Neural indicators of articulator-specific sensorimotor influences on infant speech perception

Dawoon Choi[a,1], Ghislaine Dehaene-Lambertz[b,c,d], Marcela Peña[e], and Janet F. Werker[a,1]

[a]Department of Psychology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; [b]Cognitive Neuroimaging Unit, INSERM, F-91191 Gif/Yvette, France; [c]NeuroSpin, Commissariat à l'Énergie Atomique et aux Énergies Alternatives, F-91191 Gif/Yvette, France; [d]Cognitive Neuroimaging Unit, University Paris-Sud, F-91191 Gif/Yvette, France; and [e]Escuela de psicología, Pontificia Universidad Catolica de Chile, Santiago 7820244, Chile

Contributed by Janet F. Werker, March 30, 2021 (sent for review December 10, 2020; reviewed by Marco Buiatti and Núria Sebastián-Gallés)

While there is increasing acceptance that even young infants detect correspondences between heard and seen speech, the common view is that oral-motor movements related to speech production cannot influence speech perception until infants begin to babble or speak. We investigated the extent of multimodal speech influences on auditory speech perception in prebabbling infants who have limited speech-like oral-motor repertoires. We used event-related potentials (ERPs) to examine how sensorimotor influences to the infant's own articulatory movements impact auditory speech perception in 3-mo-old infants. In experiment 1, there were ERP discriminative responses to phonetic category changes across two phonetic contrasts (bilabial–dental /ba/-/ɗa/; dental–retroflex /ɗa/-/ɖa/) in a mismatch paradigm, indicating that infants auditorily discriminated both contrasts. In experiment 2, inhibiting infants' own tongue-tip movements had a disruptive influence on the early ERP discriminative response to the /ɗa/-/ɖa/ contrast only. The same articulatory inhibition had contrasting effects on the perception of the /ba/-/ɗa/ contrast, which requires different articulators (the lips vs. the tongue) during production, and the /ɗa/-/ɖa/ contrast, whereby both phones require tongue-tip movement as a place of articulation. This articulatory distinction between the two contrasts plausibly accounts for the distinct influence of tongue-tip suppression on the neural responses to phonetic category change perception in definitively prebabbling, 3-mo-old, infants. The results showing a specificity in the relation between oral-motor inhibition and phonetic speech discrimination suggest a surprisingly early mapping between auditory and motor speech representation already in prebabbling infants.

infancy | sensorimotor | speech perception | EEG

Infants rapidly acquire robust representations of the native phonetic repertoire from the natural multisensory speech input of their environment. Multimodal speech signals are generated by a common underlying source—the vocal tract and the articulatory movements used during production (1, 2). Adult speech perception is influenced by synchronously occurring multimodal speech cues, including auditory, visual, motor, and sensorimotor signals (3). Recent advances reveal that speech production relies on both auditory and sensorimotor signals (4, 5), but also, sensorimotor input can affect the perception of auditory (6) and visual (7) speech. Indeed, neural evidence indicates bidirectional interaction between the speech perception and production systems in the adult brain (8). It has been widely assumed that the interactions between the articulator-specific sensorimotor information and acoustic phonetic perception would appear later in development after infants begin to babble and to produce speech themselves. This assumption is not surprising given that motor coordination is immature early in life and appears to have a protracted development. However, to fully understand how infants acquire their native speech sound repertoire, it is critical to examine whether sensorimotor/motoric dimensions of speech are relevant for auditory speech perception even in infants who are prebabbling. If so, then sensorimotor influences on speech perception may be part of the foundation that sets the stage for language acquisition in general and babbling in particular, rather than production experience driving the eventual auditory-sensorimotor/motor speech interaction.

While the speech signal that infants experience and learn from is multimodal, speech perception research during the acquisition period has focused mainly on auditory speech perception, and to a modest extent, on audiovisual speech perception. Infants reliably match heard and seen speech at 2 mo of age by looking longer to the face that is articulating the syllable being played (9, 10). Remarkably, infants are also able to match audio and visual speech even for nonnative consonants and vowels, which they have not encountered in their linguistic environment (11). While some have suggested that audiovisual speech perception abilities in infants reflect a domain-general preference for synchronously occurring stimuli (12), there is neural evidence of multimodal phonetic representation already at 2 mo of age (13). In ref. 13, a phonetic mismatch response (MMR) was observed to the category change of an auditory vowel, both when the preceding stimuli were repetitions of visemes (a face articulating the same or a different vowel) or speech sounds. The consistency of the MMRs to the phonetic category change regardless of modality suggests that infants have access to an integrated intermodal representation (13).

There is less experimental work investigating sensorimotor interactions with speech perception; however, several recent behavioral studies have addressed this question by experimentally

Significance

In the adult brain, multimodal speech perception that interfaces with a bidirectional interaction of perception and production speech systems is increasingly accepted. Speech perception in infancy is already highly multisensory, suggesting an early emerging representation for speech across sensory modalities. We provide electrophysiological evidence for sensorimotor influences on auditory speech discrimination responses in 3-mo-old infants who are several months away from producing canonical babbling. Auditorily, infants discriminated both contrasts tested. However, a tongue-tip articulatory inhibition diminished the /ɗa/-/ɖa/ discrimination (both phones involve the tongue-tip movement as a place of articulation), whereas accentuating the /ba/-/ɗa/ discrimination (different articulators involved in the articulation of the two phones). These findings suggest that prebabbling infants' speech perception is more robustly multisensory than previously considered.

Author contributions: D.C., G.D.-L., M.P., and J.F.W. designed research; D.C. performed research; D.C., G.D.-L., M.P., and J.F.W. analyzed data; and D.C., G.D.-L., M.P., and J.F.W. wrote the paper.

Reviewers: M.B., University of Trento; and N.S.-G., Universitat Pompeu Fabra.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: sheri.d.choi@gmail.com or jwerker@psych.ubc.ca.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2025043118/-/DCSupplemental.

Published May 12, 2021.

manipulating infants' own oral-motor movements. In the first such study, 4-mo-old infants' labial configuration was manipulated (by gently holding an appropriately shaped object in their mouth) to either resemble the shape made for producing /i/ or /u/ vowels while they were tested in an audiovisual matching task. Results showed that infants' matching of these same vowels was changed by the manipulation (14). The influence of sensorimotor cues on auditory-only speech perception was more recently tested, this time with infants aged 6 mo, who do not typically produce well-formed consonant–vowel (CV) syllables. Replicating previous work (15), English-learning infants this age discriminated a dental /ɗa/–retroflex /ɖa/ phonetic contrast that is nonnative to English speakers, but native to Hindi speakers' contrast. These two consonants differ, in adult Hindi production, only on the placement of the tongue tip during articulation: The dental involves placement of the tongue tip behind the back front teeth, whereas retroflex production involves curling the tongue tip back and placing it against the roof of the mouth. However, when an infant's tongue-tip movement was inhibited by having a caregiver gently hold a teether on the tongue, discrimination of this nonnative /ɗa/-/ɖa/ contrast was disrupted (16, 17). A control experiment showed that discrimination of this contrast was maintained when a different teether that does not interfere with tongue-tip movement was used, indicating that it was not the mere presence of a teething toy but rather the inhibition of the relevant articulator that accounted for the disruption of discrimination (16).

These specific sensorimotor influences on auditory and audiovisual speech perception provide evidence that the relation between sensorimotor information and auditory speech perception is present in infants who have not had extensive speech production or babbling experience. Although preverbal infants this young have yet to gain the full articulatory control required to generate speech-like sounds, behavioral studies reviewed above suggest that a sensorimotor mapping of the articulators may be available to infants before babbling begins, possibly through spontaneously generated movement patterns during prenatal development (18). These patterns may be progressively refined through orofacial movements (e.g., sucking movements and nonspeech vocalizations) that help to shape the motor articulatory space that must be aligned with the phonetic perceptual space to ensure correct productions.

Anatomically, the core neural pathways for speech including the cortical connections between the frontal (productive) and temporal (receptive) speech areas are in place before term birth (19). While the ventral pathway is more mature at birth, the dorsal pathway (i.e., the arcuate fasciculus) that functionally transforms auditory and motor speech codes rivals in maturity by 10 wk (20, 21). In ref. 21, the authors concluded that the functional connectivity, or cross-talk, between the suprasylvian part of the arcuate fasciculus, the posterior part of the superior temporal sulcus, and area 44 in the left inferior frontal region is established within the first few postnatal months based on a unique correlational pattern in the maturational indices across these regions, which also collectively form key nodes of the adult phonological loop. The early maturation and functional engagement of the arcuate fasciculus, which is a bidirectional tract between the productive and receptive areas, suggest that the necessary connectivity that subserves the sensorimotor influence on auditory perception is in place within several months after birth.

Current Study

The aim of the current study was to examine whether auditory speech discrimination is affected by sensorimotor influences at an age when the productive and receptive regions of the brain are functionally connected but when infants are still several months away from beginning to babble CV syllables. CV syllable production begins around 7 to 9 mo of age (22); thus, testing infants at 3 mo of age ensures that babbling would not have begun. We used electroencephalogram (EEG) to investigate how sensorimotor input could influence 3-mo-old infants' ability to auditorily discriminate phonetic contrasts that minimally differ in the place of articulation, and examined the neural dynamics underlying auditory-sensorimotor integration in preverbal infants. We measured infants' event-related potential (ERP) responses during the speech perception task—without (experiment 1) and with (experiment 2) sensorimotor influences—using a mismatch paradigm designed to assess phonetic category discrimination. This ERP paradigm has been validated by previous studies that show that young infants and even prematurely born newborns detect phonetic category change as evidenced by a phonetic MMR (23, 24)

In experiment 1 (N = 22), English-learning infants passively listened to speech syllables presented in sequences of four syllables with an isochronous onset. In experiment 2 (N = 22), infants' tongue-tip movements were inhibited using a teething toy (Tomy Learning Curve Fruity Teethers) that was gently held in the infant's mouth by the caregiver while infants passively listened to the syllables (Fig. 1). In each experiment, we measured the ERP discriminative responses to two phonetic contrasts: an English bilabial /ba/ vs. dental /ɗa/ contrast and a non-English (Hindi) dental /ɗa/ vs. retroflex /ɖa/ contrast. Previous behavioral and EEG studies demonstrate that prelingual infants auditorily discriminate both the /ba/-/ɗa/ and the /ɗa/-/ɖa/ phonetic contrasts (25, 26); therefore, in experiment 1 (auditory discrimination), we hypothesized that 3-mo-old infants would discriminate both contrasts. Behaviorally, tongue-tip movement suppression disrupted 6-mo-old infants' discrimination of the /ɗa/-/ɖa/ contrast (16), but no prior studies examined its effects on the /ba/-/ɗa/ discrimination. Furthermore, sensorimotor influences on auditory discrimination had not previously been examined in infants as young as 3 mo of age. In experiment 2, we hypothesized that if sensorimotor-auditory speech relations are present and functional even in 3-mo-old infants, then a similar disruption in the /ɗa/-/ɖa/ contrast discrimination may be expected. While both the dental /ɗa/ and the retroflex /ɖa/ require tongue-tip movement during articulation, /ba/ production requires bilabial movement during articulation. If this articulator distinction is a salient feature of discrimination, then we may expect tongue-tip inhibition to differentially influence the ERP responses to the /ba/-/ɗa/ and the /ɗa/-/ɖa/ contrast. Alternatively, because /ɗa/ is present in both contrasts, the tongue-tip inhibition may result in similar disruption across both contrasts. The goal of the current study was to examine the specificity of the auditory-sensorimotor relation that reflects the underlying articulatory code in prebabbling infants.

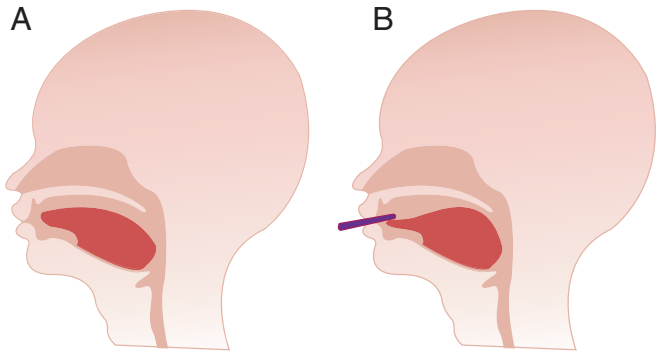

Fig. 1. Midsagittal views of infant vocal tract. (*A*) The tongue is in its natural state. (*B*) The teething toy held by the caregiver depresses the tongue tip. The tongue is muscular hydrostat, modeled as a solid muscle cylinder, which maintains a constant volume under pressure (50); thus, changes to the height result in an antagonistic force to the tongue root to maintain constant volume.

Results

Cluster Analyses.

Experiment 1. We examined the main effect of Condition (i.e., the difference between standard and deviant trials; see *Materials and Methods*) to assess the ERP discriminative responses to the phonetic category changes. We observed a significant difference in a cluster of left-frontal electrodes between 450 and 710 ms following the onset of the fourth syllable ($P_{\text{Monte Carlo cluster corrected}} = 0.023$; Cohen's $d = 0.74$; Fig. 2). We then tested for a Condition (Standards vs. Deviants) by Phonetic Contrast (/ba/-/ɗa/ contrast vs. /ɗa/-/ɖa/ contrast) interaction, which was not significant (the smallest cluster P value was 0.63). The sensors and time windows identified from the first main analysis were extracted and averaged per subject and experimental condition.

Experiment 2. When infants' tongue-tip movement was inhibited, we did not observe a main effect of Condition for a comparison of the ERPs between the standard and deviants across both phonetic contrasts (the smallest cluster P value was 0.16). However, there was a significant interaction effect of Condition by Phonetic Contrast ($P_{\text{Monte Carlo cluster corrected}} = 0.0052$); thus, we conducted additional cluster-based permutation paired t tests comparing standard vs. deviant trials for the /ba/-/ɗa/ contrast and the /ɗa/-/ɖa/ contrast separately. We observed a significant cluster to the /ba/-/ɗa/ contrast, over a cluster of central-posterior electrodes 290–490 ms following the onset of the fourth syllable ($P_{\text{Monte Carlo cluster corrected}} = 0.020$; Cohen's $d = 0.93$; Fig. 3). However, no significant cluster was observed to the /ɗa/-/ɖa/ contrast (the smallest cluster P value was 0.36). The sensors and time windows from the test of Condition from the /ba/-/ɗa/ contrast were extracted and averaged per subject and experimental condition.

Control Analyses.

Experiment 1. To assure that the difference revealed above is in response to the phonetic category change, rather than an experimental artifact, we compared the ERP responses in the same spatiotemporal cluster (450–710 ms) after each syllable in a three-way ANOVA with Condition (Standard and Deviant), Syllable position (Repetitions [1 to 3] and Fourth), and Phonetic Category (/ba/-/ɗa/ and /ɗa/-/ɖa/) as factors. There were no main effects of Phonetic Category [$F_{(1,21)} < 1$, $\eta_p^2 = 0.039$] or Syllable Position [$F_{(1,21)} < 1$, $\eta_p^2 = 0.012$], indicating that, overall, the responses did not vary depending on the Phonetic Category nor the Syllable Position. We observed a significant main effect of Condition [$F_{(1,21)} = 13.678$, $P = 0.001$, $\eta_p^2 = 0.39$], and a significant Condition by Syllable Position interaction [$F_{(1,21)} = 9.461$, $P = 0.006$, $\eta_p^2 = 0.311$]. No other interaction effects were significant (values of $P > 0.29$). Simple main effects showed that the ERP response following the Fourth syllable significantly differed between the standard and the deviant trials for both phonetic contrasts [/ba/-/ɗa/: $F_{(1,21)} = 6.937$, $P = 0.016$; /ɗa/-/ɖa/: $F_{(1,21)} = 20.520$, $P < 0.001$]; however, this difference was not detected for the preceding repeated syllables in neither phonetic contrasts [/ba/-/ɗa/: $F_{(1,21)} < 1$, and /ɗa/-/ɖa/: $F_{(1,21)} < 1$].

Bayesian multilevel regression modeling further supported these results: There was an effect of Condition in the amplitude

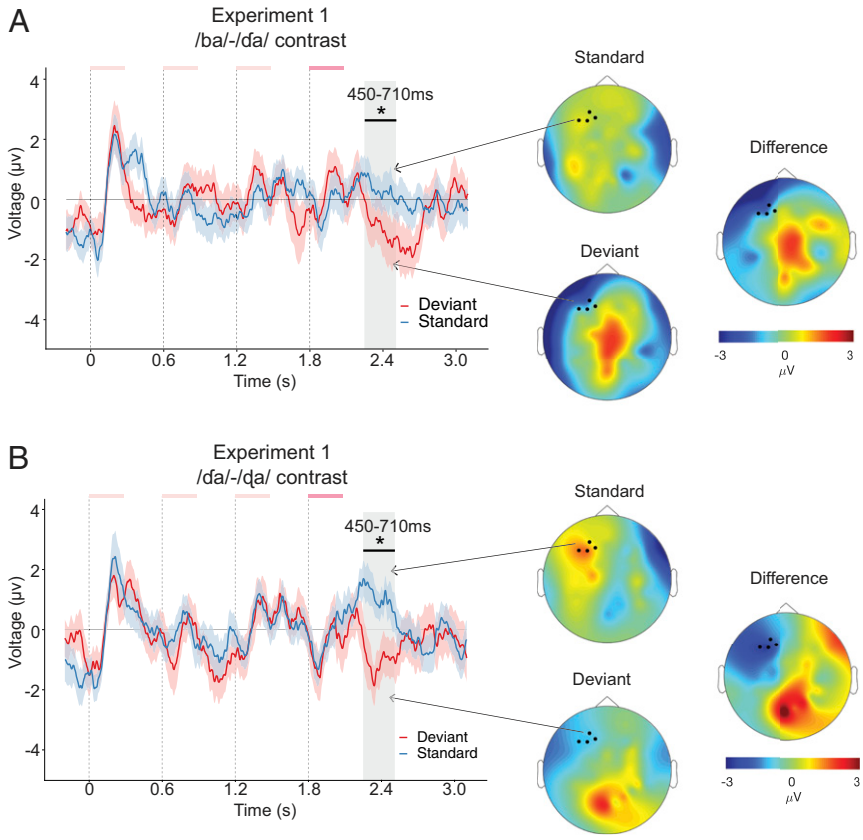

Fig. 2. Experiment 1: ERP responses. (*A*, *Left*) The grand averaged ERP time course of the deviant (red line) and standard (blue line) trials to the /ba/-/ɗa/ contrast. The mean voltage and the SEs for each condition are plotted for the left-anterior cluster of sensors. The vertical dotted lines indicate syllable onset (1 to 4). The gray bar indicates the time window of the significant spatiotemporal cluster (2.25–2.51 s; i.e., 450–710 ms post fourth syllable onset) over the electrodes 9, 11,12, and 13. (*A*, *Right*) Voltage topographies for the Deviant and Standard trials, and the difference (deviant – standard) averaged across the time window of the cluster. (*B*, *Left*) The grand averaged ERP time courses to the /ɗa/-/ɖa/ contrast. (*B*, *Right*) Voltage topographies for the standard and deviant trials, and the difference (deviant – standard).

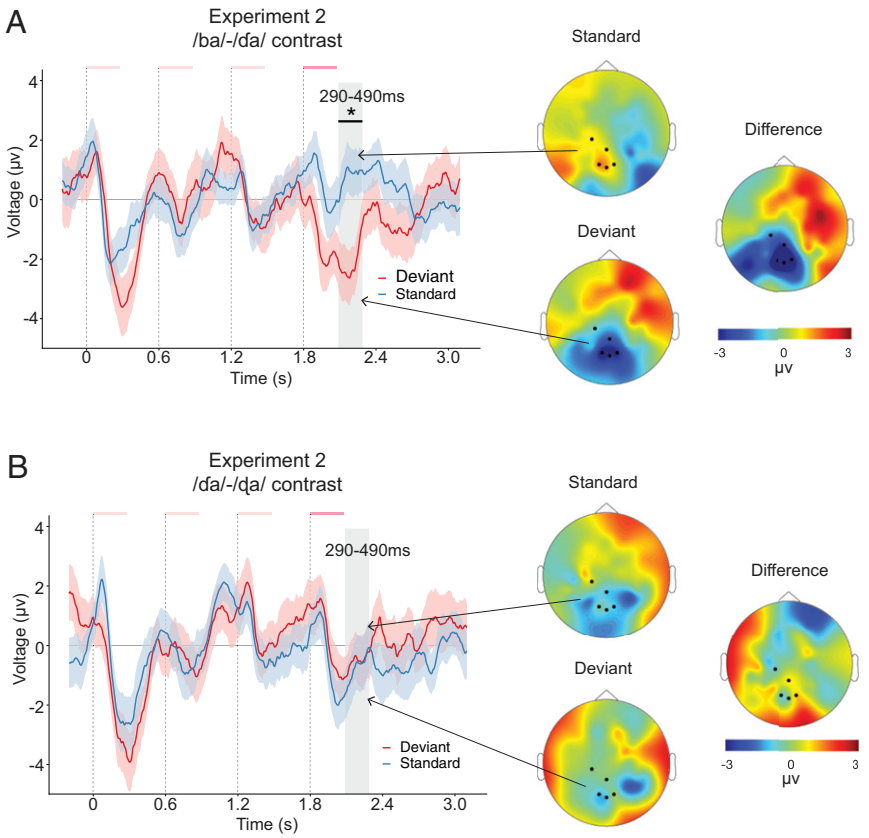

Fig. 3. Experiment 2: ERP responses. (*A*, *Left*) The grand averaged ERP time course of the deviant (red line) and the standard (blue line) trials to the /ba/-/da/ contrast. The mean voltage and the SEs for each condition are plotted for the left-anterior cluster of sensors. The vertical dotted lines indicate each syllable onset (1 to 4). The gray bar indicates the time window of the significant spatiotemporal cluster (2.09–2.29 s, i.e., 290–490 ms post fourth syllable onset) over the electrodes 21, 33, 34, 36, and 38. (*A*, *Right*) Voltage topographies for the deviant and standard trials, and the difference (deviant − standard) averaged across the time window of the cluster. (*B*, *Left*) The grand averaged ERP time courses to the /da/-/ɖa/ contrast from the same spatiotemporal cluster as in the above, /ba/-/da/ contrast. An ERP discriminative response was not observed to the /da/-/ɖa/ contrast in experiment 2. The gray bar indicates the time window of the significant cluster observed to the /ba/-/da/ contrast in experiment 2. (*B*, *Right*) Voltage topographies for the Standard and Deviant trials, and the difference (deviant − standard).

response to the /ba/-/da/ ($\gamma = 0.55$, $CI_{95\%} = [0.01, 1.08]$) and to the /da/-/ɖa/ ($\gamma = 0.94$, $CI_{95\%} = [0.39, 1.48]$) contrasts, indicated by the difference between the standard and deviant trials that did not significantly overlap with zero. The Bayesian confidence intervals indicated that the effect of Condition was stronger to the /da/-/ɖa/ contrast than to the /ba/-/da/ contrast (Fig. 4*A*).

Experiment 2. A three-way ANOVA (Condition by Syllable position by Phonetic category) was conducted on the identified spatiotemporal cluster (290–490 ms) from experiment 2. The main effect was not significant for Phonetic Contrast [$F_{(1,21)} < 1$, $\eta_p^2 = 0.008$] nor for Syllable Position [$F_{(1,21)} < 1$, $\eta_p^2 = 0.018$], but there was a significant main effect of Condition [$F_{(1,21)} = 14.844$, $P = 0.001$, $\eta_p^2 = 0.414$]. There were multiple significant interaction effects including a Condition by Syllable Position interaction [$F_{(1,21)} = 4.389$, $P = 0.048$, $\eta_p^2 = 0.173$], a Condition by Phonetic Contrast interaction [$F_{(1,21)} = 10.681$, $P = 0.004$, $\eta_p^2 = 0.337$], and a significant three-way interaction [$F_{(1,21)} = 6.284$, $P = 0.020$, $\eta_p^2 = 0.230$]; only the Phonetic Contrast by Syllable Position interaction was not significant [$F_{(1,21)} < 1$, $\eta_p^2 = 0.006$]. Follow-up analyses indicated a significant difference in Condition (Deviant and Standard) only at the level of the Fourth syllable to the /ba/-/da/ contrast [$F_{(1,21)} = 29.606$, $P < 0.001$]. There were no significant effects of Condition at any other levels [$F_{(1,21)} < 1$ for the repeated syllables for both contrasts and for the response to the Fourth syllable for the /da/-/ɖa/ contrast].

The Bayesian multilevel regression model further supported these results. There was an effect of Condition in the amplitude response to the /ba/-/da/ contrast ($\gamma = 1.16$, $CI_{95\%} = [0.62, 1.70]$). However, there was no effect of Condition to the /da/-/ɖa/ contrast ($\gamma = 0.003$, $CI_{95\%} = [-0.55, 0.55]$); the Bayesian CIs overlapped with zero, indicating strong evidence in support of no difference between the standard and the deviant trials (Fig. 4*B*).

Discussion

The current study investigated infants' neural responses to phonetic category changes with and without oral-motor influence across two experiments. Each experiment targeted two distinct phonetic contrasts: a /ba/-/da/ contrast and a /da/-/ɖa/ contrast. In experiment 1, we observed a significant effect of Condition (i.e., differences in the neural response between standard and deviant trials) to both contrasts in a cluster of left-anterior electrodes. In experiment 2, when infants' tongue-tip movements were suppressed, the data-driven approach revealed that there was no overall effect of Condition, but a Condition by Phonetic Contrast interaction was significant. To the /ba/-/da/ contrast, the standard and deviant trials showed distinct responses in a cluster of posterior electrodes, but no differences were observed to the /da/-/ɖa/ contrast. Follow-up analyses on the spatiotemporal cluster as observed in the /ba/-/da/ contrast, experiment 2, showed that the difference between the standards and the deviant trials overlapped with zero for the /da/-/ɖa/ contrast ($CI_{95\%}$ [−0.55, 0.55]). In other

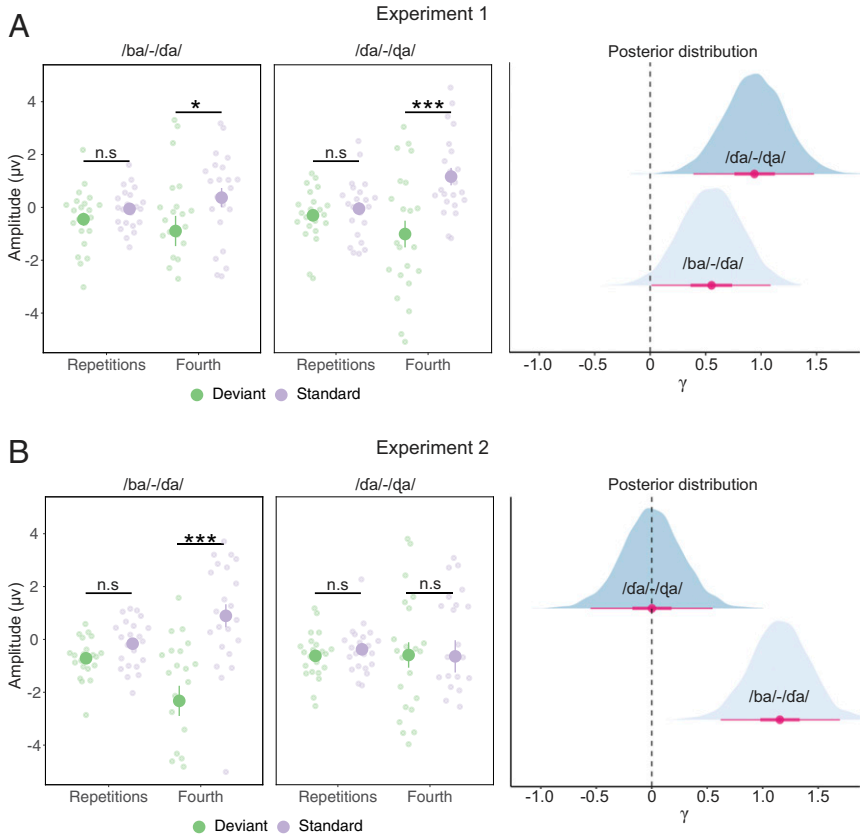

Fig. 4. (*A*) Experiment 1. (*Left*) A comparison across Syllable Position (Repetitions and Fourth), Phonetic Contrast (/ba/-/ɗa/ and /ɗa/-/ɖa/), and Condition (Standard and Deviant). The mean voltage of the Fourth syllable (450–710 ms following syllable onset) and Repetitions (average of 450–600 ms following the onset of first, second, and third syllables) are plotted separately to the /ba/-/ɗa/ and the /ɗa/-/ɖa/ contrasts. Mean and SEs are plotted. (*Right*) Experiment 1: Bayesian multilevel regression model. Posterior distributions of the γ value of the Condition parameter, which indicates the degree of change from the Deviant to the Standard condition. The central dot indicates the highest density posterior mean, and the line indicates the 95% HPDI. (*B*) Experiment 2. (*Left*) The mean voltage of the Fourth syllable (290–490 ms following syllable onset) and Repetitions (average of 290–490 ms following the onset of first, second, and third syllables) are plotted separately to the /ba/-/ɗa/ and the /ɗa/-/ɖa/ contrasts. (*Right*) Experiment 2: Bayesian multilevel regression model. Posterior distributions of the γ value of Condition parameter indicating the change in slope from the Deviants to the Standards across subjects. The central dot indicates the highest density posterior mean, and the line indicates the 95% HPDI.

words, when infants' tongue-tip movement was restricted, while the neural responses to the /ba/-/ɗa/ discrimination was clearly observed, we did not find evidence for /ɗa/-/ɖa/ discrimination in this same spatiotemporal cluster. These findings suggest that sensorimotor influences on the speech articulator (i.e., the tongue) modulated phonetic processing, but with a degree of specificity rather than broadly disrupting speech processing.

Our finding that the neural responses to phonetic category discrimination are modulated by articulatory sensorimotor influences in prebabbling infants is illuminating, because at 3 mo of age, speech experience is limited and infants have not yet attuned to the native consonants (27, 15). Moreover, infants at this age are not producing well-formed CV syllables characteristic of canonical babbling (22). Although the motor repertoire at this age is relatively immature, the tongue-tip inhibition in experiment 2 had a different impact on phonetic perception whether or not the phones in the contrast require a tongue-tip movement, during production by adults, indicating that the sensorimotor input is simultaneously integrated with the auditory speech signal during phonetic processing. Although the articulator inhibition was present (by the parent holding a teething toy) throughout the testing of both phonetic distinctions and the trial conditions were randomly presented, an early ERP discrimination response was present to the /ba/-/ɗa/ contrast whereas there was no ERP discrimination response to the /ɗa/-/ɖa/ contrast. Thus, sensorimotor articulatory dimensions relevant to the heard speech sounds interact with auditory speech perception in prebabbling infants.

The pattern of the MMR was earlier and larger (290–490 ms) to the /ba/-/ɗa/ contrast in experiment 2 than to both phonetic contrasts in experiment 1 (450–710 ms). It is possible that the latency and magnitude of the ERP response was biased due to a shift in the probabilistic distribution of the phonetic categories perceived. MMRs are elicited following a violation of auditory regularity and represent prediction errors based on a probabilistic model of the environment (28, 29). Thus, the magnitude of the difference between the standard and deviant stimuli, but also the statistical distributions of the trials themselves can either accelerate (30) or amplify the MMN response in adults (28), and the MMR in infants (31). Therefore, one potential explanation for across-experiment differences in the MMR is that the articulatory inhibition biased the statistical regularities in the input in experiment 2 compared to experiment 1. In our design, infants heard an equal distribution of the syllables (/ba/, /ɗa/, and /ɖa/), equal instances of standard and deviant trials, and heard both directions of category change during the deviant changes (*Materials and Methods*). In experiment 1, because the probabilistic distribution across these three dimensions was constant, no given syllable has a greater predictive value over the others. In experiment 2, however, if the

sensorimotor input informs auditory speech perception in infants and the /ɗa/-/ɖa/ discrimination is impaired by the tongue-tip inhibition, then /ba/ becomes singularized as the only bilabial sound. Thus, /ba/ is less frequent within this perceptual space, which means that it is also less predicted. This change in the probability distribution could account for the faster and larger MMR to the /ba/-/ɗa/ contrast observed in experiment 2.

Although we cannot rule out the possibility that infants may be attempting to imitate the auditory stimuli as they listen during the experiment, the randomized presentation of the trials and the short interstimulus intervals likely prevent any imitative efforts. In experiment 2, the tongue-tip restriction might prevent overt imitation, in particular, for /ɗa/ and /ɖa/, but much less so for /ba/. Thus, if imitative attempts were to have taken place in experiment 2, it would only accentuate the difference between /ba/ and the other two syllables and further reinforce any auditory-motor loop. Another potential alternative is that infants were more attentive or alert when the teething toy was held in their mouth (experiment 2), compared to when they were passively listening to the sounds (experiment 1); however, such an account does not explain the dissociative effects observed between the two phonetic contrasts heard within experiment 2. Lastly, it is possible that movement-related artifacts generated by the infant's interaction with the teething toy may not have been sufficiently accounted for. Yet, this account also fails to explain the significant difference between the two contrasts within experiment 2. A comparison of the standard trials across the two experiments showed that the standard trials in experiments 1 and 2 were not significantly different from each other (*SI Appendix*, Fig. S3), suggesting that the data were comparable despite the articulator inhibition in experiment 2.

Young infants process auditory speech in a highly sophisticated manner. Infants only a few months old, and even newborn preterm infants, show a MMR specific to phonetic category change normalizing across voice quality changes within and across genders (32, 24). Furthermore, infants as young as 2 mo of age detect phonetic invariance across coarticulation (33, 34), revealing stable phonetic representations despite acoustic variability. It has been proposed that the infant brain achieves invariant representations of speech sounds through a vectorization of the acoustic input along orthogonal dimensions corresponding to phonetic features, that are subsequently integrated into a phonetic representation (35). The finding from ref. 35 that phonetic features defined relative to articulatory dimensions are pertinent to describe infants' as adults' speech perception space, converges with the current study demonstrating that direct manipulations of the sensorimotor information to the articulators can modulate perception of the auditory speech signal.

The current evidence suggests that, at the earliest stages of language acquisition, the sensorimotor system may be relevant for the perception of auditory speech signals. Critically, we do not find evidence that motor programs are necessarily referenced for speech perception as is predicted by the standard motor theory of speech perception (36). Rather, we find only that perception can be modulated by relevant sensorimotor input in a way that reflects an underlying multimodal speech representation that is shared across the sensory signals in the infant brain. In adults, articulatory suppression has only a modest effect on speech perception (37), and phonetic perception impairments are observed after strokes involving the left superior temporal region and the left parietal sulcus (38) coherent with brain imaging studies in healthy adults (39). However, recent work with direct cortical recordings using electrocorticography shows that during auditory speech perception, the superior and inferior regions of the ventral motor cortex are activated and follows a structure along acoustic features similar to the auditory cortex (40). Thus, even if the motor system is not required for speech perception, speech representations might be coded along similar dimensions between the auditory and the motor cortices. In adults, the predicted sensory consequences to a speech motor program are conveyed to the auditory cortex from the vSMC (lateral sensorimotor cortex); as well, the auditory and somatosensory error signals are conveyed to the vSMC such that corrective motor movements could take place (4). The pathways for communicating predicted auditory and sensorimotor patterns, as well as altering the motor program based on the feedback that exist in the adult brain, could already be present in the preverbal infant brain. These same circuits could be involved in potentially modulating the audio-motor circuits within the motor cortex, which in turn could mediate auditory perception.

Across sensory systems, substantial initial organization is established before postnatal experience through the mechanisms of spontaneously generated patterns of neural activity in early and prenatal development (41). While the bulk of available empirical evidence is based on animal models, these principles plausibly extend to the development of sensory systems in humans (42). Here, we propose that activity-dependent processes may critically shape the initial motor and sensorimotor foundations for speech production, and the sensorimotor system calibrated in this way interacts with the early emerging speech network and experience. Rhythmic stereotypies such as tongue protrusion and retraction observed until about 3 mo of age, have been suggested to be a form of self-generated rhythmic activations that induce activity-dependent development of the aerodigestive system (43). Thus, movement-induced sensory feedback in the earliest days of development could lead to the initial formation of the sensorimotor maps of the speech articulators (e.g., lips and tongue). To the extent that the motor primitives to speaking are shared with other functions of the articulators, such as aerodigestion, the initial refinement of these motor primitives may be shared early on in development. A critical link between the early sensorimotor mapping relevant for the articulatory space for speech and the human speech and language network, could be established starting prenatally during the third trimester when both sensorimotor organization and the emergence of the cortical language network are underway.

What evidence is there to suggest that these sensorimotor mappings are linguistically meaningful? The cortical language network canonical in the adult brain is present already from the third trimester of gestation (20). Neuroimaging evidence implies substantial prenatal and early postnatal development and organization of the human language network that forms the basis for functioning speech perception and production systems (19, 44). Interaction between the calibrated motor and sensorimotor systems achieved through these activity-dependent and spontaneous processes described above and the phonetic system supported by the early emerging language network could abet the acquisition of the correspondence between articulatory and acoustic dimensions of speech. This presents a more efficient and general learning mechanism than the alternative, which is to define a precise combination of articulatory actions for each phoneme that the infant must learn.

Implications and Conclusions. In summary, the current results that sensorimotor speech information is relevantly integrated with auditory speech processing in infants who are prebabbling provide insights into the sensorimotor–auditory speech interactions prior to production or extensive perceptual experience. In turn, this reveals that the human language system is robustly multisensory not only following full acquisition but already early in development and during the acquisition period. These results have implications for congenital oral-motor dysmorphologies and disorders. Contrary to the view that interventions will be impactful following babbling (7–9 mo) but not before, if a bidirectional perception–oral-motor link is present already at a younger age, and speech representation in infants is already multisensory, then a disrupted motor system could impact speech acquisition from early on. It remains to be

examined whether there are long-term influences from conditions that more fully limit oral-motor movements in young infants.

Materials and Methods

Participants. Thirty-two English-learning infants (19 males, 13 females; mean age, 112 d; SD, 7.83 d) recruited from the greater Vancouver area, Canada, were included in the study. An additional 25 infants were tested but excluded due to excessive movement artifacts, technical issues, or insufficient data (*SI Appendix*). Of the 32 infants in the sample, 12 infants completed both the passive listening (experiment 1) and the oral-motor inhibition during passive listening (experiment 2) experiments; the testing order was counterbalanced across infants such that six infants first completed experiment 1. Ten additional infants completed experiment 1 and an additional 10 infants completed experiment 2, such that 22 infants were included in each experiment (*SI Appendix*, Sample size estimation). The infants' primary caregivers provided informed consent prior to the experiment. The research was approved by the University of British Columbia Behavioral Research Ethics Board (Certificate H95-80023).

Stimuli. Three sound tokens were selected from a synthesized 16-step continuum (26): a voiced bilabial stop (/3ba/), a voiced dental stop (/9ɗa/), and a voiced retroflex stop (/15ɖa/). The stimuli were equal in duration (275 ms) and were precisely matched for low-level acoustic features (*SI Appendix*, Fig. S1).

Experimental Paradigm. We used a similar auditory mismatch design to previous infant speech perception studies (13, 34). Each trial consisted of four consecutive syllables; the first three syllables were repetitions of the same syllable, and the fourth syllable was either a repetition of the preceding three syllables (standard trial) or a different syllable that crossed the phonetic category boundary (deviant trial). The syllable-to-syllable stimulus-onset asynchrony was 600 ms and intertrial interval was 4 s. Infants were exposed to a maximum of 120 trials (60 standard trials and 60 deviant trials) per experiment. Standard trials were repetitions of syllables from a single phonetic category (/ba/, /ɗa/, or /ɖa/), and deviant trials consisted of a phonetic category change on the fourth syllable in both directions for each phonetic contrast (ba/ to /ɗa/, /ɗa/ to /ba/, /ɗa/ to /ɖa/, and /ɖa/ to /ɗa/). The number of the standard /ɗa/ trials was doubled to achieve a balanced cumulative number of each of the three syllables heard across the experiment, while maintaining an equal number of standard and deviant trials, since even 3- to 4-mo-old infants are sensitive to the probabilistic regularities of global and local changes (31). The trial presentation was randomized across all possibilities, ensuring that each infant was exposed to all seven trial types in a randomized and balanced manner.

Procedure. The infant was seated on the caregiver's lap while wearing an EEG cap and facing a computer monitor in an acoustically shielded room. The screen was placed ∼60 cm from the seated infant and displayed a dynamic visual animation for the infant to watch. Speech sounds were presented at 70 dB from an audio speaker (Fostex 6301NX) placed behind the screen. The experimenter monitored the infant from outside the acoustically shielded room through a camera mounted inside the room, and presented the stimuli using a custom-written program on Psychophysics Toolbox (45) in Matlab (2016b). If the infant began to show discomfort or if the caregiver signaled to stop, the experimenter terminated the study.

EEG Acquisition. EEG data were collected at a sampling rate of 1,000 Hz with a 64-electrode geodesic sensor net (EGI; N400 amplifier) referenced to the vertex (Cz). The net was placed on the infant's head relative to the anatomical markers while the infant sat on the caregiver's lap. The maximal impedance was kept under 40 kΩ.

EEG Preprocessing. EEG preprocessing analyses were conducted using functions from EEGLAB (46). First, the continuous EEG data were bandpass filtered from 0.5 to 20 Hz. The filtered data were segmented into 4-s epochs starting from −0.2 to 3.8 s from the onset of the first syllable. The length of the epoch included a 200-ms prestimulus period prior to the onset of the first syllable and ended 2 s after the onset of the fourth syllable. Following artifact rejection (*SI Appendix*), the data were re-referenced to the mean voltage. The trials were further collapsed based on Phonetic Contrast (/ba/-/ɗa/ and /ɗa/-/ɖa/) and Condition (Standard and Deviant) (*SI Appendix*). To minimize the potential effects of slow drifts on the fourth syllable analyses, we applied a baseline correction considering as baseline the mean voltage of the entire time window preceding the fourth syllable onset (−0.2 to 1.8 s from the trial onset); during this time window, the stimuli are identical between the standard and deviant trials.

EEG Data Analysis.

Data-driven analyses. The ERP differences between the standard and deviant trials were examined using a cluster-based nonparametric statistic for each experiment. The nonparametric cluster-based permutation combines clustering and randomization procedures to identify the spatiotemporal clusters (electrodes and time points) that showed statistically distinct responses (47). We first examined whether there is a main effect of Condition (Standard vs. Deviant) by collapsing across the two phonetic contrasts; a cluster-based permutation paired *t* test was conducted between standard and deviant trials with a cluster-alpha threshold of 0.1, a minimal cluster size of two electrodes, and 5,000 permutations over the 800-ms period following the fourth syllable (1.8 to 2.6 s from the trial onset) (48). We also tested for an interaction of Condition by Phonetic contrast; to conduct this analysis in Fieldtrip, we calculated the difference between the standard and deviant trials, and compared this difference in a paired *t* test between the /ba/-/ɗa/ and the /ɗa/-/ɖa/ contrasts. If the interaction cluster-based permutation test was significant but the test of main effect was not, follow-up tests were conducted on the two phonetic contrasts separately. We averaged the voltage values from the sensors and the time window selected following this procedure, per subject and for each experimental condition for further analyses.

Control analyses. To ensure that the response was specific to the last syllable and not due to any systematic noise, we also averaged the voltage in the same cluster of sensors, and the same time window following each of the first three syllables (i.e., Repetitions) and compared them against the responses following the fourth syllable. If the distinct ERP responses between the standard and deviant trials reflected a response to phonetic category change, then a difference is expected following the fourth syllable but not on the Repetitions.

ANOVA. For each experiment, we conducted a three-way repeated-measures ANOVA with Condition (Standard and Deviant), Syllable Position (Repetitions [1 to 3] and Fourth), and Phonetic Contrast (/ba/-/ɗa/ and /ɗa/-/ɖa/) as factors. Because none of the factors within the repeated-measures ANOVA exceeded more than two levels, a Mauchly's test of sphericity was not required.

Bayesian regression analysis. We used Bayesian multilevel regression models to quantify the strength of evidence using Bayesian confidence intervals on the main effect of interest (Condition by Phonetic Contrast). Model fit was implemented using the package *brms* (49) v2.12 within the R computing environment. Standardized (*z*-scored) data were fit to a varying intercept model with Condition (Standard and Deviant), Phonetic Contrast (/ba/-/ɗa/ and /ɗa/-/ɖa/), and Condition by Phonetic Contrast interaction specified as fixed effects. Individual participants were modeled as random intercepts. Weakly informative priors were selected for each parameter. The mean and the highest posterior density interval (HPDI) of the β of the fixed effects and the interaction effect were estimated. Four independent chains, each with 1,000 warmup samples and 2,000 iterations were run, resulting in a total of 4,000 draws from the posterior. Model fit was assessed for good convergence as indicated by Gelman–Rubin $\hat{R} < 1.01$. To examine whether the $\beta_{\text{Condition}}$ is modulated by different levels of the Phonetic Contrast, a linear model of slope $\gamma = \beta_{\text{Condition}} + \beta_{\text{Condition*Phonetic Contrast}}$ is specified and reported (*SI Appendix*).

Data Availability. Anonymized EEG data have been deposited in the Open Science Framework (https://osf.io/my496/).

ACKNOWLEDGMENTS. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2015-03967) and the Canada Foundation for Innovation John R. Evans Leaders Fund (33096) awarded to J.F.W., from Global Research Alliance in Language–Pontificia Universidad Católica de Chile awarded to M.P., and European Research Council under the European Union's Horizon 2020 Research and Innovation Program (Grant Agreement 695710) awarded to G.D.-L.

1. C. Scholes, J. I. Skipper, A. Johnston, The interrelationship between the face and vocal tract configuration during audiovisual speech. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 32791–32798 (2020).
2. H. C. Yehia, T. Kuratate, E. Vatikiotis-Bateson, Linking facial animation, head motion and speech acoustics. *J. Phonetics* **30**, 555–568 (2002).
3. M. Keough, D. Derrick, B. Gick, Cross-modal effects in speech perception. *Annu. Rev. Linguist.* **5**, 49–66 (2019).
4. F. H. Guenther, *Neural Control of Speech* (MIT Press, 2016).
5. G. Hickok, J. Houde, F. Rong, Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron* **69**, 407–422 (2011).
6. T. Ito, M. Tiede, D. J. Ostry, Somatosensory function in speech perception. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1245–1248 (2009).
7. M. Masapollo, F. H. Guenther, Engaging the articulators enhances perception of concordant visible speech movements. *J. Speech Lang. Hear. Res.* **62**, 3679–3688 (2019).

8. D. Poeppel, M. F. Assaneo, Speech rhythms and their neural foundations. *Nat. Rev. Neurosci.* **21**, 322–334 (2020).
9. P. K. Kuhl, A. N. Meltzoff, The bimodal perception of speech in infancy. *Science* **218**, 1138–1141 (1982).
10. M. L. Patterson, J. F. Werker, Infants' ability to match dynamic phonetic and gender information in the face and voice. *J. Exp. Child Psychol.* **81**, 93–115 (2002).
11. F. Pons, D. J. Lewkowicz, S. Soto-Faraco, N. Sebastián-Gallés, Narrowing of intersensory speech perception in infancy. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10598–10602 (2009).
12. G. Hollich, R. S. Newman, P. W. Jusczyk, Infants' use of synchronized visual information to separate streams of speech. *Child Dev.* **76**, 598–613 (2005).
13. D. Bristow *et al*., Hearing faces: How the infant brain matches the face it sees with the speech it hears. *J. Cogn. Neurosci.* **21**, 905–921 (2009).
14. H. H. Yeung, J. F. Werker, Lip movements affect infants' audiovisual speech perception. *Psychol. Sci.* **24**, 603–612 (2013).
15. J. F. Werker, R. C. Tees, Phonemic and phonetic factors in adult cross-language speech perception. *J. Acoust. Soc. Am.* **75**, 1866–1878 (1984).
16. A. G. Bruderer, D. K. Danielson, P. Kandhadai, J. F. Werker, Sensorimotor influences on speech perception in infancy. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13531–13536 (2015).
17. D. Choi, A. G. Bruderer, J. F. Werker, Sensorimotor influences on speech perception in pre-babbling infants: Replication and extension of Bruderer et al. (2015). *Psychon. Bull. Rev.* **26**, 1388–1399 (2019).
18. M. S. Blumberg, H. G. Marques, F. Iida, Twitching in sensorimotor development from sleeping rats to robots. *Curr. Biol.* **23**, R532–R537 (2013).
19. G. Dehaene-Lambertz, The human infant brain: A neural architecture able to learn language. *Psychon. Bull. Rev.* **24**, 48–55 (2017).
20. J. Dubois *et al*., Exploring the early organization and maturation of linguistic pathways in the human infant brain. *Cereb. Cortex* **26**, 2283–2298 (2016).
21. F. Leroy *et al*., Early maturation of the linguistic dorsal pathway in human infants. *J. Neurosci.* **31**, 1500–1506 (2011).
22. D. K. Oller, *The Emergence of the Speech Capacity* (Lawrence Erlbaum Associates Publishers, 2000).
23. G. Dehaene-Lambertz, S. Dehaene, Speed and cerebral correlates of syllable discrimination in infants. *Nature* **370**, 292–295 (1994).
24. M. Mahmoudzadeh, F. Wallois, G. Kongolo, S. Goudjil, G. Dehaene-Lambertz, Functional maps at the onset of auditory inputs in very early preterm human neonates. *Cereb. Cortex* **27**, 2500–2512 (2017).
25. M. Peña, J. F. Werker, G. Dehaene-Lambertz, Earlier speech exposure does not accelerate speech acquisition. *J. Neurosci.* **32**, 11159–11163 (2012).
26. J. F. Werker, C. E. Lalonde, Cross-language speech perception: Initial capabilities and developmental change. *Dev. Psychol.* **24**, 672–683 (1988).
27. P. K. Kuhl, K. A. Williams, F. Lacerda, K. N. Stevens, B. Lindblom, Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* **255**, 606–608 (1992).
28. M. I. Garrido, M. Sahani, R. J. Dolan, Outlier responses reflect sensitivity to statistical structure in the human brain. *PLoS Comput. Biol.* **9**, e1002999 (2013).
29. C. Wacongne *et al*., Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20754–20759 (2011).
30. H. Tiitinen, P. May, K. Reinikainen, Attentive novelty detection in humans is governed by pre-aHentive sensory memory. *Nature* **372**, 90–92 (1994).
31. A. Basirat, S. Dehaene, G. Dehaene-Lambertz, A hierarchy of cortical responses to sequence violations in three-month-old infants. *Cognition* **132**, 137–150 (2014).
32. G. Dehaene-Lambertz, M. Peña, Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport* **12**, 3155–3158 (2001).
33. J. Bertoncini, R. Bijeljac-Babic, P. W. Jusczyk, L. J. Kennedy, J. Mehler, An investigation of young infants' perceptual representations of speech sounds. *J. Exp. Psychol. Gen.* **117**, 21–33 (1988).
34. K. Mersad, G. Dehaene-Lambertz, Electrophysiological evidence of phonetic normalization across coarticulation in infants. *Dev. Sci.* **19**, 710–722 (2016).
35. G. Gennari, S. Marti, M. Palu, A. Fló, G. Dehaene-Lambertz, Orthogonal neural codes for phonetic features in the infant brain. *bioRxiv* [Preprint] (2021). https://doi.org/10.1101/2021.03.28.437156. Accessed 28 March 2021.
36. A. M. Liberman, F. S. Cooper, D. P. Shankweiler, M. Studdert-Kennedy, Perception of the speech code. *Psychol. Rev.* **74**, 431–461 (1967).
37. R. C. Stokes, J. H. Venezia, G. Hickok, The motor system's [modest] contribution to speech perception. *Psychon. Bull. Rev.* **26**, 1354–1366 (2019).
38. K. Kim *et al*., Neural processing critical for distinguishing between speech sounds. *Brain Lang.* **197**, 104677 (2019).
39. G. Dehaene-Lambertz *et al*., Neural correlates of switching from auditory to speech perception. *Neuroimage* **24**, 21–33 (2005).
40. C. Cheung, L. S. Hamiton, K. Johnson, E. F. Chang, The auditory representation of speech sounds in human motor cortex. *eLife* **5**, e12577 (2016).
41. R. Khazipov, H. J. Luhmann, Early patterns of electrical activity in the developing cerebral cortex of humans and rodents. *Trends Neurosci.* **29**, 414–418 (2006).
42. Z. Molnár, H. J. Luhmann, P. O. Kanold, Transient cortical circuits match spontaneous and sensory-driven activity during development. *Science* **370**, eabb2153 (2020).
43. N. Keven, K. A. Akins, Neonatal imitation in context: Sensorimotor development in the perinatal period. *Behav. Brain Sci.* **40**, e381 (2017).
44. M. A. Skeide, A. D. Friederici, The ontogeny of the cortical language network. *Nat. Rev. Neurosci.* **17**, 323–332 (2016).
45. D. H. Brainard, The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
46. A. Delorme, S. Makeig, EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).
47. E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
48. R. Oostenveld, P. Fries, E. Maris, J. M. Schoffelen, FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).
49. Bürkner, Paul-Christian, brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw.* **80**, 10.18637/jss.v080.i01 (2017).1548-7660.
50. W. M. Kier, K. K. Smith, Tongues, tentacles and trunks: The biomechanics of movement in muscular-hydrostats. *Zool. J. Linn. Soc.* **83**, 307–324 (1985).

PNAS
www.pnas.org

Supplementary Information for
Neural Indicators of articulator-specific sensorimotor influences on infant speech perception.

Dawoon Choi, Ghislaine Dehaene-Lambertz, Marcela Peña & Janet F. Werker

Corresponding author: Dawoon Choi and Janet F. Werker
Email: sheri.d.choi@gmail.com, jwerker@psych.ubc.ca

This PDF file includes:

Supplementary Information Text

Methods

Sample size estimation. The sample size was determined based on data from the study by (1) who used a similar experimental design with 3-month-old infants (n=25). Based on their reported mean and standard error to the effect of condition (Standard and Deviant) at the level of the test syllable, we estimated an effect size Cohen's d of 0.83. To achieve a comparable effect size, a minimum sample size of 21 is required for power of 0.95 as estimated using G*Power 3.1 (2).

Stimuli. We selected synthesized speech sounds from three consonant categories (Bilabial /b/, Dental /ɗ/, Retroflex /ɖ/) paired with the vowel /a/ synthesized as a 16-step continuum, whereby F2 and F3 varied in the starting frequency and transition with the Mattingly synthesizer on the VAX 11/780 (Haskins Laboratories, New Haven, Connecticut) (3). See Figure S1

Artifact rejection. Channels placed around the periphery have been reported to frequently contain artifacts (4), therefore, we excluded 11 periphery channels around the scalp including the two mastoid channels (Figure S2). For each epoch, a channel was marked if it exceeded a local deviation of $150\ \mu v$, exceeded the absolute threshold of $100\ \mu v$, or if the amplitude was larger than 5 times the standard deviation of the mean of the data in this channel overall. Trials that contained more than 25% marked channels and channels that were marked in more than 40% of the trials were removed from further analyses. Following this artifact rejection criteria, 10.68 trials (SD = 0.45) and 1.18 channels (SD = 1.65) were excluded in Experiment 1, and 11.36 trials (SD = 8.23) and 0.5 channels (SD = 1.65) were excluded in Experiment 2 on average, per infant. While the maximum number of trials per experiment was 120, not all infants completed all trials, such that after artifact detection, Experiment 1 yielded an average of 95.68 trials (SD = 28.20) per infant and Experiment 2 yielded an average of 79.27 trials (*SD* = 24.00) per infant for further analyses. The data were re-referenced to the mean voltage (5), and averaged based on the trial type (each unique type of standard and deviant trials). Following trial type averaging, if there were channels that exceeded $40\ \mu v$ they were excluded from further analysis. The trials were further collapsed into Phonetic Contrast (/ba/-/ɗa/ and /ɗa/-/ɖa/) and Condition (Standard and Deviant). Epochs with long time-windows may contain slow drifts that are difficult to detect with automatic pre-processing. Thus, to minimize the potential effects of slow drifts on the 4th syllable analyses, we applied a baseline correction considering as baseline the mean voltage of the entire time-window preceding the 4th syllable onset (-0.2 s to 1.8 s from the trial onset), this period being the same in standard and deviant trials.

Trial numbers per experimental condition. As described in the Methods section of the paper, there were overall 7 unique trial types (3 standard trialtypes and 4 deviant trialtypes). Each trialtype was presented up to a maximum of 15 trials per experiment, with the exception of standard /ɗa/ condition which was presented up to 30 trials. This was to achieve an equal distribution of the cumulative number of the standard and deviant trials, and to achieve an equal number of each of the /ba/, /ɗa/, and /ɖa/ sound tokens throughout the experiment. Previous work has shown that infants, like adults, are sensitive to the 'global' regularities in the standard vs. deviant trials (6), and also perceptually learn relevant phonetic categories through distributional learning (7). The average number of epochs included in the averaged ERPs per trialtype across participants for Experiment 1 is presented in Table 1, and for Experiment 2, is presented in Table 2. Further, Table 3 shows the Trialtypes that were averaged as the standard and deviant trials for each of the phonetic contrasts (/ba/-/ɗa/ contrast, and /ɗa/-/ɖa/ contrast).

Comparison of standard trials across Experiments. As a control analysis to examine whether the data are comparable across Experiments 1 and 2, we assessed whether the standard trials showed statistical difference across experiments. A cluster-based permutation paired t-test was conducted between the averaged standard trials in Experiment 1 vs. Experiment 2, with a cluster-alpha threshold of 0.1, a minimal cluster-size of two electrodes, and 5,000 permutations over the

800 ms period following the 4th syllable (1.8 s to 2.6 s from the trial onset). We did not detect any clusters that reached statistical significance (the smallest cluster p-value was 0.93). Figure S3 shows the data plotted for each standard trials (/ba/, /ɗa/, and /ɖa/) from Experiment 1 and 2 within the same plots: the first column of Figure S3 shows the grand-average of sensors from a left posterior cluster of sensors (as identified from the cluster-based permutation analysis with data from Experiment 1), and the second column of Figure S3 shows the grand-average of sensors from a posterior cluster of sensors (as identified from the non-parametric cluster-based permutation test with data from Experiment 2).

Bayesian Regression Modelling

Model specification. The likelihood is described as a Gaussian with two parameters mu and sigma. Under the Gaussian distribution it is assumed that the outcome y_i is normally distributed around the mean μ_i with error σ_e.

$$y_i = Normal(\mu_i, \sigma_e)$$
$$u_i = \alpha + \beta x_i$$

Under the multilevel framework, this can be extended to a varying intercept model such that in addition to the overall grand intercept α, each individual within the cluster j is given a unique intercept $\alpha_{j[i]}$. Each member is assumed to also have a normal distribution, and as a result, an additional variance component σ_α known as a hyperprior (8) is also estimated. Under the Bayesian framework, a prior distribution is specified for each parameter modelled. In the current experiment, we modelled each subject as a random varying-intercept from the overall grand intercept, and fixed effects of Condition, Phonetic Contrast and Condition by Phonetic Contrast Interaction. The slope of the Condition to the /ba/-/da/ contrast is specified as $\gamma_{ba-da} = \beta_{Condition}$, and to the /da/-/Da/ contrast is specified as $\gamma_{ba-da} = \beta_{Condition} + \beta_{Condition*Phonetic\ Contrast}$

The parameters used in this model are as follows

$$amplitude_i = Normal(\mu_i, \sigma_e)$$
$$\mu_i = \alpha + \alpha_{subject[i]} + \gamma_{condition} + \beta_{Phonetic\ Contrast}$$
$$\gamma_{condition} = \beta_{Condition} + \beta_{Condition*Phonetic\ Contrast}$$
$$\alpha_j \sim Normal(\alpha, \sigma_{subject})$$
$$\alpha \sim Normal(0, 2)$$
$$\beta_{Condition} \sim Normal(0, 2)$$
$$\beta_{Phonetic\ Contrast} \sim Normal(0, 2)$$
$$\beta_{Condition*Phonetic\ Contrast} \sim Normal(0, 2)$$
$$\sigma_{subject} \sim HalfCauchy(1)$$
$$\sigma_e \sim HalfCauchy(1)$$

There are two sources of variation: The standard deviation of the residual error of the overall model σ_e and the standard deviation of the varying intercepts by subject $\sigma_{subject}$. The inclusion of the varying intercept term means that the standard deviation of the population of varying intercepts will also inform the estimation of the overall intercept. Such a partial pooling strategy is considered an advantage of a multilevel model which contributes to better estimation than a single level models, in particular for repeated measures (8).

Experiment 1. The posterior estimation for all parameters in the Bayesian multilevel model is presented in Table 4. The model includes Condition, Phonetic Contrast, and Condition by Phonetic Contrast interaction as fixed effects. Subjects are modelled as random effects. The Mean, the Standard Error (SE), and lower and upper values of the 95% confidence intervals of each $\beta_{Condition}$, $\beta_{Phonetic\ Contrast}$, and $\beta_{Condition*Phonetic\ Contrast}$ are reported, as well as the overall model intercept, Sigma, and standard deviations of the random effects. The Rhat values indicate the overall model fit; Good convergence is indicated by Gelman-Rubin $\hat{R}$ <1.01.

The change in Condition to the /ba/-/da/ and the /da/-/Da/ phonetic contrasts separately is indicated in Table 5. The table presents the mean and the upper and lower 95% CI values to γ_{ba-da} and γ_{da-Da} in Experiment 1.

Experiment 2. The posterior estimation for all parameters in the Bayesian multilevel model to data from Experiment 2 is presented in Table 6. The model includes Condition, Phonetic Contrast, and Condition by Phonetic Contrast interaction as fixed effects. Subjects are modelled as random effects. The Mean, the Standard Error (SE), and lower and upper values of the 95% confidence intervals of each $\beta_{Condition}$, $\beta_{Phonetic\ Contrast}$, and $\beta_{Condition*Phonetic\ Contrast}$ are reported, as well as the overall model intercept, Sigma, and standard deviations of the random effects. The Rhat values indicate the overall model fit; Good convergence is indicated by Gelman-Rubin $\hat{R}$ <1.01.

The change in Condition to the /ba/-/da/ and the /da/-/Da/ phonetic contrasts separately is indicated in Table 7. The table presents the mean and the upper and lower 95% CI values to γ_{ba-da} and γ_{da-Da} in Experiment 2.

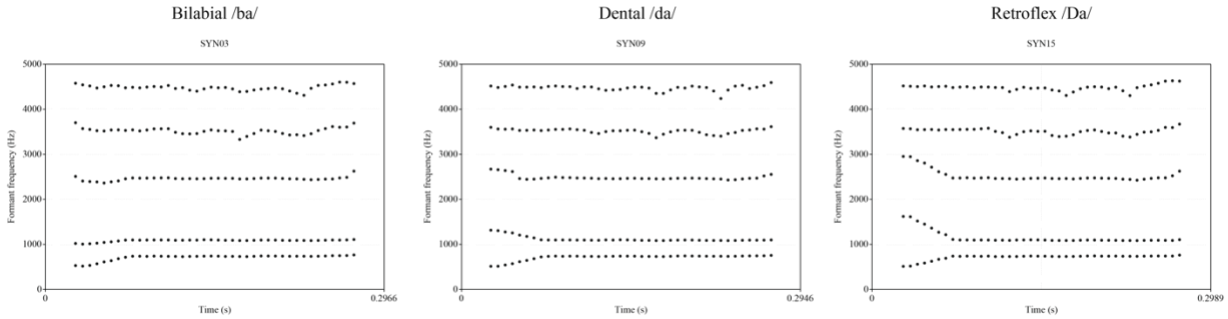

Fig. S1. Center frequency of formant transitions of bilabial /ba/ (left), dental /ɗa/ (center) and retroflex /ɖa/ (right) synthesized syllables.

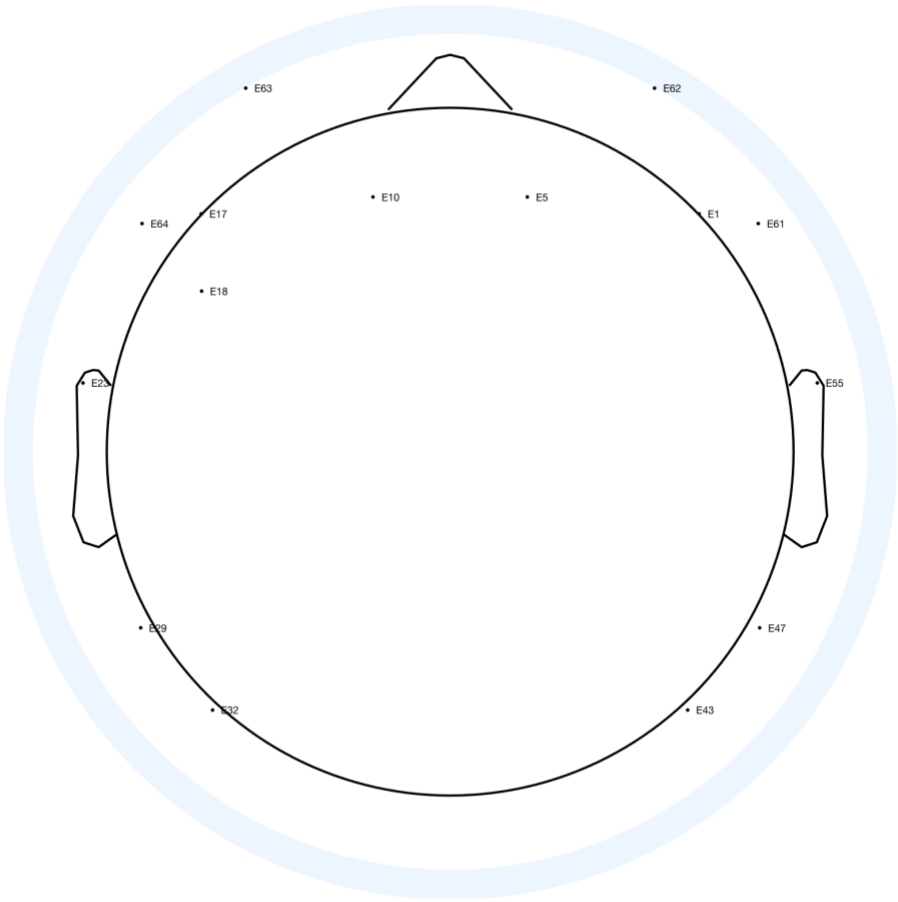

Fig. S2. The location and labels of sensors in the periphery excluded from analyses.

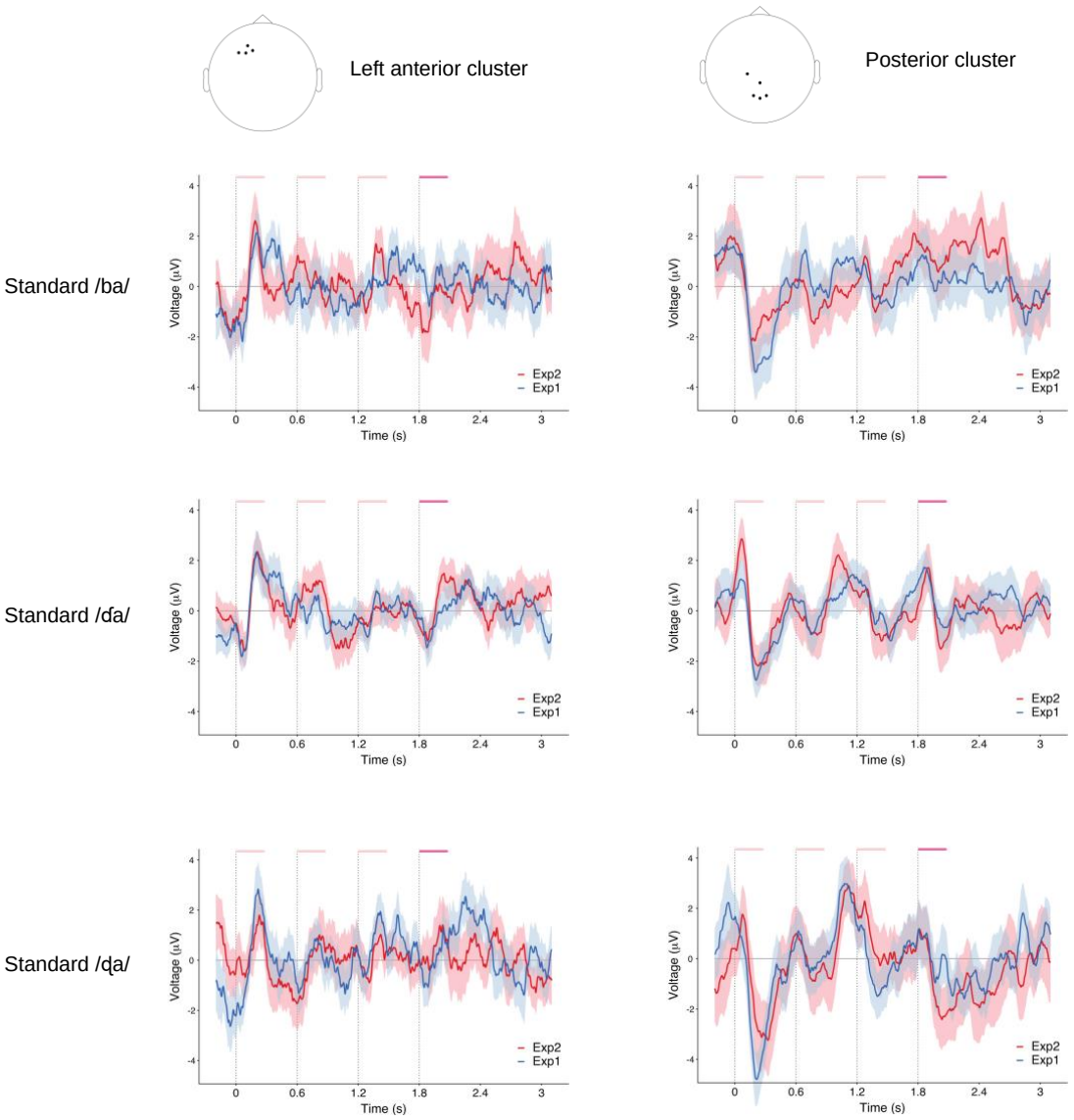

Fig. S3. The data from each of the three standard trials (/ba/ on the first row, /ɗa/ on the second row, and /ɖa/ on the third row). Each plot depicts the data from Experiment 1 (in blue) and Experiment 2 (in red). The first column shows grand-averages of data for each trialtype from the left anterior cluster of sensors. The second column shows grand-averages from the posterior cluster of sensors.

Table S1. Summary table of average number of epochs per trial type in Experiment 1

Trialtype	Mean	Standard Deviation
/ba/ standard	10.64	3.85
/ɗa/ standard	21.27	7.33
/ɋa/ standard	10.50	3.71
/ɗa/-/ba/ change	10.50	3.47
/ba/-/ɗa/ change	10.77	3.60
/ɋa/-/ɗa/ change	10.95	3.32
/ɗa/-/ɋa/ change	10.36	3.77

Table S2. Summary table of average number of epochs per trial type in Experiment 2

Trialtype	Mean	Standard Deviation
/ba/ standard	9.09	3.25
/ɗa/ standard	16.64	5.74
/ɖa/ standard	8.55	3.07
/ɗa/-/ba/ change	8.41	3.03
/ba/-/ɗa/ change	8.59	3.46
/ɖa/-/ɗa/ change	8.14	2.98
/ɗa/-/ɖa/ change	8.55	3.02

Table S3. Summary of the trial types that were averaged for each condition and phonetic condition that were submitted for comparison.

	Standard condition	**Deviant condition**
/ba/-/ɗa/ contrast	/ba/ standard /ɗa/ standard	/ba/-/ɗa/ change /ɗa/-/ba/ change
/ɗa/-/ɖa/ contrast	/ɗa/ standard /ɖa/ standard	/ɗa/-/ɖa/ change /ɖa/-/ɗa/ change

Table S4. Summary of posterior density estimation in Experiment 1.

Parameter	Mean	SE	Q2.5	Q97.5	Rhat
Intercept	-0.349	0.199	-0.741	0.04	1
Condition	0.552	0.271	0.011	1.085	1
Phonetic Contrast	-0.044	0.276	-0.582	0.498	1.002
Condition: Phonetic Contrast	0.388	0.383	-0.363	1.124	1
Sd random effects	0.194	0.13	0.009	0.48	1.001
Sigma	0.925	0.076	0.788	1.086	1.003

Table S5. Gamma estimations in Experiment 1.

Parameter	Mean	Q2.5	Q97.5
γ_{ba-da}	0.94	0.39	1.477
γ_{da-Da}	0.552	0.011	1.085

Table S6. Summary of posterior density estimation in Experiment 2.

Parameter	Mean	SE	Q2.5	Q97.5	Rhat
Intercept	-0.608	0.2	-1.002	-0.212	1.002
Condition	1.155	0.271	0.621	1.696	1.001
Phonetic Contrast	0.631	0.264	0.117	1.151	1
Condition: Phonetic Contrast	-1.152	0.381	-1.889	-0.426	1.001
Sd random effects	0.278	0.155	0.016	0.586	1.003
Sigma	0.887	0.076	0.751	1.048	1.002

Table S7. Gamma estimations in Experiment 2.

Parameter	Mean	Q2.5	Q97.5
γ_{ba-da}	0.003	-0.552	0.548
γ_{da-Da}	1.155	0.621	1.696

SI References

1. Mersad, K., & Dehaene-Lambertz, G. (2016). Electrophysiological evidence of phonetic normalization across coarticulation in infants. *Developmental Science*, *19*(5), 710–722. https://doi.org/10.1111/desc.12325
2. Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149-1160.
3. Werker, J. F., & Lalonde, C. E. (1988). Cross-Language Speech Perception: Initial Capabilities and Developmental Change. *Developmental Psychology*, *24*(5), 672–683. https://doi.org/10.1037/0012-1649.24.5.672
4. Fujioka, T., Mourad, N., He, C., & Trainor, L. J. (2011). Comparison of artifact correction methods for infant EEG applied to extraction of event-related potential signals. *Clinical Neurophysiology*, *122*(1), 43-51.
5. Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, & Computers*, *30*(1), 34-43.
6. Basirat, A., Dehaene, S., & Dehaene-Lambertz, G. (2014). A hierarchy of cortical responses to sequence violations in three-month-old infants. *Cognition*, *132*(2), 137-150.
7. Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111.
8. McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.