Bilingualism: Language and Cognition

cambridge.org/bil

Matching the Mismatch: The interaction between perceptual and conceptual cues in bilinguals' speech perception

Noelle Wig and Adrián García-Sierra

University of Connecticut

Research Article

Cite this article: Wig N, García-Sierra A (2021). Matching the Mismatch: The interaction between perceptual and conceptual cues in bilinguals' speech perception. *Bilingualism: Language and Cognition* **24**, 467–480. https://doi.org/10.1017/S1366728920000553

Received: 15 July 2019
Revised: 29 July 2020
Accepted: 21 August 2020
First published online: 4 November 2020

Keywords:
bilinguals; speech perception; event-related potentials; perceptual-conceptual interaction

Address for correspondence:
Adrián García-Sierra,
E-mail: adrian.garcia-sierra@uconn.edu

Abstract

Speech perception involves both conceptual cues and perceptual cues. These, individually, have been shown to guide bilinguals' speech perception; but their potential interaction has been ignored. Explicitly, bilinguals have been given perceptual cues that could be predicted by the conceptual cues. Therefore, to target the perceptual-conceptual interaction, we created a restricted range of perceptual cues that either matched, or mismatched, bilinguals' conceptual predictions based on the language context. Specifically, we designed an active speech perception task that concurrently collected electrophysiological data from Spanish–English bilinguals and English monolinguals to address the extent to which this cue interaction uniquely affects bilinguals' speech sound perception and allocation of attentional resources. Bilinguals' larger MMN-N2b in the mismatched context aligns with the Predictive Coding Hypothesis to suggest that bilinguals use their diverse perceptual routines to best allocate cognitive resources to perceive speech.

Introduction

Bilinguals' speech perception

Being bilingual mandates the ability to store, balance, and switch between two native languages in an effortless manner. On that note, bilinguals have been shown to use conceptual cues and perceptual cues provided by an immediate language context to influence their speech perception. Specifically, conceptual cues refer to the top-down knowledge of the language being spoken at a given moment in time, while perceptual cues refer to the bottom-up properties of the speech signal. Although conceptual cues and perceptual cues, independently, have been shown to influence bilinguals' speech perception in a linguistic manner, their inherent interaction remains to be understood.

Double phonemic boundary

This perceptual-conceptual interaction is possible to observe in Spanish–English bilinguals given the overlap among the phonetic structures of these languages. Namely, both Spanish and English can use voice onset time (VOT), a phonetic quality of speech defined by the time (in milliseconds) between the articulatory occlusion and the vibration of the vocal folds (i.e., voicing), to distinguish voiced (i.e., /b, d, g/) and voiceless (i.e., /p, t, k/) stop consonants; but, these languages do not phonemically categorize stop consonants in the same way (Abramson & Lisker, 1967). Specifically, Spanish draws its categorical boundary dividing voiceless and voiced stop consonants between negative VOT (-100 to 0 ms) and short lags (0–25 ms), whereas English draws its categorical boundary between short lags (0–25 ms) and long lags (25 ms – 100 ms). Thus, Spanish and English boundaries overlap such that short lags (0–25 ms) are perceived as voiceless stop consonants in Spanish, but as voiced stop consonants in English.

This leads Spanish–English bilinguals to perceptually shift their boundary towards negative VOT in Spanish contexts and towards long lags in English contexts when language contexts are established before, and throughout, behavioral tasks (Casillas & Simonet, 2018; Elman, Diehl & Buchwald, 1977; Flege & Eefting, 1987a; García-Sierra, Diehl & Champlin, 2009; Gonzales & Lotto, 2013; Gonzales, Byers-Heinlein & Lotto, 2019), and passive listening tasks that measure the neural activity associated with speech sound discrimination (García-Sierra, Ramírez-Esparza, Silva-Pereyra, Siard & Champlin, 2012). Bilinguals whose languages share similar phonetic structures to that of Spanish and English also have shown a similar pattern (Antoniou, Tyler & Best, 2012; Hazan & Boulakia, 1993). This perceptual shift in accordance with the immediate language context has been coined as bilinguals' double phonemic boundary (García-Sierra et al., 2009), and provides an ideal scenario to investigate cue interaction in speech perception. Namely, Spanish–English bilinguals can perceive the

© The Author(s), 2020. Published by Cambridge University Press

CAMBRIDGE UNIVERSITY PRESS

same perceptual cue (i.e., short lags) differently across conceptual cues (i.e., as voiceless /t, p, or k/ in Spanish contexts and as voiced /b, d, or g/ in English contexts).

Literature review

Perceptual cues and conceptual cues in speech perception

Previous studies have used perceptual cues, including conversations, videos, and/or magazines to create influential language contexts in bilinguals' speech perception (Antoniou et al., 2012; Casillas & Simonet, 2018; Elman et al., 1977; Flege & Eefting, 1987a; García-Sierra et al., 2009; 2012; Gonzales & Lotto, 2013; Hazan & Boulakia, 1993). However, this design cannot disambiguate if the bottom-up properties, or rather the top-down conceptual knowledge, provided by the perceptual cues themselves primarily influenced the observed double phonemic boundary effect.

Accordingly, bilinguals' perception has proven to go beyond the incoming, bottom-up input. Namely, Gonzales et al. (2019) investigated if the conceptual cues provided by a language context itself, alone, could promote a perceptual shift. To do so, the "bafri - pafri" pseudoword VOT continuum used in Gonzales and Lotto (2013) was first stripped from any language-specific perceptual cue to create the language-neutral pseudoword VOT continuum "baf- to paf-." Second, all instructions were given in English and told bilinguals they would hear a native speaker in the language of interest begin, but not finish, saying one of two rare words in that respective language. Simply, both contexts contained identical perceptual cues, such that any perceptual difference across contexts could be attributed to the conceptual expectation of hearing one language or another. Still, bilinguals' phonetic boundary shifted in accordance with the language context. Thus, this observed double phonemic boundary effect led the researchers to hypothesize that bilinguals' speech perception can be driven by conceptual cues alone.

Now, the next natural step is to investigate the perceptual-conceptual interaction. Explicitly, can bilinguals detect a mismatch between perceptual and conceptual information? This interaction has been widely hidden by measuring bilinguals' perceptual shift along a range of perceptual cues that spanned the phonetic boundaries of both native languages (i.e., negative VOT to long lags). Simply, the perceptual cues never mismatched the conceptual cues, and allowed bilinguals to identify an appropriate phonemic boundary for a given language context.

Bilinguals' perceptual sensitivity

However, one investigation has tested bilinguals' speech perception across restricted phonetic ranges. Namely, García-Sierra, Schifano, Duncan, and Fish (n.d., unpublished manuscript) individually presented VOT ranges that represented the Spanish contrast (i.e., negative VOT to short lags) or the English contrast (i.e., short lags to long lags). The results showed that Spanish–English bilinguals', but not English monolinguals', perception shifted in accordance with the phonemic contrast provided by a specific phonetic range. Thus, VOT ranges can perceptually cue bilinguals to implement different phonetic criteria when perceiving speech sounds. Still, no conceptual information was established prior to the presentation of the VOT ranges, once again leaving the interaction hidden.

Other investigations have shown that bilinguals remain sensitive to the linguistic properties of perceptual cues and preferentially tailor speech processing towards the most appropriate language (Casillas & Simonet, 2018; Gonzales & Lotto, 2013; Ju & Luce, 2004; Lagrou, Hartsuiker & Duyck, 2011). In an eye-tracking paradigm targeting bilinguals' lexical access, Ju and Luce (2004) presented highly proficient Spanish–English bilinguals four pictures of objects along with a spoken Spanish target word that always began with a voiceless stop consonant (i.e., /p, t, k/). Accordingly, Spanish was the only conceptual cue. However, perceptual cues varied: the Spanish target word could be produced with word-initial VOT appropriate for a voiceless stop consonant in Spanish (i.e., short lag) or English (i.e., long lag). Thus, the perceptual and conceptual cues either matched (i.e., Spanish words produced with Spanish VOT) or mismatched (i.e., Spanish words produced with English VOT). Results showed that only Spanish targets with English-like VOT (i.e., mismatch) led bilinguals to fixate longer on pictures whose English names were phonologically similar to the spoken Spanish target word (i.e., *pliers* vs. *playa*) than those whose names were phonologically dissimilar (*eyes/ruler* vs. *playa*). Simply, perceptual cues (i.e., word-initial VOT), even in a cue mismatch, produced language-specific effects in bilinguals' lexical access.

The present study uses bilinguals' double phonemic boundary to observe how bilinguals' speech perception may vary across conceptual cues when presented language-specific perceptual cues that match, or mismatch, these conceptual cues. Simply, this study targets the interaction between perceptual and conceptual cues in bilinguals' speech perception.

Electrophysiological evidence in speech perception

This study also expands prior investigations of cued speech perception by concurrently collecting electrophysiological data throughout a behavioral task to assess the brain's role in such tasks. This is important given how different groups can make the same behavioral response, but arrive at such response in different ways (Kirsner, Brown, Abrol, Chadha & Sharma, 1980; Lauro & Schwartz, 2017). Specifically, the Event-Related Potential (ERP) known as the Mismatch Negativity (MMN) is suggested to represent the brain's automatic processes, or underlying neural mechanisms, involved in encoding a stimulus difference or change in reference to the bottom-up, acoustic input (Näätänen, 1982; 1992; Näätänen, Gaillard & Mantysalo, 1978; Näätänen & Michie, 1979).

Speech perception studies have observed the MMN in response to a sequence of two stimuli that differ in phonetic properties, and thus has been thought to reflect the perception of two different speech sounds (syllables or words) (Aaltonen, Niemi, Nyrke & Tuhkanen, 1987; Diesch & Luce, 1997a, 1997b; Näätänen, 2001; Winkler, Kujala, Tiitinen, Sivonen, Alku, Lehtokoski, Czigler, Csépe, Ilmoniemi & Näätänen, 1999). Yet, the MMN has also been observed when the phonetic differences between two stimuli have not represented two different speech sounds in the language of interest (Rivera-Gaxiola, Csibra, Johnson & Karmiloff-Smith, 2000a; Rivera-Gaxiola, Johnson, Csibra & Karmiloff-Smith, 2000b; Sharma & Dorman, 1998), which suggests listeners used the general acoustic properties of speech to perceive the acoustically fixed difference between stimuli (Bohn & Flege, 1993; Brady & Darwin, 1978; Bohn & Flege, 1993; Rivera-Gaxiola et al., 2000a, b; Sharma & Dorman, 1998).

https://doi.org/10.1017/S1366728920000553 Published online by Cambridge University Press

Yet, Garrido, Kilner, Stephan and Friston (2009) propose the Predictive Coding Hypothesis to explain how top-down, conceptual expectations can affect the bottom-up, perceptually driven MMN response. Like previous research, an MMN is expected when listeners detect an infrequent sound along a sequence of frequent sounds. This outlines the bottom-up processes underlying the MMN. However, listeners additionally use a conceptual model, or top-down processes, relative to the immediate context to predict this infrequent sound. Specifically, listeners give more perceptual weight to top-down processes to reduce the chance of prediction error. Consequently, when listeners perceive a perceptual contrast among the frequent and infrequent sounds that the conceptual model did not predict, prediction error increases and listeners' perceptual weight now shifts to bottom-up input. This larger amount of prediction error, as a result of a perceptual-conceptual mismatch, is hypothesized to elicit a larger MMN relative to a perceptual-conceptual match. Even further, listeners can adjust their conceptual model to now match the perceived bottom-up contrast, if appropriate. This conceptual adjustment is also hypothesized to increase the MMN. Altogether, the MMN has the ability to reflect the perceptual-conceptual interaction in speech perception.

Another deviance-related ERP component, frequently elicited with the MMN, is the N2b (Novak, Ritter & Vaughan Jr., 1992; Sussman, Kujala, Halmetoja, Lyytinen, Alku & Näätänen, 2004; Sussman & Steinschneider, 2009). But unlike the MMN, the N2b is thought to reflect attentive, controlled detection. Thus, the N2b can reflect the attention directed towards top-down processes in perception, apart from the bottom-up driven MMN. This distinction is crucial to consider when investigating the perceptual-conceptual interaction cues in speech perception as two different sounds within the same native phonemic category can still elicit the MMN (Rivera-Gaxiola et al., 2000a, b; Sharma & Dorman, 1998). These negative-going components often overlap in time (MMN: 100–250 ms; N2b: 200–300 ms); however, they remain distinguishable by their scalp topography and mastoid polarities (Näätänen, 1982; Novak et al., 1992; Sussman et al., 2004; Sussman & Steinschneider, 2009). The MMN is observed in fronto-central sites (Fz electrode) accompanied by a polarity inversion at mastoid sites (i.e., more positive), while the N2b is observed in centro-parietal sites (Cz electrode) without a polarity inversion at mastoid sites.

Another ERP component elicited in oddball paradigms, after the MMN and N2b, is the P300, and is often broken down into 2 subcomponents: the P3a and the P3b (Picton, 1992). The P3a is thought to represent an attentional awareness towards the infrequent stimuli (|Polich, 2003; 2007). However, this subcomponent lacks a consistent appearance in auditory oddball paradigms with typical young adults (Polich, 1988), and frequently occurs in the same time window as the MMN (Datta, Shafer, Morr, Kurtzberg & Schwartz, 2010; Polich, 2003; Sutton, Braren, Zubin & John, 1965). Therefore, scalp topography is often used to distinguish the P3a (centro-parietal) and MMN (fronto-central) (Polich, 2007). Importantly, the P3b is thought to reflect the amount of attentional resources used to make a decision about the infrequent stimuli, such that a larger amplitude reflects a larger amount of attentional resources (Bonala & Jansen, 2012; Donchin, 1981; Linden, 2005; Johnson, 1988; Picton, 1992; Polich, 2007). Given the lack of exploration in how attentional resources are spent in cued speech perception across language groups, which can provide valuable insight into the unique patterns of bilinguals' language processing, we focus on the P3b. Accordingly, all further mentions of the P300 refer to the P3b subcomponent.

The present study

We investigate the interaction between conceptual cues and perceptual cues in speech perception by asking Spanish–English bilinguals and English monolinguals to identify the voiceless stop consonant /ta/ along a restricted range of perceptual cues in Spanish and English language contexts. Given that the English phonetic distinction (i.e., short lags to long lags) has been shown to be more psychoacoustically salient than the Spanish phonetic distinction (i.e., negative VOT to short lags) (Bohn & Flege, 1993; Abramson & Lisker, 1972; Keating, Mikos & Ganong, 1981; Pastore, Ahroon, Baffuto, Friedman, Puleo & Fink, 1977; Streeter, 1976; Williams, 1977, 1979), we restrict our range of perceptual cues to only present the less salient Spanish contrast in an attempt to balance the influence of perceptual cues and conceptual cues in speech perception. However, it is important to note that some English speakers also produce voiced stops (i.e., /b, d, g/) with negative VOT; albeit, they still perceive these productions as they would stops produced with short-lag VOT (Flege, 1982; Hay, 2005; Keating et al., 1981; Dmitrieva, Llanos, Shultz & Francis, 2015; Fish, García-Sierra, Ramírez-Esparza & Kuhl, 2017). Yet, Spanish–English bilinguals can perceive negative VOT as either Spanish phonetic variation that distinguishes between its phonemic categories, or as English phonetic variation within a single phonemic category like English monolinguals.

Further, we use an active odd-ball paradigm, which requires a series of frequent stimuli (i.e., standards) and infrequent stimuli (i.e., deviants), to concurrently collect electrophysiological responses. To provide an appropriate Spanish contrast, the range of standard stimuli spanned -20–0 ms VOT (i.e., negative VOT; Spanish /da/), while the range of deviant stimuli spanned 5–25 ms VOT (i.e., short lags; Spanish /ta/ & English /da/).

Different from other ERP paradigms, we present the same sequence of perceptual cues after establishing, and maintaining, two different conceptual expectations corresponding to Spanish and English language contexts. Consequently, the Spanish context models a cue match, while the English context models a cue mismatch appropriate for exploring the interaction between perceptual and conceptual cues in speech perception. Although this design prevents perceptual updating across contexts, we expect only bilinguals to show differences given how their diverse perceptual routines allow them to create appropriate conceptual expectations, and promote perceptual sensitivity.

Further, it has been shown that the diversity within the bilingual population, such as the perceptual overlap between languages, can make comparisons to monolinguals unproductive in characterizing the unique bilingual experience (Luk & Bialystok, 2013). Simply, collapsing a heterogenous bilingual sample into a homogeneous group, to then compare to monolinguals, eliminates the possibility to explore how this diversity can explain bilinguals' patterns. As such, we focus on 2 pairwise comparisons of interest: 1) bilinguals' responses in the Spanish context vs. the English context and 2) monolinguals' responses in the Spanish context vs. the English context. We test these pairwise comparisons of interest for all data including: 1) perceptual sensitivity (i.e., cumulative *d'*), 2) MMN-N2b, and 3) P300.

Explicitly, when using Signal Detection Theory (i.e., cumulative *d'*) to quantify speech perception of a voiceless stop

https://doi.org/10.1017/S1366728920000553 Published online by Cambridge University Press

consonant /ta/, we expect both monolinguals' and bilinguals' sensitivity to increase along our voiced-voiceless VOT continuum in both contexts. However, we expect only bilinguals to show sensitivity differences across contexts, such that they are more sensitive to perceptual cues in the cue match (i.e., Spanish context), relative to the cue mismatch (i.e., English context). Simply, we expect conceptual cues to predictively code, and thus enhance, bilinguals' perceptual sensitivity in a cue match.

But, when conceptual cues do not predictively code for the perceptual cues, as in the cue mismatch (i.e., English context), we expect bilinguals' heightened perceptual sensitivity to drive a conceptual update. In other words, the cue mismatch can reveal the previously hidden interaction in bilinguals' speech perception. Following the Predictive Coding Hypothesis (Garrido et al., 2009), we expect this cue interaction to be evidenced by bilinguals showing a larger MMN and/or N2b in the English context compared to the Spanish context.

Given that this is the first investigation to observe monolinguals in response to nonnative phonetic contrasts presented in nonnative language contexts, P300 analyses will provide insight into the attentional demands of speech perception in diverse language contexts, as well as clarify MMN-N2b analyses.

Methods

General procedure

All participants were University of Texas at Austin students, recruited by means of flyers. Participants answered language questionnaires to assess their level of exposure to both English and Spanish, and completed a hearing evaluation. Qualified participants participated in one, 2-hour experimental session. First, participants watched a video to establish a given language context, and then engaged in the perceptual task. Here, participants were asked to press a button each time they heard the voiceless consonant, /ta/. Participants completed this task in two language contexts: English and Spanish. Electrophysiological recordings were done concurrently with the behavioral task. The methods and recruitment material were approved by the institution IRB.

Participants

Twenty-seven Spanish–English bilinguals (15 women and 12 men; mean age = 22.07; SD = 3.55), and 27 English monolinguals (13 women and 14 men; mean age = 22.55; SD = 3.69) were retained for analyses (i.e., showed clear phonemic boundaries, clear ERP responses, and passed the hearing test). Fourteen bilinguals were born in the U.S., and 12 bilinguals were born in Spanish-speaking countries (i.e., Mexico = 7, Chili = 4, Uruguay = 1) but reported to have been living in the U.S. for 15.12 years on average (SD = 8.00). Hence, some bilinguals began learning English at an early age (i.e., U.S. born), while others began learning English later in life (i.e., Spanish-speaking country born; around 7 years old on average). This is important, given that phonetic categories are shaped by linguistic exposure as early as 4 years old (Karmiloff-Smith, 2010; Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992; Werker, 2012).

Level of bilingualism

Participants' level of bilingualism was assessed using language questionnaires previously used in a host of studies that have evaluated the impact of bilingualism on speech perception, language development, and beyond (García-Sierra et al., 2009; García-Sierra, Rivera-Gaxiola, Percaccio, Conboy, Romo, Klarman, Ortiz & Kuhl, 2011; García-Sierra et al., 2012; García-Sierra, Ramírez-Esparza & Kuhl, 2016; Ramírez-Esparza, García-Sierra & Kuhl, 2014; 2017). These questionnaires use a variety of Likert scales to assess monolinguals' and bilinguals' exposure, confidence, and everyday use of Spanish and English across the lifespan.

Relevant to this study, bilinguals were only invited to participate who had: received exposure to both Spanish and English during childhood; used both languages to communicate in their daily lives (i.e., Spanish with parents; English with teachers); and reported to be 75% (or above) confident in reading, speaking, and listening in both languages. Bilinguals were additionally interviewed in Spanish to confirm self-reported fluency. Only monolinguals who received exposure to only English during childhood, and reported to be 25% (or less) confident in reading, speaking, and listening in Spanish were invited to participate.

Screening

Potential participants completed a hearing evaluation. Participants with auditory thresholds in either ear that exceeded 20 dB at any frequency tested (250, 500, 1000, 2000, 4000, 6000, and 8000 Hz) were dismissed. Qualified participants were asked to participate in one, 2-hour experimental session.

Experimental task

Participants were instructed to press a button when they heard /ta/ in two language contexts, while electrophysiological responses were recorded. A single pseudo-random presentation of perceptual cues along our VOT continuum, or recording block, lasted 75 seconds. Participants had 1s to press the response button. A 60 second pause, or relaxation block, followed each recording block to avoid participant fatigue. This perceptual task, and stimuli, remained the same across language contexts.

A language context provided conceptual cues before (i.e., video), and after (i.e., Big Five Inventory) the perceptual cues (i.e., VOT continuum). A 10 second time interval (i.e., rest) was inserted between the presentation of the BFI questions and the subsequent recording block to prevent spectral information within the questions themselves from influencing the behavioral and/or electrophysiological responses (contrast effects) (Holt & Lotto, 2002; Holt, 2005; Lotto, Sullivan & Holt, 2003). This sequence (recording block, relaxation block, rest) was repeated 10 times in each language context.

Detailed descriptions of the conceptual and perceptual cues follow.

Conceptual cues: Language contexts

Participants watched Spanish or English video-clips (4 min/each) on a computer monitor prior to the perceptual task, once, to establish the conceptual cue. Participants then answered four questions in the language of interest about specific events occurring in the video to engage with the conceptual cue. Participants had approximately 12 seconds to answer each question.

Participants now heard a pre-recorded male voice giving the instructions in either Spanish or English, in accordance with the given language context. However, monolinguals were always

https://doi.org/10.1017/S1366728920000553 Published online by Cambridge University Press

given the English context first to ensure their understanding of the task, whereas the order was counterbalanced for bilinguals.

The conceptual cue was maintained throughout the perceptual task by having participants write answers to 2 of 18 selected questions from the BFI in the language of interest (Spanish – Benet-Martínez & John, 1998; English – John & Srivastava, 1999) in between the presentations of perceptual cues (i.e., during the relaxation block). Questions were presented simultaneously through headphones and a computer monitor. Both BFI questionnaires have 44 items with a 5-point-Likert scale, that ranges from 1 (disagree strongly) to 5 (agree strongly). An option to answer with a question mark (i.e., meaning "I don't know") was included as it was expected that monolinguals would not understand the questions in Spanish.

Perceptual cues: VOT continuum

Ten synthetic speech stimuli, each with different VOT intended to represent a change from /d/ to /t/ (voiced to voiceless) coronal stop consonants, were generated using the cascade method described by Klatt (1980). All speech stimuli were 210 ms in duration with a 10 ms burst, 30 ms formant transition and 115 ms of steady-state (vowel). Since the place of articulation for coronal stops in English (i.e., alveolar) and Spanish (i.e., dental) is discriminated differently based on age of second language acquisition (Casillas, Díaz & Simonet, 2015; Sundara, Polka & Baum, 2006; Sundara & Polka, 2008), we kept the burst properties consistent across all stimuli. We also kept the vowel properties consistent to isolate VOT as the only perceptual cue that differed across stimuli. Specifically, we used appropriate English-like burst and vowel properties, so both monolinguals and bilinguals would be familiar with these perceptual cues. However, our VOT range (i.e., negative VOT to short-lags) only provides a familiar phonemic contrast to bilinguals (i.e., Spanish). Therefore, our results can be explained as a function of VOT in cue matching (i.e., Spanish) and cue mismatching (i.e., English) language contexts, as well as compare across groups for influences of linguistic experience.[1]

Accordingly, all stimuli were first synthesized with five formants starting at appropriate /d/, or short-lag, onset frequency values (i.e., F1 = 220 Hz; F2 = 1800 Hz; F3 = 3000 Hz; F4 = 3600 and F5 = 4500 Hz) to keep the burst consistent. A turbulent noise source (Amplitude of Frication or AF) of 5 ms duration with 75 dB amplitude was applied to simulate the consonant release. The amplitude of frication exciting F2 was 30 dB (A2F) and the amplitude of frication exciting F3 was 50 dB (A3F).

To manipulate VOT, appropriate formant transitions (i.e., negative VOT or short lag) were interpolated linearly over a time range of 30 ms. Transitions to negative VOT (i.e., stimuli tokens: -20, -15, -10, -5 ms of VOT) were created by manipulating three parameters: fundamental frequency (F0), amplitude of voicing (AV), and amplitude of voice-excited parallel F1 (A1V) (Flege & Eefting, 1987b). F0 was set to 85 Hz, AV to 55 dB, and A1V to 45 dB throughout the pre-voicing period. Transitions to short-lag VOT (i.e., stimuli tokens: 0, 5, 10, 15, 20 & 25 ms VOT) were created by 1) delaying the energy in F1 relative to the onset of higher formants and 2) by applying a noise source in F2 and F3 (amplitude of aspiration or AH = 65) during the F1 cutback period. Then, to keep the vowel consistent, all 5 formants in the stimuli were ramped to suitable /a/ frequency values (F1 = 720 Hz; F2 = 1200 Hz; F3 = 2770 Hz; F4 = 3600 and F5 = 4500 Hz).

An insert earphone (EAR Tone, model 3A 10 kΩ) presented the stimuli at the comfortable listening level of 85 dB peak-equivalent SPL, measured by a sound-level meter connected to a 2-cc coupler. Stimuli were delivered at a rate of 1/s. The inter-stimulus time (ISI) varied from 1 to 1.2 seconds randomly.

Perceptual cues: Active oddball paradigm

To concurrently collect electrophysiological data, we used an active oddball paradigm in which each language context presented standard sounds 80% of the time (i.e., 600 sounds) and deviant sounds 20% of the time (i.e., 150 sounds). To create a Spanish contrast, standard stimuli represented the negative VOT category (-20–0 ms VOT), while deviant stimuli represented the short lag VOT category (5–25 ms VOT). An individual 75 second recording block delivered each standard sound 12 times and each deviant sound 3 times, in random order. Importantly, all 10 recording blocks per language context began with a standard sound, as speech sounds with ambiguous category membership (i.e., deviants; Spanish /ta/ or English /da/) are most vulnerable to contrast effects (Diehl, Elman & McCusker, 1978; Eimas & Corbit, 1973).

Explicitly, four rules were considered in each pseudo-random sequence. 1) The same standard sound could not occur consecutively. 2) At least 3 standard sounds must separate two deviant sounds. 3) At least 3 different standard sounds must separate the same standard sound. 4) Each standard sound preceded each deviant sound 6 times.

Electrophysiological recording

The electroencephalogram was recorded with gold-plated surface electrodes, NeuroScan SynAmp amplifiers, and Scan software that included 6 inverting electrodes (Cz, Fz, Fp1, Fp2, M1, M2), one non-inverting electrode (tip of the nose), and one ground electrode (Fpz). All leads were placed according to the 10–20 International System. The M1 and M2 electrodes were used to assess MMN polarity inversion at the supra-temporal auditory cortex. Eye blinks were monitored with Fp1 and Fp2 electrodes. Average data was referenced to the non-inverting, nose tip electrodes when processing data.

The recorded electroencephalogram 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. Epochs of 1000 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 0.1 to 30 Hz. Epochs with voltage changes exceeding + 100 mV were omitted from the final average.[2] For the English context, on average we retained 339.22 (SD = 67.7) standard epochs (monolinguals = 350.63 (SD = 69.57), bilinguals = 327.81 (SD = 62.42)), and 62.89 (SD = 14.25) deviant epochs (monolinguals = 63.78 (SD = 14.46)), bilinguals = 62 (SD = 13.71)). For the Spanish context, on average we retained 329.67 (SD = 75.45) standard epochs

[1]It is important to address the mismatch in phonetic makeup of the speech sounds (i.e., English-like burst + Spanish-like VOT changes + English-like vowel). Surrounding Spanish-like VOT with English phonetic properties provided our monolinguals with familiar speech cues, and let us delineate perceptual cue weighting in bilinguals' speech perception (i.e., VOT) that would not be possible using Spanish phonetic properties (i.e., Spanish burst + Spanish VOT + Spanish vowel).

[2]BESA Statistics 2.0 did not allow us to do Principal Component Analysis (PCA) for ocular artifact rejection with the number of electrodes used. Therefore, we applied stringent artifact rejection parameters, leading to lower proportions of accepted epochs, to avoid any potential disruptive blinks or noise.

https://doi.org/10.1017/S1366728920000553 Published online by Cambridge University Press

(monolinguals = 335.26 (SD = 91.58), bilinguals = 324.07 (SD = 52.2)), and 86.46 (SD = 20.39) deviant epochs (monolinguals = 88.78 (SD = 22.89), bilinguals = 84.14 (SD = 16.79)). The final ERP waveforms were filtered using a 0.1 Hz forward low cutoff filter with 6 dB/oct slope and a 40 Hz zero phase cutoff filter with 24 dB/oct slope.

Data analyses

Behavioral responses

We used an active oddball paradigm in a speech perception task to collect behavioral and electrophysiological responses concurrently, which required frequent and infrequent categories of stimuli (i.e., 120 times per 5 standards; 30 times per 5 deviants). This imbalanced delivery of stimuli, coupled with our objective to assess the interaction between perceptual and conceptual cues, motivated our decision to use Signal Detection Theory to evaluate participants' labeling performance (see Supplementary Analyses for results and discussion, Supplementary Materials).

Signal Detection Theory (SDT) relates a listener's choice behavior to a psychological decision space in 4 ways: Hit, Miss, False-Alarm, and Correct Rejection (Macmillan & Creelman, 2005). As such, any given button press response can be defined as a Hit or False Alarm (an absence of a button press response would be defined as a Miss or Correct Rejection, accordingly). Button presses for deviants (i.e., 5 ms VOT – 25 ms VOT) were scored as Hits, whereas button presses for standards (i.e., -20 ms VOT – 0 ms VOT) were scored as False-Alarms (see Figure S1, Supplementary Materials).

Importantly, SDT allows us to calculate the sensitivity measure *d'*, which can assess the perception of physically equal spaced stimuli along an ascending VOT continuum using the formula: *d'* = z (% identification item x) – z (% identification x + 1) (Macmillan & Creelman, 2005). In other words, *d'* is the difference between overt responses (False-Alarms for standards and Hits for deviants; in z-scores) given to one stimulus and the next adjacent stimulus along the VOT continuum. Therefore, *d'* can reflect sensitivity across stimuli of varying Hit and False-Alarm proportions, which is a crucial consequence of behavioral responses collected in an active oddball paradigm. Further, any given stimulus has a cumulative *d'* value equal to the sum of each preceding stimuli's individual *d'* values. Thus, the cumulative *d'* value reflects a perceptual cue's likelihood to be perceived as /ta/, such that larger values represent higher likelihoods.

Cumulative *d'* was chosen to represent behavioral responses (as opposed to logistic regression or % Hit responses) because of its ability to represent perceptual-conceptual interaction in speech perception. Explicitly, neither logistic regression nor % Hit responses provide sensitivity measures across the VOT continuum, whereas cumulative *d'* models the likelihood of perceiving /ta/ as a function of perceptual sensitivity to the distance between each perceptual cue. Thus, comparing cumulative *d'* across contexts in each group (i.e., paired *t*-tests) can best capture any interaction that occurs between conceptual and perceptual cues in speech perception.

Electrophysiological responses

Preliminary ERP waveform comparisons were conducted with BESA Statistics 2.0 (BESA GmbH, Gräfelfing, Germany) using point-by-point analyses (Groppe, Urbach & Kutas, 2011). This procedure compares standard and deviant waveforms at all time points of interest, and accordingly identifies areas of significant differences. Thus, point-by-point analyses take advantage of the characteristic high temporal resolution of EEG data that is partially overridden in traditional analyses (i.e., repeated measures ANOVAs) that rely on point values (i.e., mean/peak amplitude, latency, etc.) extracted from an a priori time interval.

Although point-by-point analyses do not require an a priori time interval to be specified, the discreet difference between standard and deviant characteristic of the MMN response can be masked by the robust difference characteristic of the P300 response when comparing along the entire waveform. Therefore, we restricted preliminary MMN-N2b point-by-point analyses to an a priori time interval between -100 ms to 300 ms (Näätänen, 1982). Preliminary P300 point-by-point analyses were conducted in the exploratory time interval between 250 ms to 1000 ms.

Next, we calculated the mean amplitude of the standard waveform, deviant waveform, and difference waveform (deviant – standard; MMN-N2b and P300) in each significant time interval indicated via point-by-point analyses. To accommodate different time intervals of significance across groups, the mean amplitudes were submitted to two-tailed *t*-tests with 10,000 permutations respective to the pairwise comparisons of interest. Importantly, permutations, or the random shuffling of labels across observation pairs a given number of times, account for the multiple comparisons problem prevalent in ERP analysis (Bullmore, Suckling, Overmeyer, Rabe-Hesketh, Taylor & Brammer, 1999; Ernst, 2004; Maris, 2012; Maris & Oostenveld, 2007). Statistical significance after permutations suggests that the observed difference cannot be sufficiently explained by random variation.

Results

Behavioral responses

As expected, Figure 1 shows that both bilinguals' and monolinguals' cumulative *d'* increased across our ascending voiced-voiceless VOT continuum in both contexts, but only bilinguals showed differences across contexts for stimuli near or at the Spanish phonemic boundary (i.e., 0, 5, 10 ms VOT).

Pairwise comparisons (see Table 1) revealed that bilinguals were significantly more sensitive to stimulus 0 ms VOT in the Spanish context compared to the English context (bilinguals, $p = .02$; monolinguals, $p = .75$). Neither bilinguals nor monolinguals showed significant differences for stimuli 5 ms and 10 ms VOT across contexts (all $p > .05$). However, it is important to note that bilinguals' sensitivity to stimulus 5 ms VOT across contexts resulted in a small to medium effect size ($d = .34$), and thus mirrored the trend observed for stimulus 0 ms VOT (i.e., increased sensitivity in the Spanish context).

These results outline the expected perceptual-conceptual interaction in speech perception. Specifically, bilinguals were more sensitive to systematic VOT changes in a cue match (i.e., Spanish context) than a cue mismatch (i.e., English context), while monolinguals showed no sensitivity differences across contexts.

Electrophysiological responses (ERPs)

Below, we summarize the findings from the Cz electrode (unless otherwise indicated) given prior research that indicates the MMN

https://doi.org/10.1017/S1366728920000553 Published online by Cambridge University Press

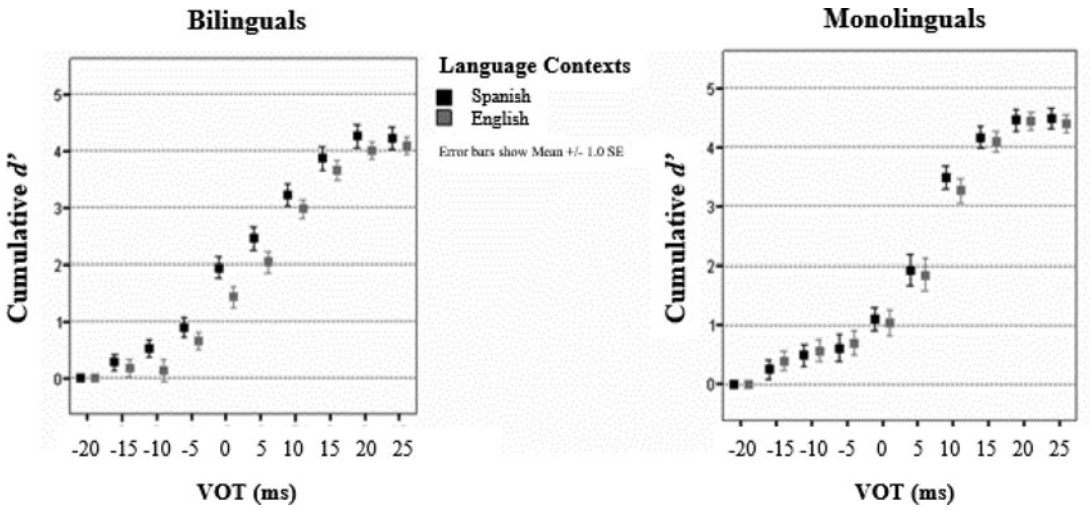


Fig. 1. Participants' cumulative *d'* scores as a function of stimuli. Stimuli -20, -15, -10, -5, and 0 ms of VOT represent standard sounds. Stimuli 5, 10, 15, 20, and 25 ms of VOT represent deviant sounds.

Table 1. *Bilinguals' and monolinguals' cumulative d' means in two language contexts.* Means were submitted to two-tailed *t*-tests with 10,000 permutations. Significant *p*-values (p < .05) and medium to large effect sizes (d > .5) are bolded.

		English cumulative *d'*		Spanish cumulative *d'*				95% Confidence Int.		
Group	VOT (ms)	Mean	SD	Mean	SD	*t* (26)	Sig.	Lower	Upper	Cohen's d
Bil.	0	1.43	*0.92*	1.95	*0.97*	2.37	***0.02***	−.923	−.090	**0.46**
	5	2.04	*0.94*	2.46	*0.94*	1.79	*0.08*	−.845	.039	0.34
	10	2.98	*0.86*	3.22	*1.00*	1.07	*0.30*	−.671	.195	0.21
Mono.	0	1.04	*1.13*	1.10	*1.02*	0.32	*0.75*	−.438	.299	0.06
	5	1.84	*1.40*	1.93	*1.34*	0.35	*0.73*	−.562	.407	0.07
	10	3.26	*1.10*	3.48	*1.00*	1.24	*0.22*	−.537	.116	0.24

Note: 95% confidence intervals are calculated from the mean differences.

as a fronto-central effect (Näätänen, 1982; 1992; Näätänen et al., 1978; Näätänen & Michie, 1979), the N2b as a centro-parietal effect (Novak et al., 1992; Sussman et al., 2004), and the P300 as a posterior effect (Bonala & Jansen, 2012; Linden, 2005; Johnson, 1988; Picton, 1992). Findings from the Cz and Fz electrodes can be found in Tables 2 and 3 (standard vs. deviant, and difference waveform, respectively).

Standard vs. deviant waveforms

First, we report the results from bilinguals' two-tailed *t*-tests with 10,000 permutations using the mean amplitudes calculated from the significant time windows indicated by the point-by-point analyses between -100 ms to 300 ms, and then 250 ms to 1000 ms.

Bilinguals in the English context showed significant differences between standard and deviant responses, such that the deviant was more negative, between 170–272 ms after stimulus onset (t (26) = 2.890, p = .008; see Figure 2, top left panel). Bilinguals in the Spanish context showed similar significant differences between standard and deviant responses between 242–294 ms (in the Fz electrode; t (26) = 2.664, p = .016; see Figure 2, top central panel). These results suggest that bilinguals showed an MMN and/or N2b in both contexts.

Bilinguals in the English context showed significant differences between standard and deviant responses, such that the deviant was more positive, between 330–724 ms after stimulus onset (t (26) = 7.010, p < 0; see Figure 2 top left panel). Bilinguals in the Spanish context showed similar significant differences between standard and deviant responses between 304–720 ms (t (26) = 6.511, p < 0; see Figure 2, top central panel). These results suggest that bilinguals showed a P300 in both contexts.

Next, we report the results from monolinguals' two-tailed *t*-tests with 10,000 permutations using the mean amplitudes calculated from the significant time windows indicated by the point-by-point analyses between -100 ms to 300 ms, and then 250 ms to 1000 ms.

Monolinguals in the English context showed significant differences between standard and deviant responses, such that the deviant was more negative, between 172–284 ms after stimulus onset (t (26) = 2.524, p = .004; see Figure 2, bottom left panel). Monolinguals in the Spanish context showed similar significant differences between standard and deviant responses between 184–248 ms (t (26) = 2.230, p = .038; see Figure 2, bottom central panel). These results suggest that monolinguals showed an MMN and/or N2b in both contexts.

https://doi.org/10.1017/S1366728920000553 Published online by Cambridge University Press

Table 2. *Bilinguals' and monolinguals' standard and deviant amplitudes in both language contexts.* Point-by-point analyses in BESA Statistics 2.0 were used to identify the time windows of significant amplitudes differences between standard and deviant in a given language contexts. Mean amplitudes within these significant time windows were submitted to two-tailed *t*-tests with 10,000 permutations. Pairwise comparisons yielding significant *p*-values ($p < .05$), and medium to large effect sizes (d > .5) are bolded.

Group	Lang. Context	Elec.	Time Window (ms)	Standard ERP (μV)		Deviant ERP (μV)				95% Confidence Int.		Cohen's d
				Mean	SD	Mean	SD	*t* (26)	Sig.	Lower	Upper	
Bil.	Eng.	Fz	166–300	−0.506	*1.501*	−1.410	*1.855*	3.354	***0.002***	0.354	1.388	**0.65**
		Cz	170–272	0.638	*1.858*	−0.362	*2.327*	2.890	***0.008***	0.325	1.669	**0.56**
		Cz	330–724	0.961	*1.447*	4.270	*3.217*	7.010	***0.000***	4.189	2.385	**1.35**
	Sp.	Fz	182–204	0.378	*1.724*	−0.316	*2.349*	2.080	*0.051*	0.026	1.317	0.40
		Fz	222–294	−0.780	*1.834*	−1.766	*2.502*	2.664	***0.016***	0.232	1.689	**0.51**
		Cz	244–268	0.121	*2.231*	−0.935	*3.720*	1.926	***0.068***	−0.008	2.092	0.37
		Cz	304–720	0.860	*1.650*	4.140	*3.465*	6.511	***0.000***	4.225	2.301	**1.25**
Mono.	Eng.	Fz	180–302	−0.051	*1.545*	−1.085	*2.349*	3.260	***0.004***	0.405	1.633	**0.63**
		Cz	172–284	1.140	*1.494*	−0.078	*2.524*	2.524	***0.004***	0.476	1.959	0.49
		Cz	314–628	0.329	1.650	3.315	*3.383*	5.749	***0.000***	3.958	1.988	**1.11**
	Sp.	Cz	184–248	1.540	*1.576*	0.697	*2.103*	2.230	***0.038***	0.147	1.610	0.43
		Cz	290–618	0.061	*1.476*	3.770	*3.460*	6.933	***0.000***	4.754	2.705	**1.33**

Note: 95% confidence intervals are calculated from the mean differences.

Monolinguals in the English context showed significant differences between standard and deviant responses, such that the deviant was more positive, between 314–628 ms after stimulus onset (t (26) = 5.749, $p < 0$; see Figure 2, bottom left panel). Monolinguals in the Spanish context showed similar significant differences between standard and deviant responses between 290–618 ms (t (26) = 6.933 $p < 0$; see Figure 2, bottom central panel). These results suggest that monolinguals showed a P300 in both contexts.

The difference waveform (MMN-N2b & P300)

To compare the MMN-N2b and P300, the mean amplitude of the difference waveform (deviant waveform – standard waveform) within the significant time intervals of interest (i.e., point-by-point analyses) in each context were submitted to two-tailed *t*-tests with 10,000 permutations.

Bilinguals showed significant MMN-N2b differences (210 ms – 228 ms) across contexts (t (26) = -2.381, $p = .025$). Specifically, bilinguals showed a larger negative response in the English context compared to the Spanish context (Figure 2; see top right panel), as expected by the Predictive Coding Hypothesis. Bilinguals showed consistent effects at the Cz electrode without polarity inversions (using the average of both mastoids) (English context: t (26) = .860, $p = .398$; Spanish context: t (26) = .518, $p = .609$) and later MMN time windows of significance, which suggests that an N2b overlapped the MMN, and thus perceptual cues influenced a conceptual update in bilinguals' speech perception.

Monolinguals also showed significant MMN-N2b differences across contexts (248 ms – 300 ms), which was not expected (t (26) = -3.050, $p = .005$) (see Figure 2; bottom right panel). This further suggests that the MMN relies on bottom-up processes, and reflects the automatic detection of an acoustic difference as opposed to a phonemic difference. Nonetheless, Figure 3 illustrates a pattern in monolinguals' deviant and MMN-N2b waveforms for the Spanish context, not seen in bilinguals' waveforms. Namely, the mean amplitudes of the Spanish deviant and MMN-N2b waveforms are positive, when the a priori expectations are for them to be negative (Näätänen, 1982). Further, monolinguals also showed no mastoid polarity inversions for the Cz electrode (English context: t (26) = .256, $p = .800$; Spanish context: t (26) = .282, $p = .780$). Together, these findings suggest that the positive P300 may have spilled over into the negative MMN-N2b response in the Spanish context.

Since this is the first study to record monolinguals' ERPs for a nonnative contrast presented in a nonnative language context (see Hisagi, Shafer, Strange & Sussman, 2010 for a nonnative contrast without language contexts), we further investigated the P300 response (i.e., amplitude and latency analyses) to clarify this unexpected difference.

Bilinguals showed a significantly more positive response in the Spanish language context than the English language context between 290 ms – 352 ms (t (26) = -2.245, $p = .042$) and 406 ms – 470 ms (t (26) = -2.541, $p = .021$) (see Figure 2; top right panel). Monolinguals also showed a significantly more positive response in the Spanish language context than the English language context, but between 245 ms – 338 ms (t (26) = -2.686, $p = .013$) and 342 ms – 462 ms (t (26) = -2.608, $p = .015$) (see Figure 2; bottom right panel). These suggest that both bilinguals' and monolinguals' P300 responses were larger in the Spanish context.

However, monolinguals' significant MMN-N2b (248 ms – 300 ms) and P300 time intervals (245 ms – 338 ms) overlapped, whereas bilinguals' did not (MMN-N2b = 210 ms – 228 ms; P300 = 290 ms – 352 ms). This again suggests that monolinguals' P300 spilled over into the MMN-N2b in the Spanish context.

To test this observation, we compared the mean latency of the P300 response for the pairwise comparisons of interest. Only monolinguals' mean latency of the P300 response differed across contexts (monolinguals: M_d = -21.48, SE_d = 10.42, $p = .05$, 95% CI [-42.00, -1.11]; bilinguals: M_d = 3.78, SE_d = 11.77, $p = .75$, 95% CI

https://doi.org/10.1017/S1366728920000553 Published online by Cambridge University Press

Table 3. *Bilinguals' and monolinguals' difference waveform amplitudes in both language contexts.* Point-by-point analyses in BESA Statistics 2.0 were used to identify the time windows of significant amplitude differences across English and Spanish contexts. Mean amplitudes in each context within these significant time windows were submitted to two-tailed *t*-tests with 10,000 permutations. Pairwise comparisons yielding significant *p*-values ($p < .05$) and medium to large effect sizes ($d > .5$) are bolded.

ERP	Group	Elec.	Time Window (ms)	English Context (mV)		Spanish Context (mV)		*t* (26)	Sig.	95% Confidence Interval		Cohen's d
				Mean	SD	Mean	SD			Lower	Upper	
MMN-N2b	Bil.	Cz	210–228	-1.208	*1.815*	-0.419	2.307	-2.381	***0.025***	-1.432	-0.146	0.46
	Mono.	Fz	245–300	-1.257	*1.970*	-0.395	1.753	-2.704	***0.012***	-1.469	-0.218	**0.52**
		Cz	248–300	-0.959	*2.228*	0.221	2.122	-3.050	***0.005***	-1.957	-0.451	**0.59**
P300	Bil.	Cz	290–352	0.949	*3.460*	1.973	3.283	-2.245	***0.042***	-1.952	-0.184	0.43
		Cz	406–470	3.876	*2.702*	4.910	3.443	-2.541	***0.021***	-1.891	-0.264	0.49
	Mono.	Fz	250–470	0.076	*2.260*	1.252	2.119	-3.439	***0.005***	-1.823	-0.482	**0.66**
		Cz	245–338	-0.046	*2.445*	1.152	2.410	-2.686	***0.013***	-2.083	-0.352	**0.52**
		Cz	342–462	3.634	*3.451*	5.100	3.260	-2.608	***0.015***	-2.564	-0.368	**0.50**

Note: 95% confidence intervals are calculated from the mean differences.

[-18.78, 27.11]), such that monolinguals' P300 was earlier in the Spanish context. This again supports the spillover effect, which would create an artificial difference in the MMN-N2b across contexts for monolinguals.

Discussion

This study aimed to understand the interaction between perceptual and conceptual cues during speech perception. This was possible to study in Spanish–English bilinguals as these languages use different phonetic ranges to phonemically distinguish voiced from voiceless stop consonants. Further, these phonetic ranges and other language-specific perceptual cues (Antoniou et al., 2012; Casillas & Simonet, 2018; Elman et al., 1977; Flege & Eefting, 1987a; García-Sierra et al., 2009; 2012; unpublished manuscript, Gonzales & Lotto, 2013; Hazan & Boulakia, 1993), in addition to conceptual expectations (Gonzales et al., 2019), have influenced bilinguals' speech perception in a linguistic manner. Thus, these investigations show that bilingual speakers rely on perceptual and/or conceptual cues to best "accommodate" phonetic information.

To better understand how perceptual cues and conceptual cues interact during bilinguals' speech perception, we conceptually cued Spanish–English bilinguals in both of their languages, but perceptually cued them in only one language. Simply, bilinguals were presented phonemically relevant perceptual cues in the Spanish context (i.e., cue match), but not in the English context (i.e., cue mismatch). We tested English monolinguals alongside Spanish–English bilinguals to assess if any observed differences across contexts could be attributed to bilinguals' diverse linguistic perceptual routines. Altogether, participants were asked to behaviorally identify /ta/ along a VOT continuum that phonemically distinguishes stop consonants in Spanish, but not English (i.e., –20 ms VOT to 25 ms VOT), in Spanish and English language contexts

Further, this is the first study to employ an active, electrophysiological odd-ball paradigm, in which multiple standard sounds (i.e., –20, –15, –10, –5 & 0 ms of VOT) were randomly presented with multiple deviant sounds (i.e., 5, 10, 15, 20 & 25 ms of VOT). Multiple stimuli, as opposed to a single representative stimulus for a respective category (i.e., standard and deviant), minimizes the fixed acoustic difference between stimuli and better mirrors the phonetic variability language users face every day (Phillips, Pellathy, Marantz, Yellin, Wexler, Poeppel, McGinnis & Roberts, 2000); thus, making the observation of the MMN more likely to reflect phonemic distinction and more applicable to the interaction between perceptual and conceptual cues.

Behavioral responses

Bilinguals were more sensitive (i.e., larger cumulative *d'*) to stimuli near or at the Spanish phonetic boundary in the Spanish context than the English context, while monolinguals showed no sensitivity differences across contexts. Although this was a small effect, it agrees with Keating et al. (1981) who found that the phonetic composition of listeners' native languages affects speech sound perception along a given VOT range. Specifically, listeners' perception is enhanced along the phonetic range typically used to make a given phonetic contrast. In the present study, Spanish–English bilinguals, but not English monolinguals, were given this opportunity and can explain monolinguals' lack of difference.

But since the same phonetic range was presented in both contexts, bilinguals' differences here can elaborate upon Keating

https://doi.org/10.1017/S1366728920000553 Published online by Cambridge University Press

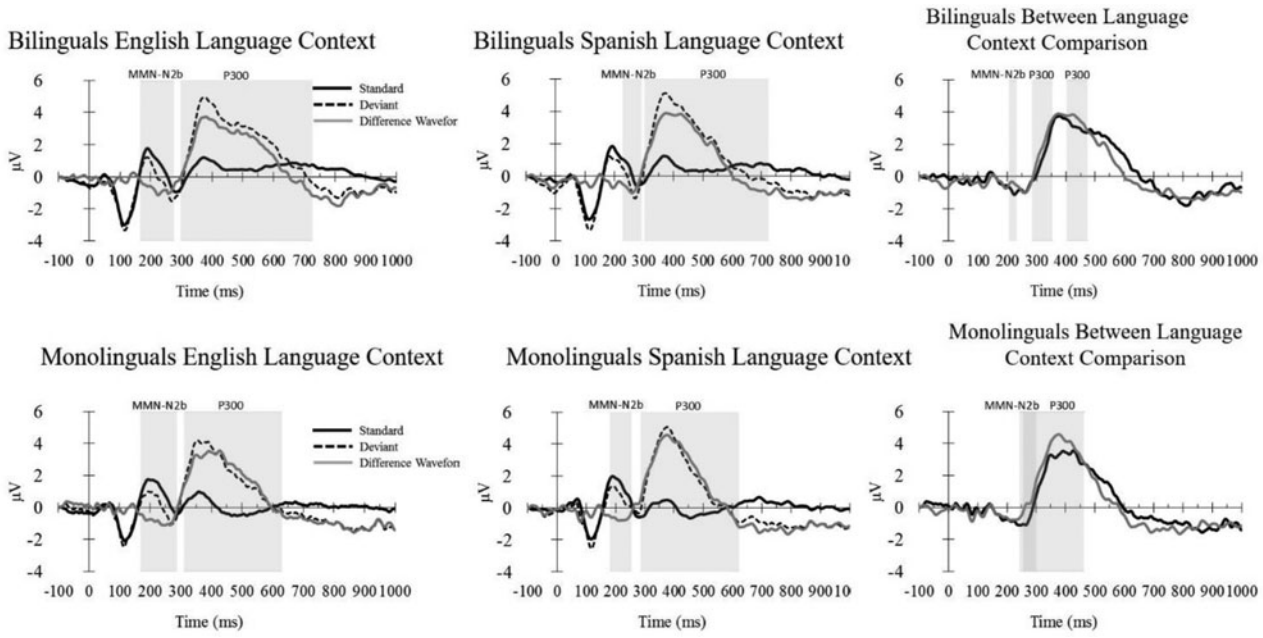


Fig. 2. Bilinguals' and monolinguals' ERPs during both language contexts (left and central panels), and difference waveform comparison across contexts (right panel) from electrode Cz. Gray shaded areas represent the statistically significant time windows as indicated by the two point-by-point analyses in BESA Statistics 2.0 (-100 ms to 300 ms & 250 ms to 1000 ms). Mean amplitudes in the significant time windows revealed when comparing standard and deviant responses were used to calculate the difference waveform (deviant – standard) within each context. Mean amplitudes in the significant time windows revealed when comparing the English context and Spanish context difference waveforms were used to compare the MMN-N2b and P300 across contexts. Overlap among these gray shaded areas, indicated by a darker shade of gray, represents the ERP spillover effect. Positive values are plotted up.

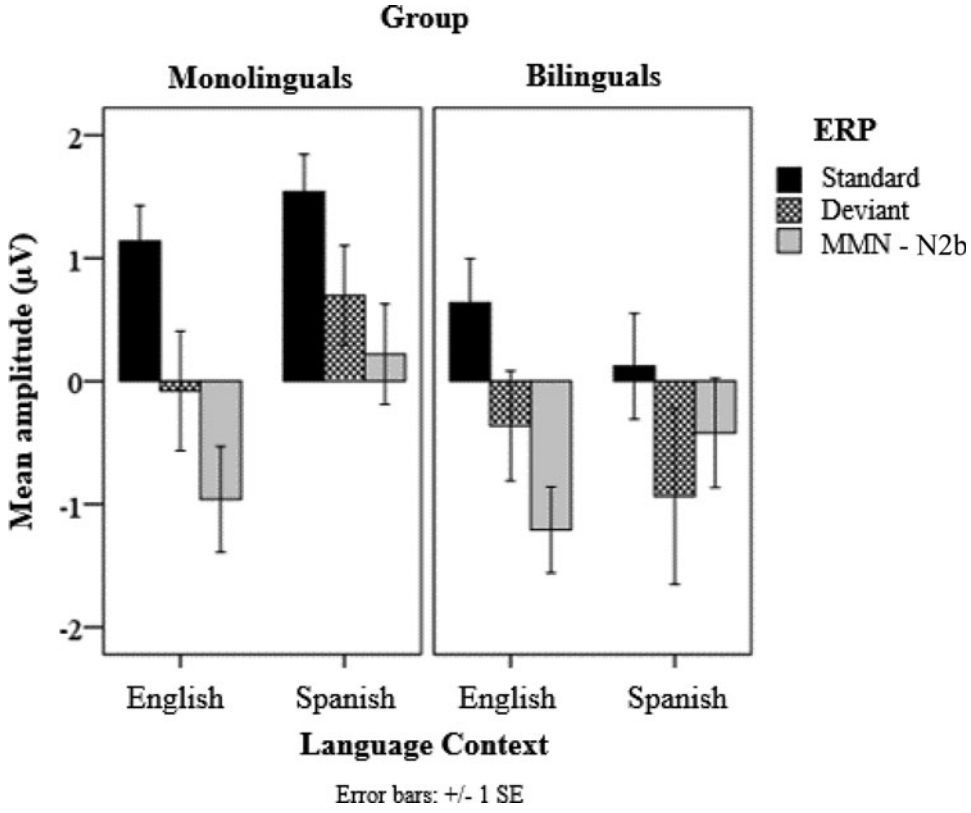


Fig. 3. Monolinguals' and bilinguals' mean amplitude (μV) of standard, deviant, and MMN-N2b responses in both language contexts. Positive amplitudes are plotted up.

et al.'s (1981) findings. Specifically, the language context establishes conceptual expectations towards which perceptual contrasts will be provided, and thus increases perceptual sensitivity of stimuli near or at the predicted phonetic boundary. This complements prior evidence that suggests only bilinguals' perception of speech sounds is influenced by the immediate language context

https://doi.org/10.1017/S1366728920000553 Published online by Cambridge University Press

in a linguistic manner (Antoniou et al., 2012; Casillas & Simonet, 2018; Elman et al., 1977; Flege & Eefting, 1987a; García-Sierra et al., 2009; 2012; Gonzales & Lotto, 2013; Gonzales et al., 2019; Hazan & Boulakia, 1993). Next, we describe how ERPs further reveal the interaction between perceptual and conceptual cues across the cue match (i.e., Spanish) and cue mismatch (i.e., English) contexts.

Electrophysiological responses (ERPs)

We found bilinguals' MMN-N2b to be larger in the English context (i.e., cue mismatch) than the Spanish context (i.e., cue match). In other words, bilinguals' MMN-N2b increased during a cue mismatch relative to a cue match. As framed by the Predictive Coding Hypothesis (Garrido et al., 2009), bilinguals' larger MMN in the English context can be attributed to a larger amount of prediction error and adjustment of conceptual expectations after perceiving the Spanish contrast provided by the perceptual cues. These significant effects were observed in a later MMN time window across both Fz and Cz electrodes without a polarity inversion at the mastoids, which suggests that the N2b overlapped the MMN.

Such an MMN and N2b overlap ideally represents the perceptual-conceptual interaction in speech perception. Explicitly, bilinguals and monolinguals alike can detect the acoustic (i.e., bottom-up) differences between standard and deviant stimuli; thus, eliciting the MMN. But only bilinguals can meaningfully attend to this difference (i.e., top-down knowledge). In line with the Predictive Coding Hypothesis of the MMN (Garrido et al., 2009), establishing a language context that predicts a particular speech contrast demands less attentional deviance detection processes (i.e., smaller N2b) than a language context that does not predict the same contrast (i.e., larger N2b). Simply, more attention is needed when top-down expectations mismatch bottom-up input (i.e., cue mismatch). This aligns with our findings of bilinguals' larger negative-going component in response to a Spanish phonemic contrast in the English context compared to the Spanish context. Further, we see the N2b as primarily responsible for bilinguals' ERP difference across contexts given how both contexts presented the same stimuli, and how top-down expectations are proposed to affect bottom-up deviance detection (i.e., Predictive Coding Hypothesis).

This contrasts previous ERP findings that suggested the language context did not interact with bilinguals' phonetic perception (Winkler, Kujala, Alku & Näätänen, 2003). However, our study differs in ways that would allow us observe such previously suppressed interaction between conceptual and perceptual cues in bilinguals. First, our phonetic contrast can be perceived as only two different sounds across both languages (i.e., /da/ in Spanish or English, and /ta/ in Spanish), whereas Winkler et al.'s (2003) phonetic contrast can be perceived as three different sounds across both languages (i.e., /æ/ or /e/ in Finnish, and /ɛ/ in Hungarian). As a result, Spanish–English bilinguals, unlike Hungarian–Finnish bilinguals, represent the same speech sounds (i.e., /da/) differently in each language (i.e., negative VOT in Spanish vs. short lags in English), and the same phonetic sounds (i.e., short lags) differently in each language (i.e., Spanish /ta/ equals English /da/). Thus, language context may help to facilitate linguistic overlaps. Also, we embedded phonetic variation within isolated CV speech sounds (i.e., /da/ or /ta/), as opposed to words (i.e., /pæti/ or /peti/), as lexical information itself has been shown to influence bilinguals' speech perception aside from the immediate language context (Kirsner et al., 1980; Lauro & Schwartz, 2017).

Spanish–English bilinguals' perceptually cued conceptual adjustment (i.e., larger MMN-N2b in cue mismatch) has been further hidden in prior research by presenting a range of perceptual cues that span the phonetic categories used to contrast voiced from voiceless stop consonants in both languages (i.e., negative VOT to long lags) (Casillas & Simonet, 2018; Elman et al., 1977; Flege & Eefting, 1987a; García-Sierra et al., 2009; 2012; Hazan & Boulakia, 1993). Simply, the perceptual range was not language-specific. As a result, bilinguals' perception could be guided by conceptual expectations provided by the language context, and not be challenged by perceptual cues. Yet, our study presents a restricted perceptual range that fulfills the conceptual expectations in one language context (i.e., Spanish), but challenges the conceptual expectations in the other to reveal the previously hidden perceptual-conceptual conflict in speech perception.

It is also important to note that García-Sierra et al. (2012) observed the opposite MMN pattern in bilinguals, such that bilinguals' MMN was larger when the conceptual cues predicted the provided phonetic contrast. However, this study differed from the current study in 2 ways: 1) the range of perceptual cues, and 2) maintenance of the language context. Given how García-Sierra et al. (2012) only presented 2 speech sounds (i.e., 1 standard and 1 deviant) per language context, their larger MMN reported may reflect how bilinguals' conceptual expectations facilitate phonemic distinction from a fixed acoustic difference, while the larger MMN observed here may instead reflect how bilinguals' perceptual sensitivity adjusts conceptual expectations to then facilitate phonemic distinction. These hypotheses are further supported by the fact that García-Sierra et al. (2012) maintained the language context throughout stimulus presentation (i.e., read a magazine in language of interest), which increases the likelihood of conceptual influence throughout speech perception, while maintenance of the language context (i.e., BFI questionnaire in the language of interest) and stimulus presentations alternated in the current study. Simply, the MMN in the García-Sierra et al. (2012) study reflects a perceptual update in bilinguals' speech perception, while the current study reflects a conceptual update as expected by The Predictive Coding Hypothesis (Garrido et al., 2009).

Further understanding of bilinguals' conceptual update as observed in this study could be benefitted by future investigations comparing data collected in the first half of the study to that of the latter half. As such, we might expect to observe enhanced effects (i.e., larger MMN-N2b in the English context) in the latter half compared to the first half. Regardless, bilinguals have been shown to continuously, and quickly switch language modes (i.e., conceptually update) as a function of language context during speech perception (Casillas & Simonet, 2018).

The current study provides new insight towards the patterns observed in bilinguals' speech perception by being the first to collect ERPs from monolinguals in an active speech perception task across native and nonnative language contexts, as opposed to across native and nonnative speech contrasts (Hisagi et al., 2010). Given how perceptual routines develop as a function of our early environmental exposure and go on to facilitate processing (Jusczyk, 2000), English monolinguals were not expected to show MMN-N2b differences across English and Spanish contexts, unlike Spanish–English bilinguals. This would suggest that bilinguals maintain perceptual sensitivity as a mechanism to manage

https://doi.org/10.1017/S1366728920000553 Published online by Cambridge University Press

the interaction between their languages; a demand monolinguals do not experience.

Although initial analyses did reveal a difference in monolinguals' MMN-N2b across contexts, this difference was the overlap among monolinguals' P300 and MMN-N2b in the Spanish context; an effect not observed in either language context for bilinguals. This prevents a direct comparison of monolinguals' true MMN-N2b trends to that of bilinguals, but still offers new insight into the different patterns that can characterize bilinguals' speech perception. Specifically, given how the P300 is thought to reflect the use of attentional resources (Bonala & Jansen, 2012; Linden, 2005; Johnson, 1988; Picton, 1992), our results suggest that monolinguals and bilinguals did not allocate attentional resources in the same manner.

Perceptual routines may offer an explanation. Explicitly, bilinguals, but not monolinguals, were provided the opportunity to perceive a native phonemic contrast along the VOT continuum. As a result, bilinguals' diverse perceptual routines may have successfully facilitated perceptual distinction (i.e., MMN), attentional awareness (i.e., N2b), and resource allocation (i.e., P300) in both contexts, while monolinguals' lack of appropriate perceptual routines may have prioritized resource allocation in a nonnative context. (i.e., earlier P300 in Spanish context).

In addition, both bilinguals' and monolinguals' larger P300 in the Spanish context can be explained in similar ways. For bilinguals, similar to how a cue match facilitates bilinguals' perceptual sensitivity (i.e., increased cumulative *d'* in the Spanish context), it may also facilitate bilinguals' allocation of attentional resources. Specifically, a cue match may increase the relevance of a respective set of perceptual routines, and thus lead bilinguals to preferentially allocate attentional resources in such contexts. On the other hand, monolinguals' lack of appropriate perceptual routines may have led them to rely more on attentional resources when perceiving speech sounds that do not provide a native phonemic contrast. Consequently, a nonnative context demands more attentional resources relative to a native context, explaining monolinguals' larger P300 in the Spanish context.

Further, the P300 has been shown to be larger in response to self-relevant stimuli, compared to self-irrelevant stimuli. (Berlad & Pratt, 1995; Fishler et al., 1987; Gray et al., 2004; Ninomiya et al., 1998; Onitsuka et al.,1997), in which it is thought that participants preferentially allocate attentional resources to stimuli perceived to be self-relevant. Accordingly, self-relevant stimuli in the current study would be those used for phonemic distinction in a participant's native language. With this perspective, the deviants (i.e., short lags) were the only self-relevant perceptual cues provided for English monolinguals. Thus, monolinguals' larger P300 in the Spanish context perhaps suggests that self-relevant stimuli attract more attentional resources when presented in an increasingly irrelevant context. Specifically, since monolinguals did not preferentially "spend" their attentional resources on self-irrelevant information, the amount of resources available to spend on self-relevant stimuli accumulated. Hence, monolinguals had more attentional resources to allocate towards self-relevant stimuli in a self-irrelevant context, which would explain the observed larger P300 in the Spanish context (Gray et al., 2004).

This contrasts prior research that suggests the immediate language context does not influence monolinguals' speech perception (Elman et al., 1977; García-Sierra et al., 2009; Gonzales & Lotto, 2013), but still in a different manner from bilinguals' speech perception. Specifically, our results suggest that perceptual routines influence the allocation of attentional resources in a speech perception task. Specifically, bilinguals' diverse range of perceptual routines facilitates attentional resources (i.e., bilinguals), whereas monolinguals' restricted range of perceptual routines prioritizes the need for attentional resources.

Conclusions

Results of this study elaborate prior findings showing how perceptual cues influence bilinguals' speech perception (i.e., larger cumulative *d'*), but importantly extend these findings by showing how perceptual and conceptual cues interact (i.e., larger MMN-N2b in English context for bilinguals). This interaction was discovered by presenting the same restricted VOT continuum (i.e., negative VOT to short lags) in one language context that phonemically contrasts these phonetic categories (i.e., Spanish), or does not (i.e., English). Our results support the Predictive Coding Hypothesis (Garrido et al., 2009) in bilinguals' speech perception, and that resource allocation is facilitated when bilinguals' conceptual expectations are initially met by the perceptual cues (i.e., larger P300 in the Spanish context).

Supplementary Material. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1366728920000553

Acknowledgements. The amount of effort put in developing this paper is comparable between the first and second author. The authors thank Dr. Nairán Ramírez-Esparza for her valuable input.Notes

References

Aaltonen O, Niemi P, Nyrke T and Tuhkanen M (1987) Event-related potentials and the perception of a phonetic continuum. *Biological Psychology* **24**, 197–207.

Abramson AS and Lisker L (1967) Discriminability along the voicing continuum: Cross language tests. Paper presented at the Proc. 6th International Congress of Phonetic Sciences, Prague.

Abramson AS and Lisker L (1972) Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics* **1**, 1–8.

Antoniou M, Tyler MD and Best CT (2012) Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics* **40**, 582–594.

Benet-Martínez V and John OP (1998) Los Cinco Grades across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology* **75**, 729–750.

Berlad I and Pratt H (1995) P300 in response to the subject's own name. *Electroencephalography and Clinical Neurophysiology* **96**, 472–474.

Brady SA and Darwin CJ (1978) A range effect in the perception of voicing. *Journal of the Acoustical Society of America* **63**, 1556–1558.

Bohn OS and Flege JE (1993) Perceptual Switching in Spanish–English Bilinguals. *Journal of Phonetics* **21**, 267–290.

Bonala B and Jansen B (2012) A computational model for generation of the P300 evoked potential component. *Journal of Integrative Neuroscience* **11**, 277–294.

Bullmore ET, Suckling J, Overmeyer S, Rabe-Hesketh S, Taylor E and Brammer MJ (1999) Global, Voxel, and Cluster Tests, by Theory and Permutation, for a Difference Between Two Groups of Structural MR Images of the Brain. *IEEE Transactions on Medical Imaging* **18**, 32–42.

Casillas JV, Díaz Y and Simonet M (2015) Acoustic of Spanish and English Coronal Stops. In The Scottish Consortium for ICPhS 2015 (ed), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow.

Casillas JV and Simonet M (2018) Perceptual categorization and bilingual language modes: Assessing the double phonemic boundary in early and late bilinguals. *Journal of Phonetics* **71**, 51–64.

Datta H, Shafer VL, Morr ML, Kurtzberg D and Schwartz RG (2010) Electrophysiological Indices of Discrimination of Long-Duration, Phonetically

https://doi.org/10.1017/S1366728920000553 Published online by Cambridge University Press

Similar Vowels in Children With Typical and Atypical Language Development. *Journal of Speech, Language, and Hearing Research* **53**, 757–777.

Diehl RL, Elman JL and McCusker SB (1978) Contrast effects on stop consonant identification. *Human Perception and Performance* **4**, 599–609.

Diesch E and Luce T (1997a) Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation. *Psychophysiology* **34**, 501–510.

Diesch E and Luce T (1997b) Magnetic mismatch fields elicited by vowels and consonants. *Experimental Brain Research* **116**, 139–152.

Dmitrieva O, Llanos F, Shultz AA and Francis A (2015) Phonological status, not voice onset time, determines the acoustic realization of onset F0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics* **49**, 77–95.

Donchin E (1981) Surprise!…Surprise? *Psychophysiology* **18**, 493–513.

Eimas PD and Corbit JD (1973) Selective adaptation of linguistic feature detectors. *Cognitive Psychology* **4**, 99–109.

Elman JL, Diehl RL and Buchwald SE (1977) Perceptual switching in bilinguals. *Journal of the Acoustical Society of America* **62**, 971–974.

Ernst MD (2004) Permutation Methods: A Basis for Exact Inference. *Statistical Science* **19**, 676–685.

Fish MS, García-Sierra A, Ramírez-Esparza N and Kuhl PK (2017) Infant-directed speech in English and Spanish: Assessments of monolingual and bilingual caregiver VOT. *Journal of Phonetics* **63**, 19–34.

Fishler I, Jin Y, Boaz T, Perry N and Childers D (1987) Brain potentials related to seeing one's own name. *Brain and Language* **30**, 245–262.

Flege JE (1982) Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics* **10**, 177–192.

Flege JE and Eefting W (1987a) Cross-Language Switching in Stop Consonant Perception and Production by Dutch Speakers of English. *Speech Communication* **6**, 185–202.

Flege JE and Eefting W (1987b) Production and Perception of English Stops by Native Spanish Speakers. *Journal of Phonetics* **15**, 67–83.

García-Sierra A, Diehl RL and Champlin C (2009) Testing the double phonemic boundary in bilinguals. *Speech Communication* **51**, 369–378.

García-Sierra A, Ramírez-Esparza N and Kuhl PK (2016) Relationships between quantity of language input and brain responses in bilingual and monolingual infants. *International Journal of Psychophysiology* **110**, 1–17.

García-Sierra A, Ramírez-Esparza N, Silva-Pereyra J, Siard J and Champlin CA (2012) Assessing the double phonemic representation in bilingual speakers of Spanish and English: An electrophysiological study. *Brain and Language* **121**, 194–205.

García-Sierra A, Rivera-Gaxiola M, Percaccio CR, Conboy BT, Romo H, Klarman L, Ortiz S and Kuhl PK (2011) Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics* **39**, 546–557.

García-Sierra A, Schifano E, Duncan GM and Fish MS (in press) An analysis of the perception of stop consonants in bilinguals and monolinguals in two different phonetic contexts: a range-base language cueing approach. *Action, Perception, & Psychophysics*.

Garrido MI, Kilner JM, Stephan KE and Friston KJ (2009) The mismatch negativity: A review of underlying mechanisms. *Clinical Neuropsychology* **120**, 453–463.

Gonzales K, Byers-Heinlein K and Lotto AJ (2019) How bilinguals perceive speech depends on which language they think they're hearing. *Cognition* **182**, 318–330.

Gonzales K and Lotto AJ (2013) A bafri, un pafri: Bilinguals' Pseudoword Identifications Support Language-Specific Phonetic Systems. *Psychological Science* **24**, 2135–2142.

Gray HM, Ambady N, Lowenthal WT and Deldin P (2004) P300 as an index to self-relevant stimuli. *Journal of Experimental Social Psychology* **40**, 216–224.

Groppe DM, Urbach TP and Kutas M (2011) Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology* **48**, 1711–1725.

Hazan VL and Boulakia G (1993) Perception and Production of a Voicing Contrast by French-English Bilinguals. *Language and Speech* **36**, 17–38.

Hay JF (2005) How auditory discontinuities and linguistic experience affect the perception of speech and non-speech in English- and Spanish-speaking listeners (Doctoral dissertation). Available from Proquest Dissertations and Theses database. (UMI No. 3203519).

Hisagi M, Shafer VL, Strange W and Sussman ES (2010) Perception of a Japanese vowel length contrast by Japanese and American English listeners: Behavioral and electrophysiological measures, *Brain Research* **1360**, 89–105.

Holt LL (2005) Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science* **16**, 305–312.

Holt LL and Lotto AJ (2002) Behavioral examinations of the level of auditory processing of speech context effects. *Hearing Research* **167**, 156–169.

John OP and Srivastava S (1999) The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In Pervin LA & John OP (eds), *Handbook of personality: Theory and research* (2nd Ed), New York, NY: Guilford Press, pp. 102–138.

Johnson R (1988) Scalp-recorded P300 Activity in Patients Following Unilateral Temporal Lobectomy, *Brain* **111**, 1517–1529.

Ju M and Luce PA (2004) Falling on Sensitive Ears: Constraints on Bilingual Lexical Activation. *Psychological Science* **15**, 314–318.

Jusczyk PW (2000) *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

Karmiloff-Smith A (2010) Multiple Trajectories to Human Language Acquisition: Domain-Specific or Domain-General? *Human Development* **53**, 239–244.

Keating PA, Mikos MJ and Ganong WF (1981) A Cross-Language Study of Range of Voice Onset Time in the Perception of Initial Stop Voicing. *Journal of the Acoustical Society of America* **70**, 1261–1271.

Kirsner K, Brown HL, Abrol S, Chadha K and Sharma K (1980) Bilingualism and Lexical Representation. *Quarterly Journal of Experimental Psychology* **32**, 585–594.

Klatt DH (1980) Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America* **67**, 971–990.

Kuhl PK, Williams KA, Lacerda F, Stevens KN and Lindblom B (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* **255**, 606–608.

Lagrou E, Hartsuiker RJ and Duyck W (2011) Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **37**, 952–965.

Lauro JG and Schwartz AI (2017) Bilingual non-selective lexical access in sentence contexts: A meta-analytic review. *The Journal of Memory and Language* **92**, 217–233.

Linden D (2005) The P300: Where in the brain is it produced and what does it tell us? *The Neuroscientist* **11**, 563–576.

Lotto AJ, Sullivan SC and Holt LL (2003) Central locus for nonspeech context effects on phonetic identification. *Journal of the Acoustical Society of America* **113**, 53–56.

Macmillan NA and Creelman CD (2005) *Detection Theory: A User's Guide* (Second ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Luk G and Bialystok E (2013) Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology* **25**, 605–621.

Maris E (2012) Statistical testing in electrophysiological studies. *Psychophysiology* **49**, 549–565.

Maris E and Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* **164**, 177–190.

Näätänen R (1982) Processing Negativity - an Evoked-Potential Reflection of Selective Attention. *Psychological Bulletin* **92**, 605–640.

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

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, Gaillard AWK and Mantysalo S (1978) Early selective attention effect on evoked potential reinterpreted. *Acta Psychologica* **42**, 313–329.

Näätänen R and Michie PT (1979) Early Selective-Attention Effects on the Evoked-Potential - a Critical-Review and Reinterpretation. *Biological Psychology* **8**, 81–136.

Ninomiya H, Onitsuka T, Chen C and Sato E (1998) P300 in response to subject's own face. *Psychiatry and Clinical Neurosciences* **52**, 519–522.

https://doi.org/10.1017/S1366728920000553 Published online by Cambridge University Press

Novak G, Ritter W and Vaughan Jr., HG (1992) Mismatch Detection and the Latency of Temporal Judgments. *Psychophysiology* **29**, 398–411.

Onitsuka T, Ninomiya H, Chen C and Sato E (1997) P300 in response to subject's own voice. *Japanese Journal of EEG EMG* **25**, 243–248.

Pastore RE, Ahroon WA, Baffuto KJ, Friedman C, Puleo JS and Fink EA (1977) Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance* **3**, 686–696.

Picton TW (1992) The P300 Wave of the Human Event-Related Potential. *Journal of Clinical Neuropsychology* **9**, 456–479.

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

Polich J (1988) Bifurcated P300 peaks: P3a and P3b revisited? *Journal of Clinical Neurophysiology* **5**, 287–294.

Polich J (2003) Theoretical overview of P3a and P3b. In: Polich J (eds) *Detection of Change*. Springer, Boston, MA.

Polich J (2007) Updating P300: an integrative theory of P3a and P3b. *Journal of Clinical Neurophysiology* **118**, 2128–2148.

Ramírez-Esparza N, García-Sierra A and Kuhl PK (2014) Look who's talking: speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science* **17**, 880–891.

Ramírez-Esparza N, García-Sierra A and Kuhl PK (2017) The Impact of Early Social Interactions on Later Language Development in Spanish–English Bilingual Infants. *Child Development* **88**, 1216–1234.

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

Rivera-Gaxiola M, Johnson MH, Csibra G and Karmiloff-Smith A (2000b) Electrophysiological correlates of category goodness. *Behavioral Brain Research* **112**, 1–11.

Sharma A and Dorman MF (1998) Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential. *Journal of the Acoustical Society of America* **104**, 511–517.

Streeter LA (1976) Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature* **259**, 39–41.

Sundara M, Polka L and Baum S (2006) Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition* **9**, 97–114.

Sundara M and Polka L (2008) Discrimination of coronal stops by bilingual adults: The timing and nature of language interaction. *Cognition* **106**(1), 234–258.

Sussman E, Kujala T, Halmetoja J, Lyytinen H, Alku P and Näätänen R (2004) Automatic and controlled processing of acoustic and phonetic contrasts. *Hearing Research* **190**, 128–140.

Sussman E and Steinschneider M (2009) Attention effects on auditory scene analysis in children. *Neuropsychologia* **47**, 771–785.

Sutton S, Braren M, Zubin J and John ER (1965) Evoked-Potential Correlates of Stimulus Uncertainty. *Science* **150**, 1187–1187.

Werker J (2012) Perceptual foundations of bilingual acquisition in infany. *Annals of the New York Academy of Sciences* **1251**, 50–61.

Williams L (1977) The perception of stop consonant voicing by Spanish–English bilinguals. *Perception & Psychophysics* **21**, 289–297.

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

Winkler I, Kujala T, Alku P and Näätänen R (2003) Language context and phonetic change detection. *Cognitive Brain Research* **17**, 833–844.

Winkler I, Kujala T, Tiitinen P, Sivonen P, Alku P, Lehtokoski A, Czigler I, Csépe V, Ilmoniemi RJ and Näätänen R (1999) Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* **36**, 638–642.

https://doi.org/10.1017/S1366728920000553 Published online by Cambridge University Press