundary because it is an early auditory brain response and is not affected by participants’ attention (Näätänen, 1986; Näätänen, Simpson, & Loveless, 1982). For example, participants can read a book or watch a silent movie while the brain data is being collected.
As a result, the MMN offers a unique way to investigate bilinguals’ double phonemic boundary because it allows establishment of language context (reading a book in the language of interest) without relying on precursor sentences that can influence speech perception (Bohn & Flege, 1993; Diehl et al., 1978; Holt, 2005; Holt & Lotto, 2002). Since the MMN precedes conscious perception (see Sussman, Winkler, Huotilainen, Ritter, & Nätänen, 2002), it can be used to explore the effects of language context early in the perceptual processes of mapping acoustic/phonetic cues to categorical mental representations. Most importantly, recent research has shown that the MMN can be modulated by lexical, semantic and syntactic information (Shtryov & Pulvermüller, 2002). Consequently the MMN can be used to explore how higher-levels of information processing interact with early levels of sensory processing to jointly determine what we perceive (see: McClelland, Murman, & Holt, 2006; McClelland & Rumelhart, 1981). Therefore, the MMN is uniquely suited for investigating bilinguals’ double phonemic representation because does not require participants’ attention, it provides a measure of phoneme categorical allocation and it is sensitive to higher order-linguistic information.

To our knowledge there is only one MMN study exploring bilinguals’ double phonemic representation. Winkler, Kujala, Alku, and Nätänen (2003) tested the effects of language contexts on the MMN. Bilingual Hungarian–Finnish speakers were assessed in their ability to discriminate two Finnish words /pæti/ (was qualified) and /pɛti/ (bed). These are perceived as two distinct words by monolingual Finnish speakers and allophones of one word, /Pɛti/ (Peter) by monolingual Hungarian speakers. Bilinguals’ word discrimination was tested in a Hungarian language context and in a Finnish language context. Language context was established by presenting a word in Finnish or Hungarian, which was different from the standard and deviant, on 1.3% of the trials. The researchers expected bilinguals to discriminate the standard and deviant words in the Finish language context, but not in the Hungarian language context. However, the results showed no MMN amplitude change as a function of language contexts, suggesting no double phonemic representation in bilinguals. The authors concluded that learning a second language establishes new phonemic categories that are used irrespective of language context.

In a recent study, Paulmann, Elston-Güttler, Gunter, and Kotz (2006) tested the effects of language context on the N400 – a later ERP component sensitive to semantic integration. Bilingual Germans speakers of English were tested in their ability to suppress L1, while doing a semantic task in L2. The task consisted in making lexical decisions of English words that are spelled the same in German but have different meaning in each of these languages (interlingual homographs, for example, “chef” means ‘cook in English and ‘boss’ in German). Contrary to the authors’ expectations, the results showed no differences in the N400 with language context. However, in a similar study with bilingual German speakers of English, Elston-Güttler, Gunter, and Kotz (2005) provided more robust language context by presenting the interlingual homographs in full grammatical sentences. For example, the prime word gift (denoting ‘poison’ in German) was presented in the sentence The woman gave her a pretty gift followed by the target word poison. Immediately after the target word, participants were asked to tell if gift and poison were related. As expected, these results showed an interaction between the amplitude of the N400 and language contexts, suggesting that it is possible to produce more robust language context effects by visually presenting sentences in the language of interest throughout the lexical task.

In summary, both ERP and behavioral studies demonstrate that results vary depending on methods used to establish language context. Three findings are of particular interest. First, short exposure to the language of interest, before the perceptual task or during the perceptual task, may not be sufficient to establish language contexts (Caramazza et al., 1973; Paulmann et al., 2006; Williams, 1977; Winkler et al., 2003). Second, if language contexts are established by presenting auditory precursor sentences, the acoustic history of the sentences can influence the perception of speech sounds regardless of language status (Bohn & Flege, 1993; Garcia-Sierra et al., 2009). Third, it is possible to establish strong language contexts by visually presenting sentences in the language of interest (Elston-Güttler et al., 2005). The success or failure in establishing language contexts can be explained by what Grosjean (2001) conceptualizes as language mode. Grosjean proposes that bilinguals function along a continuum that reflects the state of activation of a given language at a given point in time in their everyday activities. At one end of the continuum, bilinguals are in monolingual mode and at the other end of the continuum, bilinguals are in bilingual mode. In the monolingual mode, bilinguals use one language while deactivating the other language to the greatest extent possible. In the bilingual mode, bilinguals choose a base language and bring the other language as needed. However and importantly, Grosjean hypothesizes that bilinguals are governed by a “base language” which controls language processing at any time. Therefore, in Grosjean’s view, in order for language context to set a “mode”, bilinguals must be engaged in the language of interest throughout the perceptual task or the base language will predominate. The present study was designed to immerse bilinguals in the language of interest throughout the perceptual task to keep them in monolingual mode (i.e., Spanish mode in the Spanish language context and English mode in the English language context), but the immersion occurs visually and does not produce biases due to acoustic input.

2. Study overview

The present investigation explores bilinguals’ double phonemic representation by means of MMN. The MMN is an ideal measure to test speech discrimination across language contexts since it does not require participants’ attention. Participants were immersed in the language of interest by silently reading magazines during the ERP sessions. Two speech contrast conditions were recorded in each of the language contexts. In the phonemic in English condition, the speech sounds tested represented two different phonemes (across category) for the English language, but represented the same phoneme (within category) for the Spanish language. In the phonemic in Spanish condition, the speech sounds tested represented two different phonemes for the Spanish language but, represented the same phoneme for the English language. We expected that bilinguals in the phonemic in English condition would show better discrimination (larger MMN amplitudes) during the English language context recordings than during the Spanish language context recordings. The opposite pattern was expected for the phonemic in Spanish condition. That is, bilinguals would better discriminate the sounds, as indicated by the MMN, during the Spanish language context recordings than during the English language context recordings.

A control group of English monolinguals was recruited to verify that the speech sounds tested would produce the expected MMN responses. For example, we expected that monolinguals would easily discriminate the sounds used in the phonemic in English condition, but not the sounds in the phonemic in Spanish condition. The monolingual group was tested only in the English language context.

3. Method

3.1. Participants’ exposure to English and Spanish

All participants responded a background language questionnaire that was designed to assess English and Spanish language
exposure from childhood to adulthood (see Garcia-Sierra et al., 2009). Nine bilingual participants were born in the US and two in Mexico. All 12 monolingual participants were Americans born in the US.

3.1.1. Amount of English and Spanish exposure

Our Spanish–English background questionnaire revealed a language pattern in bilinguals that changed from Spanish dominant (ages 3–6) to English dominant (ages 18–21), transitioning through a period of relatively balanced exposure to both languages (ages 9–12).

From ages 3 to 6, 73% of bilinguals were exposed to Spanish 100% of the time. From ages 9 to 12, 64% of the bilinguals were exposed equally to Spanish and English. From ages 18 to 21, 82% of the bilinguals were exposed more to English than to Spanish. Monolinguals reported being exposed only to English from ages 3 to 21.

3.1.2. Amount of English and Spanish use

From ages 3 to 6, 64% of bilinguals reported using (speaking) Spanish 100% of the time. From ages 9 to 12, 64% of bilinguals reported using equal amount of English and Spanish. From ages 18 to 21, 64% of the bilinguals reported using more English than Spanish. Monolinguals reported speaking only English from ages 3 to 21.

3.2. Participants’ English and Spanish level of bilingualism

Proficiency in English and Spanish was assessed by self-reports. Participants were asked to rate themselves on four factors: proficiency in speaking, reading, writing, and listening comprehension. We report only speaking and listening comprehension. For speaking proficiency a Likert scale from 1 to 5 was used (1 = “I cannot speak the language, I have a few words or phrases and, I cannot produce sentences,” and 5 = “I have a native-like proficiency with few grammatical errors and I have good vocabulary”). The means for bilinguals were 4.73 (SD = .47) for English and 4.36 (SD = .67) for Spanish. For monolinguals, the overall means showed that English was indeed their dominant language (mean = 5.00, SD = 0.0), while Spanish proficiency was poor (mean = 2.22, SD = 1.10).

For listening comprehension a Likert scale from 1 to 5 was used (1 = “I only understand few words of what is being said,” and 5 = “I understand all of what is being said”). Bilinguals showed equivalent mean scores for English and Spanish (M = 4.73, SD = .47; M = 4.37, SD = .47; respectively). Monolinguals showed a mean score of 5.00 (SD = 0.0) for listening comprehension in English and a mean score of 3.00 (SD = 1.125) for listening comprehension in Spanish. Only those bilinguals who reported in the Likert scale to be at least 75% confidence in speaking and comprehending English and Spanish were invited to participate in the experiment. Monolinguals, on the other hand, were invited only if they reported to be 25% (or less) confidence in speaking and listening in Spanish.

3.3. Participants

The sample was intended to be as homogeneous as possible. For this reason only right handed females were recruited to participate in the investigation. Participants with auditory thresholds that exceeded 20 dB at any frequency,.25, .5, 1, 2, 4, 6, and 8 kHz received $5.00 and were excused from the experiment. Eleven female bilingual speakers of Spanish and English, and 12 female monolingual speakers of English participated in the study. Nineteen participants were retained for subsequent analysis because they showed clear evoked potentials, passed the hearing screen, and attended to both ERP sessions (i.e., four participants were excluded). The final sample was 10 bilinguals (mean age = 19.73, SD 1.10), and nine monolinguals (mean age = 20.33, SD 1.97). All participants were undergraduate students from the University of Texas at Austin. Participants were recruited by means of flyers and word of mouth and were paid $70.00 for two experimental sessions of approximately 120 min each. Subjects gave written informed consent after the procedures of the experiment were explained to them.

3.4. General procedure

3.4.1. Language contexts

A single language context was established in each experimental session and context sessions were counterbalanced. Language contexts were established by exposing participants to the language of interest before and during the experimental task. Experimental sessions consisted of 30–45 min of Quick-cap placement and 1 h of data collection. The interactions that occurred between the experimenter and the participants during electrode-cap placement were in the language of interest. In the Spanish language context, the principal investigator (Spanish native speaker) spoke to participants in Spanish. In the English language context, the research assistant JS (native English speaker) spoke to them in English. During data collection, participants were instructed to read a magazine in the language of interest. Each language context session was divided into two ERP recordings (two speech contrast conditions; see below). Each ERP recording consisted of five data collection blocks (3 min long) and four relaxation blocks (5 min long). Data collection blocks and relaxation blocks were alternated throughout the experiment. During the relaxation blocks, the researcher asked the participants to describe in the language of interest, the amount of Spanish and English they used during everyday activities in a given weekday and during the weekend. At the end of the 5 min conversation, the experimenter resumed ERP data collection. The time between the end of a relaxation block and the start of data collection was 1 min. According to previous studies a 1 min silence is sufficient to minimize the influence of acoustic history (conversation with researcher) on speech perception (Holt, 2005; Holt & Lotto, 2002).

3.4.2. Stimuli

The stimuli selected were taken from a VOT speech continuum synthesized by Garcia-Sierra et al. (2009). The speech continuum ranging from /ga/ to /ka/ was synthesized in 27 VOT steps, and varied from -100 ms to +100 ms of VOT in 10 ms steps except from 0 to 60 ms were it varied in 5 ms VOT steps. The continuum was generated using the cascade method described by Klatt (1980). Each stimulus was 500 ms long including 10 ms burst. Vowel duration varied from 490 ms (0 VOT) to 390 ms (+100 VOT) depending on VOT duration. Only three stimuli were used to collect ERP data (see below). For the purpose of the present investigation, the vowel duration of the stimuli selected was shortened 123 ms. The F0 contour was kept constant across the stimuli. The total duration for the stimuli used was 376 ms.

3.4.3. Speech contrast conditions

Garcia-Sierra et al.’s (2009) continuum was tested in a pilot study of monolingual speakers of English living in the US (N = 15) and monolingual speakers of Spanish living in Mexico (N = 18). Fig. 1 shows that both monolingual groups perceived stimulus -20 ms and stimulus +50 ms as ‘ga’ and ‘ka’, respectively. In contrast, stimulus +15 ms was perceived as ‘ka’ by Spanish monolingual speakers and it was perceived as ‘ga’ by English monolingual speakers. Accordingly, these three stimuli were selected to establish the speech contrast conditions. Two speech contrast conditions were tested: the phonemic in English condition and the phonemic in Spanish condition. The phonemic in English condition used stimulus +50 as standard and stimulus +15 as deviant. We expected that during the English language context, bilinguals would perceive stimuli +50 and +15 as sounds belonging to
Presslich (1986) was applied to the data off-line. The recorded correction method developed by Semlitsch, Anderer, Schuster, and approximately 2 cm later to either eye. An electro-oculogram artifact corrected with electrode placed on the left and right outer canthi, approximately 2 cm later to either eye. Lateral and horizontal eye movements were monitored with electrodes placed on the orbis ocularis muscle above and below the left eye. An electro-oculogram artifact correction method developed by Semlitsch, Anderer, Schuster, and Presslich (1986) was applied to the data off-line. The recorded different speech categories (‘ka’ and ‘ga’; respectively). However, we expected that during the Spanish language context, bilinguals would perceive stimuli +50 and +15 as sounds belonging to the same speech category (‘ka’ and ‘ka’). In the phonemic in Spanish condition we used stimulus –20 as standard and stimulus +15 as deviant. We expected that during the Spanish language context, bilinguals would perceive stimuli –20 and +15 as sounds belonging to different speech categories (‘ga’ and ‘ka’; respectively). However, we expected that during the English language context, bilinguals would perceive stimuli –20 and +15 as sounds belonging to the same speech category (‘ga’ and ‘ga’).

3.5. Electrophysiological procedure

3.5.1. Stimuli presentation

A classic odd-ball paradigm was used to collect the evoked potentials. This paradigm consists in delivering infrequent stimuli (physically deviant) within a repetitive homogeneous stimulus sequence (standard stimulus). The standard sounds occurred with a probability of 0.85 (850 stimulus repetitions) and the deviant sounds occurred with a probability of 0.15 (150 stimulus repetitions). Two rules governed stimuli presentation: (1) The deviant sound could not occur two times consecutively and, (2) there were at least three standard sounds between deviant sounds. The time between the offset of a stimulus and the onset of the next stimulus (inter stimulus interval) was 500 ms. An insert earphone (EAR Tone, model 3A 10 kHz) was used to present the speech sounds. The peak sound intensity (dB SPL) of each stimulus was measured with a sound-level meter that was connected to a 2-cc coupler. All stimuli were delivered at 85 dB peak-equivalent SPL, which is considered a comfortable listening level.

3.5.2. Electrophysiological recording

The electroencephalogram was recorded with Ag/AgCl sintered surface electrodes using NeuroScan SynAmp amplifier (16 bit A/D converter), and Scan4.3 software from 32 inverting electrodes (Quik-Cap). The resolution of the SynAmp amplifier was set to 0.168 μV (A/D accuracy). Electrodes were referenced to linked earlobes, and the ground electrode was placed 1.5 cm anterior to the central frontal electrode (F2). All leads were placed according to the 10–20 International System. Eye blinks were monitored with electrodes placed on the orbis ocularis muscle above and below the left eye. Lateral and horizontal eye movements were monitored with electrode placed on the left and right outer canthus, approximately 2 cm later to either eye. An electro-oculogram artifact correction method developed by Semlitsch, Anderer, Schuster, and Presslich (1986) was applied to the data off-line. The recorded electromyogram was digitized at 500-Hz sampling rate and filtered using a band-pass filter with low and high cut-off frequencies at 0.05 Hz and 40 Hz, respectively. Electrode impedances were maintained below 5 kΩ across language contexts and conditions. Epochs of 700 ms with a 100 ms pre-stimulus interval were derived from the continuous electroencephalographic recording after off-line filtering the data with a band-pass filter from 1 to 30 Hz. The ERP responses to the first ten stimuli in each block as well as epochs showing voltage changes exceeding +100 mV were omitted from the final average. The ERP averaging process was done by presenting a 1 ms trigger that was time-locked to the presentation of each stimulus (Stim2 Neuroscan Compumedics).

3.6. Number of ERP trials accepted

3.6.1. Phonemic in English condition

The number of trials accepted for the standard sound – excluding the standards following deviants – during the English language context was 324.0 (SD = 123.0) and 316.4 (SD = 125.5) during the Spanish language context in bilinguals. No significant difference was found across language contexts (t(8) = .104, p = .92). The number of trials accepted for the deviant sound during the English language context was 105.4 (SD = 22.1) and 98.1 (SD = 22.1) during the Spanish language context in bilinguals. No significant difference was found between language contexts (t(8) = .82, p = .44). ERPs were collected only in the English language context for monolinguals. Therefore, no statistical comparisons were done across language context. The number of standard trials – excluding the standards following deviants – accepted for monolinguals was 652 (SD = 101) and 102 (SD = 3) for the deviant sound.

3.6.2. Phonemic in Spanish condition

The number of trials accepted for the standard sound – excluding the standards following deviants – during the English language context was 337.1 (SD = 151.3) and 332 (SD = 131.5) during the Spanish language context in bilinguals. No significant difference was found across language contexts (t(8) = −.063, p = .95). The number of trials accepted for the deviant sound during the English language context was 95.0 (SD = 25.0) and 93.1 (SD = 14.1) during the Spanish language context in bilinguals. No significant difference was found across language contexts (t(8) = .18, p = .86). ERPs were collected only in the English language context for monolinguals. Therefore, no statistical comparisons were done across language context. The number of standard trials – excluding the standards following deviants – accepted for monolinguals was 618.5 (SD = 136) and 100 (SD = 11) for the deviant sound.

3.7. Data analysis

3.7.1. Standard and deviant ERP amplitude analysis

The onset of the MMN was expected at 220–235 ms after stimulus onset because the acoustic differences of the speech contrasts tested occurred in the first 20–35 ms (Näätänen, 1992). For both monolingual and bilinguals, the electric brain activity associated with phonemic discrimination was analyzed from 220 to 280 ms after stimulus onset. The amplitude for the standard and deviant response was calculated by averaging the voltage values from the ERP time window (i.e., mean-amplitude). Mean-amplitude was used because the measured amplitude is not biased by the number of trials accepted in each of ERP-responses (Luck, 2005). The mean-amplitude of the deviant ERP response was compared with the mean-amplitude of the standard ERP response. Baseline correction was computed by subtracting the average pre-stimulus voltage (100 ms) from the average voltage occurring in the time window of interest.
Eight regions of interest were computed from the 32 electrodes, each containing the average of a group of three or four electrodes. The regions were Left-Frontal (F7, F3, FT7 and, FC3), Left-Central (T7, C3, TP7 and, CP3), Left-Parietal (P7, P3 and, O1), Mid-Fronto-central (FZ, FCZ and, CZ), Mid-Centroparietal (CPZ, PZ, and OZ), Right-Frontal (F8, F4, FT8 and, FC4), Right-Central (T8, C4, TP8 and, CP4) and, Right-Parietal (P8, P4 and, FT8). In order to test the effects of language context on speech perception in bilinguals, we observed changes between standard and deviant (ERP type) across the eight electrode regions of interest for the phonemic to English condition (i.e., standard +50 ms of VOT; deviant +15 ms of VOT) and the phonemic to Spanish condition (standard −20 ms of VOT; deviant +15 ms of VOT). A 2 (language context; English and Spanish) × 2 (phonemic condition: phonemic in English and phonemic in Spanish) × 2 (ERP type: standard and deviant) × 8 (Electrode regions: Left-Frontal, Left-Central, Left-Parietal, Mid-Fronto-central, Mid-Centroparietal, Right-Frontal, Right-Central and, Right-Parietal) repeated measures ANOVA was performed. Greenhouse-Geisser epsilon (ε) was used for non-sphericity correction when necessary.

3.7.2. MMN Amplitude analysis

Research has shown that the scalp distribution of the MMN has its maximal amplitude in fronto-central electrodes (Schröger, 1998). Also, studies show that there is a MMN polarity inversion at electrodes positioned over the opposite side of the Sylvian fissure, such as the mastoid leads (Alho, 1995; see Deouell, 2007 for a review). Accordingly, in this study the MMN polarity inversion was investigated to verify that the ERP-effects were the consequence of MMN generators. In particular, the fronto-central MMN response (FCZ) was compared with the average of MMN response at leads T7 and T8 (nearest electrodes to the mastoids). The MMN was calculated by subtracting the standard response from the deviant response (deviant-minus-standard). Dependent T-tests were done independently for each condition and each language context.

3.7.3. Additional analyses

Two ERP time windows were analyzed. An early ERP time window (160–220 ms from stimulus onset) investigated the early stages in stimulus-specific sensory integration. This ERP-response is known as the N1 ERP-component and it precedes conscious representation of the stimulus (Nätänen & Picton, 1987; Nätänen & Winkler, 1999). The later ERP time window (280–480 ms from stimulus onset) assessed an ERP positivity that in some instances is followed by the MMN. The positive component is known as the P3a and occurs about 280 ms after stimulus onset. Typically, individuals will show a P3a response when they detect a difference between standard and deviant while attending to another task (Escera, Alho, Schröger, & Winkler, 2000; Friedman, Cycowicz, & Gaeta, 2001; Squires, Squires, & Hillyard, 1975).

4. Results

Our first goal was to investigate the effects of language contexts and speech contrast condition on the amplitude of standard and deviant responses in bilinguals. The repeated measure ANOVA showed a main effect for ERP type (F(1,8) = 7.5, p = .026 ηp² = .50) with no significant main effect for language context or phonemic condition. The main effect for ERP type showed that the deviant sound produced a significant more negative ERP response than the standard sound (Standard Mean = −20 µV SD = .76; Deviant Mean = −51 µV SD = .75). Also, and more relevant to our research questions, the results showed a significant interaction between language context, speech contrast condition and, ERP type (F(1,8) = 10.0, p = .014 ηp² = .55). This interaction suggested that bilinguals’ deviant response differed from the standard response as a function of language context and speech contrast condition. This finding was further investigated by a 2 (ERP type: standard and deviant) × 8 (Electrode regions: Left-Frontal, Left-Central, Left-Parietal, Mid-Fronto-central, Mid-Centroparietal, Right-Frontal, Right-Central and, Right-Parietal) repeated measures ANOVAs. The ANOVAs were done independently for group, condition and language context.

4.1. Do bilinguals show an amplitude change depending on the language context?

A double phonemic representation would result in differences in the standard and deviant responses in the phonemic in English condition and in the phonemic in Spanish condition across language contexts.

4.1.1. Phonemic in English condition

Our expectation was that bilinguals in the English language context would perceive standard (+50 ms VOT) and deviant (+15 ms VOT) as belonging to different phonemic categories (/ka/ and /ga/; respectively). On the other hand, bilinguals in the Spanish language context would perceive both speech sounds as belonging to the same phonemic category (/ka/). Specifically, a significant amplitude difference between standard and deviant ERP-responses in the English language context was expected, but not in the Spanish language context.

4.1.1.1. English language context. Fig. 2A shows standard and deviant ERP-responses from the eight electrode regions. The overall deviant-ERP response (Mean = −.94 SD = 1.20) was more negative than the overall standard-ERP response (Mean = .10 SD = .80). This difference was statistically significant (F(1,8) = 18.1, p = .003, ηp² = .70) with no significant interaction between standard and deviant responses as a function of electrode regions (F(1,8, 15.0) = 2.0, p = .20, ηp² = .20). There was a significant MMN amplitude difference between electrode FCZ and the averaged mastoids (FCZ Mean = −.15 SD = 1.0; Mastoids’ Mean = −.44 SD = .60; t(8) = −6.0, p = .0003).

4.1.1.2. Spanish language condition. Fig. 2B shows standard and deviant ERP responses for the eight electrode regions. The overall deviant-ERP response (Mean = −.37 SD = 1.23) was similar to the overall standard-ERP response (Mean = −.23 SD = .71). Differences between standards and deviants were not significant (F(1,8) = .30, p = .60, ηp² = .036) and there were no significant interactions between standard and deviant responses and electrode regions (F(3,0,24.1) = 1.5, p = .23, ηp² = .16). The MMN amplitude inversion between electrodes FCZ and averaged mastoids was not significant (FCZ Mean = −.20 SD = 1.24; Mastoids’ Mean = −.22 SD = .74; t(8) = .11, p = .92).

4.1.2. Phonemic in Spanish condition

Our expectation was that bilinguals in the Spanish language context would perceive standard (−20 ms VOT) and deviant (+15 ms VOT) as belonging to different categories (/ga/ and /ka/; respectively). On the other hand, during the English language context, both speech sounds were perceived as belonging to the same phonemic category (/ga/). Specifically, a significant amplitude difference between standard and deviant ERP-responses in the Spanish language context was expected, but not in the English language context.

4.1.2.1. Spanish language context. Fig. 3A shows the standard and deviant ERP-responses from the eight electrode regions. Frontal regions produced a more negative response for the deviant than for
the standard, and the overall deviant-ERP response (Mean = −.26 SD = .90) was more negative than the overall standard-ERP response (Mean = .21 SD = 1.12) in a significant way (F(1,8) = 10.0, p = .01, η²p = .55). There was no significant interaction between standard and deviant responses as a function of electrode regions (F(2.3,18.3) = 1.0, p = .41, η²p = .11). There was a significant MMN amplitude difference between electrode FCZ and the averaged mastoids (FCZ Mean = −.86 SD = .72; Mastoids’ Mean = −.33 SD = .36; t(8) = 3.0, p = .02).

4.1.2.2. English language context. Fig. 3B shows standard and deviant ERP-responses for the eight electrode regions. The overall deviant-ERP response (Mean = −.46 SD = .90) was similar to overall standard-ERP responses (Mean = −.90 SD = .94). Differences between standards and deviants were not significant (F(1.8) = 2.6, p = .14, η²p = .25) and there was no significant interaction between standard and deviant responses and electrode regions (F(2.6,21.0) = .73, p = .53, η²p = .08). The MMN amplitude inversion between electrodes FCZ and averaged mastoids showed to be no significant (FCZ Mean = .47 SD = 1.3; Mastoids’ Mean = .15 SD = .40; t(8) = .90, p = .40).

4.2. Do monolinguals perceived the stimuli as belonging to different phonemic categories?

We compared standard and deviant responses in the phonemic in English condition in the English language context in monolinguals to verify that stimuli were discriminated in the expected way.

4.2.1. Phonemic in English condition

Our expectation was that monolinguals in the English language context would perceive the standard (+50 ms of VOT) and the deviant (+15 ms of VOT) as belonging to different phonemic categories (/ka/ and /ga/; respectively). That is, a significant amplitude difference between standard and deviant ERP-responses was expected for the phonemic in English condition. Please note that ERPs were obtained only during the English language context.

Fig. 4A shows the standard and deviant ERP responses from the eight electrode regions. The overall deviant-ERP response (Mean = −.70 SD = .67) was more negative than overall standard-ERP response (Mean = −.10 SD = .74). The results showed a significant main effect for ERP type (F(1,8) = 14.2, p = .006, η²p = .64) with no significant interaction between standard and deviant responses and electrode regions (F(2,8,22.44) = 2.65, p = .08, η²p = .25). There was a significant MMN amplitude difference between electrode FCZ and the averaged mastoids (FCZ Mean = −.93 SD = .70; Mastoids’ Mean = −.33 SD = .60; t(8) = −2.7, p = .02).

4.2.2. Phonemic in Spanish condition

Our expectation was that monolinguals in the English language context would perceive the standard (−20 ms of VOT) and the deviant (+15 ms of VOT) as belonging to the same phonemic category (/ga/ and /ga/; respectively). Namely, no significant amplitude
difference between standard and deviant ERP-responses were expected for the phonemic in Spanish condition during the English language context.

Fig. 4B shows the standard and deviant ERP responses from the 8 electrode regions. The overall deviant-ERP responses (Mean = −36 SD = 1.13) was similar when the overall standard-ERP responses (Mean = −10 SD = .70). The results showed no main effect for ERP type (F(1, 8) = 1.6, p = .23, $\eta^2_p = .17$) and, no interaction between standard and deviant responses across electrode regions (F(1.6, 12.8) = .92, p = .4, $\eta^2_p = .10$). The MMN amplitude inversion between electrodes FCZ and averaged mastoids were not significantly different (FCZ Mean = −.57 SD = .85; Mastoids' Mean = −.24 SD = .60; t(8) = −1.8, p = .11).

The stimuli tested represented clear phonemic categories, as indicated by the MMN, for the monolingual participants. Results obtained from the bilinguals indicate that language contexts can influence the early stages of speech representation as indicated by the changes observed in the MMN amplitude. Specifically, identical acoustic–phonetic information was perceived as belonging to different phonemic categories (presence of the MMN), or as belonging to the same phonemic category (absence of the MMN), depending on language context.

4.3. Additional analyses

For the bilingual group a 2 (language context; English and Spanish) × 2 (phonemic condition; phonemic in English and phonemic in Spanish) × 2 (ERP type; standard and deviant) × 8 (Electrode regions: Left-Frontal, Left-Central, Left-Parietal, Mid-Frontocentral, Mid-Centroparietal, Right-Frontal, Right-Central and, Right-Parietal) repeated measures ANOVA was done to test the N1 ERP response and P3a ERP response. For the monolingual group, a 2 (phonemic condition; phonemic in English and phonemic in Spanish) × 2 (ERP type; standard and deviant) × 8 (Electrode regions: Left-Frontal, Left-Central, Left-Parietal, Mid-Frontocentral, Mid-Centroparietal, Right-Frontal, Right-Central and, Right-Parietal) repeated measure ANOVA was done to test the N1 ERP response and P3a ERP response. In both groups’ analyses (bilingual and monolingual) a Greenhouse-Geisser epsilons (ε) was used for non-sphericity correction when necessary.

4.3.1. Bilinguals’ N1 response

No significant main effects or interactions were found for ERP, language context or speech contrast conditions or electrode regions.

4.3.2. Bilinguals’ P3a response

No main effects for ERP type, language context, speech contrast condition or electrode regions were found. A significant interaction between ERP type, language context, speech contrast condition and electrode regions was found (F(7, 19.2) = 11.4, p = .0003 $\eta^2_p = .58$). Further analyses revealed that in the phonemic in English condition during the English language context the deviant response was more positive than the standard response at frontal electrode regions (i.e., Left-Frontal, Mid-Frontocentral and, Right-Frontal showed a significant P3a response). No significant P3a effects were
found in the phonemic in English condition during the Spanish language context. In the phonemic in Spanish condition during the Spanish language context the deviant response was more positive than the standard response at frontal electrode regions (i.e., Left-Frontal showed a significant P3a response). No significant P3a effects were found in the phonemic in Spanish condition during the English language context.

4.3.3. Monolinguals’ N1 response

No significant main effects or interactions were found for ERP, language context or speech contrast conditions or electrode regions.

4.3.4. Monolinguals’ P3a response

No main effects for ERP type, language context, speech contrast condition or electrode regions were found. A significant interaction between ERP type and speech contrast condition was found ($F(1,8) = 7.7, p = .024, \eta^2_p = .50$). Further analyses revealed that in the phonemic in English condition the deviant response was more positive than the standard response at frontal electrode regions (i.e., Mid-Frontocentral, Right-Frontal and Right-Central electrode regions showed a significant P3a response). No significant P3a effects were found in the phonemic in Spanish condition.

The results showed a predominant frontal distribution for the P3a ERP component. Bilinguals’ and monolinguals’ P3a response was significant only for the speech contrast conditions that matched the language contexts (e.g., Phonemic in English condition during the English language context). These results suggested that language context can enhance involuntarily detection of speech cues among standard and deviant sounds.

5. Discussion

The present investigation explored the double phonemic boundary in Spanish–English bilinguals. Specifically, our goal was to test bilinguals’ ability to discriminate speech sounds that share acoustic–phonetic information in English and Spanish, but have different categorical mental representations in each of these languages. Behavioral research of the double phonemic boundary has yielded contradictory findings. Some studies show that monolingual and bilingual speakers develop fixed but distinct phonemic boundaries (i.e., bilinguals merge the phonetic rules of their two languages into one boundary, Caramazza et al., 1973; Williams, 1977). Other investigations have shown that phonemic boundaries in bilinguals shift depending on the language in use (Elman et al., 1977; Flege & Eefting, 1987; García-Sierra et al., 2009; Hazan & Boulakia, 1993). The previously reported shifts in the phonemic boundary suggest that bilinguals can develop independent phonetic rules for speech sounds that share acoustic information in their two languages, but have different meaning in each of these languages. However, it has been suggested that bilinguals’ double phonemic representation is not linguistic in nature but rather is
a consequence of the methods used to establish language contexts (Bohn & Flege, 1993).

In the present investigation we assessed the effects of language context on brain activity associated with speech discrimination in bilingual and monolingual participants. Specifically, we expected that bilinguals’ categorical mental representation would change from /g/ to /k/ as a function of language context. Unlike previous research, we established language context by asking participants to read silently in the language of interest during ERP recordings. This procedure allowed us to set bilinguals in a specific language mode (Grosjean, 2001) throughout the perceptual task.

Our results showed significant amplitude changes between standard and deviant responses as a function of language contexts in bilinguals. Differences between standard and deviant ERP responses were significant only for the expected stimulus condition and language context combinations: in the phonemic in English condition during English language context and in the phonemic in Spanish condition during Spanish language context. Our results suggest that our data collection and language context methods were successful and that bilingual speakers of English and Spanish possess a double phonemic representation. These results are consistent with behavioral studies supporting the double phonemic boundary in bilinguals (Elman et al., 1977; Flege & Eefting, 1987; García-Sierra et al., 2009; Hazan & Boulaaki, 1993) but conflict with previous behavioral and ERP investigations of double phonemic representation in bilinguals (Bohn & Flege, 1993; Caramazza et al., 1973; Williams, 1977; Winkler et al., 2003).

Winkler and colleagues (2003) explored double phonemic representation in Hungarian–Finnish bilinguals using the MMN. Their results showed no significant MMN amplitude changes as a function of language contexts. The authors concluded that bilinguals apply L2-phonetic-distinctions irrespectively of the language in use. In this study we argue that in order to create strong language context effects, participants should be immersed in the language of interest throughout the perceptual task. Winkler et al. established language context by presenting a word in Finnish or Hungarian, which was different from the standard and deviant, on 1.3% of the trials, which may not have produced a strong language context effect. In fact, in a set of studies using language context to prime bilinguals, the primer-effect was shown only when language contexts were presented consistently throughout the entire task (Elston-Güttler et al., 2005).

The inclusion of a monolingual group served to verify that our speech sounds represented clear phonemic categories. Indeed, monolingual English speakers categorized the speech sounds, as indicated by the MMN, in the expected way. Standard and deviant responses differed significantly for the stimuli pair used in the phonemic in English condition during the English language context. In contrast, standard and deviant did not differ significantly for the stimuli pair used in the phonemic in Spanish condition, during the English language context. Although our findings suggested that language context influenced speech perception, it is essential to test our methodology in other bilingual and monolingual populations that show an overlap in their voicing boundaries (e.g., French vs. English, German vs. Spanish).

Bilinguals’ and monolinguals’ difference waveforms show a positive peak around 300 ms after stimulus onset. We interpreted this positivity as the P3a ERP component. The P3a response has been described to be elicited when attention is being switched from the background sound (standard) to the target sound (Escera et al., 2000; Friedman et al., 2001; Squires et al., 1975). The statistical analyses for the P3a response showed that it was frontally distributed and significant for the conditions where the phonetics of the speech contrasts matched those of the language contexts (phonemic in English condition during English language context and phonemic in Spanish condition during Spanish language context) in bilinguals and monolinguals. The presence of P3a component suggests that the phonetic differences between standard and deviant captured participants’ attention in an involuntary way. In fact, both groups showed the strongest P3a response in the speech contrast condition were the duration of aspiration (+VOT) was the main cue (i.e., phonemic in English condition). This suggests that our participants perceived aspiration more saliently than pre-voicing. The fact that bilinguals’ attention to phonetic differences (P3a) changed as a function of language context, supports our interpretation that bilinguals possess a double phonemic representation.

We are aware that our findings regarding the MMN response may not necessarily reflect a language effect (i.e., be based on knowledge of a second language) but rather a familiarity effect. For example, it has been shown that target words and non-speech sounds are better discriminated in familiar contexts (familiar standard words/non-speech sounds) than in unfamiliar contexts (unfamiliar standard words/non-speech sounds) (Jacobsen, Schröger, Winkler, & Horváth, 2005; Jacobsen, Horváth, Schröger, Lattner, Widmann, & Winkler, 2004). The fact that familiarity-effects were found in non-speech sounds suggests that enhanced discrimination in familiar context may reflect a more general feature of auditory processing that is not linguistic in nature. Our language context methods could have produced a familiarity-effect instead of a language-effect per se. Bilinguals showed significant MMN responses when the phonetics of the speech sounds match the language context (i.e., familiar context) but, no MMN was recorded when the phonetics of the speech sounds did not match the language context (i.e., unfamiliar context). In our opinion, testing speech discrimination of monolinguals in two language contexts could disambiguate language effects and familiarity effects. If familiarity effects were influencing our results, then we would expect that monolinguals would show the same speech discrimination pattern as bilinguals. Unfortunately, in this investigation our method did not allow us to test monolinguals’ brain responses in two language contexts since monolinguals reported no confidence in reading Spanish.

However, there is some evidence suggesting that MMN amplitude changes as a function of language contexts are not the consequence of familiarity effects. Specifically, García-Sierra (2007) established two language contexts by exposing participants to video-clips and pre-recorded audio samples in the language of interest between ERP recording blocks. This method allowed data collection from monolinguals and bilinguals in two language contexts since previous knowledge of a second language was not required. Results from this study showed evidence of stronger MMN responses for bilinguals, but not monolinguals, as a function of language context, indicating double phonemic representation rather than familiarity effects.

In the present investigation, we observed robust MMNs for speech sounds belonging to different phonetic categories (between-category) and reduced MMN amplitudes for speech sounds belonging to the same phonetic category (within-category). Namely, our results indicate that the MMN to speech is sensitive to specific linguistic processes, such as categorical perception of speech (Cheour et al., 1998; Dehaene-Lambertz, 1997; Diaz, Baus, Escera, Costa, & Sebastian-Galles, 2008; Näätänen et al., 1997; Peltola, Kujala, Tuomainen, Ek, Aaltonen, & Näätänen, 2003; Rivera-Gaxiola, Csibra et al., 2000; Sharma & Dorman, 2000; Winkler et al., 1999a, 1999b). However, it has also been postulated that the MMN represents a more general acoustic process rather than a linguistic one. For example, equivalent MMN amplitudes have been reported for speech sounds belonging to the same phonetic category and speech sounds belonging to different phonetic categories (Maiste, Wiens, Hunt, Scherg, & Picton, 1995; Pettigrew et al., 2004; Sams, Aulanko, Aaltonen, & Näätänen, 1990; Sharma, Kraus, McGee, Carrell, & Nicol, 1993; Tampas, Harkrider, et al., 2005).
The contradictory findings regarding MMN sensitivity to acoustic vs. phonetic representations of speech can be explained by methodological differences rather than neural speech processing sensitivity (Dehaene-Lambertz, 1997). Research showing within-category sensitivity tested monolingual speakers’ abilities to discriminate sounds occurring in their native language (e.g., two stop consonants with different VOT durations but representing the same sound; Maitse et al., 1995; Pettigrew et al., 2004; Sams et al., 1990; Sharma et al., 1993; Tampas et al., 2005); whereas research showing no within-category sensitivity tested monolinguals’ abilities in discriminating non-native speech contrasts (e.g., two vowels sounds, one present in their native language and the other not present in their language: Cheour et al., 1998; Dehaene-Lambertz, 1997; Diaz et al., 2008; Nätänen et al., 1997; Peltola et al., 2003; Sharma & Dorman, 2000; Winkler et al., 1999a, 1999b). These findings can be explained by Best’s perceptual assimilation theory (Best, 1993, see also Best, McRoberts, & Goodell, 2001) that proposes that monolingual speakers will show poor within-category discrimination for sounds not present in their native language due to “phonetic perceptual assimilation”. In other words, non-native speech sounds are perceptually grouped to the “nearest” native-phonetic category. Our results are consistent with Best’s assimilation theory (1993) in the sense that monolingual participants showed reduced within-category sensitivity for the speech contrast condition consisting of a native and a non-native speech sound for the English language (phonemic in Spanish condition). That is, monolinguals grouped the standard (−20 ms VOT(non-native)) and the deviant (+15 ms VOT(native)) into the same phonemic category. Negative VOT has been shown to be a temporal cue that English monolingual speakers learn to detect only after rigorous auditory training (Tremblay, Kraus, Carrell, & McGee, 1997; Tremblay, Kraus, & McGee, 1998). Overall, it seems that the MMN is especially sensitive to phonetic differences between native and non-native languages (Nätänen, 2001).

Bilinguals showed a change in within-category sensitivity as a function of language contexts. To be precise, bilinguals showed less within-category sensitivity in the phonemic in English condition during the Spanish language context than during the English language context. The same pattern was found in the phonemic in Spanish condition. These results cannot be compared with the monolingual results because the target speech sounds (+15 ms of VOT) is ambiguous for the bilingual speaker but not for the monolingual speaker. More research is necessary in order to compare bilinguals’ and monolinguals’ sensitivities to sounds belonging to the same phonetic categories (within-category sensitivity). For example, bilinguals could be tested with speech sounds that are not ambiguous for either of their languages but belong to the same phonetic category (e.g., − 50 vs. −30 ms of VOT). Also, bilinguals could be tested in their ability to discriminate speech contrasts that are not present in either of their native languages.

Our findings support the assumption of bilinguals’ double phonemic representation; this finding is in accordance with the idea that perception is a parallel, rather than a serial process. McClelland and Rumelhart (1981) proposed that perception is the result of an interaction between previous information and sensory input. Our results showed that in the case of bilinguals’ speech perception, the early stages of speech representation can be affected by the language being used in a given moment (language mode). More research is needed in this field, especially to incorporate this idea with current and new models on bilingualism.

There are two limitations associated with this study. One limitation is that data was not collected from mastoid electrode sites. Ideally, the polarity inversion associated with the MMN is quantified by comparing the MMN response form fronto-central electrodes and the mastoid electrodes (Alho, 1995). In the present investigation we used electrodes T6 and T7 as substitutes for mastoid electrodes. Although the T6 and T7 electrodes are positioned more frontally and higher than the mastoid electrodes, they are located below the Sylvian fissure as indicated by the 10–20 International System and can be used to explore MMN brain generators. Our comparisons between the fronto-central electrode (FCZ) and the mastoid substitutes (average of T6 and T7) showed a significant amplitude change in the expected conditions. For example, the MMN response from FCZ and the averaged of T6/T7 differed in a significant way in the phonemic in English condition during English language context and in the phonemic in Spanish condition during Spanish language context. Therefore, these results suggest that the enhanced negativity associated with the deviant response does reflect the MMN generators.

Another limitation in this study is that the deviant response was not compared against a deviant control. For example, in Jacobsen and colleagues (2004, 2005), the amplitude of the deviant response is compared against the amplitude of the same deviant sound when presented in isolation. This method controls for the acoustic differences between standard and deviant sounds. Further research associated with language contexts should consider introducing a deviant control condition.

In summary, our results indicate that language context can influence the formation of phonemic memory traces involved in speech discrimination. These findings suggest that bilingual speakers of English and Spanish can assign identical acoustic–phonetic information to different phonemic categories depending on the language they are using at the moment. Nevertheless, our results should be interpreted with caution since monolingual speakers of English were tested only in the English language context. New studies exploring double phonemic representation should test bilingual and monolingual speakers in identical conditions and should control for the acoustic differences between deviant and standard sounds.

Acknowledgments

This study was supported by the US-Mexico/Borderlands Research Award, College of Liberal Arts, The University of Texas at Austin, by the Jesse H. Jones Fellowship, College of Communication, The University of Texas at Austin, and by The National Science Foundation Science of Learning Program (SBE-0354453). We thank Denise Padden for her valuable contribution in reviewing this manuscript. We also thank Dr. Javier Alvarez Bermudez at the Autonomous University of Nuevo Leon, Mexico for making the behavioral testing possible.

References


