Sensorimotor influences on speech perception in pre-babbling infants: Replication and extension of Bruderer et al. (2015)

Dawoon Choi¹ · Alison G. Bruderer² · Janet F. Werker¹

Published online: 29 April 2019
© The Psychonomic Society, Inc. 2019

Abstract

The relationship between speech perception and production is central to understanding language processing, yet remains under debate, particularly in early development. Recent research suggests that in infants aged 6 months, when the native phonological system is still being established, sensorimotor information from the articulators influences speech perception: The placement of a teething toy restricting tongue-tip movements interfered with infants' discrimination of a non-native contrast, /Da/-/da/, that involves tongue-tip movement. This effect was selective: A different teething toy that prevented lip closure but not tongue-tip movement did not disrupt discrimination. We conducted two sets of studies to replicate and extend these findings. Experiments 1 and 2 replicated the study by Bruderer et al. (Proceedings of the National Academy of Sciences of the United States of America, 112 (44), 13531–13536, 2015), but with synthesized auditory stimuli. Infants discriminated the non-native contrast (dental /da/ - retroflex /Da/) (Experiment 1), but showed no evidence of discrimination when the tongue-tip movement was prevented with a teething toy (Experiment 2). Experiments 3 and 4 extended this work to a native phonetic contrast (bilabial /ba/ - dental /da/). Infants discriminated the distinction with no teething toy present (Experiment 3), but when they were given a teething toy that interfered only with lip closure, a movement involved in the production of /ba/, discrimination was disrupted (Experiment 4). Importantly, this was the same teething toy that did not interfere with discrimination of /da/-/Da/ in Bruderer et al. (2015). These findings reveal specificity in the relation between sensorimotor and perceptual processes in pre-babbling infants, and show generalizability to a second phonetic contrast.

Keywords Speech perception · Infancy · Language acquisition · Multisensory

Introduction

Decades of research have investigated speech sound discrimination in infancy, but only recently have researchers begun to explore how oral-motor processes might interact with, or contribute to, phoneme discrimination. Phonemes are the elementary building blocks of speech, and their combinatorial power

is one of the key contributors to the rich complexity of language. Establishing the native phoneme inventory is thus an essential first step toward acquisition of the native language. Infants, like adults, perceive phonemes in a categorical-like manner, with better discrimination across than within phonetic category boundaries (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Werker & Lalonde, 1988). Initially, pre-verbal infants are sensitive to the phonetic categories of both native and non-native languages (Werker, Gilbert, Humphrey, & Tees, 1981). Within the first year of life, infants' perceptual sensitivities simultaneously become sharpened to native phonetic distinctions (Kuhl et al., 2006; Narayan, Werker, & Beddor, 2010) and diminished to many non-native ones (e.g., Best & McRoberts, 2003; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Lalonde, 1988; Werker & Tees, 1984; see Maurer & Werker, 2014, for a review, including interesting exceptions to this common pattern).

The acquisition of a proficient oral communication system requires not only a richly structured auditory linguistic system,

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13423-019-01601-0>) contains supplementary material, which is available to authorized users.

✉ Dawoon Choi
dchoi@psych.ubc.ca

¹ Department of Psychology, The University of British Columbia, 2136 West Mall, Vancouver, BC V6T 1Z4, Canada

² North Island College, 2300 Ryan Road, Courtenay, BC V9N 8N6, Canada

with phonemes as the basic units, but also the kinematic control for production. During speech production, the acoustics of speech that carry linguistically meaningful information occur synchronously with salient visible and felt facial movements; each form of information is the product of an underlying source – the gestures used to produce speech. The dynamic visual features associated with different facial configurations and the accompanying head movements are correlated in time with the dynamic auditory signals of speech, such as the speech envelope and the fundamental frequency (Chandrasekaran, Trubanova, Stillitano, Caplier, & Ghazanfar, 2009; Munhall & Vatikiotis-Bateson, 2004). Shared amodal attributes (e.g., intensity, duration, tempo, and rhythm) in both visual and auditory speech signals reflect articulator actions, and such multimodal information enhances speech perception (e.g., Erber, 1975; Grant & Seitz, 2000; Sumby & Pollak, 1954). Neurally, concurrent visual stimuli modulate activity in auditory pathways (Reale et al., 2007; Van Wassenhove, Grant, & Poeppel, 2004), and both auditory and audiovisual speech activate the speech motor system in the perceiver (Okada & Hickok, 2009; Skipper, van Wassenhove, Nusbaum, & Small, 2007). Further, there is evidence that self-produced articulatory movements can alter adult perception of speech produced by others (Sams, Möttönen, & Sihvonen, 2005; Scott, Yeung, Gick, & Werker, 2013), as can disrupting articulator-specific areas in the premotor cortex (Möttönen, Dutton, & Watkins, 2013). However, not every study replicates the influence of self-produced movements on perception (see Matchin, Groulx, & Hickok, 2014). Moreover, the extent to which auditory-motor interactions influence the processing of acoustic speech in a feedforward manner is actively debated (Hickok, Houde, & Rong, 2011; Rauschecker & Scott, 2009).

The structural development of adult-like language networks – including the links between speech perception, word comprehension, and motor production – occurs over many years, but the neural organization present at birth may facilitate the rapid acquisition of language. Neuroimaging studies of young infants show that the structural connectivity supporting language processing is present at or shortly after birth, including very early appearing pathways linking the auditory cortices to the premotor areas (Leroy et al., 2011; Perani et al., 2011), and rapidly developing links to the frontal language areas (e.g., Broca's area) (Dehaene-Lambertz et al., 2006; Dubois et al., 2015). In line with the anatomical data, in 6-month-old infants, the left inferior frontal region is active during speech perception, which suggests that an early link between the auditory and motor processes is available in infants before the onset of babbling (Imada et al., 2006). Further research using MEG has shown that speech motor areas (including the inferior frontal region) and auditory areas (superior temporal region) activate equally during

both native and non-native syllable discrimination at 7 months. By 11–12 months, recognizing non-native speech results in greater activation of motor areas compared to native speech, while recognizing native speech results in greater activation in auditory areas compared to non-native speech (Kuhl, Ramírez, Bosseler, Lin, & Imada, 2014). These results suggest that motor activation accompanies speech perception even in preverbal infants; in particular, such activation may be most important for processing less familiar sounds, and less involved in perceiving phones for which there are well-established, “native” representations (Kuhl et al., 2014). Thus, early-appearing oral-motor influences on infant speech perception may be enabled by the structural and functional connectivity that is in place between sensory and motor areas, even prior to the onset of canonical babbling.

In fact, growing behavioral evidence suggests that speech perception is multisensory in infancy. Very young infants, with minimal sensory experience, can match the sounds of vowels (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999) and consonants (MacKain, Best, & Strange, 1981; Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009) with visual mouth movements. At the neural level, habituation to visually articulated vowels influences the evoked responses to auditory vowels in a phoneme-specific manner in 3-month-old infants (Bristow et al., 2009). These data indicate that infants are prepared to detect, or rapidly learn, the coupling between seen and heard speech.

Because the shared information supporting this coupling comes from the underlying gestures used to produce speech (see Best, 1995), it is of interest to investigate whether auditory speech perception is influenced by feedback from oral-motor movements even in early infancy.

Support for a relation between speech perception and oral-motor movements can be found in research on infant imitation. When tested in auditory-visual matching tasks at 4–5 months of age, nearly half of the infants make mouth movements – particularly during matching trials (Kuhl & Meltzoff, 1982; see also Patterson & Werker, 1999) – and many produce vocalizations that perceptually match the vowels produced by adult speakers (Kuhl & Meltzoff, 1996). Imitation attempts, as evident in relevant mouth movements, have more recently been reported in newborn infants (Chen, Striano, & Rakoczy, 2004; Coulon, Hemimou & Streri, 2013). Furthermore, recent studies explored the reverse relation, of production influencing infants' perception, and demonstrated a correlational relationship between infants' own babbling and their consonant discrimination (DePaolis, Vihman & Keren-Portnoy, 2011; Majorano, Vihman & DePaolis, 2014). In auditory-visual vowel-matching studies, infants were better able to match heard and seen vowels that they could themselves produce (Streri, Coulon, Marie, & Yeung, 2016). Such studies

suggest that even preverbal infants may have an “intermodal representation” of speech information across the auditory, visual, and motor domains (Kuhl & Meltzoff, 1982; Guellaï, Streri, & Yeung, 2014).

Recently, experimental studies directly manipulating infant oral-motor movements and testing the impact on speech perception have been reported. In one of only two published empirical behavioral experiments (to our knowledge), manipulating infant lip configurations were found to influence vowel-matching of auditory and visual speech at 4.5 months of age (Yeung & Werker, 2013). Specifically, infants were tested in an “ooo” – “eee” auditory-visual matching task, while parents held a finger or teething toy in their infant’s mouth. These manipulations led to a pursing of the lips (as would be used in an “ooo” production) or a stretching of the lips (as would be used in an “eee” production), respectively. The researchers found specific influences from each articulator manipulation on “ooo” versus “eee” auditory-visual matching. This seminal study introduced the method of manipulating oral-motor movements, but left open the possibility that the motor influence on speech perception was mediated by visual speech. Moreover, it involved testing infants on a native contrast for which the infants had months of listening and watching experience, and – because it involved vocalic sounds – involved speech sounds that infants may already have begun to produce themselves. Thus, the Yeung and Werker (2013) evidence of a link between oral-motor movements and speech perception could be explained by experiential-based learning.

In a more recent study of direct relevance to the current experiments, it was shown that interfering with infants’ own oral-motor movements can interfere directly with auditory speech discrimination of a non-native and hence unfamiliar consonant distinction (Bruderer, Danielson, Kandhadai, & Werker, 2015). In the 2015 paper “Sensorimotor influences on speech perception in infancy,” Bruderer and colleagues investigated whether limiting tongue-tip movement would alter English-learning 6-month-old infants’ (who are not yet producing consonant-vowel syllables) discrimination of a non-native phonetic contrast that, during mature production of the two sounds, differs only in the placement of the tongue-tip. Several studies have shown this Hindi (non-English) dental /da/ versus retroflex /Da/ phonetic contrast to be discriminable by English infants at 6–8 months, but not at 10–12 months (e.g., Best, McRoberts, LeFleur, & Silver-Isenstadt, 1995; Danielson, Bruderer, Kandhadai, Vatikotis-Bateson, & Werker, 2017; Peña, Werker, & Dehaene-Lambertz, 2012; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005; Werker & Tees, 1984). In Bruderer et al. (2015), 6-month-old infants did not discriminate the minimally-different /da-/ /Da/ phonetic contrast when their

tongue-tip movement was limited. Importantly, there was specificity in the sensorimotor influence: only a flat teether that interfered with tongue-tip movement attenuated discrimination (Experiment 2, Bruderer et al., 2015), while discrimination was maintained when a gummy teether – which allowed tongue-tip movement, but interfered with lip closure – was held in the infant’s mouth (Experiment 3, Bruderer et al., 2015), or when no sensorimotor manipulation was present (Experiment 1, Bruderer et al., 2015). The Bruderer et al. (2015) study, therefore, implicates sensorimotor influences on auditory speech perception, even in the absence of the potentially moderating information available in visual speech. By using non-native stimuli, the study also controlled for the possibility of the infants having learned the link between perception and production of these specific syllables.

The experiments in the current study were designed to test the robustness, specificity, and generalizability of the findings of Bruderer et al. (2015). The first set of studies tested discrimination of the same phonetic contrast (dental /da/ vs. retroflex /Da/) and used the same method of limiting tongue-tip movement (see Fig. 1a) as in Experiment 2 of Bruderer et al. (2015), but used single exemplars of each phone category from a synthetic continuum, rather than multiple natural exemplars. The second set of studies, Experiments 3 and 4, explored the generalizability of the phenomenon and the specificity of the relation between articulator movement and phonetic discrimination. Infants were tested on a bilabial /ba/ versus dental /da/ place of articulation contrast while inhibiting bilabial closure of the infant’s mouth with the same gummy teether (see Fig. 1b) that did not interfere with /da-/ /Da/ discrimination in Experiment 3 of Bruderer et al. (2015). Moreover, the /ba-/ /da/ distinction is native to English-learning infants, thus one they would have heard and seen produced. Therefore, we tested the question of whether the motor influence persisted even for familiar phoneme discrimination. To parallel the first set of studies, synthetic auditory stimuli were taken from the same continuum used in Experiments 1 and 2 for Experiments 3 and 4 (Werker & Lalonde, 1988).

Experiments 1 and 2: Replication of the influence of limiting tongue-tip movement on pre-babbling infants’ discrimination of dental and retroflex plosives

Rationale

In this study set, two separate experiments (Experiment 1 and Experiment 2) were conducted to replicate the first two

Fig. 1 (a) An infant with a flat teething toy placed in his mouth by a caregiver. The teething toy is placed just far enough to impede the movement of the tongue tip. (b) An infant with a gummy teething toy placed in his mouth by a caregiver. The soft, u-shaped mould matches an infant's gum contour and is placed in between the gums while the hard cover rests on his lips

studies in Bruderer et al. (2015). The testing procedure, the visual stimuli, the paradigm, and the demographics of the infant sample (e.g., age range, minimum gestation, language background) were identical to the original studies. Only the auditory stimuli differed in that they were synthesized rather than naturally produced. Two separate experiments were conducted. Experiment 1 was conducted to validate that infants can discriminate the synthesized dental and retroflex phonetic categories using the Alternating (Alt)/ Non-Alternating (NAlt) looking time procedure. Experiment 2 was conducted to examine infants' performance on the same speech discrimination task while caregivers held a teething toy in the infant's mouth that limited tongue-tip movement.

Experiment 1

Methods

Participants Participants were 24 English-learning infants (12 females; mean age, 5 months 27 days; range, 5 months 15 days to 6 months 13 days) who had a minimum of 90% exposure to English and no exposure to languages that contain the non-native contrast under investigation (see [Supplementary Material](#), Section 1 for excluded infants).

All participants were full term (> 38 weeks) with no known hearing impairments or developmental disorders. Based on estimated effect sizes from comparable studies, the pre-determined sample size was 24 infants for Experiments 1–4. Previous studies that have used the

Alt-NAlt procedure with a sample size of 24 per group reported medium effect sizes (Maye, Werker, & Gerken, 2002; Teinonen, Aslin, Alku, & Csibra, 2008; Tyler et al., 2014; Yeung, Chen, & Werker, 2013). One study that used the Alt-NAlt method to study tonal discrimination in infants used a sample size of 16 infants per age group, and demonstrated large effect sizes for 4- and 6-month-olds, but a small effect size for 9-month-olds (Mattock et al., 2008). The original study by Bruderer et al. (2015) used a sample size of 24 infants in their test of auditory-only perceptual discrimination (Experiment 1, Bruderer et al., 2015) and showed a medium effect size ($\eta_p^2=.16$). Taking the estimated population effect size value from the sample effect size of $\eta_p^2=.16$, an *a priori* power analysis for a repeated-measures ANOVA with two repeated measures (Trial Type and Pair), under the assumption of a conservative correlation of 0.5 between variables, showed that a sample size of 24 would result in a power of 0.78 (using G*Power version 3.1; Faul, Erdfelder, Lang, & Buchner, 2007).

Stimuli The auditory stimuli were selected from a synthetically created five formant, place of articulation voiced-plosive 16-step continuum, where F2 and F3 were varied. The stimuli were originally created for Werker and Lalonde (1988) with the Mattingly synthesizer on the VAX 11/780 (Haskins Laboratories, New Haven, CT, USA). Werker and Lalonde (1988) reported that native Hindi speakers perceived tokens from the continuum as three phoneme categories – bilabial /ba/, dental /da/, and retroflex /Da/ – whereas English speakers perceived two phoneme categories – bilabial /ba/ and alveolar /da/. For the current studies, we confirmed with a 16-step continuum, and with new participants, that adult Hindi speakers perceive the three distinct aforementioned phonemic categories (see [Supplementary Material](#), Section 2 for stimuli details). For this experiment, stimuli /9da/ and /15Da/ were used to test Hindi dental /da/ versus retroflex /Da/, a pairing shown to be discriminable by infants this age (e.g., Werker & Lalonde, 1988).

For each trial, participants heard a 14-s sound sequence composed of 14 individual tokens with a stimulus onset asynchrony (SOA) of 1 s (interstimulus interval (ISI) of 725 ms), paired with a neutral visual stimulus (static black and white checkerboard). In between the test trials, a moving object appeared and remained on the screen until the infants oriented their attention to the screen, at which point the experimenter initiated the next trial, and the checkerboard was shown again.

Apparatus The visual stimuli were presented on a 22-in. LCD monitor via the SMI dedicated software “Experimenter

Center” from a Dell PC. The infants were seated on their parent’s or caregiver’s lap, on a chair distanced approximately 55 cm away from the display. An eye-tracker (SensoriMotor Instruments Inc.) mounted beneath the external monitor was used to record the infant’s gaze to the visual stimulus. A split screen of the eye-tracking information was recorded, as well as a video of the participant’s face via iMovie onto a MacBook Pro, from which the experimenter monitored the participant. The auditory stimuli were presented at 70 dB through a single-drive speaker (Fostex 6301NX) located directly behind the screen. Caregivers’ hearing was masked through a pair of headphones.

Design and procedure The design followed a standard alternating (Alt)/non-alternating (NAlt) looking time procedure, whereby infants’ looking times in two trial types (Alt and NAlt) are compared (Best & Jones, 1998b; Feldman, Myers, White Griffiths, & Morgan, 2013; Mattock, Molnar, Polka, & Burnham, 2008; Maye, Werker, & Gerken, 2002; Yeung, Chen, & Werker, 2013). Specifically, Alt trials contained sequences of both phones from the contrast, and NAlt trials contain sequences of only one of the phones. Each infant, sitting on their caregiver’s lap, was exposed to eight trials – four Alt trials and four NAlt trials – which were presented in an alternating fashion such that no two trials of the same type were presented in a row. Eight different order types were developed; half of the orders started with an Alt trial and half of the orders started with a NAlt trial (see [Supplementary Material](#), Sections 3 and 4 for counterbalancing and order details). The main independent variable was the Trial Type (Alt or NAlt), and the dependent variable was Total Infant Looking Time to the checkerboard for each trial (see [Supplementary Material](#), Section 5 for procedural details).

Data were analyzed using two different analysis approaches, as proposed in our pre-registration. In the first analysis, we conducted a repeated-measures within-subject ANOVA with Trial Type (Alt/NAlt) and Pair (first, second, third, or fourth sequence) as within-subject factors. Each trial Pair consisted of a single Alt and a single NAlt trial, and each infant heard four Pairs sequentially. If Mauchly’s test of sphericity was significant using a liberal alpha criterion ($\alpha = 0.15$), we used the Greenhouse-Geisser correction for epsilon < 0.75 , and the Huynh-Feldt correction for epsilon > 0.75 to adjust the degrees of freedom. In the second analysis, we conducted a Bayesian t-test on the average looking time data across the Pairs with a default Cauchy prior scale set to $r = 0.707$ (Wagenmakers et al., 2018), and calculated the Bayes Factor (BF) using JASP (2017, version 0.9.1) (see [Supplementary Material](#), Section 6 for details on Bayesian analysis).

Longer looking time to either Alt or NAlt trials has been interpreted as indicative of discrimination in previous research (see Best & Jones, 1998; Bruderer et al., 2015; Tyler, Best, Goldstein, & Antoniou, 2014, for longer looking to Alt sequences, and Maye, Werker, & Gerken, 2002; Teinonen, Aslin, Alku, & Csibra, 2008; Yeung & Werker, 2009, for longer looking to NAlt sequences), meaning a significant difference in looking time to either Trial Type is taken as evidence of discrimination. However, given the results from Bruderer et al. (2015) with the natural tokens of this same phonetic contrast, it was predicted for this experiment that infants would look longer at Alt than at NAlt trials if they can discriminate the Hindi /da-/Da/ distinction.

Results

Repeated-measures ANOVA analysis A 2 (Trial Type) $\times$ 4 (Pair) repeated-measures ANOVA was conducted on the looking time data. The analysis revealed a significant main effect of Pair, $F(1.94, 44.66) = 5.12$, $P = 0.011$, $\eta_p^2 = 0.18$, which indicates that infants’ looking times declined as the trials progressed, a behavior typical in looking time paradigms with multiple trials. Mauchly’s test showed that the assumption of sphericity had been violated ($\chi^2(5) = 17.11$, $P = 0.004$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = 0.71$). The quadratic contrast was significant, $F(1, 23) = 14.200$, $P = 0.001$, $\eta_p^2 = 0.382$, indicating an initial increase in looking time followed by a decline. There was also a significant main effect of Trial Type, $F(1, 23) = 5.82$, $P = 0.024$, $\eta_p^2 = 0.20$; infants looked significantly longer to the Alt trials ($M = 9.55$ s, $SD = 0.37$) compared to the NAlt trials ($M = 9.10$ s, $SD = 0.40$) as in Bruderer et al. (see Fig. 2). The Pair by Trial Type interaction was not significant $F(3,69) = 2.46$, $P = 0.070$, $\eta_p^2 = 0.097$.

Bayesian analysis Bayesian paired-samples t-tests were conducted on the average looking times of each Trial Type (Alt and NAlt). In the first analysis, we calculated the Bayesian Factor (BF) for the H_0 (equal average looking times between Alt and NAlt) against the two-sided H_a (different average looking times between Alt and NAlt). We report that BF_{10} was 2.32, indicating that the data are 2.32 times more likely under the alternative hypothesis. In the second analysis, based on the previously reported results (Experiment 1, Bruderer et al., 2015) that infants looked longer to Alt trials, we conducted a one-sided H_a that the average looking time during Alt trials is greater than the average looking time during NAlt trials ($BF_{+0} = 4.57$). The BF shows that the observed data are 4.57 more likely under H_a , and

Fig. 2 Average looking time to each of the trial types (Alt in red and NAlt in blue) across the four pairs of trials during auditory discrimination of the non-native contrast (Experiment 1)

following the conventions of Lee and Wagenmakers (2014), this indicates moderate evidence that looking time during Alt trials is greater than during NAlt trials (see [Supplementary Material](#), Section 6, Fig. 1).

Discussion

Results from Experiment 1 replicate the previous findings that English-learning infants discriminate the non-native dental /da/-retroflex /Da/ phonetic contrast using the Alt-NAlt procedure (Experiment 1, Bruderer et al., 2015), and do so with synthesized stimuli. These results are also in line with previous studies demonstrating infant discrimination of this non-native phonetic contrast using other methods (e.g., Werker & Lalonde, 1988; Werker & Tees, 1984).

Experiment 2

This experiment was designed to be a replication of Experiment 2 in Bruderer et al. (2015). Specifically, we asked whether the insertion of a teether that interferes with tongue-tip movement would prevent discrimination of the Hindi dental /da/-retroflex /Da/ phonetic contrast. Evidence consistent with this hypothesis would be provided by a non-significant difference in looking time to the Alt versus the NAlt trials, whereas a significant difference in looking time between the two trial types would falsify the hypothesis.

Methods

The age and language criteria of the sample were identical to Experiment 1. Twenty-four English-learning infants (12 females; mean age, 5 months 28 days; range, 5 months 16 days to 6 months 14 days) participated in the experiment (see [Supplementary Material](#), Section 1 for excluded infants).

The stimulus set, setup, and apparatus were identical to Experiment 1, with the exception that infants in Experiment 2 were given a flat teething toy (Tomy Learning Curve Fruity Teethers, Oak Brook, IL, USA) that limited movement of the tongue-tip. The caregiver placed the teether gently in the infant's mouth, holding it above the tongue and just far enough into the mouth to impede full tongue-tip movement (see Fig. 1a). The design and procedure were also identical to Experiment 1, with the exception that in Experiment 2, the caregiver held a teething toy in the infant's mouth for the duration of the study.

Results

Repeated-measures ANOVA analysis A 2 (Trial Type) $\times$ 4 (Pair) repeated-measures ANOVA was conducted on the looking time data. There was a significant main effect of Pair, $F(2.64, 60.79) = 7.21$, $P = 0.001$, $\eta_p^2 = 0.24$, which indicated that looking time declined across the four Pairs of trials. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(5) = 10.33$, $P = 0.067$), therefore degrees of freedom were corrected using Hyunh-Feldt estimates of sphericity ($\epsilon = 0.79$). The linear contrast was significant, $F(1,23) = 13.04$, $P = 0.001$, $\eta_p^2 = 0.36$. There was no main effect of Trial Type, $F(1,23) < 0.001$, $P = 0.98$, $\eta_p^2 < 0.001$, indicating that the average looking time to the Alt and NAlt trials did not significantly differ. Overall, when the infants had a flat teething toy in their mouth, there was no significant difference in looking time to Alt ($M = 9.61$ s, $SD = 0.49$) compared to NAlt ($M = 9.60$ s, $SD = 0.48$) trials (see Fig. 3). There was no interaction of Trial Type and Pair $F(3,69) = 1.38$, $P = 0.26$, $\eta_p^2 = 0.057$, indicating a similar pattern of looking to both trial types across the four Pairs of trials.

Bayesian analysis A Bayesian paired-samples t-test was conducted on the average looking times of each trial type for H_0 (equal average looking times for Alt and NAlt trials) against the two-sided H_a that the looking times during Alt and NAlt trials differ. We report BF_{01} of 4.66, indicating that the data were 4.66 times more likely to occur under the null hypothesis. The moderate evidence for H_0 (Lee & Wagenmakers, 2014) suggests that infants did not discriminate between the

Fig. 3 Average looking time to each of the trial types (Alt in red and NAlt in blue) across the four pairs of trials during auditory discrimination of the non-native contrast while infants hold a tongue-tip movement inhibiting teething toy (Experiment 2)

two phones in Experiment 2 (see [Supplementary Material](#), Section 6, Fig. 2).

Discussion

In Experiment 2, we showed that when movement of the infants' tongue-tip was inhibited, infants were not able to discriminate the non-native dental /da/ - retroflex /Da/ contrast, which differs mainly in the placement of the tongue-tip. Infants' lack of discrimination, shown here with synthesized stimuli, replicates the findings reported in Experiment 2 of Bruderer et al. (2015). This provides further support for the hypothesis that there are sensorimotor influences on phonetic discrimination in young infants.

Experiments 3 and 4: An extension study examining the influence of limiting bilabial closure on infants' discrimination of bilabial and dental stop consonants

Rationale

In the second set of experiments, we investigated whether sensorimotor influences on speech perception are specific to the manipulation of the tongue-tip or extend to other oral-motor articulators, such as the lips. Accordingly, we tested whether infants' discrimination of the bilabial plosive /ba/ versus dental plosive /da/ is influenced by a teething toy that prevents movement of the lips. Experiment 3 was conducted

to validate that infants can discriminate the synthesized bilabial and dental plosive phonetic categories in the alternating/non-alternating procedure. Experiment 4 was conducted to examine infants' performance on the identical phonetic distinction (/ba/ vs. /da/) when the oral-motor movements are altered by the presence of a teething toy that selectively engages the lips and prevents a bilabial closure.

Experiment 3

Methods

Twenty-four 6-month-old English-learning infants (12 females; mean age, 5 months 29 days; range, 5 months 16 days to 6 months 10 days) participated in the experiment (see [Supplementary Material](#), Section 1 for excluded infants). The participant criteria, sample size, apparatus, design, and method procedure were identical to Experiment 1; only the auditory stimuli differed. Infants completed an Alt/NAlt phonetic discrimination task of the bilabial /ba/ versus dental /da/ place-of-articulation contrast with stimuli /3ba/ and /9da/, respectively.

Results

Repeated-measures ANOVA analysis A 2 (Trial Type) $\times$ 4 (Pair) repeated-measures ANOVA was conducted on the looking time data. The analysis revealed a significant main effect of Pair, $F(2.53, 58.17) = 7.82$, $P < 0.001$, $\eta_p^2 = 0.26$, indicating that infants' looking time declined across the trials. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(5) = 10.62$, $P = 0.060$), therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\epsilon = 0.76$). A follow-up indicated a significant linear contrast, $F(1,23) = 14.74$, $P = 0.001$, $\eta_p^2 = 0.39$. Further, there was a significant main effect of Trial Type, $F(1,23) = 4.58$, $P = 0.043$, $\eta_p^2 = 0.166$, indicating longer looking to NAlt trials ($M = 9.35$ s, $SD = 0.47$) compared to Alt trials ($M = 8.68$ s, $SD = 0.42$) (see Fig. 4). There was no significant interaction between Trial Type and Pair $F(3,69) = 0.22$, $P = 0.89$, $\eta_p^2 = 0.009$.

Bayesian analysis Bayesian paired-samples t-tests were conducted on the average looking times of each Trial Type (Alt and NAlt) for the two-sided H_a (different average looking times between Alt and NAlt trials) against the H_0 (equal average looking times for Alt and NAlt trials). We report $BF_{10} = 1.46$, indicating that the data are 1.46 more likely to occur under the H_a . However, one subject had a difference score that exceeded 2.5 standard deviations away from the mean. Removing this outlier, the analysis showed $BF_{10} = 8.92$, indicating moderate evidence (Lee & Wagenmakers, 2014) for H_a .

Fig. 4 Average looking time to each of the trial types (Alt in red and NAlt in blue) across the four pairs of trials during auditory discrimination of the native contrast (Experiment 3)

that looking times differ between Alt and NAlt trials (see [Supplementary Material](#), Section 6, Fig. 3).

Discussion

These data indicate that 6-month-old English-learning infants can discriminate the native bilabial /ba/ versus dental /da/ phonetic contrast as revealed by differences in looking time to the two trial types. While infants looked longer to Alt trials when hearing non-native dental /da/ versus retroflex /Da/ (Experiment 1 above; Bruderer et al., 2015), here, infants looked longer to NAlt trials when hearing the native /ba/ versus /da/ contrast. The distinct directionality of looking-preference was surprising, but previous studies (as reviewed in Experiment 1) have taken as evidence differential looking in either direction in the Alt/NAlt paradigm.

Experiment 4

This experiment was designed to test the generalizability of the finding that preventing articulator-specific movement interferes with phonetic discrimination. As such, Experiment 4 examined infants' ability to discriminate the native bilabial /ba/ versus dental /da/ distinction when the infants' lip closure was prevented with a gummy teething toy. If discrimination of this native phonetic contrast can be disrupted by a relevant oral-motor restriction, then we expected infants to not show discrimination. Evidence consistent with this hypothesis would be provided by a non-significant difference in looking time to the Alt versus the NAlt trials, whereas a significant

difference in looking time between the two trial types would falsify this hypothesis.

Methods

Twenty-four English-learning infants (12 females; mean age, 6 months 0 days; range, 5 months 16 days to 6 months 15 days) participated in the experiment (see [Supplementary Material](#), Section 1 for excluded infants). The participant criteria, sample size, apparatus, design, and procedure were identical to Experiment 2. Only the auditory stimuli and the sensorimotor manipulations differed. Infants completed an Alt/NAlt phonetic discrimination task with the bilabial plosive /ba/ and dental plosive /da/ contrast (as used in Experiment 3) while the caregiver held a teething toy that limited bilabial closure (Nuby Gum-eez, Monroe, LA) in the infant's mouth. The caregiver placed the gummy teether gently in the infant's mouth such that the gum-shaped silicone mould rested over the gums, and the hard shell rested over the lips, impeding lip closure (see Fig. 1b).

Results

Repeated-measures ANOVA analysis A 2 (Trial Type) $\times$ 4 (Pair) repeated-measures ANOVA was conducted on the looking time data. The analysis revealed a non-significant main effect of Pair, $F(2.15, 49.51) = 2.95$, $P = 0.058$, $\eta_p^2 = 0.11$, indicating that the infants' looking time did not significantly decline across the trials. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(5) = 13.02$, $p = 0.023$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = 0.717$). There was no main effect of Trial Type $F(1,23) = 0.46$, $P = 0.50$, $\eta_p^2 = 0.020$, indicating that the pattern of looking time to the Alt and NAlt trials did not significantly differ. Overall, when the infants had a gummy teething toy inserted in their mouths, their looking time to Alt trials ($M = 9.30$ s, $SD = 0.44$) compared to NAlt ($M = 9.08$ s, $SD = 0.46$) trials did not differ (see Fig. 5). There was no interaction of Trial Type and Pair $F(3,69) = 1.32$, $P = 0.27$, $\eta_p^2 = 0.054$, indicating similar patterns of looking to both trial types across the four Pairs of trials.

Bayesian analysis A Bayesian paired-samples t-test was conducted on the average looking times of each trial type for H_0 (equal average looking times for Alt and NAlt trials) against the two-sided H_a that the looking times during Alt and NAlt trials differ. We report that BF_{01} was 3.78, which indicates that the data were 3.78 times more likely (moderate evidence, Lee & Wagenmakers, 2014) to occur under H_0 than under H_a . Thus, we conclude that infants did not discriminate between the two phones (see [Supplementary Material](#), Section 6, Fig. 4).

Fig. 5 Average looking time to each of the trial types (Alt in red and NAlt in blue) across the four pairs of trials during auditory discrimination of the native contrast while infants hold a lip movement inhibiting teething toy (Experiment 4)

Discussion

The results from Experiment 4 show that when full lip closure is prevented, infants are not able to discriminate the native bilabial /ba/ versus dental /da/ phonetic contrast. This result indicates that even when the teething toy prevents achievement of the oral-motor movement involved in only one of the phones in the contrast (in this case, the bilabial /ba/), discrimination can be disrupted. Notably, the same gummy teething toy did not disrupt the Hindi dental /da/ versus retroflex /Da/ discrimination in (Experiment 3, Bruderer et al., 2015), demonstrating that preventing movement of an articulator that is not involved in the production of the phones in the contrast does not influence perception.

General discussion

The results from these four experiments advance our understanding of how, in infancy, oral-motor processes might interact with, or influence, speech sound discrimination. In particular, they reveal that impeding oral-motor movements associated with the production of specific phones interferes with infants' discrimination of those phones. A flat teether inhibiting tongue-tip movement prevented discrimination of a place of articulation distinction involving tongue-tip movements, and a gummy teether inhibiting a bilabial closure prevented discrimination of a distinction involving a bilabial versus other (dental) phone. As noted above, this gummy

teether did not interfere with discrimination of the dental /da/ versus retroflex /Da/ distinction in (Experiment 3, Bruderer et al., 2015), thus providing the strongest evidence to date that there is specificity in the relation between disrupted oral-motor movements and speech sound discrimination.

The fact that these findings were observed at an age when infants do not yet produce well-formed consonant-vowel syllables suggests that infants' existing neural circuitry is sufficient to support this behavior without specific experience. Whether the relation stems from information transfer between motor and auditory areas, or whether there is a single neural circuit supporting both, can only be answered by targeted neuroimaging work. Additionally, it remains to be determined whether the motor and auditory representations are precise enough to support experience-independent linking of specific oral-motor gestures to specific phonetic properties, or whether that specification is rapidly achieved with experience, drawing attention to the relevant articulators by inhibiting their movement.

We hypothesized (see also Streri, Coulon, Marie, & Yeung, 2016) that motor movements might most profoundly impact speech perception in young infants before the native phoneme repertoire becomes well established. Consistent with this, in Kuhl et al. (2014), while neural activation of motor areas was evident in response to native and non-native phonemes at 7 months, motor area activation was only evident at 11 months when infants heard non-native speech. Our behavioral results in Experiments 3 and 4 with 6-month-old infants similarly indicate that – at this young age – it is not only non-native, but also native phonetic discrimination that is disrupted by preventing movement of the relevant articulator. Further behavioral work with infants 10–12 months or older, who have already attuned to the sounds of the native language, could determine whether discrimination of the native contrast /ba/-/da/ is no longer disrupted even with motor interference. If disruption is no longer evident, then the hypothesis that motor feedback is most influential prior to attunement would be strengthened. Finally, given that the gummy teether only disrupts lip closure but does not prevent tongue-tip movement, Experiment 4 provides evidence that preventing the oral-motor movement associated with the production of only one of the phones in a minimal pair contrast may be sufficient to disrupt discrimination. Further research is required to confirm the generalizability of this finding.

While it is tempting to interpret these results as showing a strong relation between perception and production in infant speech perception, there are several caveats. First and foremost, we have no evidence that infants spontaneously move their articulators when listening to speech. Thus, we do not know whether infants use their own motor movements to support discrimination. We only demonstrate that discrimination is disrupted if the relevant articulators are prevented from moving. In future work, it would be informative to image

the articulators, perhaps through ultrasound, while infants are listening to speech to determine if relevant oral-motor movements can be observed. An additional avenue for future work will be to test infants who, by way of oral-motor dysmorphologies, are unable to move their articulators in particular ways and to test whether discrimination of related sounds is lacking in these infants. There are some studies to suggest this may be the case; children 7 years of age with a repaired cleft palate who show posterior placement of alveolar targets could not distinguish /t^h/ and /k^h/ in an identification task with a synthesized continuum with those two sounds as end points (Whitehill, Francis, & Ching, 2003).

It is of interest that infants in Experiment 1 showed discrimination by exhibiting longer looking to Alt over NAlt trials and that infants in Experiment 3 showed discrimination by longer looking to NAlt over Alt trials (see Fig. 6). In our pre-registration, we had indicated that discrimination would be inferred irrespective of the directionality. While the direction was the same in Experiment 1 (dental /da/ - Retroflex /Da/) as in the earlier work with naturally produced versions of these syllables (Bruderer et al., 2015), the fact that it differed, albeit with a different (and in this case, “native”) phonetic contrast for Experiment 3 (bilabial /ba/ - dental /da/), is unexplained. It is possible that because English-learning infants are familiar with the contrast and are used to hearing both syllable types, repetition of only one sound is more unusual to them, therefore leading to longer listening time to NAlt trials. Some support for this possibility comes from the fact that when a familiarization phase precedes Alt versus NAlt testing, longer looking to NAlt has been reported (e.g., Maye, Werker, & Gerken, 2002; Teinonen, Aslin, Alku, & Csibra, 2008; Yeung & Werker, 2009). However, this hypothesis is yet

unconfirmed, and further work with more native and non-native distinctions is required.

In summary, across a series of four experiments, we provided evidence that inhibiting relevant oral-motor movements can interfere with speech sound discrimination in young infants. These results add to the growing evidence (Bruderer et al., 2015; Kuhl et al., 2014; Yeung & Werker, 2013) of a specific link between production and perception, even in infants who are not yet able to produce well-formed consonant-vowel syllables.

Author Note This research is funded by grants from the National Institutes of Health (1R21HD079260-01) and the Natural Sciences and Engineering Research Council of Canada (RGPIN-2015-03967) to Janet F. Werker.

Acknowledgements We thank Mikayla Blumenthal for her assistance in data collection, Savannah Nijeboer and Jackie Hart Smith for their support in recruiting infants and providing feedback on various drafts, and Valter Ciocca for his insight on Bayesian analysis.

References

- Best, C. (1995). A Direct Realist View of Cross-Language Speech Perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium: York Press.
- Best, C., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development*, 21, 295. [https://doi.org/10.1016/S0163-6383\(98\)91508-9](https://doi.org/10.1016/S0163-6383(98)91508-9)
- Best, C., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, 46, 183–216. <https://doi.org/10.1177/00238309030460020701>
- Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J.-F. (2009). Hearing faces: how the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, 21, 905–921. <https://doi.org/10.1162/jocn.2009.21076>
- Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, 112(44), 13531–13536. <https://doi.org/10.1073/pnas.1508631112>
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., & Ghazanfar, A. A. (2009). The Natural Statistics of Audiovisual Speech. *PLOS Computational Biology*, 5(7), e1000436. <https://doi.org/10.1371/journal.pcbi.1000436>
- Chen X., Striano T., Rakoczy H. (2004). Auditory-oral matching behavior in newborns. *Developmental Science*, 7, 42–47. <https://doi.org/10.1111/j.1467-7687.2004.00321.x>
- Coulon M., Hemimou C., Streri A. (2013). Effects of seeing and hearing vowels on neonatal facial imitation. *Infancy*, 18, 782–796. <https://doi.org/10.1111/infa.12001>
- Danielson, D. K., Bruderer, A. G., Kandhadai, P., Vatikiotis-Bateson, E., & Werker, J. F. (2017). The organization and reorganization of

Fig. 6 Average difference scores between Alt and NAlt trials. The mean difference in looking times between the Alt and NAlt trial types for each of the four experiments. A score greater than zero indicates a preference to look longer during the presentation of Alt trials, and a score below zero indicates a NAlt preference

- audiovisual speech perception in the first year of life. *Cognitive Development*, 42, 37–48. <https://doi.org/10.1016/j.cogdev.2017.02.004>
- Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., & Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences*, 103(38), 14240–14245. <https://doi.org/10.1073/pnas.0606302103>
- DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior & Development*, 34, 590–601. <https://doi.org/10.1016/j.infbeh.2011.06.005>
- Dubois, J., Poupon, C., Thirion, B., Simonnet, H., Kulikova, S., Leroy, F., ... Dehaene-Lambertz, G. (2015). Exploring the early organization and maturation of linguistic pathways in the human infant brain. *Cerebral Cortex*, 26(5), 2283–2298. <https://doi.org/10.1093/cercor/bhv082>
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303–306. <https://doi.org/10.1126/science.171.3968.303>
- Erber, N. P. (1975). Auditory-visual perception of speech. *The Journal of Speech and Hearing Disorders*, 40(4), 481–492. <https://doi.org/10.1044/jshd.4004.481>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191.
- Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., & Morgan, J. L. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition*, 127(3), 427–438. <https://doi.org/10.1016/j.cognition.2013.02.007>
- Grant, K. W., & Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, 108(3), 1197. <https://doi.org/10.1121/1.1288668>
- Guellai, B., Streri, A., & Yeung, H. H. (2014). The development of sensorimotor influences in the audiovisual speech domain: Some critical questions. *Frontiers in Psychology*, 5(812), 1–7. <https://doi.org/10.3389/fpsyg.2014.00812>
- Hickok, G., Houde, J., & Rong, F. (2011). Perspective Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron*, 69(3), 407–422. <https://doi.org/10.1016/j.neuron.2011.01.019>
- Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). Infant speech perception activates Broca's area: a developmental magnetoencephalography study. *Neuroreport*, 17(10), 957–962. <https://doi.org/10.1097/01.wnr.00000223387.51704.89>
- JASP Team (2017). JASP (Version 0.9.1)[Computer software]. Retrieved from <https://jasp-stats.org/>
- Kuhl, P., & Meltzoff, A. (1982). The bimodal perception of speech in infancy. *Science*, 218(4577), 1138–1141. <https://doi.org/10.1126/science.7146899>
- Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606–608. <https://doi.org/10.1126/science.1736364>
- Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, 100(4), 2425–2438. <https://doi.org/10.1121/1.417951>
- Kuhl, P. K., Ramírez, R. R., Bosseler, A., Lin, J.-F. L., & Imada, T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(31), 11238–11245. <https://doi.org/10.1073/pnas.1410963111>
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2), F13–F21. <https://doi.org/10.1111/j.1467-7687.2006.00468.x>
- Lee, M. D., & Wagenmakers, E. J. (2014). Bayesian cognitive modeling: A practical course. Cambridge university press.
- Leroy, F., Glasel, H., Dubois, J., Hertz-Pannier, L., Thirion, B., Mangin, J. F., & Dehaene-Lambertz, G. (2011). Early maturation of the linguistic dorsal pathway in human infants. *Journal of Neuroscience*, 31(4), 1500–1506. <https://doi.org/10.1523/JNEUROSCI.4141-10.2011>
- MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2(4), 369. <https://doi.org/10.1017/S0142716400009796>
- Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The Relationship Between Infants' Production Experience and Their Processing of Speech. *Language Learning and Development*, 10(2), 179–204. <https://doi.org/10.1080/15475441.2013.829740>
- Matchin, W., Groulx, K., & Hickok, G. (2014). Audiovisual speech integration does not rely on the motor system: evidence from articulatory suppression, the McGurk effect, and fMRI. *Journal of Cognitive Neuroscience*, 26(3), 606–620. https://doi.org/10.1162/jocn_a_00515
- Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, 106, 1367–1381. <https://doi.org/10.1016/j.cognition.2007.07.002>
- Maurer, D., & Werker, J. F. (2014). Perceptual Narrowing During Infancy: A Comparison of Language and Faces. *Developmental Psychobiology*, 56, 154–178. <https://doi.org/10.1002/dev.21177>
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101–11. [https://doi.org/10.1016/S0010-0277\(01\)00157-3](https://doi.org/10.1016/S0010-0277(01)00157-3)
- Möttönen, R., Dutton, R., & Watkins, K. E. (2013). Auditory-Motor Processing of Speech Sounds. *Cerebral Cortex*, 23(5), 1190–1197. <https://doi.org/10.1093/cercor/bhs110>
- Munhall, K. G., & Vatikiotis-Bateson, E. (2004). Spatial and Temporal Constraints on Audiovisual Speech Perception. In C. Spence & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 177–188). Cambridge: MIT Press. Retrieved from <http://psycnet.apa.org/psycinfo/2004-17469-011>
- Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Developmental Science*, 13(3), 407–420. <https://doi.org/10.1111/j.1467-7687.2009.00898.x>
- Okada, K., & Hickok, G. (2009). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. *Neuroscience Letters*, 452(3), 219–223. <https://doi.org/10.1016/j.neulet.2009.01.060>
- Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, 22(2), 237–247. [https://doi.org/10.1016/S0163-6383\(99\)00003-X](https://doi.org/10.1016/S0163-6383(99)00003-X)
- Peña, M., Werker, J. F., & Dehaene-Lambertz, G. (2012). Earlier Speech Exposure Does Not Accelerate Speech Acquisition. *Journal of Neuroscience*, 32(33), 11159–11163. <https://doi.org/10.1523/JNEUROSCI.6516-11.2012>
- Perani, D., Saccuman, M. C., Scifo, P., Anwander, A., Spada, D., Baldoli, C., ... Friederici, A. D. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences of the United States of America*, 108(38), 16056–16061. <https://doi.org/10.1073/pnas.1102991108>
- Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United*

- States of America*, 106(26), 10598–10602. <https://doi.org/10.1073/pnas.0904134106>
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724. <https://doi.org/10.1038/nn.2331>
- Reale, R. A., Calvert, G. A., Thesen, T., Jenison, R. L., Kawasaki, H., Oya, H., ... Brugge, J. F. (2007). Auditory-visual processing represented in the human superior temporal gyrus. *Neuroscience*, 145(1), 162–184. <https://doi.org/10.1016/j.neuroscience.2006.11.036>
- Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, 8(2), 162–172. <https://doi.org/10.1111/j.1467-7687.2005.00403.x>
- Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, 23(1–3), 429–435. <https://doi.org/10.1016/j.cogbrainres.2004.11.006>
- Scott, M., Yeung, H., Gick, B., & Werker, J. F. (2013). Inner speech captures the perception of external speech. *The Journal of the Acoustical Society of America*, 133(4), EL286–EL292. <https://doi.org/10.1121/1.4794932>
- Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing Lips and Seeing Voices: How Cortical Areas Supporting Speech Production Mediate Audiovisual Speech Perception. *Cerebral Cortex*, 17(10), 2387–2399. <https://doi.org/10.1093/cercor/bhl147>
- Streri, A., Coulon, M., Marie, J., & Yeung, H. H. (2016). Developmental change in infants' detection of visual faces that match auditory vowels. *Infancy*, 21(2), 177–198. <https://doi.org/10.1111/infa.12104>
- Sumbly, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 26(2), 212–215. <https://doi.org/10.1121/1.1907309>
- Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, 108, 850–855. <https://doi.org/10.1016/j.cognition.2008.05.009>
- Tyler, M. D., Best, C. T., Goldstein, L. M., & Antoniou, M. (2014). Investigating the role of articulatory organs and perceptual assimilation of native and non-native fricative place contrasts. *Developmental Psychobiology*, 56(2), 210–227. <https://doi.org/10.1002/dev.21195>
- Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2004). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, 102(4), 1181–1186. <https://doi.org/10.1073/pnas.0408949102>
- Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... Matzke, D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1), 35–57. <https://doi.org/10.3758/s13423-017-1343-3>
- Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 52(1), 349–355. <https://doi.org/10.2307/1129249>
- Werker, J. F., & Lalonde, C. (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>
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49–63. [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)
- Whitehill, T. T., Francis, A. L., & Ching, C. K. (2003). Perception of place of articulation by children with cleft palate and posterior placement. *Journal of speech, language, and hearing research*, 46(2), 451–461.
- Yeung, H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-Month-olds use distinct objects as cues to categorize speech information. *Cognition*, 113(2), 234–243. <https://doi.org/10.1016/j.cognition.2009.08.010>
- Yeung, H., & Werker, J. F. (2013). Lip Movements Affect Infants' Audiovisual Speech Perception. *Psychological Science*, 24(5), 603–612. <https://doi.org/10.1177/0956797612458802>
- Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68(2), 123–139. <https://doi.org/10.1016/j.jml.2012.09.004>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.