Linguistic effect on speech perception observed at the brainstem

T. Christina Zhao[a] and Patricia K. Kuhl[a,1]

[a]Institute for Learning & Brain Sciences, University of Washington, Seattle, WA 98195

Contributed by Patricia K. Kuhl, May 10, 2018 (sent for review January 5, 2018; reviewed by Bharath Chandrasekaran and Robert J. Zatorre)

Linguistic experience affects speech perception from early infancy, as previously evidenced by behavioral and brain measures. Current research focuses on whether linguistic effects on speech perception can be observed at an earlier stage in the neural processing of speech (i.e., auditory brainstem). Brainstem responses reflect rapid, automatic, and preattentive encoding of sounds. Positive experiential effects have been reported by examining the frequency-following response (FFR) component of the complex auditory brainstem response (cABR) in response to sustained high-energy periodic portions of speech sounds (vowels and lexical tones). The current study expands the existing literature by examining the cABR onset component in response to transient and low-energy portions of speech (consonants), employing simultaneous magnetoencephalography (MEG) in addition to electroencephalography (EEG), which provide complementary source information on cABR. Utilizing a cross-cultural design, we behaviorally measured perceptual responses to consonants in native Spanish- and English-speaking adults, in addition to cABR. Brain and behavioral relations were examined. Results replicated previous behavioral differences between language groups and further showed that individual consonant perception is strongly associated with EEG-cABR onset peak latency. MEG-cABR source analysis of the onset peaks complimented the EEG-cABR results by demonstrating subcortical sources for both peaks, with no group differences in peak locations. Current results demonstrate a brainstem–perception relation and show that the effects of linguistic experience on speech perception can be observed at the brainstem level.

speech perception | brainstem response | linguistic experience | MEG/EEG

Early language experience profoundly affects how speech sounds are perceived throughout life (1). Transient consonants present the most challenging and intriguing case as their perception relies on neural encoding and processing of extremely brief, highly dynamic, and yet often weak acoustic signals. The smallest change in the signal (e.g., a few milliseconds increase in a pause) at a critical point in the stimulus can result in a drastic change in perception (e.g., from "bat" to "pat"), a phenomenon referred to as "categorical perception" (CP) for speech perception (2). For decades, research has repeatedly demonstrated that these critical positions for acoustic change (i.e., categorical boundaries) differ for speakers of different language backgrounds (3–6).

The influence of linguistic experience begins very early in development (7). A rich literature has documented that by 12-mo of age, infants' ability to discriminate nonnative consonants already begins to decline while the ability to discriminate native consonants improves (8–10). For adults, discrimination of nonnative consonants remains far from "native-like," even after periods of intensive training (11–13).

Research focusing on the cortical level of neural processing has found evidence to corroborate perceptual data. For example, the mismatch response (MMR), indexing neural sensitivity to speech sound changes at the cortical level, has been observed to be larger in speakers of a language in which the speech contrast is native than in speakers of a language in which the contrast was nonnative (14). In infants, MMRs to nonnative contrasts at 11 mo of age were observed to be significantly reduced compared with MMRs observed at 7 mo age (15).

The current study set out to examine whether the effects of linguistic experience on consonant perception already manifest at a deeper and earlier stage than the cortex, namely, at the level of auditory brainstem. The auditory brainstem encodes acoustic information and relays them to the auditory cortex. Its activity can be recorded noninvasively at the scalp using a simple setup of three electroencephalographic (EEG) electrodes (16). Using extremely brief stimuli such as clicks or tone pips, the auditory brainstem response (ABR) can be elicited and has been used clinically to assess the integrity of the auditory pathway (17). More recently, researchers began recording auditory brainstem responses to complex sounds, such as speech and musical tones [i.e., complex ABR (cABR)] with the same EEG setup (18). The cABR consists of an onset component (onset) and a frequency-following response (FFR) component (19). The onset reflects the encoding of transient changes in the signal (e.g., the consonant) while the FFR reflects the brainstem's tracking of sustained periodic information (i.e., fundamental frequency and higher harmonics in vowels and tones) (20). Thus far, most studies have focused on the FFR to vocalic portions of speech sounds, which contain high levels of acoustic energy (e.g., lexical tones and vowels). Early language experience has been observed to enhance FFR to lexical tones and vowels

Significance

Early linguistic experience affects perception and cortical processing of speech, even in infants. The current study examined whether linguistic effects extend to brainstem speech encoding, where responses are rapid, automatic, and preattentive. We focused on transient consonants and measured perception behaviorally and the corresponding complex auditory brainstem response (cABR) onset using simultaneous electroencephalography/magnetoencephalography (EEG/MEG) in native Spanish and English speakers. We demonstrate that the latency of an EEG-cABR onset peak predicts consonant perception and that the perception differs according to language background. MEG-cABR analysis demonstrates deep sources for the onset, providing complimentary support for a brainstem origin. Effects of early linguistic experience on speech perception can be observed at the earlier stage of speech encoding at the brainstem.

Author contributions: T.C.Z. and P.K.K. designed research; T.C.Z. performed research; T.C.Z. analyzed data; and T.C.Z. and P.K.K. wrote the paper.

Reviewers: B.C., University of Texas at Austin; and R.J.Z., Montreal Neurological Institute, McGill University.

The authors declare no conflict of interest.

Data deposition: Data and analysis pipeline are available through Open Science Framework (osf.io/6fwxd).

[1]To whom correspondence should be addressed. Email: pkkuhl@u.washington.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1800186115/-/DCSupplemental.

Published online August 13, 2018.

(21–23). Similarly, researchers have demonstrated that early music training experience enhances adults' FFR to nonnative lexical tones (24–26). To date, only one study examines directly the relations between vowel perception and the FFR component, suggesting that FFRs to vowels reflect the acoustic signal of the vowels, rather than the perception of them (27). However, the underlying source of FFR remains a topic of active debate (19, 28).

To examine transient consonant encoding at the brainstem level, we differed from previous studies and focused on the onset component of cABR as a function of perception and linguistic experience. This approach allows us to address several important gaps in the literature and therefore can advance our theoretical understanding of the effects of linguistic experience on speech perception, reflected deep at the brainstem level. First, we increased the confidence that any effects observed in the current study are predominantly subcortical in nature through two means: (*i*) by examining the EEG-cABR onset to transient stop consonants (<20 ms) and (*ii*) by using simultaneous magnetoencephalography (MEG)/EEG recordings. The early nature of the onset response ensures a predominant subcortical source. Simultaneous MEG recordings complimented the EEG-cABR by providing source information for the onset peaks. Dipole modeling methods revealed subcortical sources for the onset peaks, corroborating previous research (19) and supporting a brainstem origin for the EEG-cABR onset. Second, by gathering both perception data and cABR from the same individual, the current study is able to examine the brainstem–perception relation.

Specifically, we tested two hypotheses that (*i*) individuals' EEG-cABR onset in response to consonants is related to their perception of the sounds, and (*ii*) both the EEG-cABR onset and perception are affected by listeners' language background. Further, using simultaneously measured MEG-cABR, we provide complimentary information to EEG-cABR regarding the sources of the onset peaks.

Monolingual English speaking adults (n = 29) and native Spanish speaking adults (n = 20) were recruited to first complete computer-based identification and discrimination tasks to assess their categorical perception of bilabial stop consonants varying on the voice-onset-time (VOT) continuum (for details see *Materials and Methods*). The VOT continuum ranged from −30 ms to +30 ms in 10-ms increments. In the identification task, participants judged whether the speech sound is /ba/ or /pa/ through speeded key presses. The percent identification of a sound on the continuum as /pa/ is calculated as the outcome measure. In the discrimination task, participants judged whether two speech sounds were the same or different. The sensitivity indices (d′) for pairs of adjacent sounds along the continuum were calculated as the outcome measure to minimize response bias (29). In line with previous research, monolingual English speakers demonstrated significantly different categorical perception for the VOT continuum compared with native Spanish speakers (Fig. 1 and *SI Appendix*, Fig. S1). Even though the language background in the native Spanish speaker group is much wider in range (*SI Appendix*, Table S1), the categorical perception was more variable in English speakers, as the VOT cue is not the only and may not be the predominant cue for the voiced-voiceless (/b/-/p/) contrast in some English speakers (e.g., aspiration) (30). To maximize the detection of differences at the brainstem level, we selected participants whose perception results suggested strong weighting of the VOT cue for further MEG/EEG recording. The selection criteria were: (*i*) strong evidence of a category boundary from the identification or discrimination task and (*ii*) a phonetic boundary aligned with previous studies (< +10 ms for native Spanish speakers and > +10 ms for monolingual English speakers).

In the simultaneous MEG/EEG recording sessions, participants' brainstem response to the stimulus with +10 ms VOT (p10 from here on, Fig. 2*A*), where maximal difference in perception between groups was observed, was recorded in two blocks (3,000 trials per block). Three EEG electrodes were placed in accordance with cABR recording conventions (CZ, ground on forehead and reference on right ear lobe). Recordings were done while participants were sitting under the MEG dewar watching silent videos. The stimuli were presented in alternating polarities through insert earphones (*Materials and Methods*). Individuals who were selected for MEG/EEG recordings and completed successful MEG/EEG recordings were included in the analysis: fourteen monolingual English speakers and fifteen native Spanish speakers. The two groups did not differ in their music training background (subject details in *Materials and Methods*).

The MEG/EEG data were preprocessed, filtered, and averaged (across trials and then blocks) to increase the signal-to-noise ratio. The cABR responses were extracted for each participant in both the EEG and the MEG channels (details in *Materials and Methods*). In Fig. 2 *B–D*, an example from an individual's recording for both MEG and EEG is shown to demonstrate the data quality and also that the brainstem responses can be observed on multiple MEG channels. In Fig. 2*E*, dipole analysis from one example individual with existing MRI shows MEG onset peaks are indeed localized to subcortical sources.

To address our research questions, two analyses were done on the EEG-cABR data: (*i*) the onset peaks of EEG-cABR (magnitudes and latencies) corresponding to the stop consonant (peak I for initial noise burst and peak II for voice onset, Fig. 3) were compared between the two groups with different linguistic backgrounds. We hypothesized that because VOT is a timing cue, a group difference would likely lie in the latency of response. (*ii*) The onset peaks in EEG-cABR were examined in relation to individuals' behavioral results on p10 perception (Fig. 4). In addition, source information for the peaks (Fig. 5*A*) was provided through the equivalent current dipole (ECD) analyses on the MEG-cABR data. Dipole locations for both peaks were compared between groups (Fig. 5*B*). We expected deep sources for both peaks (in support of EEG brainstem response) for both groups and that the dipole source locations would not differ between groups.

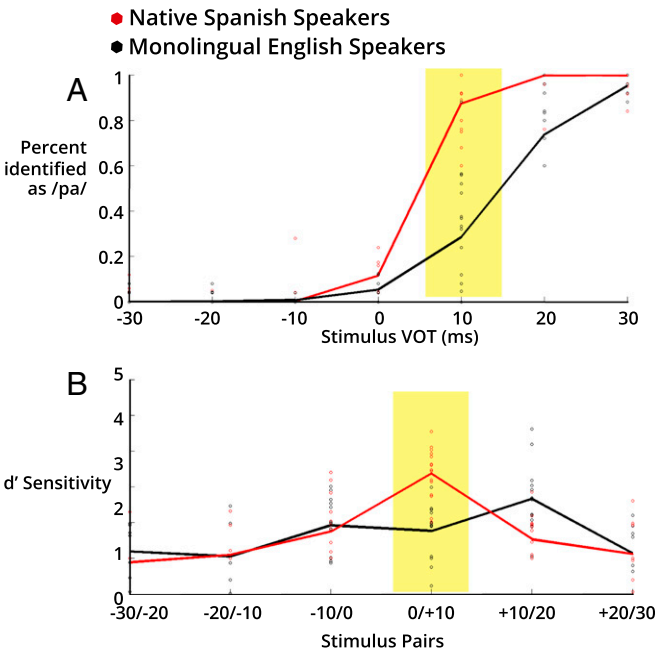

Fig. 1. Categorical perception of bilabial stop consonants on a VOT continuum in monolingual English speakers (black) and native Spanish speakers (red) from the final sample. (*A*) Data from the identification task. Shaded area for stimulus with +10 ms VOT (p10), highlights the range of behavior across the two groups. (*B*) Data from the discrimination task. Shaded area for stimulus pair (0/+10 ms VOT) highlights the range of behavior across the two groups.

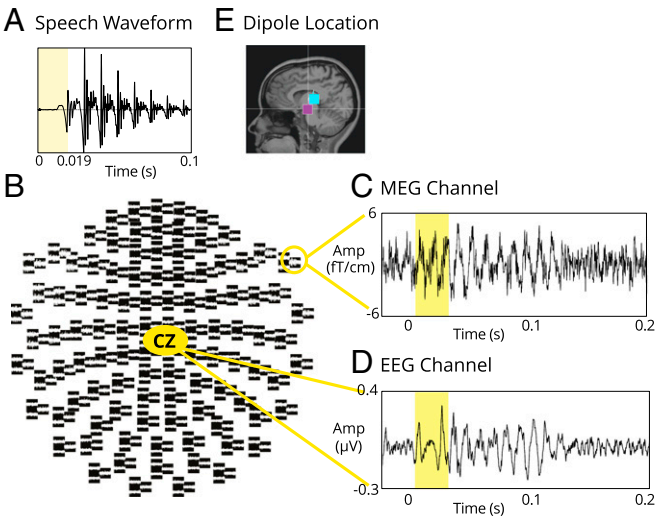

Fig. 2. (*A*) Waveform of the p10 stimulus. Shaded area represents the stop consonant portion. (*B*) MEG channel distribution with the EEG-CZ channel. (*C*) cABR to the p10 stimulus from one representative MEG channel from one individual. Shaded area corresponds to the onset component to the stop consonant. (*D*) cABR to the same stimulus simultaneously recorded by the EEG channel in the same individual. Shaded area corresponds to the onset component to the stop consonant. (*E*) Dipole locations from one individual demonstrated brainstem sources for onset peaks (magenta, peak I; cyan, peak II).

Results

Behavioral results from all participants are shown in *SI Appendix*, Fig. S1. Significant group differences on the perception of p10 were confirmed. Similarly, behavioral results from the final sample included in the study are shown in Fig. 1. As expected, the effects in the final sample were the same but larger due to the selection process. Native Spanish speakers identified the p10 syllable significantly more strongly as "pa" (mean pa identification = 0.85 ± 0.13) than monolingual English speakers [mean pa identification = 0.35 ± 0.24], t (1, 27) = 6.96, $P < 0.001$, Bonferroni correction applied, d = 2.59 (Fig. 1*A*, shaded area). Spanish speakers also discriminated stop consonants with 0 and +10 ms VOT more accurately (mean d′ = 2.37 ± 0.72) than English speakers [mean d′ = 0.76 ± 0.74], t (1, 27) = 5.90, $P < 0.001$, Bonferroni correction applied, d = 2.21 (Fig. 1*B*, shaded area).

For the EEG-cABR data, we first extracted the magnitude and latency values for the first two peaks (Fig. 3*A*, peak I and peak II) in the response. Peak I is in response to the initial noise burst while peak II is in response to the voice onset of the stop consonant (Fig. 2*A*, shaded area). Independent t tests were conducted to compare the magnitudes and latency values between the two groups.

For peak I, independent t tests revealed no significant difference in either magnitude or latency between the two groups [magnitude: t (1, 27) = −1.13, latency: t (1, 27) = 0.117] (Fig. 3). For peak II, independent t tests revealed no significant difference in magnitude, t (1, 27) = 1.69, but a significant difference in latency, t (1, 27) = −3.98, $P < 0.001$, d = 1.45 (Fig. 3). That is, and as predicted, native Spanish speakers exhibited a later peak II (latency = 30.6 ms ± 0.36 ms) than English speakers (latency = 29.69 ms ± 0.81 ms).

Further, we conducted correlation analyses between individuals' peak II latencies and their behaviorally measured perceptual data to examine the brain–behavior relation. Significant regression models were observed between peak II latency and p10 identification (R^2 = 0.323, β = 0.569, P = 0.001, Fig. 4*A*), as well as between peak II latency and d′ scores measuring discrimination ability for the +10-ms/0-ms syllable pair (R^2 = 0.209, β = 0.457, P = 0.013, Fig. 4*B*). That is, the longer the peak II latency in a participant's cABR onset, the more strongly the participant identified the stimulus as /pa/ and the higher the participant's discrimination of the stimulus from its neighboring stimulus on the continuum (VOT = 0 ms).

From the MEG-cABR data, we obtained source information on the onset peaks to compliment the EEG-cABR by using the ECD method (*Materials and Methods*). A spherical head model was used with all magnetometers for dipole fitting without MRI structural scans. Specifically, the fitted ECD coordinates for each peak were extracted and examined to address: (*i*) whether the sources of these two peaks suggest deep sources, approximating brainstem origins; and (*ii*) whether there were differences in the sources of these peaks between groups. The group averaged topographies around Peak I and Peak II time points and ECD locations for both peaks can be visualized in Fig. 5. The topographies demonstrate field patterns with deep sources (i.e., activities along the edge, Fig. 5*A*) instead of cortical sources (i.e., dipolar patterns in each hemisphere, *SI Appendix*, Fig. S2), identified in previous studies (19, 31). For peak I, no significant differences were observed for x, y, or z coordinates [x: t (1, 27) = 0.21, y: t (1, 27) = −0.27, z: t (1, 27) = −0.06]. For peak II, no significant differences were observed either for x, y, or z coordinates [x: t (1, 27) = 0.16, y: t (1, 27) = −0.65, z: t (1, 27) = −0.86].

Discussion

The current study demonstrated that the effects of early linguistic experience on transient consonant perception can be

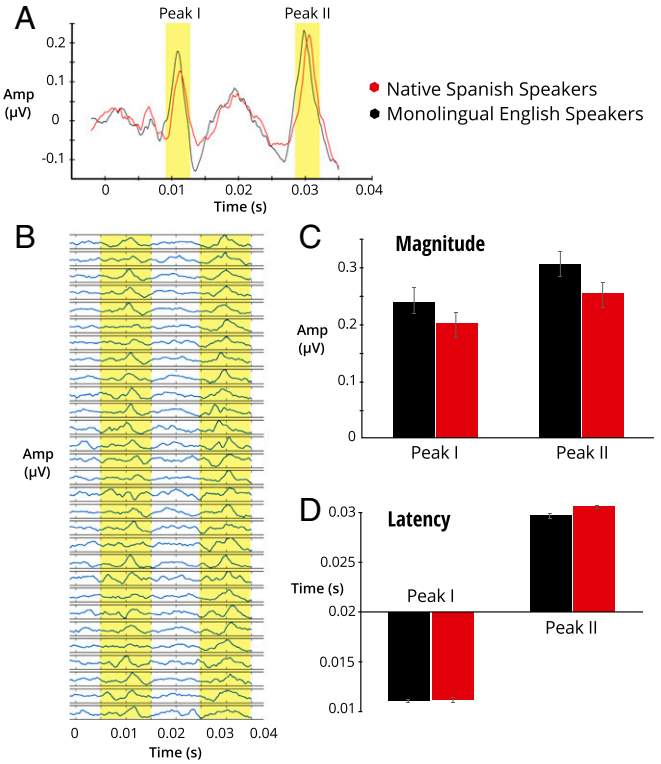

Fig. 3. (*A*) Group average of the onset response from the EEG channel for p10 stimulus. Shaded area indicates the peaks of interest. Peak I represents brainstem response to the initial noise burst. Peak II represents brainstem response to the voice onset. (*B*) Individual onset responses from the final sample; shaded area indicates time window for peak extraction. (*C*) Group averages for peak I and peak II magnitude. No statistical difference was observed between groups for either peak. (*D*) Group averages for peak I and peak II latency. Native Spanish speakers exhibited a significantly longer latency for peak II, but not peak I.

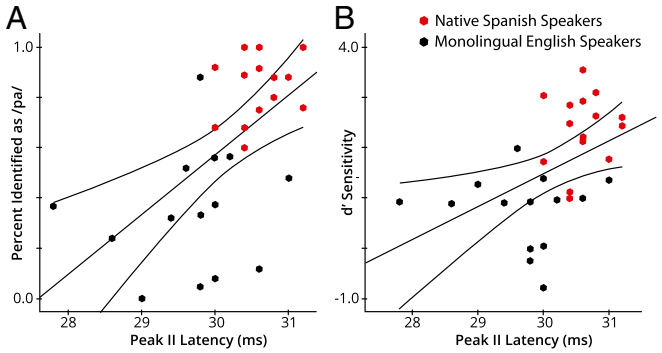

Fig. 4. Relations between peak II latency and perception. (*A*) Scatterplot between peak II latency for p10 stimulus in cABR and percent identified p10 as /p/. (*B*) Scatterplot between peak II latency for p10 stimulus in cABR and d′ sensitivity to discriminate between p10 and stop consonant with 0 ms VOT (+0/+10 ms VOT pair). Significant models were observed for both regressions.

observed at the auditory brainstem encoding level. By targeting transient consonant encoding, reflected by the onset of cABR, we addressed important gaps in the literature. Most studies have focused on examining the FFR component of cABR in response to sustained, high-energy portions of speech (e.g., vowels). Direct brainstem–perception relations have been rarely investigated. In the current study, by examining the EEG-cABR onset to transient consonants, we minimized the effects of cortical contributions, and by simultaneously recording MEG data, we complimented the EEG-cABR results by providing source information of the onset peaks through dipole modeling methods.

Adult speakers from different linguistic backgrounds (monolingual English speakers vs. native Spanish speakers) were examined both behaviorally, using classic speech perception tasks, and at the auditory brainstem level using simultaneous MEG/EEG recordings of cABR. We tested the idea that individuals' EEG-cABR onsets in response to consonants are related to their perception, and that both onset responses and perception are clearly differentiated by the language background of the listener. Further, we provided complimentary information to EEG-cABR regarding the source of the onset peaks using MEG-cABR data.

Native Spanish speakers exhibited a significantly later peak II in the EEG-cABR onsets in response to a bilabial stop consonant (p10), compared with the monolingual English speakers. This peak corresponds to the onset of voicing in this stop consonant. Significant relations were observed between the individual latency value for peak II and listeners' perception of this stop consonant (p10). That is, across language backgrounds, the longer the latency for peak II, the more strongly the speaker identified the stop consonant as /p/, and the better the speaker discriminated the stimulus from a neighboring stop consonant on the VOT continuum (VOT = 0 ms). Simultaneous MEG recordings allowed access to the source information through ECD modeling methods. The onset peaks were localized to deep sources, complementing the EEG-cABR and providing further support for a brainstem origin. As predicted, no differences between groups in the source locations were observed for either peak.

Taken together, our current results strongly support the idea that the effects of early linguistic experience on transient consonant perception can be observed at a very early stage of auditory processing (i.e., the brainstem level). Specifically, brainstem encoding is strongly associated with the perception of the consonants measured behaviorally. Both onset responses and perception can be separated by the language background of the listener, suggesting brainstem speech encoding is experience related and behaviorally relevant.

The current study is consistent with previous research suggesting early experiential effects on the FFR component of cABR, a component that reflects the encoding of periodic information in complex sounds, such as fundamental frequency of voice or tones (18, 20). However, the origin of the FFR has been debated (19, 28). By examining the early EEG-cABR onsets to a transient stop consonant in relation to perception, along with the ECD modeling from MEG-cABR data, the current study extends the existing literature and provides additional and stronger evidence to support the theoretical framework that the early stage of speech sound encoding at the brainstem level may be modulated by long-term experience through a descending corticofugal system (25). The corticofugal pathway originates from cortex and can extend to as early as the auditory periphery (cochlea) (32). For example, it has been demonstrated in animals that short-term laboratory-controlled experience can affect auditory frequency tuning curves at the auditory cortex, brainstem, midbrain, as well as in the hair cells in the cochlea (33, 34). In humans, at the onset of hearing around 24–25 wk of gestation, complex and dynamic speech sounds become audible (35, 36), and short-term in utero auditory experience is already manifested by the time of an infant's birth (37–39). Language experience is therefore one of the longest and most profound experiences humans have, making the current results compatible with a corticofugal mechanism. At the same time, the current study provides a rare opportunity to examine a brain–behavior relation when both measurements can be transformed onto the same scale (i.e., time). While numerous studies have established that a small change in the neural signal (e.g., microvolts in EEG amplitude) can be predictive of a behavioral shift at a larger scale (e.g., perception and/or cognitive skills) (e.g., refs. 40 and 41), it is difficult to speculate the exact "amplification mechanism." In the current study, the "cross-over" point in category perception can also be calculated (i.e., where identification is 50% on a psychometric function fitted on each individual's data), indicating the stimulus position where perception changes from /ba/ to /pa/. The result shows that a range of 3.4 ms in the brainstem peak II latency across subjects corresponds to a range of 18.9 ms in their category boundaries, providing some information of this brain-to-perception amplification mechanism.

Several points should be taken into consideration regarding the current results. First, group effects on the cABR onsets remain limited to the subgroup that showed strong weighting of the VOT cue. It remains to be further examined whether cABR

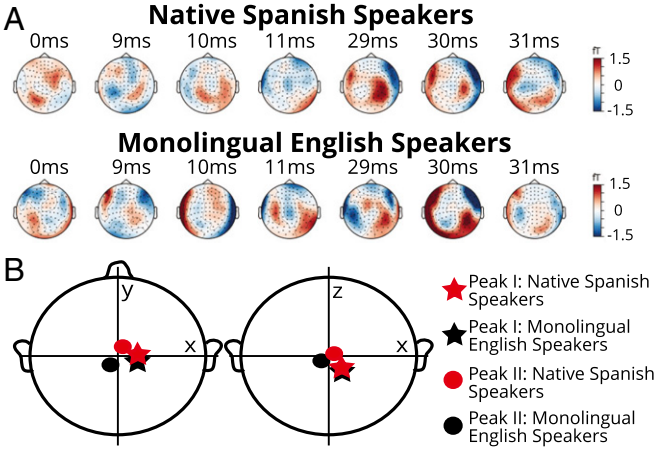

Fig. 5. (*A*) Group average topographies in the peak I and peak II time windows. (*B*) Group average dipole (ECD) locations for peak I and peak II. All peaks have deep sources and no statistical differences were observed between groups in the ECD coordinates.

onset is also related to perception in individuals who do not rely on the VOT cue in the voiced/voiceless contrast. Discovering this will help us further elucidate the mechanism through which cABR is related to perception and experience. Second, while the cABR onset peaks can be predominantly considered subcortical in nature, given the high level of repetition in stimulus presentation, other sources (e.g., auditory cortex) may contribute as a result of online adaptation and learning. More sensitive MEG recording methods (e.g., more advanced denoising methods to allow analysis with fewer trials) with individual MRIs in the future would allow distributed source localization methods and could potentially help elucidate dynamic changes in the underlying sources during the actual recording session.

The current study also generates important future research directions regarding this experiential effect; for example, what is the earliest point in development that an experiential effect can be observed at the brainstem level and how does this effect interact with perceptual development as well as cortical processing? Speech categories emerge as distributional patterns in speech input and social experiences occur, and different cortical processing patterns have been observed for different speech categories (42, 43). During the "sensitive period" between 6 and 12 mo of age, infants rapidly learn to discriminate native speech contrasts while their ability to discriminate nonnative speech contrasts start to decline (9). Theories posit that the neural system starts a process of "neural commitment" to the processing of native speech categories by 12 mo of age (8, 44). To understand the extent of such neural commitment, future research is warranted to examine the relations between cortical and brainstem levels of phonetic processing in cross-linguistic adult studies as well as infants at the beginning and the end of the sensitive period for phonetic learning.

Materials and Methods

Behavioral Assessment on Categorical Perception.

Participants. Monolingual English speakers ($n = 29$, male = 9, age = 21.48 ± 2.15) and native Spanish speakers ($n = 20$, male = 6, age = 25.40 ± 4.76) were first recruited to complete the behavior assessment on their categorical perception of bilabial stop consonants varying on the VOT dimension. All participants were healthy adults with no reported speech, hearing, or language disorders. All participants were right-handed (Edinburgh handedness quotient = 0.98 ± 0.04). All procedures were approved by the Institutional Review Board of the University of Washington and informed consents were obtained from all participants.

Stimuli. Bilabial stop consonants with varying VOTs were synthesized by Klatt synthesizer in Praat software (45). The VOT values ranged from −30 ms to +30 ms with 10-ms increments. The syllable with 0 ms VOT was first synthesized with a 2-ms noise burst and vowel /a/. The duration of the syllable was 90 ms. The fundamental frequency of the vowel /a/ began at 95 Hz and ended at 90 Hz. Silent gaps or prevoicing were added after or before the initial noise burst to create syllables with positive or negative VOTs. The fundamental frequency for the prevoicing portion was 100 Hz. The waveform of the syllable p10 is shown in Fig. 2*A*.

Equipment and procedure. Upon arrival, all participants were consented and completed a short survey on their language and music backgrounds. Then they proceeded to a sound attenuated booth for the computer-based identification and discrimination tasks to assess categorical perception. All sounds were delivered through Sennheiser HDA 280 headphones at 72 dB SPL. Both tasks were completed on a Dell XPS13 9333 computer running Psychophysical Toolbox (46) in MATLAB version 2016a (MathWorks, Inc.).

In an identification trial, a syllable was played and the participant was instructed to identify the sound as either /ba/ or /pa/ through key presses within 1 s. All seven syllables on the continuum were repeated 25 times in a randomized order. In a discrimination trial, two syllables, either identical or adjacent on the continuum, were played with a 300-ms interstimulus interval. Participants were instructed to respond whether the two sounds were the same or different through a key press within 1 s. All 24 (6 pairs × 4 sequences) possible combinations (e.g., +10/0, 0/+10, +10/+10, and 0/0) were repeated 20 times in a randomized order. Practice trials were first administered and participants had to reach 80% to proceed. One participant was excluded during this phase.

Data analysis. For the identification task, the percentage of identification of the syllable as /pa/ was calculated for each sound. For the discrimination task, d′ values for each pair of syllables (e.g., +30/+20 ms VOT) were calculated to index the sensitivity of discrimination. The d′ measure takes into consideration both hit (response of "different" when sounds were different) and false alarm (response of "different" when sounds were the same) responses, and therefore addresses the issue of response bias (29).

M/EEG Brainstem Recording.

Participants. A subgroup of the participants who completed the behavioral assessments was further selected for M/EEG recording of their brainstem responses (monolingual English speakers = 16, native Spanish speakers = 18).

Among the participants invited back, two participants opted out for M/EEG recording and three participants' M/EEG recordings were rendered too noisy (i.e., impendence of EEG channels >10 kΩ). Final sample for group analyses (i.e., complete behavioral and good M/EEG recording) included 14 monolingual English speakers (male = 5, age = 21.21 ± 1.76) and 15 native Spanish speakers (male = 5, age = 26.00 ± 5.04). There was no difference between the two groups in their music training background [years of training: monolingual English = 3.66 ± 2.80, native Spanish = 2.12 ± 2.24, $t(1, 27) = 1.64$, $P = 0.112$]. The native Spanish speakers reported speaking at least one foreign language (English) at high proficiency (*SI Appendix*, Table S1); however, none were raised as simultaneous bilinguals. The English speakers reported to be monolingual with minimal proficiency in other languages.

Stimuli. Bilabial stop consonant with +10 ms VOT (p10 from here on) was selected for the M/EEG recording given the range in perception data (Fig. 1). The waveform of the synthesized syllable is shown in Fig. 2*A*. The 10-ms silent gap starts from the offset of the noise burst and ends at the onset of voicing, making the first peak of onset of voicing at 19 ms (edge of the shaded area).

Equipment and procedure. Simultaneous MEG and EEG recordings were completed inside a magnetically shielded room (MSR) (IMEDCO). For EEG, a standard three-electrode setup was used: CZ electrode on a 10–20 system, ground electrode on the forehead and the reference electrode on right earlobe. Impendence of all electrodes was kept under 10 kΩ. For MEG, an Elekta Neuromag system was used with 204 planar gradiometers and 102 magnetometers. Five head-position-indicator (HPI) coils were attached to identify head positions under the MEG dewar at the beginning of each block. Three landmarks [left preauricular (LPA) and right preauricular (RPA), and nasion] and the HPI coils were digitized along with 100 additional points along the head surface (Isotrak data) with an electromagnetic 3D digitizer (Fastrak, Polhemus). The sounds were delivered from a TDT RP 2.7 (Tucker-Davis Technologies Real-Time Processor), controlled by custom Python software on a Hewlett-Packard workstation, to insert earphones (Nicolet TIP-300). The stimulus was processed such that the rms values were referenced to 0.01 and it was further resampled to 24,414 Hz for the TDT. The sounds were played at the intensity level of 80 dB with alternating polarities. The interstimulus intervals were 150 ms with jitters within a ±10-ms window. Two blocks of recordings (3,000 sounds per block) were completed at the sampling rate of 5,000 Hz. The participants listened passively and watched silent videos during recording.

Data analysis. All data analysis was done using the MNE-Python software (47) in conjunction with the Elekta Neuromag source modeling software (XFit-5.5). For EEG data, after referencing the CZ channel, the data from each block was first notch filtered at 60 Hz and at its harmonics up to 2,000 Hz to remove any power line interference. It was further band-pass filtered between 80 Hz and 2,000 Hz using a fourth order Butterworth filter. Epochs between −10 ms and 150 ms (in relation to sound onset) were extracted and averaged within the block and then across the blocks. Epochs with peak-to-peak values exceeding 100 μV were rejected. An example from one participant is shown in Fig. 2*B*. Because the two peaks are visible in all participants with consistent timing (Fig. 3*B*), we extracted the magnitude and latency values for the two peaks related to the stop consonant for each participant by identifying the maximum value and its corresponding time in the windows between 5 and 15 ms for peak I and between 25 and 35 ms for peak II (18).

MEG data were first preprocessed using the oversampled temporal projection (OTP) method (48) and the temporally extended spatial-signal-separation (tSSS) method (49, 50) to suppress sensor noise and magnetic interference originating from outside of the MEG dewar. Both algorithms are implemented in the MNE-Python software. The preprocessed data were then subjected to the same notch filter (60 Hz and harmonics) and band-pass filters (80–2,000 Hz) as for the EEG data. The same epochs (−10–150 ms) were extracted and averaged. The epochs with peak-to-peak amplitude exceeding 4 pT/cm for gradiometers or 4.0 pT/cm for magnetometers were rejected.

ECDs were fitted separately for the two peaks related to the stop consonant using all magnetometers in the XFit (51). A spherical head model was used as individual MRI scans were not available. The center of the head was adjusted on the head coordinates, based on the Isotrak data containing the head shape and fiducial points (nasion, left and right preauricular points), which was collected at the beginning of the recording. For peak 1, ECDs were fitted between 9 ms and 13 ms with 0.1-ms steps while for peak 2, the time window was between 29 ms and 33 ms. ECDs inside the sphere model with goodness of fit over 65% were considered (peak I, mean goodness of fit ± SD = 73.25 ± 6.66; peak II, mean goodness of fit ± SD = 74.03 ± 5.28). An ECD was selected as the best-fitting dipole for each peak if the current density (Q) was the largest with the best goodness-of-fit value.

Between groups, the current density (Q) and the goodness-of-fit values were comparable for both peaks [peak I, Q: t (1, 27) = −0.86 and g: t (1, 27) = −0.65; peak II, Q: t (1, 27) = 0.18 and g: t (1, 27) = −0.62]. The ECD coordinates (x, y, and z) for each peak were extracted for each participant as a measure reflecting source location.

ACKNOWLEDGMENTS. The manuscript has been greatly improved by comments from colleagues, including Dr. Samu Taulu (M/EEG methods) and Dr. Matthew Masapollo and Dr. Alexis Bosseler (speech perception). The research described here was supported by the Ready Mind Project as well as a generous contribution by Mr. Bard Richmond to the University of Washington Institute for Learning & Brain Sciences.

1. Strange W, Jenkins JJ (1978) Role of linguistic experience in the perception of speech. *Perception and Experience*, eds Walk RD, Pick HL (Springer, Boston), pp 125–169.
2. Holt LL, Lotto AJ (2010) Speech perception as categorization. *Atten Percept Psychophys* 72:1218–1227.
3. Abramson AS, Lisker L (1970) Discriminability along the voicing continuum: Cross language tests. *Proceedings of the Sixth International Congress of Phonetic Sciences* (Academia, Prague), pp 569–573.
4. Abramson AS, Lisker L (1972) Voice-timing perception in Spanish word-initial stops. *Haskins Laboratories Status Report on Speech Research* 29:15–25.
5. Hay J (2005) How auditory discontinuities and linguistic experience affect the perception of speech and non-speech in English- and Spanish-speaking listeners. PhD dissertation (University of Texas at Austin, Austin).
6. Miyawaki K, et al. (1975) An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Percept Psychophys* 18:331–340.
7. Kuhl PK (2004) Early language acquisition: Cracking the speech code. *Nat Rev Neurosci* 5:831–843.
8. Kuhl PK, et al. (2008) Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philos Trans R Soc Lond B Biol Sci* 363:979–1000.
9. Kuhl PK, et al. (2006) Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev Sci* 9:F13–F21.
10. Werker JF, Tees RC (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav Dev* 7:49–63.
11. Logan JS, Lively SE, Pisoni DB (1991) Training Japanese listeners to identify English/r/ and/l/: A first report. *J Acoust Soc Am* 89:874–886.
12. Flege JE, Yeni-Komshian GH, Liu S (1999) Age constraints on second-language acquisition. *J Mem Lang* 41:78–104.
13. Lim SJ, Holt LL (2011) Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cogn Sci* 35:1390–1405.
14. Sharma A, Dorman MF (2000) Neurophysiologic correlates of cross-language phonetic perception. *J Acoust Soc Am* 107:2697–2703.
15. Rivera-Gaxiola M, Silva-Pereyra J, Kuhl PK (2005) Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Dev Sci* 8:162–172.
16. Jewett DL, Romano MN, Williston JS (1970) Human auditory evoked potentials: Possible brain stem components detected on the scalp. *Science* 167:1517–1518.
17. Mason JA, Herrmann KR (1998) Universal infant hearing screening by automated auditory brainstem response measurement. *Pediatrics* 101:221–228.
18. Skoe E, Kraus N (2010) Auditory brain stem response to complex sounds: A tutorial. *Ear Hear* 31:302–324.
19. Coffey EB, Herholz SC, Chepesiuk AM, Baillet S, Zatorre RJ (2016) Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat Commun* 7:11070.
20. Johnson KL, Nicol TG, Kraus N (2005) Brain stem response to speech: A biological marker of auditory processing. *Ear Hear* 26:424–434.
21. Bidelman GM, Gandour JT, Krishnan A (2011) Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J Cogn Neurosci* 23:425–434.
22. Intartaglia B, et al. (2016) Native language shapes automatic neural processing of speech. *Neuropsychologia* 89:57–65.
23. Krishnan A, Xu Y, Gandour J, Cariani P (2005) Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res* 25:161–168.
24. Wong PCM, Skoe E, Russo NM, Dees T, Kraus N (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 10:420–422.
25. Kraus N, Chandrasekaran B (2010) Music training for the development of auditory skills. *Nat Rev Neurosci* 11:599–605.
26. Kraus N, et al. (2014) Music enrichment programs improve the neural encoding of speech in at-risk children. *J Neurosci* 34:11913–11918.
27. Bidelman GM, Moreno S, Alain C (2013) Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage* 79:201–212.
28. Coffey EBJ, Musacchia G, Zatorre RJ (2017) Cortical correlates of the auditory frequency-following and onset responses: EEG and fMRI evidence. *J Neurosci* 37:830–838.
29. Macmillan NA, Creelman CD (2008) *Detection Theory: A User's Guide* (Psychology Press, New York).
30. Lisker L, Abramson AS (1964) A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20:384–422.
31. Krishnaswamy P, et al. (2017) Sparsity enables estimation of both subcortical and cortical activity from MEG and EEG. *Proc Natl Acad Sci USA* 114:E10465–E10474.
32. Suga N (2008) Role of corticofugal feedback in hearing. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 194:169–183.
33. Suga N, Ma X (2003) Multiparametric corticofugal modulation and plasticity in the auditory system. *Nat Rev Neurosci* 4:783–794.
34. Xiao Z, Suga N (2002) Modulation of cochlear hair cells by the auditory cortex in the mustached bat. *Nat Neurosci* 5:57–63.
35. Hepper PG, Shahidullah BS (1994) Development of fetal hearing. *Arch Dis Child* 71:F81–F87.
36. Birnholz JC, Benacerraf BR (1983) The development of human fetal hearing. *Science* 222:516–518.
37. Moon CM, Lagercrantz H, Kuhl PK (2013) Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatr* 102:156–160.
38. DeCasper AJ, Fifer WP (1980) Of human bonding: Newborns prefer their mothers' voices. *Science* 208:1174–1176.
39. Partanen E, et al. (2013) Learning-induced neural plasticity of speech processing before birth. *Proc Natl Acad Sci USA* 110:15145–15150.
40. Krizman J, Marian V, Shook A, Skoe E, Kraus N (2012) Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proc Natl Acad Sci USA* 109:7877–7881.
41. Ress D, Backus BT, Heeger DJ (2000) Activity in primary visual cortex predicts performance in a visual detection task. *Nat Neurosci* 3:940–945.
42. Maye J, Werker JF, Gerken L (2002) Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82:B101–B111.
43. Chang EF, et al. (2010) Categorical speech representation in human superior temporal gyrus. *Nat Neurosci* 13:1428–1432.
44. Bosseler AN, et al. (2013) Theta brain rhythms index perceptual narrowing in infant speech perception. *Front Psychol* 4:690.
45. Boersma P, Weenink D (2009) Praat: Doing Phonetics by Computer, Version 5.1.05.
46. Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436.
47. Gramfort A, et al. (2014) MNE software for processing MEG and EEG data. *Neuroimage* 86:446–460.
48. Larson E, Taulu S (2018) Reducing sensor noise in MEG and EEG recordings using oversampled temporal projection. *IEEE Trans Biomed Eng* 65:1002–1013.
49. Taulu S, Kajola M (2005) Presentation of electromagnetic multichannel data: The signal space separation method. *J Appl Phys* 97:124905.
50. Taulu S, Hari R (2009) Removal of magnetoencephalographic artifacts with temporal signal-space separation: Demonstration with single-trial auditory-evoked responses. *Hum Brain Mapp* 30:1524–1534.
51. Parkkonen L, Fujiki N, Mäkelä JP (2009) Sources of auditory brainstem responses revisited: Contribution by magnetoencephalography. *Hum Brain Mapp* 30:1772–1782.