ECOLOGICAL PSYCHOLOGY
2016, VOL. 28, NO. 4, 216–261
http://dx.doi.org/10.1080/10407413.2016.1230372

Routledge
Taylor & Francis Group

Articulating What Infants Attune to in Native Speech

Catherine T. Best[a,b,c], Louis M. Goldstein[c,d], Hosung Nam[c,e], and Michael D. Tyler[a,f]

[a]MARCS Institute, Western Sydney University; [b]School of Humanities and Communication Arts, Western Sydney University; [c]Haskins Laboratories; [d]Department of Linguistics, University of Southern California; [e]Department of English Language and Literature, Korea University; [f]School of Social Sciences and Psychology, Western Sydney University

ABSTRACT

To become language users, infants must embrace the integrality of speech perception and production. That they do so, and quite rapidly, is implied by the native-language attunement they achieve in each domain by 6–12 months. Yet research has most often addressed one or the other domain, rarely how they interrelate. Moreover, mainstream assumptions that perception relies on *acoustic* patterns whereas production involves *motor* patterns entail that the infant would have to translate incommensurable information to grasp the perception–production relationship. We posit the more parsimonious view that both domains depend on commensurate *articulatory* information. Our proposed framework combines principles of the Perceptual Assimilation Model (PAM) and Articulatory Phonology (AP). According to PAM, infants attune to articulatory information in native speech and detect similarities of nonnative phones to native articulatory patterns. The AP premise that gestures of the speech organs are the basic elements of phonology offers articulatory similarity metrics while satisfying the requirement that phonological information be discrete and contrastive: (a) distinct articulatory organs produce vocal tract constrictions and (b) phonological contrasts recruit different articulators and/or constrictions of a given articulator that differ in degree or location. Various lines of research suggest young children perceive articulatory information, which guides their productions: discrimination of between- versus within-organ contrasts, simulations of attunement to language-specific articulatory distributions, multimodal speech perception, oral/vocal imitation, and perceptual effects of articulator activation or suppression. We conclude that articulatory gesture information serves as the foundation for developmental integrality of speech perception and production.

> … each blind man felt a part of the animal in his reach, reporting that it was like a wall; a snake; a tree; a fan. … "Each was partly in the right, and all were in the wrong." (Saxe, 1873, pp. 135–136)

By the last quarter of the 1st year, infants have become perceptually attuned to many aspects of native speech. They have also begun to recognize and understand spoken words. Building a comprehension vocabulary, that is, a lexicon, requires that children first recognize

CONTACT Catherine T. Best c.best@westernsydney.edu.au MARCS Institute, Western Sydney University, Locked Bag 1797, Penrith NSW 2751, Australia.

words as familiar spoken patterns. But for learning and recognition of words to become optimally efficient, the perceiver must apprehend their component phonological structure, that is, as specific sequences of consonants and vowels (e.g., Cutler, 2008). Languages differ in the phonological inventories they use to build and distinguish words as well as in their specific phonetic realizations even for consonants and vowels that they hold in common. Thus, fast and accurate word recognition depends on language-specific attunement of speech perception, and this entails that attunement to native speech is central to language acquisition.

But the infant becomes a single unified *native user* of a specific language, that is, concurrently becomes a speaker-perceiver. Clearly, native tuning of perception and production must be integral for language acquisition to proceed effectively and efficiently. Perceptual attunement must guide the child's learning of how to both recognize and produce native words. In turn, caregiver responses to the child's utterances—comprehending her intended words, or failing to—provide feedback regarding how words must be articulatorily shaped to be at least minimally recognizable and, eventually, to be perceived as native-like for the child's community. Yet despite the obvious centrality of the perception–production crossroads for acquiring a language, this intersection has received scant empirical or theoretical attention. On the one side, much research has addressed how language experience modifies infants' initial speech perception abilities, with scarce consideration as to how exactly language-specific perceptual tuning guides infants' vocalizations toward native-like speech productions. And on the other side, work on preverbal vocal development has addressed universal properties of early speech-like productions and their shift toward language-specific biases but again with barely any thought as to how those skills rely on perceptual attunement to that language.

Our aim is to create a coherent framework for understanding and investigating infants' co-attunement of perception and production to native speech. The core question for this endeavor is as follows: What information do infants tune in to in native speech that could support contingent development in both perceiving and speaking the language? The answer must be compatible with what we know about infant language learning in its normal, natural context: (a) highly engaged *face-to-face* interactions in which (b) caregivers produce dynamically correlated *multimodal speech* that displays (c) *hyperarticulation* and/or *expanded variability* in target consonants and vowels, to which (d) infants are *vocally responsive*, often engaging in (e) dyadic *vocal matching* with their communicative partners (i.e., reciprocal infant–caregiver vocal "imitation," which can be full or partial, immediate or delayed). Our integrated perspective on language-specific developmental tuning of speech perception and production takes those observations into account. We combine and adapt the principles of the Perceptual Assimilation Model (PAM; Best, 1993, 1994, 1995; PAM-L2; Best & Tyler, 2007) and Articulatory Phonology (AP; e.g., Browman & Goldstein, 1989, 1992; L. M. Goldstein, Byrd, & Saltzman, 2006), with particular attention to a proposed extension of AP principles to early speech development, the Articulatory Organ Hypothesis (AOH; Best & McRoberts, 2003; L. M. Goldstein & Fowler, 2003; Studdert-Kennedy & Goldstein, 2003). According to PAM, infants attune to articulatory information in native speech and detect similarities of nonnative phones to native articulatory patterns (see Figure 1 for a schematic of predicted assimilation patterns). AP posits that phonological distinctions are conveyed by articulatory gestures, that is, constrictions of specified degrees of closure at specific locations that are achieved by one or more of the set of active vocal tract articulatory organs (Figure 2 is an updated schematic diagram of the articulatory organs, their spatial/functional organization, and their constriction parameters). The core premise of our proposed PAM-AOH

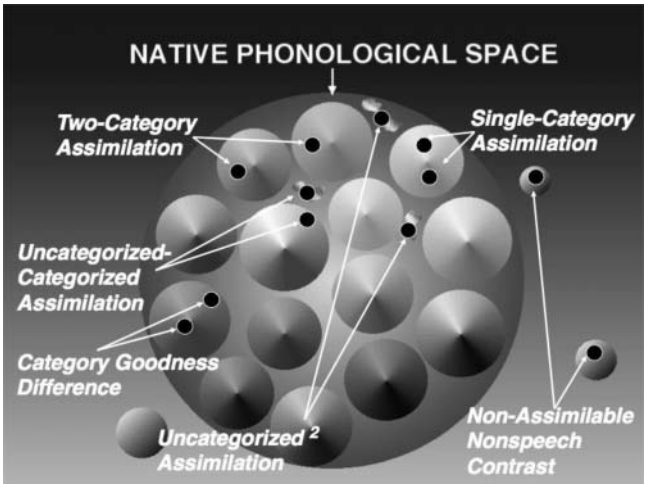

Figure 1. Schematic diagram of the Perceptual Assimilation Model (PAM; Best, 1995; Best & Tyler, 2007), illustrating an adult's native language phonological space, in which the conical "islands" represent native consonant categories that have been delineated and sharpened by experience with perceiving and producing native speech, and the major predicted patterns of perceptual assimilation of nonnative consonant contrasts to the native phonological system. Pairs of black circles represent nonnative consonant contrasts, with the various predicted contrast assimilation patterns indicated by arrows and labels.

approach is that optimal co-attunement of the two domains relies on the same information in native speech: articulatory information generated by the coordinated gestures of the speech organs that generate multimodal speech signals.

A coherent framework for perception–production relations in early development should ideally accommodate existing findings on native-language attunement in both domains. Therefore, we begin with a critical review of evidence on speech perception and production across the infant's 1st year. From that foundation, we consider whether and how current theoretical approaches could account for that range of findings across the two sides of native language speech attunement. We then present our proposed account and supporting evidence, some of which involves our reinterpretation of others' findings.

We focus on consonant contrasts for several reasons. Consonants and vowels clearly play markedly different phonological roles in a spoken language (e.g., Mehler, Peña, Nespor, & Bonatti, 2006). Consonants have received wider and more systematic perceptual investigation in both adults and infants. They are perceived more categorically than vowels; involve narrower and more rapidly produced articulatory constrictions; and are more likely than vowels to serve as syllable onsets, which appear to serve as the primary basis of organization of the mental lexicons of languages (e.g., Marslen-Wilson & Zwitserlood, 1989; Vitevitch, Armbrüster, & Chu, 2004).

Infant attunement to native speech: What do we know?

Becoming a native perceiver

A remarkable array of developmental patterns has been observed in infants' perception of minimal contrasts between consonants, both those that are employed in the language they

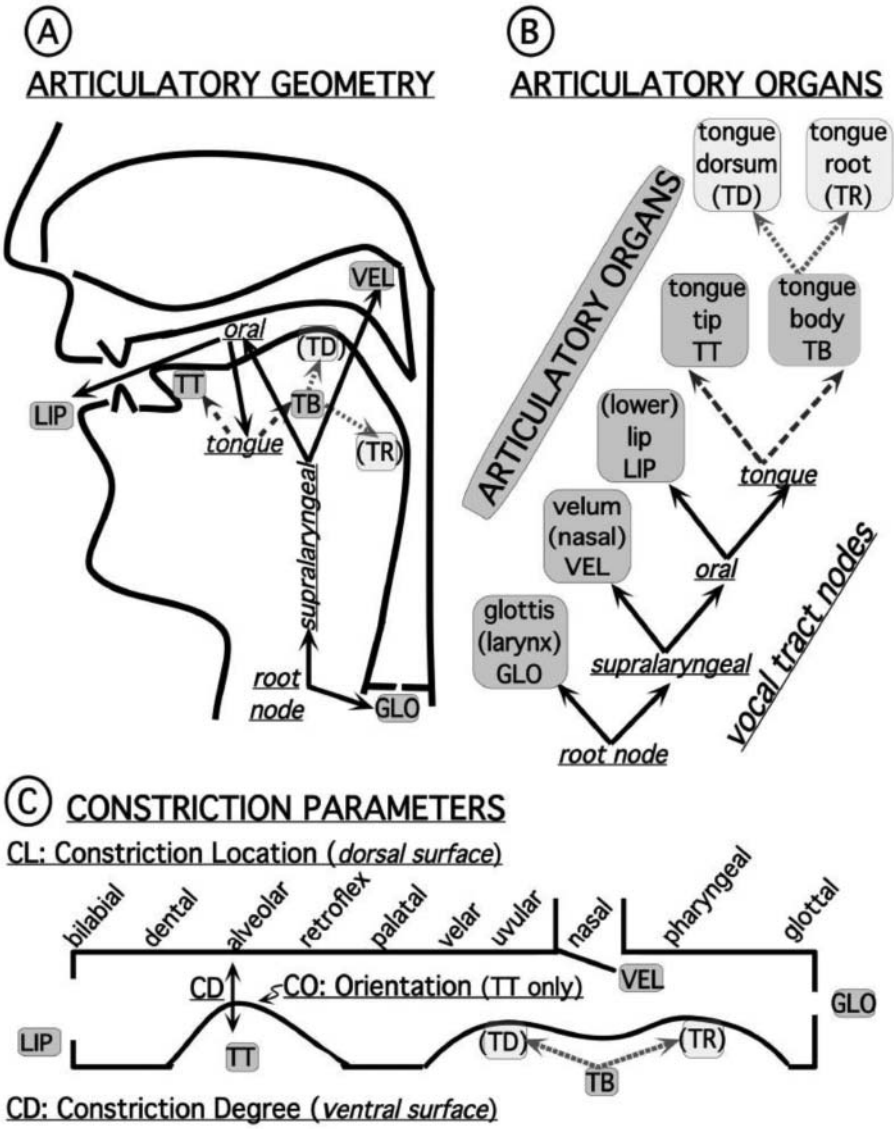

Figure 2. Schematics of the three dimensions of articulatory gestures as proposed for the revised Perceptual Assimilation Model with Articulatory Organ Hypothesis (PAM-AOH): (A) articulatory geometry (modeled after Browman & Goldstein, 1989, 1992); (B) articulatory organ hierarchy (active articulators and their nested nodes), which is an unfolded, straightened version of A; (C) articulatory actions (constriction degrees) represented along a straightened side view of the vocal tract's ventral (lower surface: active articulators) and dorsal (upper surface: passive articulators/locations) surfaces.

are learning—native contrasts—and those that are not native but are used in other languages, that is, nonnative contrasts (e.g., Best, 1994; Werker, 1989, 1991; Werker & Curtin, 2005). A minimal contrast refers to a pair of consonants that share all phonetic features except one crucial distinguishing feature; for example, English /p/ vs. /t/ are both stop consonants and voiceless but are distinguished by different places of articulation: /p/ has a bilabial

place (closure between the lower and upper lip), whereas /t/ has an alveolar place (tongue tip closure against the alveolar ridge, behind the upper front teeth).

As we review the perceptual findings, we discuss them in light of two underlying assumptions about the nature of information perceivers detect in speech that are held in common by the two most widely accepted theoretical hypotheses about early perceptual tuning to native speech. Those theoretical hypotheses are (a) that there is a *critical period* (or optimal period) in infancy during which neural responsiveness to specific speech contrasts is either maintained/enhanced by exposure to those contrasts (i.e., neural commitment) or is lost through lack of early exposure to them due to a decline in neural plasticity (see, e.g., Weikum, Oberlander, Hensch, & Werker, 2012) and (b) that perceptual attunement to native speech involves *statistical learning* of the frequencies and distributions of phonetic features in native speech input (see, e.g., Pierrehumbert, 2003; cf. models combining the two hypotheses, e.g., Kuhl, 2004; Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Werker & Tees, 2005). Our alternative hypothesis is that humans remain capable of attending to and learning certain articulatory properties of nonnative speech contrasts across the life span rather than losing that ability entirely outside of some early critical period.

The assumptions of concern, which are held by both of those theoretical viewpoints, are (a) that the features on which exposure-based changes in speech perception rest are *auditory* in nature and (b) that infants' *differential exposure* to specific auditory features in ambient speech determines how perception of the corresponding phonetic contrasts will change developmentally. Our alternative assumptions are (a) that both infants and adults detect *articulatory* rather than auditory information in speech and (b) that certain types of articulatory information remain easily *detectable across the life span* even when perceivers lack exposure to them in native speech. Although many of the infant speech perception findings we review next are compatible with the auditory features assumption of the critical period and/or statistical learning hypotheses, other findings are incompatible with that assumption. Amount of auditory exposure falls short of a perfect fit with a number of observed developmental changes in infants' perception of nonnative as well as native speech contrasts. Moreover, as we discuss in more detail later, findings that *are* compatible with auditory exposure premises are also consistent with an articulatory basis for perceptual attunement, whereas the converse is not always the case.

As a heuristic for organizing our presentation of the full range of findings, we adopt Gottlieb's (1976, 1981) proposed epigenetic trajectories in perceptual development as they have been extended to infant speech perception (Aslin & Pisoni, 1980; Werker & Tees, 1992, 1999): MAINTENANCE, FACILITATION, INDUCTION, DECLINE, and FLAT.[1] Note that we use the terms only as descriptors of direction of developmental changes in speech perception. We are not committed to those authors' assumptions that developmental changes in infants' perception of nonnative speech contrasts reflect the operation of a critical period in early development during which exposure to specific speech features *must* occur in order for them to shape the neural mechanisms underlying speech perception. We note here some key evidence that runs counter to such a strict critical period hypothesis premise, evidence

[1] Aslin and Pisoni (1980) used the terms LOSS and NO EFFECT for the latter two patterns. LOSS was revised to DECLINE by Werker and Tees (1992); we adopt their term to avoid the connotation of *permanent* loss of ability. As the term NO EFFECT implies that exposure occurs but does not improve performance, we instead use the term FLAT to indicate that initially poor performance remains poor due to lack of exposure.

suggesting that neuroplasticity for language learning continues to operate across the life span. For example, some early developmental changes in speech perception can be greatly or even fully reversed by second language (L2) learning/training in adulthood (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tokhura, 1997; Flege, 1984; Flege & MacKay, 2004; Guion, Flege, Akahane-Yamada, & Pruitt, 2000; MacKain, Best, & Strange, 1981), and adults discriminate a number of unfamiliar nonnative contrasts, as well as certain native contrasts, significantly better than infants (e.g., Best & McRoberts, 2003; Polka, Colantonio, & Sundara, 2001).

Research on experience-related changes in infant speech perception has focused more on nonnative distinctions than on native ones (see Table 1 for a summary of developmental speech perception findings according to the posited epigenetic trajectories). A DECLINE from initially good discrimination of nonnative consonant contrasts has been frequently reported, which appears compatible with the prediction that lack of early exposure should lead to a decrement in initially good discrimination. In these cases, discrimination is good prior to 6–7 months of age, but performance then begins to decrease around 8 months of age, and discrimination becomes nonsignificant by ~10 months (e.g., English-learning infants on Hindi dental vs. retroflex stops /t̪/-/ʈ/; Werker & Lalonde, 1988; Werker & Tees, 1984 and Nthlakampx velar vs. uvular ejectives /kʼ/-/qʼ/; Werker & Tees, 1984).

A comprehensive account of experiential effects in infant speech perception, however, must address both nonnative and native consonant contrasts. By statistical learning or critical period principles, infant perception of native contrasts should show one of the following developmental trajectories: MAINTENANCE of initially good discrimination, or FACILITATION of initially moderate discrimination, or INDUCTION from initially poor discrimination levels. However, perception of native contrasts should never show a developmental DECLINE. Conversely, for nonnative consonant contrasts, that is, distinctions that are not used contrastively in the infant's environment, both of those theoretical approaches predict only developmental DECLINE or flat trajectories. Neither predicts MAINTENANCE of initially good discrimination nor improvement of initially moderate or poor discrimination, that is, no FACILITATION or INDUCTION, for nonnative consonant contrasts the infant has not been exposed to (e.g., see Kuhl, 2004; Kuhl et al., 2008; Kuhl et al., 2006; Werker, 1989).

Studies on infants' discrimination of native consonant contrasts have indeed often found the developmental patterns that are expected on the basis of exposure, that is, simply because the contrasts are present in native speech input. MAINTENANCE of good discrimination from 6–8 months through 10–12 months has been observed in numerous studies. For example, English-learning infants show maintenance of good discrimination across the 1st year for English /b/-/d/ (Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Werker & Lalonde, 1988; Werker & Tees, 1984), and both English- and French-learning infants maintain good discrimination for English/French /b/-/v/ (Polka et al., 2001). A recent cross-language comparison found one of the other expected patterns, FACILITATION in native English-learning infants and DECLINE in nonnative Japanese-learning infants, for English /r/-/l/, which showed modest discrimination in both groups at 6–8 months. By 10–12 months, the native English infants' discrimination improved significantly, but the nonnative Japanese infants' discrimination dropped to nonsignificant (Kuhl et al., 2006). INDUCTION of discrimination by native English-learning infants has also been observed for the English /d/-/ð/ contrast that was not discriminated by either native English-learning or nonnative French-learning infants at 6–8 months (Polka et al., 2001). That contrast was discriminated by English but not French

Table 1. Summary of prior findings on infants' perception of native and nonnative consonant contrasts at 6–8 versus 10–12 months of age, as interpreted in terms of the epigenetic trajectories posited for infant speech perception development (Aslin & Pisoni, 1980) and suborganized according to Articulatory Organ Hypothesis predictions for between-organ [unshaded] versus within-organ [light shading] contrasts (AOH; Best & McRoberts, 2003; L. M. Goldstein, 2003; L. M. Goldstein & Fowler, 2003; Studdert-Kennedy & Goldstein, 2003) and our newly introduced privative contrasts [medium shading] (+/− gesture of a given articulator).

Consonant contrast	Stimuli/Infant language	AOH contrast	Organ gesture distinction	References
Maintenance				
English /b/-/d/ *bilabial vs. alveolar stops*	Native/English	Between-organ	Lips vs. tongue tip closure	Best et al. (1995); Werker & Lalonde (1988); Werker & Tees (1984)
English /b/-/g/ *bilabial vs. velar stops*	Native/English	Between-organ	Lips vs. tongue dorsum closure	Moffitt (1971); Morse (1972)
English /s/-/ʃ/ *alveolar vs. palatal fricatives*	Native/English	Between-organ	Tongue tip vs. dorsum critical	Eilers & Minifie (1975); Holmberg et al. (1977)
English /f/-/θ/ *labiodental vs. dental fricatives*	Native/English	Between-organ	Lip vs. tongue tip critical	Holmberg et al. (1977); Levitt et al. (1988); Tyler et al. (2014); cf. Eilers et al. (1977)
Zulu /ǀ/-/ǁ/ *dental vs. lateral clicks*	Nonnative/ English	Between-organ	Tongue tip vs. dorsum closure	
Tigrinya /pʼ/-/tʼ/ *bilabial vs. alveolar ejectives*	Nonnative English	Between-organ	Lips vs. tongue tip closure	Best & McRoberts (2003); Best et al. (1995); Best et al. (1988)
!Xóõ /ʘˣ/-/ǀˣ/ *velar-fricated bilabial vs. dental clicks*	Nonnative English	Between-organ	Lips vs. tongue tip closure	Tyler et al. (2014)
Nuu-Chah-Nulth /x/-/χ/ *velar vs. uvular fricatives*	Nonnative English	Between-organ	Tongue dorsum vs. root critical	Tyler et al. (2014)
Nuu-Chah-Nulth /χ/-/ħ/ *uvular vs. pharyngeal fricatives*	Nonnative English	Between-organ	Tongue root vs. aryepiglottis critical	Tyler et al. (2014)
English /b/-/v/ *bilabial stop vs. labiodental fricative*	Native English	Within-organ	Lip: bilabial closure vs. dental critical	Polka et al. (2001)
English /tʃ/-/ʃ/ *palatoalveolar affricate vs. stop*	Native English	Within-organ	Tongue tip closure vs. critical	Tsao et al. (2006)
English /s/-/θ/ *alveolar vs. dental fricatives*	Native English	Within-organ	Tongue tip critical alveolar vs. dental	Tyler et al. (2014)
English /b/-/m/ *bilabial oral vs. nasal stops*	Native English	Privative	+/− velum-lowering gesture	Eimas & Miller (1980)
English /b/-/p/ *bilabial voiced vs. voiceless stops*	Native English	Privative	+/− glottal-opening gesture phased with release of lip closure	Eilers et al. (1979); Eimas et al. (1971)
Spanish /b/-/p/ *bilabial prevoiced vs. unaspirated stops*	Native Spanish	Privative	+/− glottal-opening gesture phased with lip closure	Eilers et al. (1979); Lasky et al. (1975)
	Native*/Kikuyu			Streeter (1976) *NOTE: Kikuyu has this voicing contrast for dental and velar but not bilabial stops*
Facilitation				
English /r/-/l/ *rhotic vs. lateral approximants*	Native English	Within-organ	Tongue tip closure vs. narrow + root narrow pharyngeal vs. uvular	Kuhl et al. (2006)
Induction				
English /d/-/ð/ *alveolar stop vs. interdental fricative*	Native English	Within-organ	Tongue tip alveolar closure vs. interdental critical	Polka et al. (2001); Sundara et al. (2006)

(*Continued on next page*)

Table 1. (*Continued*)

Consonant contrast	Stimuli/Infant language	AOH contrast	Organ gesture distinction	References
English /s/-/z/ voiced vs. voiceless *alveolar fricatives*	Native English	Privative	+/− glottal-opening gesture	Best & McRoberts (2003); Eilers (1977); Eilers & Minifie (1975); Eilers et al. (1977); cf. Best et al. (2001)
Decline				
Hindi /t̪/-/ʈ/ *dental vs. retroflex stops*	Nonnative English	Within-organ	Tongue tip closure dental vs. retroflex	Anderson et al. (2003); Werker & Lalonde (1988); Werker & Tees (1984)
Czech /z/-/r̝/ (ř) *alveolar voiced fricative vs. fricated trill*	Nonnative English	Within-organ	Tongue tip alveolar critical vs. (loose) closure	Trehub (1976)
Nthlakampx /kʼ/-/qʼ/ *velar vs. uvular ejectives*	Nonnative English	Within-organ	Tongue dorsum closure velar vs. uvular	Best et al. (1995); Werker & Tees (1984); cf. Anderson et al. (2003)
Zulu /k/-/kʼ/ *voiceless vs. ejective velar stops*	Nonnative English	Within-organ	Glottal closure vs. opening gestures	Best & McRoberts (2003)
Mandarin /tɕ/-/ɕ/ *palatal affricate vs. fricative*	Nonnative English	Within-organ	Tongue dorsum closure vs. critical	Tsao et al. (2006)
English /r/-/l/ *rhotic vs. lateral approximants*	Nonnative/ Japanese	Within-organ	Tongue tip closure vs. narrow + root narrow pharyngeal vs. uvular	Kuhl et al. (2006)
Zulu /b/-/ɓ/ *plosive vs. implosive bilabial stops*	Nonnative English	Privative	+/− larynx-lowering gesture	Best & McRoberts (2003)
Zulu /ɬ/-/ɮ/ *voiceless vs. voiced lateral fricatives*	Nonnative English	Privative	+/− glottal opening	Best & McRoberts (2003)
English /tʃ/-/ʃ/ *palatoalveolar affricate vs. stop*	Nonnative/ Mandarin	Privative	Tongue tip closure vs. critical	Tsao et al. (2006)
Flat				
English /d/-/ð/ *alveolar stop vs. interdental fricative*	Nonnative/ French	Within-organ	Tongue tip alveolar closure vs. interdental critical	Polka et al. (2001); Sundara et al. (2006)
Spanish /b/-/p/ *bilabial pre-voiced vs. unaspirated stops*	Nonnative/ English	Privative	+/− glottal-opening gesture phased with lip closure	Eilers et al. (1979); Lasky et al. (1975)

adults, and by 4 years it was finally discriminated by English monolingual children but not by French monolingual children, who showed a FLAT trajectory of continued poor discrimination. Interestingly, success was delayed even further in bilingual children who were learning both English and French (Sundara, Polka, & Genessee, 2006; see also Bosch & Ramon-Casas, 2011; Curtin, Byers-Heinlein, & Werker, 2011; Shafer, Yan, & Datta, 2011; cf. exceptions to developmental delays, e.g., Sundara & Scutellaro, 2011).

It is important to note, however, several findings on infants' perception of both native and nonnative consonant contrasts are inconsistent with predictions based on differential early auditory exposure. Two studies have found DECLINE for discrimination of phonetic distinctions that *are* present in native input. In both cases, good performance at 6–8 months gave way to significantly lower discrimination at 10–12 months. Specifically, by 10–12 months English-learning infants showed a decline in discrimination of the English /s/-/z/ contrast (Best & McRoberts, 2003) and of the aspirated [t^h] versus unaspirated [t] contextual allophones of English /t/ (Pegg & Werker, 1997). In the former study, a comparable DECLINE was also found for a phonetically comparable

nonnative consonant contrast that does not occur even allophonically in English, the voiced versus voiceless lateral fricatives /ɬ/-/ɮ/ of Zulu (Best & McRoberts, 2003). That is, English-learning infants showed similar developmental decline for native English and nonnative Zulu coronal fricative voicing contrasts despite marked differences in exposure to the two contrasts.

Also unexpected by auditory exposure-based accounts are four reports of the converse pattern. Three studies found MAINTENANCE of initially good discrimination from 6 through 12–14 months for nonnative consonant place of articulation contrasts that are completely lacking even as allophones in the infants' language environment. Specifically, English-learning infants show no decline across the 1st year in discrimination of the Zulu dental versus lateral click consonants /ǀ/-/ǁ/ (Best et al., 1995; Best, McRoberts, & Sithole, 1988) and the Nuu Chah Nulth uvular versus pharyngeal voiceless fricatives /χ/-/ħ/ (Tyler, Best, Goldstein, & Antoniou, 2014). The fourth study reported MAINTENANCE of English-learning infants' good discrimination of the Tigrinya ejective stop contrast /pʼ/-/tʼ/, which uses a voicing manner that does not occur in native speech (Best & McRoberts, 2003).

These findings all run counter to the hypothesis that the patterns of developmental change in perception of native versus nonnative distinctions are determined by differential early exposure to specific auditory features in speech. As commonsense as that assumption may seem, it fails to handle the full range of infant speech perception findings. An alternative account is needed that coherently addresses both sets of perceptual results, the ones that are expected as well as the ones that are unexpected, according to auditory exposure.

In addition to accounting for infant speech perception findings that are discrepant to the differential acoustic exposure predictions, however, achieving the goal of understanding how infants develop as integrated perceivers and speakers of a native language requires that we also consider the emergence of native language biases in infants' speech-like productions. Unfortunately, the literature on experiential effects in perception has not directly addressed how they relate to developmental changes in speech-like production in infancy. Therefore, we turn now to the literature on the appearance of native language biases in speech production during the 1st year to see if we can discern any parallels to the picture we have observed for perceptual development. We again focus mainly on consonants.

Becoming a native speaker

Classical propositions (Jakobson, 1968) were that infants' prelinguistic speech-like vocalizations (a) reflect biologically driven motoric patterns that are functionally independent of language, (b) are thus universal across language environments, (c) and contain virtually all phonetic elements found across languages as well as some unattested in any known language (d) as well as being temporally and substantively discontinuous with the child's early spoken words. There have been no direct evaluations of the first hypothesis, that infant vocalizations are biologically driven motoric patterns unconnected to language; indeed, this premise may be impossible to test. Empirical studies have, however, evaluated the third and fourth tenets, that is, segmental exuberance and discontinuity from early words, and have clearly refuted both. The range of consonant-like elements in infant babbling, across numerous language environments, is quite restricted rather than broadly inclusive (e.g., Cruttenden, 1970; Irwin, 1947, 1948; Kent & Murray, 1982; Matyear, MacNeilage, & Davis, 1998). In addition, empirical evidence clearly shows continuity and similarity, rather than discontinuity, between

babbling and true words, regardless of whether group patterns or individual idiosyncrasies are tracked over early development. The two types of productions co-occur across the months surrounding the child's first birthday, and even more important, the same segmental biases are found in late babbling and early words, for example, vocal motor routines, word templates (Blake & de Boysson-Bardies, 1992; Keren-Portnoy, Majorano, & Vihman, 2009; Velleman & Vihman, 2007; Vihman & Croft, 2007; Vihman, Macken, Miller, Simmons, & Miller, 1985).

But the second premise, that infant vocalizations display universal traits, remains uncontested, especially for the first 9 months of life. It has played a central role in research on early vocal development (e.g., English: Stark, 1980; English and Spanish: Oller, 1980, 2000; French: Konopczinski, 1990; Swedish: Roug, Landberg, & Lundberg, 1989). Moreover, it serves as the foundation for theoretical models of the biological mechanisms that have been proposed as the primary drivers of vocal development patterns, for example, the Frame/Content theory proposes oscillatory cycles of jaw opening/closing as the "frame" for the emergence of syllabic speech structure and differentiation (e.g., Davis & MacNeilage, 1995; MacNeilage, Davis, Kinney, & Matyear, 2000).

However, it is the complement to that classic premise that is our primary interest in this article: How does native language input come to *modify* the infant's initial, apparently predetermined, vocal behavior patterns? Roger Brown (1958) addressed this issue in his proposal that babbling is functionally related to and continuous with true language development, counter to the aforementioned classical premise four, and thus serves as the child's groundwork for learning a native language. He posited that this should be evident in "babbling drift," that is, a tendency for preverbal vocalizations to increasingly reflect segmental and prosodic properties of the language environment, with a concomitant decrease in the properties that that language lacks, rather than to continue to show only universal patterns. This provides, at last, a hypothesis about early productive development that implies it must be grounded in infants' perception of native speech. However, Brown's focus on lexical and morphological development implies that the nature of information most relevant to the child's attunement in producing native speech patterns is *not* auditory but rather is more abstract, that is, phonological.

Empirical studies have examined Brown's (1958) babbling drift prediction with mixed results for certain methods and clearer outcomes for others. Transcription-based observational case studies differ as to whether they found support for babbling drift in inventory of consonantal productions near the end of the 1st year or failed to find evidence in favor of such drift. For example, evidence of drift toward the proportions of consonants seen in adult native speech was found in the babbling English-learning twin infants (Cruttenden, 1970), a French-learning infant (de Boysson-Bardies, Sagart, & Bacri, 1981) and an English- versus a French-learning infant (Levitt & Aydelott Utman, 1992). Conversely, no evidence of babbling drift was found in another English-learning infant (Davis & MacNeilage, 1995) or in four Swedish-learning infants (Roug et al., 1989). But generalizability is a concern with case studies or small *n*'s, especially those involving only one or two languages. And phonetic transcriptions of infant vocalizations show relatively low interrater reliability (Stockman, Woods, & Tishman, 1981); additional concerns arise from native language speech perception biases that affect transcriptions even by well-trained observers (e.g., Oller, 2000).

Subsequent researchers have examined larger numbers of infants from contrasting language environments and assessed interlanguage reliabilities between transcribers of different native languages. Still, findings on language-specific drift in the distributions of consonant place and manner produced by 12 months have been mixed, some positive but others mixed or negative. For example, the distributions of babbled consonants across consonant manner and place in English, French, Japanese versus Swedish infants were found to be differentiated by 12 months in ways that correspond to the differences among those langauges (de Boysson-Bardies & Vihman, 1991), but conversely it has been reported that consonantal distributions are similar rather than showing language-specific differences in English versus Korean infants (Lee, Davis, & MacNeilage, 2010) and in French, Romanian, Dutch, and Arabic infants (Kern, Davis, & Zink, 2009), whereas the voice onset time [VOT] of stops in babbling failed to differ between English versus Spanish infants despite the VOT differences in adult speakers of those languages (Oller & Eilers, 1982).

Another approach has been to assess adults' abilities to identify or discriminate sections of recorded babbling from infants being reared in different language environments. Discrimination or identification of the home language is interpreted as evidence that the infants' babbling has already drifted toward their native languages. Again, some results have been positive, others negative. For example, one study found that listeners could distinguish the babbling of French, Arabic, and Mandarin Chinese infants (de Boysson-Bardies, Sagart, & Durand, 1984) whereas other studies have failed to find reliable discrimination of babbling by English versus Swedish infants (Engstrand, Williams, & Lacerda, 2003) and by English versus Spanish infants (Thevenin, Eilers, Oller, & Lavoie, 1985). But those types of judgments also have limitations. Most important, even when adults recognize language-specific features of recorded babble, that does not provide unique evidence for *segmental* drift because even filtered babbling retains prosodic patterning for which native biases *are* present from early on: English versus French 6-month-olds display language-specific prosodic differences in their babbling (Levitt & Wang, 1991; Whalen, Levitt, & Wang, 1991). Indeed, some German versus French prosodic differences are already evident in newborn vocalizations (Mampe, Friederici, Christophe, & Wermke, 2009). Whether or not untrained listeners can distinguish babble from different language environments, their performance likely reflects greater sensitivity to prosodic than to consonantal differences (e.g., Engstrand et al., 2003).

Instrumental studies, on the other hand, have revealed reliable, although modest, shifts toward more native-like acoustic properties and/or statistical distributions of consonants by the final quarter of the 1st year. Several studies have used phonetic transcription by trained listeners to examine the relative frequencies of occurrence of a range of consonants (e.g., /b/, /d/, /g/, /n/, /m/) in the recorded babbling of infants being raised in English, French, Swedish, Japanese, and Mandarin language environments and have found language-specific distributional differences by 10- to 13-month-olds (L. M. Chen & Kent, 2010; de Boysson-Bardies et al., 1981; de Boysson-Bardies & Vihman, 1991; de Boysson-Bardies et al., 1992; Levitt & Aydelott Utman, 1992; Levitt & Wang, 1991). Moreover, a recent acoustic examination of VOTs in the syllable-initial stops of infants from two language environments found native language-consistent differences in proportion of prevoiced stops produced by French versus English infants at 9 months (Whalen, Levitt, & Goldstein, 2007). However, the native VOT biases appeared only for voiced stops, and the magnitude of prevoicing did not yet

match that of either target language. Although this indicates emerging control of language-specific production parameters for prevoicing, it is far from adult-like. Therefore, it is no surprise that transcribers might miss such subtle, partial correspondence to the adult language (as they failed in, e.g., Oller & Eilers, 1982).

Perception–production relations in infancy?

Thus, it appears that native-language biases in consonant production do emerge in the 1st year, around 10–13 months, which coincides with the 10- to 12-month decline in discrimination of many, though not all, nonnative consonant contrasts. However, the findings in the two domains make it difficult to evaluate more specific developmental interrelations between perception and production of consonants, as the research has not examined attunement to the same aspects of consonants. The work on infant consonant *perception* has focused largely on discrimination of language-specific minimal contrasts, or preferences for native phonotactic patterns, that is, language-specific constraints on which consonants may be combined, and where they may appear in a word (Friederici & Wessels, 1993; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk & Luce, 1994). Studies that have examined infant consonantal *production*, on the other hand, have examined language-relevant shifts in proportions of various consonants produced in comparison to their proportions in the native inventories, or to manner or place subclasses, or to acoustic features of given consonant types. One cannot address changes in infants' production of minimal contrasts per se, as there is no way to know which consonant she intended to produce in babbling (or even whether she had an intended target!). Moreover, even if the two research domains could examine the same aspects of consonants, identifying developmental parallels would not clarify causal direction. Such parallels could give only a hint that perception and production may be working together.

Nonetheless, attunement of perception and production obviously must be interrelated for efficient development of a specific native language. Infants do show drift toward the consonantal makeup of the native language in their babbling by the last quarter of their 1st year, and they also imitate various aspects of speech presented to them, as we summarize later. From there, they soon begin learning to produce native words (though initially the correspondence between their production and the target word may be difficult for adults to recognize!) and eventually to utter words recognizably to all listeners and according to the regional accent of their language environment. The developmental relationship between perception and production would be most direct if both domains relied on the same type of information.

Most current theories of infant speech development, however, have not explicitly identified a common type of information across the two domains. Instead, most have accounted for experience-based native language attunement in only one or the other domain without even considering how they (must) interact. Most accounts have focused on identifying the mechanisms or processes that effect experience-related changes in speech perception *or* production without critically considering what type of speech information those mechanisms or processes should operate on in either domain, let alone across them. Moreover, as noted earlier, most models have assumed that speech perception rests on *auditory* representations and/or that speech production involves *motoric* representations, that is, commands to move specific vocal tract muscles/groups in specified ways. Not even the few

models that have touched on perception–production relations in development have offered a well-articulated account of how those two disparate types of information—auditory and motoric—could get translated between the two domains. Yet the information infants rely on in perception and production is central to how they co-attune to native speech in the two domains.

Informational basis of infants' attunement to native speech

Statistical learning … of what?

Numerous existing models of infants' attunement to the native language argue that it reflects infants' detection of statistical regularities in speech input (e.g., Kuhl et al., 2008; Vihman, DePaolis, & Keren-Portnoy, 2009; Werker & Curtin, 2005). The premise is that infants are able to track or calculate distributional frequencies and/or co-occurrence probabilities in native speech (e.g., transitional probabilities), which thereby adjust their speech representations to be consistent with the input statistics (see the statistical accounts presented by, e.g., Aslin, Saffran, & Newport, 1998; Pierrehumbert, 2003; Saffran, Aslin, & Newport, 1996; Swingley, 2005). Statistical procedures can be invaluable for identifying stable (or quasi-stable) patterns within a data set that deviate reliably from random "data noise."

However, such statistics are informative only to the extent that the data over which they are computed are suitable indices of the phenomena the observer needs to understand or learn. The information on which a learner conducts statistical computations determines the possible meanings she can grasp regarding the source phenomena. But this informational issue is rarely if ever addressed by proponents of statistical learning accounts; they simply assume unquestioningly that the statistics are tracked/computed on *auditory* properties. This leaves a gaping hole in existing statistical learning accounts because in order for an infant to become a native perceiver/speaker, the nature of information in speech over which statistics are computed constrains what is learnable about the spoken language (e.g., transitional probabilities are learned much more easily for consonant than vowel sequences; Mehler et al., 2006). Our point here is not to dispute or confirm whether sensitivity to statistical regularities contributes to perceptual attunement to native speech. Rather, our aim is to reach a better understanding of the nature of information in speech infants may be learning the statistics of, such that their perception and production of speech become co-attuned.

Perceptual tuning models: What information do infants attune to in native speech?

Most models of language-specific influences on infant speech perception assume that differential exposure to specific *acoustic properties* in speech is solely or primarily responsible for differences in the direction and degree of developmental change in auditory perception of the corresponding phonetic contrasts (NLM [Native Language Magnet]: Kuhl, 1993; NLNC [Native Language Neural Commitment]: Kuhl, 2004; NLMe [Native Language Magnet-expanded]: Kuhl et al., 2008). Some posit that this auditory tuning is central to acquiring recognition of spoken words (WRAPSA [Word Recognition and Phonological Structure Acquisition]: Jusczyk, 1997; DRIBLER [Dimensionally Reduced Item-Based LExical Recognition]: Anderson, Morgan, & White, 2003; see PRIMIR [Processing Rich

Information from Multidimensional Interactive Representations] for inclusion of articulatory and visual information as well as auditory information: Curtin et al., 2011; Werker & Curtin, 2005).

The auditory focus is consistent with an overriding tendency in speech perception research to employ unimodal acoustic-only target stimuli in studies of both adults and infants. However, as we noted early on in this article, infants acquire language largely in face-to-face interactions, which provide a multimodal complex of dynamic visual-facial (talking face), kinesthetic-proprioceptive (self-production), and haptic/tactile information (touching the caregiver's talking face, feeling the breath of her aspirated consonants) in addition to acoustic consequences of the caregiver's speech, all of which are highly intercorrelated over time. For example, there is evidence that adults automatically integrate haptic/tactile speech information with auditory information when perceiving consonants (Derrick & Gick, 2013; Fowler & Dekle, 1991; Gick & Derrick, 2009).

Some models do acknowledge this multidimensionality of speech, indicating that infants track the statistical distributions of articulatory as well as auditory features of speech and incorporate them into their phonetic representations (e.g., Pierrehumbert, 2003; PRIMIR: Werker & Curtin, 2005). Even models that acknowledge a role for articulatory information, however, tend to assume that associative processes are required to establish links between nonacoustic properties of speech and auditory representations, which are implied or explicitly posited to be more basic (see FLMP [Fuzzy Logical Model of Perception]: Massaro, 1984; see also Kuhl & Meltzoff, 1982, 1984, 1996; Vihman, DePaolis, & Keren-Portnoy, 2009). But such assumptions have simply been stated with neither direct evidence nor logical evaluation to support them.

The Perceptual Assimilation Model (PAM; Best, 1993, 1994; Best et al., 1988; see schematic diagram, Figure 1) posits that infants perceive articulatory information in speech, but its premises differ from those of Pierrehumbert (2003) and Werker and Curtin (2005) in two key ways: PAM assumes that (a) articulatory rather than auditory information is fundamental and that (b) articulatory information is amodal rather than specific to any single modality of energy or representation. PAM is not a Motor Theory of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985), that is, it does not espouse the principle that speech is perceived via reference to the motor commands the perceiver would need to issue in order to produce a match of the target signal (see Best, 1995). PAM instead explicitly assumes that articulatory information is amodal, that it encompasses perception and production, and that it is instantiated across multiple modalities of energy in both domains. It is amodal on the production side because it does not reside solely in motoric representations or neuromotor commands and is amodal on the perception side because it does not reside solely in unimodal proprioceptive, kinesthetic or haptic, or auditory or visual signals. Nor is articulatory information a set of learned associations among modality-specific features that have formed over time through experiencing the co-occurrence of patterns between modalities (e.g., Kuhl & Meltzoff, 1982, 1984, 1996; Massaro, 1984). Articulatory information instead refers to the location(s) and degree of constriction of a constriction action (gesture), which shape the dynamically correlated changes in the multiple energy modalities. Whereas the auditory/acoustic approach focuses on information in a single energy modality, articulatory information is *amodal* as it refers to the multimodal energy fluctuations that the constriction gestures of vocal tract articulators give rise

to. And although a constriction gesture results in dynamically correlated changes across multiple signal modalities, the constriction gestures are themselves amodal actions.[2]

Although PAM does not specifically invoke statistical mechanisms to account for perceptual attunement to native speech as the other models do, it does share their underlying assumption that infants seek and learn informational regularities in native speech input. This observation leads back to our core question about the nature of information utilized in speech perception and production, somewhat rephrased: If regularities in native speech are the basis for developmental changes in infant speech perception, then regularities in *what type* of information could lead to the array of developmental patterns that have been observed across nonnative and native contrasts? How could the nature of information help us to understand why both native and nonnative consonant contrasts show varying patterns of perceptual change across infancy, with some of each type showing MAINTENANCE whereas others show FACILITATION or INDUCTION, and still others show some DECLINE?

Two possible auditory factors have been proposed by other models to account for those variations in developmental patterns: (a) differences in acoustic salience and (b) differences in statistical distribution of acoustic features in speech. However, neither of these acoustic factors can explain why English-learning infants discriminate native /b/-/v/ at both 6 and 10 months (MAINTENANCE) and yet fail at both ages (delayed INDUCTION) with native /d/-/ð/ (Polka et al., 2001). Both are native English contrasts, and the consonant differences in the two contrasts are relatively comparable to each other acoustically and phonetically, as well as with respect to the numbers of lexical items they appear in, and the frequency of usage of those lexical items. Both contrasts distinguish between a voiced stop and a fricative. Each contrast uses a single active articulator to achieve contrasting constrictions differing in degree and place of articulation (POA). Specifically, /b/-/v/ use the lips, /d/-/ð/ the tongue tip, to achieve complete closure for the bilabial stop /b/ and alveolar stop /d/ but critical closure resulting in turbulent airflow for the labiodental fricative /v/ and dental fricative /ð/. Both /b/ and /d/ appear in many English lexical items (e.g., > 3,000 words begin with /d/, nearly 3,200 words begin with /b/; Celex database: Baayen, Piepenbrock, & Gulikers, 1995), including many early vocabulary items, and both can occur in a range of phonotactic contexts (e.g., consonant clusters in both onsets and codas). By comparison, /v/ and /ð/ each occur in many fewer words (respectively, < 800 words begin with /v/, < 60 begin with /ð/; Celex: Baayen et al., 1995) and phonotactic contexts: neither can occur in consonant clusters. On the other hand, both /v/ and /ð/ occur in highly frequent grammatical words, auxiliaries, and/or modifiers (e.g., ***v**ery, ne**v**er and ha**v**e; **th**e, that, and o**th**er*).

Nor can statistical learning of the frequency distribution of acoustic features in infants' input account for the DECLINE in discrimination observed with some distinctions that do occur in native speech or on the other hand for the developmental MAINTENANCE of initially good discrimination that has been observed with some nonnative consonants that are entirely lacking from native speech. For example, the observed MAINTENANCE of good discrimination of Zulu click consonants by English perceivers throughout infancy and into

[2] An actual constriction gesture results from activation of a coordinative structure, which is an abstract control structure that achieves a given constriction type by coordinating the collective action of sets of articulators, the individual contributions of which can vary across contexts (Fowler, Rubin, Remez, & Turvey, 1980). For example, the upper and lower lips and jaw work together to effect bilabial closure, and if there is a brief constraint (contextual or externally imposed) on the motion of one of the articulators, the others immediately compensate, such that bilabial closure is still achieved (e.g., Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984; see also Auban & Ménard, 2006).

adulthood (Best et al., 1995; Best et al., 1988) cannot be explained by an acoustic-exposure hypothesis, as no consonants acoustically or phonetically similar enough to clicks occur in English speech.

PAM offers a viable explanation of the Zulu click findings: clicks deviate so dramatically from any English articulations that they are perceived as nonspeech by English-speaking adults, who discriminate click consonant contrasts on that basis. However, neither the acoustic accounts, nor indeed PAM by itself, predict either the observed DECLINE between 6–8 and 10–12 months for certain native contrasts (Best & McRoberts, 2003; MacKain, 1982) or the DECLINE by 10–12 months for certain nonnative contrasts that subsequently show FACILITATION by adulthood despite lack of exposure. For example, the Zulu voiced versus voiceless lateral fricatives /ɬ/-/ɮ/ and ejective versus aspirated voiceless stops /kʼ/-/k/ show a decline in discrimination by 10–12 months in English-learning infants (Best & McRoberts, 2003), but English-speaking adults discriminate them quite well (Best, McRoberts, & Goodell, 2001).

Combining PAM with the Articulatory Organ Hypothesis (AOH; Best & McRoberts, 2003; L. M. Goldstein & Fowler, 2003), however, offers ways to account for those and other findings that pose challenges for the models described earlier. The AOH posits that between-organ distinctions (e.g., a given constriction made by the tongue tip vs. the same constriction type made by the tongue body, e.g., !Xóõ dental vs. lateral clicks /ǀ/-/ǁ/) should be easily detectable from the start, that is, even by newborns, and should remain easily discriminated throughout development, whether or not they occur contrastively in the perceiver's environment. In contrast, the AOH predicts that within-organ distinctions in constriction degree and/or location (i.e., dental vs. retroflex tongue tip constriction locations, e.g., the Hindi stops /d̪/-/ɖ/) may be initially difficult to discriminate even when they do occur in the language environment (see Figure 2 for a schematic diagram of the articulatory organs, their spatial/functional organization, and the parameters for their constriction gestures). That is, between-organ contrasts are universal, detectable by newborns, but within-organ contrasts differ across languages and often require experiential tuning. Combining these AOH principles with PAM's premises (PAM-AOH) leads to the prediction that discrimination of within-organ articulatory gesture contrasts should improve with experience if they are employed in the native language but should fail to improve developmentally if they are nonnative (see empirical tests of PAM-AOH; Best & McRoberts, 2003; Tyler et al., 2014; cf. Kuhl et al., 2006).

We examine in detail here how well the full range of findings on developmental change in infants' perception of native and nonnative consonant contrasts is coherently accounted for by the current version of PAM-AOH and offer an extension of the model to account for contrasts that may not fit the within- versus between-organ dichotomy. But first we return to a core question raised earlier, one that is relevant to our consideration of existing perceptual findings from the PAM-AOH perspective: How could perceptual attunement to native speech help to inculcate the phonetic and phonological properties of the native language into infant vocal *productions*?

Infant speech production attunement: How does babbling drift?

If babbling drifts toward native language consonantal properties (Brown, 1958), as the reviewed evidence indicates, this necessarily entails that perceptual attunement to native speech must shape those language-specific developmental changes in infants' speech-like productions. Yet most research on babbling drift fails to address either the nature of

information that infants may perceive in the speech of caregivers and other adults or how they relate that perceived information to their own productions. Both issues are discussed, however, in a recent report (L. M. Chen & Kent, 2010). Those authors posit that (a) perceptual tuning to native speech occurs via adjustments in sensitivity to the "external auditory patterns" of the infant's language environment, (b) perception–action links are then generated that associate internal representations of those auditory patterns to representations of specific articulatory muscular actions the infant has produced (see Kuhl & Meltzoff, 1982, 1996, for a similar argument re: infant vocal imitation), and (c) those learned perception–action associations are effected by *mirror neurons* (see Vihman, 2002, for a similar account of the similarity between segmental preferences in late babbling and in early words). Mirror neurons are brain cells that actively fire not only when an individual animal/human performs a particular action but also when the same individual watches another animal/human perform the same action or an analogous action that achieves the same goal. Thus, the other's behavior is "mirrored" in the observer's neural activity. The mirror neuron hypothesis for linking speech perception and production in infancy extends from broader speculation that mirror neurons underlie human (adult) speech perception (e.g., Fadiga, Craighero, & Olivier, 2005; Rizzolatti & Arbib, 1998; Rizzolatti & Craighero, 2004). But could mirror neurons actually handle the job of shaping infants' early speech-like productions toward the characteristics of the native speech they are exposed to?

Mirror neurons have been directly observed via intracellular recordings of neurons in monkey premotor cortex (frontal region F5) but have been investigated differently and more indirectly in humans. Instead of using intracellular recordings, some researchers have applied transcranial magnetic stimulation (TMS) to the surface of the head overlying the human brain region/network that is homologous to the monkey mirror neuron region, F5 (ventral premotor cortex, inferior frontal and parietal cortex, and superior temporal cortex). They have then assessed whether the TMS stimulation modulates, via corticospinal (CS) connections, the excitability of the relevant muscle groups in observers who are viewing or hearing other humans' actions. Such CS modulation is taken as evidence that the system contains mirror neurons. It has been observed in observers' hand muscles while they are watching others' manual actions (e.g., Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Strafella & Paus, 2000) and in their speech-relevant facial muscles while they watch others' visible speech articulations (e.g., Sundara, Namasivayam, & Chen, 2001). It is important to note that left hemisphere-only TMS also modulates hand muscle excitation during *auditory* perception of sounds generated by human bimanual actions (Aziz-Zadeh, Iacoboni, Zaidel, Wilson, & Mazziotta, 2004) as well as in speech-related facial muscle excitation during auditory perception of speech (Watkins & Paus, 2004; Watkins, Strafella, & Paus, 2003). Keep in mind that human mirror neurons are indirectly inferred from CS modulation of peripheral muscle groups via activation of a large neural network, however, rather than from direct cellular recording of actual mirror neurons.

The premise that a human mirror neuron region supports a direct *speech perception–production link* in humans has been criticized on both logical (e.g., Lotto, Hickok, & Holt, 2009) and empirical grounds; for example, speech perception is *not* disrupted in patients with lesions in the proposed mirror neuron region but is disrupted in patients with auditory and Wernicke's cortex lesions (Rogalsky, Love, Driscoll, Anderson, & Hickok, 2011). Moreover, mirror neurons have been posited in infants, but no actual

TMS or other purported evidence for their existence has been collected with infants. Nor have explanations been offered as to how exactly a mirror neuron network could link the auditory representation of a consonant (or vowel) in caregivers' speech to a motoric command representation or muscular action pattern for the "matching" infant articulation. Neither the acoustic nor the articulatory properties of infant vocalizations match those of adult speech due to adult–infant differences not only in articulatory speed and consonant–vowel co-articulation but also in vocal tract proportions and articulatory organ sizes, and those articulatory differences affect the specific patterns of formant structure and temporal change in the acoustic signal. For example, infant vocalization stimuli that adult listeners perceive as low vowels (e.g., /a/) cover broader acoustic areas than the corresponding adult vowels due to the developmental differences in interarticulator coupling of jaw and tongue body positions that result from vocal tract growth and reconfiguration. Thus, the "same" vowel reflects different articulatory configurations for an infant versus an adult (Ménard, Davis, Boë, & Roy, 2009).

More fundamentally, even if mirror neurons do exist and can account for human speech perception–production relations, they appear to reflect common amodal information for perception and production rather than forged linkages between disparate auditory and motoric modality-specific representations. The most comparable investigations of directly recorded monkey mirror neurons indicate that they respond to a given action that has significance for the animal regardless of which type of actor (self, other conspecific, other species) produces the action or which energy medium conveys the evidence. That is, they fire to the target action whether the proximal sensory signal is proprioceptive/kinesthetic (self-action), optical (seeing another perform the action), or acoustic (hearing oneself or another individual perform the action; Kohler et al., Rizzolatti, 2002). Those authors conclude that mirror neurons code the *meaning* of actions expressed in terms of goals, not in terms of linking or translation of modality-specific information. Thus, mirror neurons are unlikely to form unimodal representations of other-produced acoustic patterns and then form links between those auditory representations and self-produced speech motor patterns, as has been proposed for infant babbling drift and speech imitation (L. M. Chen & Kent, 2010; Kuhl & Meltzoff, 1982, 1996; Vihman, 2002). Instead, even monkey mirror neurons respond *amodally* to a specific meaningful action, whatever its distal source event or proximal medium of transmission. This is not compatible with claims that mirror neurons specifically link auditory representations to motoric ones. It is, instead, more compatible with our argument for amodal articulatory gestures as the common metric for developmental coattunement of speech perception and production.

Compatible with this view is complementary evidence that peripheral as well as central levels of the auditory system can, and do, extract articulatory information from speech. At the peripheral level, Ghosh, Goldstein, and Narayanan (2011) found that filtering acoustic speech signals using a system designed to match the critical band properties of human cochlea provided maximum mutual information with articulatory data of the same utterances, which had been co-recorded using electromagnetic articulography. That filter provided significantly better fit with the articulatory data than did filters employing any other theoretically possible critical band representations. At the central auditory level, Mesgarani, Cheung, Johnson, and Chang (2014) found electrocortiography evidence of spatiotemporal response fields

in the superior temporal gyrus (STG), a specialized speech-processing region of cortext, corresponding to phonetic features including both place and manner of articulation of consonants, place being obviously articulatorily defined. Although manner classes, such as stop versus fricative versus approximant, which can be related to either constriction degree properties or relatively simple acoustic properties, were more robustly represented, the place features of labial, coronal, and dorsal were nonetheless also distinguished in the spatiotemporal response fields of the STG.

Amodal articulatory information: Keeping the language-learning baby intact

All the king's horses and all the king's men couldn't put Humpty together again. (Opie, 1951)

The premise that speech perception and speech production are both based on amodal articulatory information in speech is more parsimonious than invoking some intermediate and yet-undefined translation or association-formation process that links up incommensurable unimodal representations for perception (auditory) and production (motoric). This is particularly important for infants to be able to attune efficiently and effectively to native speech in both perception and production. In the remainder of the article, we present the logic for our view and discuss several converging lines of evidence consistent with our premise that the foundation for infant speech perception and production is amodal articulatory information, that is, information about constriction gestures made by vocal tract articulators. We conclude with further discussion of PAM-AOH as a framework for examining the role of articulatory information in infants' integrated attunement of perception and production to the properties of native speech and its impact on perception of nonnative consonant contrasts.

Communicative benefits of a common articulatory metric

Speech is an ideal medium for language because it meets all of the fundamental requirements for sharing linguistically structured messages between communicative partners. First, speech as a system of articulatory gestures maintains *parity* between perceivers and talkers, that is, between perception and production (Fowler, 2004; see also the Motor Theory perspective; Liberman, 1996). That is, articulatory gestures count as the same information for both speaker and perceiver (L. M. Goldstein & Fowler, 2003; Studdert-Kennedy, 2002). Articulatory gestures give rise to dynamically correlated acoustic, optical, haptic, and proprioceptive consequences, any or all of which carry articulatory information. Second, articulatory gestures are *discrete* (categorical), which is essential to linguistic structure. A small set of distinct articulators is employed in creating vocal tract constrictions delineated by a small number of functionally distinct aperture sizes at a limited range of distinct locations along the vocal tract (L. M. Goldstein & Fowler, 2003). That is, they operate according to the same "particulate principle" that is also exemplified by the role of DNA in genetic transmission (Studdert-Kennedy, 1998, 2002; Studdert-Kennedy & Goldstein, 2003). Third, articulatory gestures are *recombinable* (combinatorial function), allowing virtually infinite numbers of multigesture constellations (e.g., meaningful words) to be formed

from that finite set of gestural elements. Fourth, articulatory gestures are themselves *intrinsically meaningless*, permitting them to be recombined into different words that can convey distinct and often unrelated meanings. Together, these characteristics of articulatory gestures meet the *informational* criteria necessary for communication between people of a given language community (Fowler & Galantucci, 2008; L. M. Goldstein & Fowler, 2003; Studdert-Kennedy, 2002). By comparison, auditory or motoric patterns fail the parity criterion, and each offers a weaker fit (than articulatory gestures) to one or more of the other informational criteria.

Articulating with the native language community

The informational criteria for communicative partners to be able to share structured linguistic messages are central to an infant's progress from newborn to becoming a well-attuned member of her language community, that is, a native perceiver and speaker. We have argued that the required set of informational criteria are better and more parsimoniously satisfied by the common metric of articulatory gestures for perception and production rather than by acoustic cues or motoric commands that necessitate translational/associational linkages between them. We further propose, following from our earlier discussion of PAM-AOH, that (a) developmental co-attunement to native consonants diverges for critically different types of articulatory distinctions (AOH), and (b) perceptual attunement (PAM) to these types of native articulatory distinctions shapes the developmental emergence of language-specific effects in consonants in babbling and early words. The two types of articulatory distinctions previously posited by the AOH are *between-organ* versus *within-organ* distinctions, as summarized earlier, in which "organs" refers to the primary oral tract articulator/s (lips, tongue tip, tongue dorsum; see Figure 2) involved in effecting the contrasting constriction gestures (see Best & McRoberts, 2003; L. M. Goldstein & Fowler, 2003; Tyler et al., 2014). In this article, we add a third type of articulatory contrast that differs from within- and between-organ distinctions, a *privative* organ distinction between the presence versus absence of a specified gesture by one of the remaining articulatory organs, which are postoral, that is, posterior to the oral tract articulators (see Figure 2: velum, tongue root, larynx, and possibly a separate pharyngeal articulator [an aryepiglottic mechanism; Esling, 1996; Moisik & Esling, 2011], which would be situated in between the larynx and the velum in Figure 2). For privative organ contrasts, the "absent" gesture reflects the default speech setting for that articulator.

As we indicated earlier, between-organ contrasts are detectable from birth and are common if not universal across languages, whereas within-organ contrasts are less often universal and may require facilitation or even induction through experience with a native language that uses them contrastively. Greater sensitivity from very early in life to between-organ contrasts, even those that do not specifically occur in the infant's language environment, is consistent with broader evidence that newborns respond to natural partitions of the orofacial system that all humans possess from birth into actions made by *discrete organs* (Meltzoff & Moore, 1977, 1997). In addition to the oral and posterior vocal tract articulators already named, this system encompasses additional organs involved in emotional and other paralinguistic communicative expressions, such as eyebrows, eyes, and neck (which effects tilting, turning and bobbing of the head). Crucially, the combined orofacial system as a whole is

central to linguistic and paralinguistic/nonlinguistic communication between infants and the people in their lives, as perceiving and reproducing others' vocal and facial actions are essential to an infant's growing capacity to articulate well with other members of their linguistic and sociocultural community. We turn next to converging evidence that we believe supports an articulatory basis for native language attunement in infant speech perception and production. We then revisit PAM-AOH, where we present our reasoning on how it can account for the range of existing findings on early attunement in perception *and* production and can offer novel predictions for early development across the two domains.

Converging evidence for articulatory-based attunement of perception and production

Perceptual attunement to different types of articulatory organ contrasts

Here we consider the extent to which the full range of infant perceptual findings can be accounted for by the PAM-AOH model. As outlined earlier, PAM-AOH posits that young infants should easily discriminate between-organ contrasts across development, that is, the same constriction type as achieved by different oral articulators. This prediction holds whether those contrasts are between native consonants such as English /b/-/d/ or nonnative consonants such as !Xóõ bilabial versus dental clicks /ʘ/-/ǀ/, both of which are distinguished by full closure gestures by the lips vs by the tongue tip. Conversely, PAM-AOH predicts that young infants are likely to have moderate to substantial difficulty discriminating most if not all within-organ contrasts, that is, different constriction locations or degrees achieved by the same articulator. Again, this prediction holds whether the contrast is between native consonants such as English /d/-/ð/ (full closure [stop] vs. critically narrow [fricative] constriction gestures by the tongue tip) or between nonnative consonants such as Hindi dental-retroflex /d̪/-/ɖ/ (tongue tip closure at the back of the upper front teeth vs. at the alveolar ridge). However, experience with native speech will lead to improvements in discrimination of poor-to-moderate initial discrimination of native within-organ contrasts by late infancy or sometime in childhood, whereas lack of experience with nonnative within-organ contrasts should result in continued poor discrimination or decline from initially moderate discrimination. Alternatively, however, some within-organ contrasts *may* be initially discriminable. Such contrasts should continue to be well discriminated if experienced in the native language but should show decline if they are nonnative.

We add to those two articulatory contrast types the *privative* contrasts, as we defined earlier. As the privative contrast type was not considered previously, it is not yet clear whether such contrasts should pattern with between-organ contrasts because they are distinguished by an articulator action versus lack thereof or instead pattern with within-organ contrasts because they involve an action difference for a single articulator. The developmental trajectory, therefore, may or may not differ between native contrasts, such as for English-learning infants on English /b/-/m/ in which the velum-lowering gesture occurs only for /m/, and nonnative privative contrasts, such as for English-learning infants on Zulu voiced versus voiceless lateral fricatives /ɮ/-/ɬ/ in which a laryngeal glottis-opening gesture occurs only for the voiceless one.

How well do existing infant speech perception findings fit with those predictions for articulatory organ types? As it turns out, nearly all consonant perception findings,

including those described earlier, can be explained according to whether the target contrasts were between- or within-organ/privative and for the latter types of contrasts also according to whether or not they occur in native speech (see Table 1). The between-organ consonant contrasts that were investigated show good discrimination from the earliest ages tested and displayed MAINTENANCE of good discrimination across the 1st year, as predicted by PAM-AOH. For native contrasts, this pattern has been observed in English-learning infants between 6 and 12 months for discrimination of many native consonant contrasts including English voiced stops /b/-/d/ (Best et al., 1995; Werker & Lalonde, 1988; Werker & Tees, 1984), which are distinguished by lip versus tongue tip closure gestures; for English /b/-/g/ (Moffitt, 1971; Morse, 1972), which are distinguished by lip versus tongue body closure gestures; for English voiceless fricatives /s/-/ʃ/ (Eilers & Minifie, 1975; Holmberg, Morgan, & Kuhl, 1977), which are distinguished by critical constrictions of tongue tip versus tongue dorsum that produce noisy turbulent airflow (frication); and for English voiceless fricatives /f/-/θ/ (Holmberg et al., 1977; Levitt, Jusczyk, Murray, & Carden, 1988; Tyler et al., 2014; cf. Eilers, Wilson, & Moore, 1977), distinguished by critical constrictions of lips versus tongue tip.

MAINTENANCE of good discrimination over the same development time frame has also been observed for all between-organ nonnative contrasts that have been investigated as well, as predicted by PAM-AOH. It has been observed for English-learning infants' discrimination of Zulu dental versus lateral clicks /ǀ/-/ǁ/ (Best et al., 1995; Best et al., 1988), which are distinguished by tongue tip versus dorsum closure gestures; of !Xóõ velar-fricated bilabial versus dental clicks /ʘˣ/-/ǀˣ/ (Best, Kroos, & Irwin, 2014), which are distinguished by lips versus tongue tip closure; of Tigrinya ejectives /pʼ/-/tʼ/ (Best & McRoberts, 2003), also distinguished by lip versus tongue tip closures; and of Nuu Chah Nulth velar versus uvular and uvular versus pharyngeal voiceless fricatives /x/-/χ/ as well as /χ/-/ħ/ (Tyler et al., 2014), which in light of the pharyngeal articulator mentioned earlier and the fact that these are fricatives (Esling, 1996; Moisik & Esling, 2011) we now believe are distinguished by critical constrictions of the tongue dorsum (/x/) versus tongue root (/χ/) versus aryepiglottis (/ħ/).

Conversely, some of the native within-organ consonant contrasts that were examined appear to require experience with native speech to support either INDUCTION from initially poor discrimination or FACILITATION from initially moderate discrimination to significantly improved discrimination by 10–12 months or later. Consistent with PAM-AOH predictions, English-learning infants show INDUCTION of discrimination from initially poor performance on English /d/-/ð/ (Polka et al., 2001; Sundara et al., 2006), which are distinguished by tongue tip constriction degree (closed for /d/, critical for /ð/) and location (alveolar for /d/, dental for /ð/). FACILITATION from initially moderate discrimination has been observed for English-learning infants on English /r/-/l/ (Kuhl et al., 2006), which are distinguished by a combination of tongue tip narrow constriction versus closure plus a location difference in narrow tongue root constriction[3] for /r/ (upper pharynx) versus /l/ (uvular).

Other native within-organ contrasts, however, have shown MAINTENANCE of good early discrimination. Although such initially high levels of discrimination may seem at first glance

[3]By Esling's (1996) analysis, the so-called pharyngeal constriction for /r/ is too high to be aryepiglottic and would instead be a tongue root gesture, as is the uvular gesture for /l/. The /r/ vs /l/ Tongue Root (TR) constrictions, then, differ in POA only. Thus this contrast is within-organ for both the Tongue Tip (TT) constriction (degree) and also for the TR constriction (location).

to be at odds with AOH expectations for within-organ contrasts, as noted earlier this is a possible scenario for within-organ contrasts (and is addressed further in the Conclusion). More important, consistent with AOH predictions, all of these native within-organ cases have shown MAINTENANCE of discrimination across development. English-learning infants display MAINTENANCE of discrimination for English /b/-/v/ (Polka et al., 2001), which are distinguished by lip constrictions that differ in location (bilabial for /b/; labiodental for /v/) and degree (closure for /b/; critical for /v/); for English postalveolar affricate versus fricative /tʃ/-/ʃ/ (Tsao, Liu, & Kuhl, 2006), which are distinguished by tongue tip closure for /tʃ/ but critical for /ʃ/; and for English /s/-/θ/ (Tyler et al., 2014), distinguished by critical tongue tip constrictions at alveolar versus dental locations.

Nonnative speech perception findings on within-organ contrasts are also consistent with PAM-AOH predictions. Two studies found a DECLINE, one for English-learning infants' Nthlakampx velar versus uvular ejective stops /k′/-/q′/ (Best et al., 1995; Werker & Tees, 1984), which are distinguished by velar versus uvular locations of tongue dorsum closure, the other for Japanese-learning infants' initially moderate discrimination at 6–8 months to poor discrimination at 10–12 months for English /r/-/l/ (Kuhl et al, 2006). Another found a FLAT trajectory of poor discrimination from 6 months through to adulthood for native French listeners tested on English /d/-/ð/ (Polka et al., 2001; Sundara et al., 2006). All other nonnative within-organ contrasts that have been examined have instead shown good initial discrimination, which is also possible according to PAM-AOH. However, unlike the developmentally maintained discrimination for initially good native within-organ contrasts, all of the nonnative cases showed a significant DECLINE in discrimination by 10–12 months, consistent with PAM-AOH predictions. Specifically, English-learning infants showed a DECLINE on Hindi dental versus retroflex voiceless stops /t̪/-/ʈ/ (Anderson et al., 2003; Werker & Lalonde, 1988; Werker & Tees, 1984), which are distinguished by tongue tip constriction locations; on Czech alveolar fricative versus alveolar fricated trill /z/-/ɹ̝/ (ř) (Trehub, 1976), which are distinguished by a critical (/z/) versus a tighter tongue tip constriction (/ɹ̝/); on Mandarin alveolo-palatal affricate versus fricative /tɕ/-/ɕ/ (Tsao et al., 2006), which are distinguished by a tighter tongue dorsum constriction for /tɕ/; and on Zulu voiceless aspirated versus ejective velar stops /k^{h}/-/k′/ (Best & McRoberts, 2003), which are distinguished by glottal abduction (opening) for /k^{h}/ versus closure for /k′/. Mandarin-learning infants have likewise shown decline from initially good discrimination of English postalveolar affricate versus fricative /tʃ/-/ʃ/ (Tsao et al., 2006).

The privative contrasts that have been tested show an array of developmental trajectories, which are more similar to within-organ than between-organ contrasts, as can be seen in Table 1 (light and medium gray shaded entries). Regarding native contrasts, English-learning infants show MAINTENANCE of good initial discrimination of English /b/-/m/ (Eimas & Miller, 1980), which is distinguished by a velum-lowering gesture only for /m/, and of English /b/-/p/ (Eilers, Gavin, & Wilson, 1979; Eimas, Siqueland, Jusczyk, & Vigorito, 1971), which is distinguished by a glottis-opening gesture (abduction of the vocal folds) only for /p/, phased with *release* of the bilabial constriction. Analogously, Spanish-learning infants show MAINTENANCE for Spanish prevoiced versus unaspirated stops /b/-/p/ (Eilers et al., 1979; Lasky, Syrdal-Lasky, & Klein, 1975; see also, for Kikuyu-learning infants, Streeter, 1976), where only /p/ has a glottis abduction gesture, and it is phased with bilabial *closure*. English-learning infants instead show INDUCTION of discrimination from 3 to 6 months for English /s/-/z/ (Eilers, 1977; Eilers & Minifie, 1975; Eilers et al., 1977; cf. modest decline at

10–12 months, then facilitation by adulthood; Best & McRoberts, 2003; Best et al., 2001), which is distinguished by a glottis-opening (abduction) gesture only for /s/.

As for nonnative privative contrasts, English-learning infants show a DECLINE from initially good discrimination of Zulu plosive versus implosive bilabial stops /b/-/ɓ/, which are distinguished by a glottal lowering gesture only for /ɓ/, and also of Zulu voiceless versus voiced lateral fricatives /ɬ/-/ɮ/ (Best & McRoberts, 2003), which are distinguished by a glottis abduction gesture only for the voiceless one; and a FLAT trajectory of poor discrimination of nonnative Spanish prevoiced versus unaspirated stops /b/-/p/ (Eilers et al., 1979; Lasky et al., 1975). Spanish-learning infants, however, show MAINTENANCE of good discrimination for English /b/-/p/ (Eilers et al., 1979; Lasky et al., 1975). Thus, privative contrasts pattern similarly to within-organ contrasts, with the exception of MAINTENANCE for nonnative Spanish infants' discrimination of English /b/-/p/, which is more consistent with a between-organ contrast.

Thus, it appears that the PAM-AOH framework (Best & McRoberts, 2003; L. M. Goldstein & Fowler, 2003) provides a good account of nearly all existing findings on both native and nonnative speech perception across infancy. However, even with the addition of privative organ contrasts that we provided here, two observed findings raise questions that are yet unaddressed by PAM-AOH or by the acoustic-exposure or PAM-only accounts: (a) MAINTENANCE of English-learning infants' initially good discrimination for the native English within-organ contrast /b/-/v/ versus delayed INDUCTION from initially poor discrimination for English /d/-/ð/ (Polka et al., 2001) and (b) MAINTENANCE of initially good discrimination by both native English-learning and nonnative Spanish-learning infants for privative English /b/-/p/ versus initially good discrimination by Spanish-learning infants but *poor* discrimination by English-learning infants for privative Spanish /b/-/p/. We return to these cases in the Conclusion, where we discuss how added articulatory considerations of PAM-AOH may accommodate these findings.

To do so, however, we must first review other sources of evidence that perception and production both rely on articulatory information in speech, which may contribute to those proposed modifications of PAM-AOH. The following subsection presents computer simulations of infant and adult perceptual learning of within-organ contrasts. Such simulations are needed to determine whether INDUCTION can occur in infants exposed to a bimodal distribution of articulatory values for within-organ contrasts. The simulations also demonstrate that the same computational approach, without any modifications, can yield differentiated results for two types of within-organ articulatory input distributions, consistent with PAM predictions that perceptual learning can be induced in adult L2 learners for nonnative L2 within-organ contrasts assimilated as a difference in goodness of fit to a given native consonant, that is, a Category Goodness difference (CG) assimilation, but not for L2 within-organ contrasts assimilated as equally poor exemplars of a given native consonant, that is, a Single Category (SC) assimilation (Best & Tyler, 2007).

Simulated attunement to within-organ constriction distributions

Statistical learning of frequency distributions of values along key dimensions of input speech is the mechanism usually proposed to account for developmental change in infants' discrimination of both native and nonnative consonant contrasts. Current statistical learning models that predict discrimination of a contrast will remain good (MAINTENANCE) or improve

(INDUCTION; FACILITATION) if there is a bimodal distribution in frequency of occurrence (two peaks) along a critical dimension in native speech, whereas discrimination will DECLINE or remain FLAT if the distribution of occurrence in the native language is either flat or has just a single peak (unimodal distribution). Only the INDUCTION hypothesis has been investigated in English-learning 6-, 8-, and 10-month-olds along an acoustic VOT continuum (prevoiced [–90 ms] to voiceless unaspirated [+90 ms]) that encompasses both native English stop voicing contrasts (0 vs. +90 ms) and nonnative stop voicing contrasts, for example, for Spanish (−90 vs. +20 ms). In articulatory terms, the distinctions along this continuum are privative (glottis adduction for /p/ only, whether phased relative to alveolar closure [Spanish] or release [English]). Different groups of infants at each age were familiarized to randomized presentations of either a unimodal or a bimodal frequency distribution of items in the continuum in which the bimodal distribution peaks were in the prevoiced versus the voiceless unaspirated ranges, that is, neither the native English nor nonnative Spanish contrast. They were then tested on discrimination of the end points. At the two younger ages, the bimodal familiarization group discriminated the end points, whereas the unimodal group failed to discriminate (Maye, Werker, & Gerken, 2002). Compatible results were obtained with 8-month-olds in another series of studies using Hindi voicing distinctions (like the Spanish prevoiced vs, voiceless unaspirated distinction) for dental stops and velar stops (Maye, Weiss, & Aslin, 2008). At 10 months, however, neither familiarization group discriminated the stop VOT contrast. They did succeeded on a second acoustic continuum representing the Hindi dental-retroflex stop place of articulation contrast but only if they received a double familiarization period, in which case the bimodal group alone discriminated the endpoints (Yoshida, Pons, Maye, & Werker, 2010).

We note that those studies all examined the learning of frequency distributions only for continua along acoustic dimensions. According to PAM-AOH, however, the relevant distributions are along articulatory dimensions. In addition, all but one of the aforementioned studies used a privative voicing distinction, with the exception of the Hindi place of articulation contrast (Yoshida et al., 2010), which is within-organ. The fact that a nonnative place of articulation contrast was learned at 10 months, but the voicing ones were not, suggests a possible difference in distribution-based learning of privative versus within-organ contrasts. Therefore, to evaluate the PAM-AOH hypothesis that infants can learn unimodal versus bimodal distributions for a within-organ contrast from input arrayed along a relevant articulatory dimension, we conducted computer simulations of English-learning infants' and English- versus Spanish-speaking adults' perceptual attunement to a tongue tip (TT) constriction location continuum.

For our simulations, a Hebbian learning model was employed in a manner similar to that used by Oudeyer (2006; see also Browman & Goldstein, 2000; Nam, Goldstein, & Saltzman, 2009; Wedel, 2004). In the model, the TT constriction was represented by a set of virtual ("neural") units, each of which represents some value of the TT constriction location. To create the two articulatory input distributions, we used existing articulometry data on TT constriction locations along the midsagittal plane between the upper front teeth and the posterior side of the alveolar ridge for productions of English /d/ (unimodal distribution; Figure 3A) and productions of Hindi dental and retroflex stops /d̪/-/ɖ/ (bimodal distribution; Figure 3B; L. M. Goldstein, Nam, Kulthreshtha, Root, & Best, 2008; see also L. M. Goldstein, 2003). The attunement of infants and adult L2 learners to these two articulatory distributions was modeled by comparing the learner's produced TT constriction location on

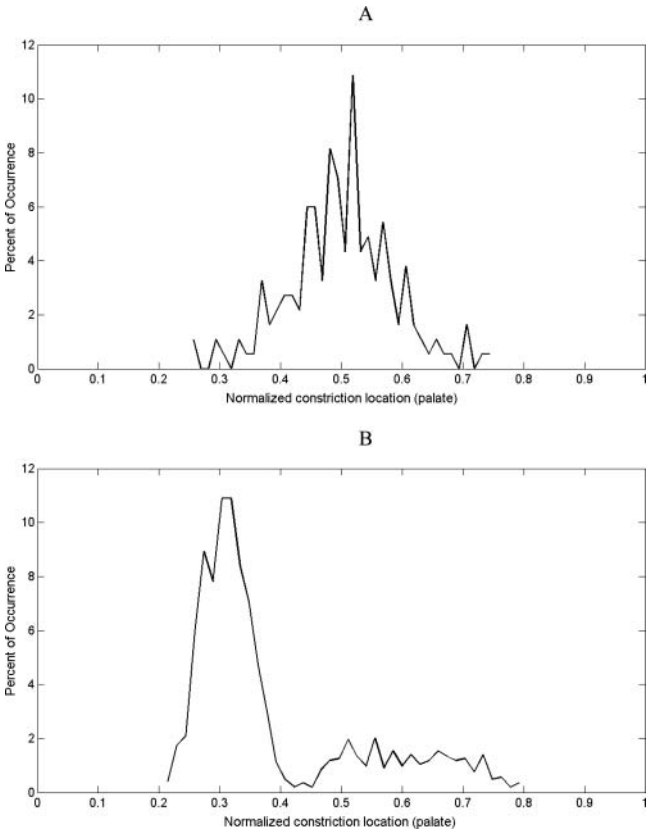

Figure 3. (A) Frequency distribution of tongue tip (TT) constriction location along hard palate (normalized to 0–1), as measured by electromagnetic articulometry (EMA), for all coronal stops in a natural spoken English passage of ~1,000 words produced by an adult female native speaker (L. M. Goldstein et al., 2008); (B) corresponding distribution for all coronal stops in a natural spoken Hindi passage of ~6,000 words produced by an adult female native speaker (L. M. Goldstein et al., 2008).

a given learning cycle (i.e., simulation iteration) with a randomly sampled TT constriction location from the external language environment. For both the learners and the language environment (parents or teachers), choosing a TT value is based on their own probability distribution of the neural units. If the TT values matched between the learners and the environment, within some quantization threshold, all the neural units responded by increasing their level of activation as a function of the proximity to the chosen value. Here the proximity is simply defined as a Gaussian distribution function. Then the probability of the learners emitting that value was increased by a small amount. For both infant and adult simulations, the model employed 10,000 neural units and ran 10,000 iterations. The standard deviation parameter of the Gaussian function was set to 0.05. The learning rate was set to 0.02 and 0.001 for the infant and the adult simulation, respectively.

The infant starting distribution was set as "blank slate," that is, a very low and flat a priori probability distribution along the TT-location dimension. On each iteration, one TT value was chosen at random from both infant and parent distributions, where the probability of choosing a given value was a function of its probability in the respective distribution. If the TT values matched within some quantization threshold, then the

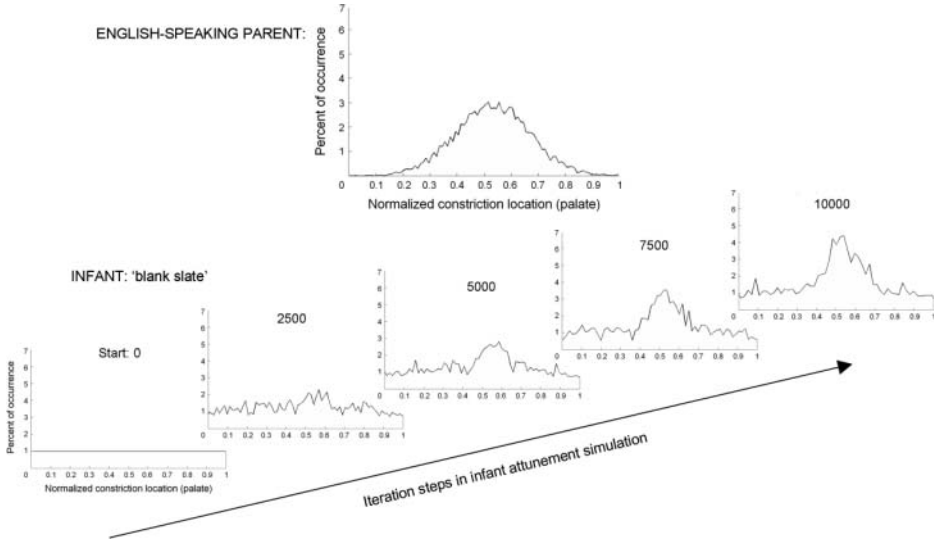

Figure 4. Simulation of "blank slate" infant (lower left) attunement to English "parent" input (top center) based on articulatory data from a native English speaker, which shows a unimodal frequency distribution of alveolar Tongue Tip (TT) constriction locations along the hard palate (normalized 0–1). The time series (lower portion of diagram) indicates successive 2,500-iteration steps in the 10,000-iteration simulation.

probability of the child emitting that value was increased by small amount. For comparison, we also conducted adult L2 learner simulations, with the starting TT location distributions for the learners set as unimodal for both the English-speaking and the Spanish-speaking learner, to reflect their lifetimes of experience with native language

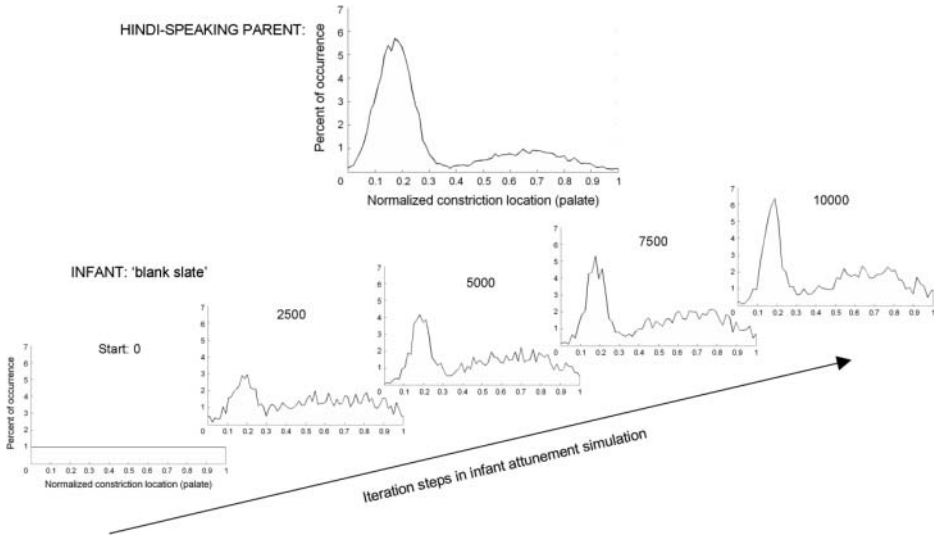

Figure 5. Simulation of "blank slate" infant (lower left) attunement to Hindi "parent" input (top center) based on articulatory data from a native Hindi speaker, which shows a bimodal frequency distribution of dental versus retroflex Tongue Tip (TT) constriction locations along the hard palate (normalized 0–1). The time series (lower portion of diagram) indicates successive 2,500-iteration steps in the 10,000-iteration simulation.

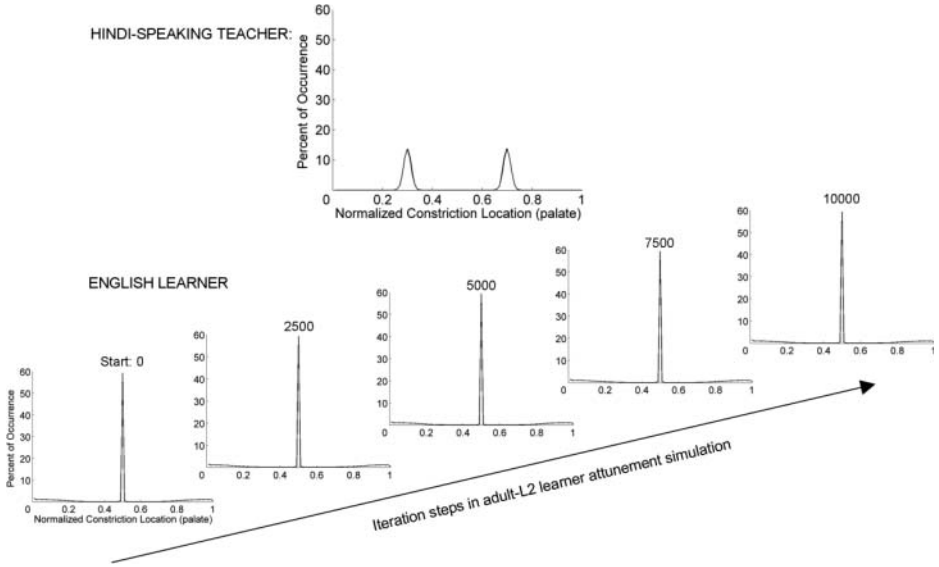

Figure 6. Simulation of an adult second language (L2) learner of Hindi showing native-language (L1) **S**ingle **C**ategory (SC) assimilation (lower left) and L2 attunement (time series) to input from an *idealized* Hindi "teacher" (top center). The L2 learner is an idealized speaker of English with a well-established unimodal distribution of English coronal stops centered at *alveolar* position, which does not line up with either Hindi mode. The time series shows the same simulation steps as in Figure 5.

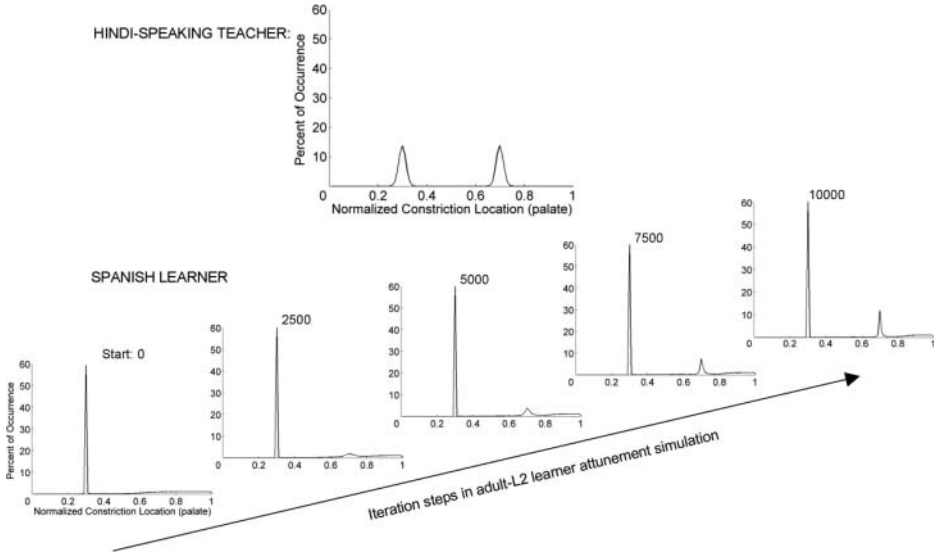

Figure 7. Simulation of a different adult L2 Hindi learner showing initial L1 **C**ategory **G**oodness difference (CG) assimilation (lower left) and L2 attunement (time series) to input from the same idealized Hindi "teacher" (top center) and simulation steps as in Figure 6. The second language (L2) learner is an idealized native language (L1) speaker of Spanish with a well-established unimodal distribution of Spanish coronal stops centered at dental position, which *does* line up with one of the two Hindi modes (dental).

coronal stops. However, the location of the peak differed between learner languages to reflect the alveolar constriction location of English /d/, that is, falling in between the two Hindi peaks, for which PAM would predict SC assimilation and poor L2 learning versus the dental location of the Spanish /d//, that is, falling in line with the Hindi dental peak, for which PAM would predict CG difference assimilation and improved L2 learning. The input distribution for their Hindi "teacher" was the same as for the Hindi "parent" distribution in the infant Hindi-learning simulation.

The simulated infant learners each gradually acquired the distribution of their "parent," that is, a unimodal distribution emerged for the infant with an English "parent" (Figure 4), whereas instead a bimodal distribution emerged for the infant with a Hindi "parent" (Figure 5). In contrast, the adult learners with the same Hindi "teacher" (bimodal input) were restricted by their initial, native-language starting distributions: the simulated English-speaking adult did not acquire the Hindi bimodal distribution (Figure 6), displaying a persistent SC assimilation of the Hindi consonant contrast, as predicted by PAM/PAM-L2 (Best, 1995; Best & Tyler, 2007). In contrast, the simulated Spanish-speaking adult did begin to acquire the Hindi contrast over iterations (Figure 7), displaying an initial CG difference assimilation of the Hindi contrast to a good versus poor match to the single Spanish dental /d/, from which a distinction between the Hindi dental and retroflex stops emerged over iterations (though they were not equal in probability as they were for the teacher), as predicted by PAM-L2 (Best & Tyler, 2007). The lack of equal probability of the modes could be interpreted as possible substitution errors of the Spanish dental stop for the Hindi retroflex stop in some words.

These simulations suggest that learning of articulatory distributions could account for early developmental changes and adult L2 learning of native versus nonnative within-organ consonant contrasts. However, it is also important to address whether infants may actually detect articulatory information in speech. To address this question, we turn next to evidence on infants' perception of speech presented across auditory, visual, and/or haptic (active tactile) modalities.

Perceiving articulatory information across speech modalities

Visible articulatory information: Seeing the talking face

As we argued earlier, infants' perception of interrelationships among speech modalities implies that they recognize the common articulatory source that gave rise to the multimodal signal. The modest literature on multimodal (simultaneous/synchronized presentation of speech in two or more modalities) and cross-modal (temporally separated presentations in the differing signal modalities) speech perception by infants has focused primarily on the dynamic visual information in talking faces as it relates to a synchronized acoustic speech signal, that is, audiovisual (AV) speech. Given our consonantal focus, we only review studies that used consonantal stimuli, omitting those that just examined vowels.

One approach has been to assess infants' detection of a simultaneous match between an audio speech stimulus and one of two video displays that are both synchronized to the audio target. Although this is the typical procedure used in AV perception studies with vowels, it has been employed in just two published consonant studies. In one, Japanese 8-month-olds were presented with audio targets and synchronized videos of a

woman producing a bilabial trill versus a whistle (Mugitani, Kobayashi, & Hiraki, 2008), both assumed by the authors to be nonspeech oral sounds. Each was a sustained bilabial constriction, producing the corresponding sound without a co-articulated vowel, thus not truly speech-like; they differed visibly in whether or not there was lip vibration. The infants looked reliably longer at the trill than the whistle video when they heard the audio trill, but those who heard the audio whistle showed no video preference. In this regard, it may be worth noting that trills, including bilabial trills, occur as consonants in some languages though not in Japanese, whereas whistles per se are not found in the consonant inventory of any known language.[4]

The other study instead presented more natural speech stimuli, consonant-vowel-consonant-vowel (CVCV) disyllables, for which the pairs of side-by-side video displays differed visibly in both consonants and vowels (MacKain, Studdert-Kennedy, Spieker, & Stern, 1983). The 5- to 6-month-old participants looked reliably longer at the video that matched the synchronized audio disyllable but only when the matching video appeared on the right side of the screen; no reliable-looking preference was found when the matching video appeared on the left. As the authors conclude, this right visual field bias implicates a left hemisphere advantage in recognizing the articulatory match between audible and visible speech, compatible with the literature on left hemisphere superiority in language processing.

Other studies have exploited the McGurk effect, by which adults perceive phonetically incongruent audio and visual consonants in synchronized vowel-identical syllables to be a unitary consonant differing from the audio signal in a way that is clearly "pulled" by the visible place of articulation; for example, audio [b] synchronized with video [g] is perceived as /d/, or audio [b] synched with video [v] as /v/ (McGurk & MacDonald, 1976). In one such study, 5-month-olds were habituated to AV-congruent [va], then tested on discrimination of an audio-only change to incongruent [ba] (which adults perceive as still being /va/) or [da] (which adults perceive as the contrasting consonant /ða/). They discriminated the change to audio [da] but not [ba], indicating the same McGurk effect as adults: they perceived audio [ba] with video [va] to be /v/, that is, same as the AV-congruent habituation /va/ (Rosenblum, Schmuckler, & Johnson, 1997). When 4.5-month-olds were instead habituated to AV-incongruent auditory [ba] with visual [ga], perceived as /d/ or /ð/ by adults, they discriminated a test-trial change to audio-only [ba] but not a change to audio-only [da] or [ða]. Conversely, the control group that was habituated to AV-congruent [ba] discriminated the changes to both audio [da] and [ða] but *not* to audio [ba], again indicating that infants show a McGurk effect like adults' (Burnham & Dodd, 2004; cf. possibly weaker effects in 4-month-olds; Desjardins & Werker, 2004). A modified mismatch-negativity (MMN) study with 5-month-olds confirmed (Kushnerenko, Teinonen, Volein, & Csibra, 2008) that they do indeed perceive AV-incongruent McGurk consonants as unitary consonants, as adults do, that is, they do not apparently detect the AV-mismatch in these cases.

Examining infants' ability to detect a *cross*-modal match between an audio-only speech stimulus and subsequently presented silent video-only articulation, however, allows us to

[4]Whistles using the mouth do not engage the vocal cords the way standard speech does, including trills and "whistled" sibilant fricatives (which are not in fact truly whistled; Lee-Kim, Kawahara, & Lee, 2014). Instead, for whistling the air inside the oral cavity is actively compressed to expel an oral airstream through narrowed rounded lips. Although so-called whistle languages such as Silbo Gomero are indeed whistled, they are not independent languages with their own phonology. Rather, they are a rarefied and restricted form of the community's spoken language in which a subset of phonemes are produced in reduced form using the whistling technique.

assess whether the same articulatory information is detected in *each* modality. Cross-modal matching of temporally separated audio and visual speech signals circumvents the possibility with synchronized AV presentations that perception may simply indicate capture by unimodal auditory or visual information rather than reflecting detection of the common articulatory information that is carried in each signal modality. Two such cross-modal studies have been conducted. One examined English- versus Spanish-learning infants on a within-organ contrast that occurs in English but not Spanish: /b/-/v/. 6- and 11-month-olds of each language environment were first familiarized with side-by-side synchronized silent videos of a woman producing /ba/ and /va/, then habituated to either audio-only /ba/ or /va/, followed by test trials with the silent video pair to assess for a visual preference for the habituated consonant. Both groups showed a visual preference for the habituated audio consonant at 6 months, but only the English infants continued to show this preference at 11 months, indicating MAINTENANCE of cross-modal A→V consonant matching for native infants but developmental DECLINE for nonnative infants (Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009). Using a similar procedure to test English-learning 4- versus 11-month-olds on native and nonnative *between*-organ contrasts (lip vs. tongue tip constrictions), we found MAINTENANCE of cross-modal congruency detection at both ages for native English /p/-/t/ as well as for nonnative Tigrinya ejective /pʼ/-/tʼ/ but not for !Xóõ bilabial versus dental clicks /ʘ/-/ǀ/ (Best, Kroos, & Irwin, 2010, 2011, 2014). Interestingly, although the 11-month-olds also showed a significant silent-video preference for each audio habituation stimulus of the Tigrinya contrast, their preference was for the incongruent video rather than for the congruent video as they showed for the English contrast. Additionally, the 4-month-olds showed this reverse incongruency preference for the !Xóõ clicks but not for the Tigrinya ejectives, whereas the 11-month-olds showed absolutely no video preference following the audio-only !Xóõ clicks. We interpreted these patterns as indicating, first, that both ages recognize an articulatory relationship between the audio and video stimuli for both native English stops and nonnative Tigrinya ejectives but that only the 4-month-olds do so for the clicks. Second, however, the reversed congruency preferences for the nonnative Tigrinya ejectives suggest that the 11-month-olds but not the 4-month-olds recognized an articulatory mismatch between the audio-only nonnative ejectives and the nonejective (English) oral stops produced in the silent videos.[5] By comparison, the 4-month-olds recognized an articulatory mismatch between the audio-only clicks (!Xóõ) and the nonclick articulations in the oral-stop English videos, whereas the 11-month-olds failed to detect *any* relationship between the audio clicks and the subsequent oral stop videos, hinting that they may have perceived the clicks as nonspeech sounds as adults do (Best et al., 1988).

Tangible articulatory information: The feeling of speech

Infants' language-learning environment includes frequent face-to-face vocal interactions with caregivers, whose faces are often in close enough proximity for infants to feel breaks in the airstream of their speech that correspond to alternations of aspirated and unaspirated phonemes as well as close enough to be within reach of the infants' hands, for example, during bottle/breast feeding, and whose facial motions and vocalizations infants often imitate (whether fully

[5] Differences in compression of the oral articulators, and possibly also in laryngeal motions in the throat, are visible between oral and ejective stop productions and influence adults' AV perception of the two types of stops (Fenwick, Davis, Best, & Tyler, 2015).

or partially, immediately or after delay). Thus, infants not only see and hear speech, they also have opportunities to *feel* speech articulations simultaneous with an utterance's audio and/or visual concomitants, whether arising from another person (tactile information from the talker's airstream and haptic [active touch] information from manual contact with the talker's face) or from the infants' own speech-like vocalizations (self-proprioceptive and kinesthetic information). The various guises of touch are clearly among the multiplicity of modalities contributing to infants' experience of native speech and must be taken into account in theoretical considerations about what nature of information infants perceive in speech.

Indeed, evidence indicates that simultaneous haptic (others' speech) or proprioceptive/kinesthetic (self) information affects the perception of audio speech. As we described earlier in this article, adults show McGurk percepts analogous to those found with AV speech when perceiving the consonant of audio-haptic (perceiver's fingers resting on an unseen talker's lips; Fowler & Dekle, 1991) or audio-tactile (air puff onto perceiver's skin; Derrick & Gick, 2013; Gick & Derrick, 2009; Gick, Ikegami, & Derrick, 2010) target items in which the haptic/tactile event is synchronized with a phonetically incongruous audio speech stimulus. These findings converge with the AV findings to support the premise that the commonality that perceivers detect among the disparate modalities of a given utterance is that all were shaped by a singular articulatory event.

Although such haptic/tactile McGurk effects have not yet been examined in infants, when proprioceptive/kinaesthetic information from their own vocal tracts is constrained, their perception of acoustically presented consonants is systematically affected. As reviewed earlier, English-learning 6-month-olds can discriminate Hindi dental versus retroflex stops /d̪/-/ɖ/, a within-organ tongue tip constriction location distinction. A recent study found that they also discriminate this contrast while sucking on a "gum-teether" pacifier that allows free tongue motion but fail when sucking on a flat pacifier that prevents tongue tip motion (Bruderer, Danielson, Kandhadai, & Werker, 2015; for pacifier/teether effects on infant vowel perception, see Yeung & Werker, 2013). This finding is compatible with evidence that adults' speech perception is also constrained or shifted by artificial or natural constraints on their speech articulator configurations/motions (Ito, Tiede, & Ostry, 2009; Sams, Möttönen, & Sihvonen, 2005; Sato, Troille, Ménard, Cathiard, & Gracco, 2013), effects that are reflected as well in that co-activation patterns of auditory and speech motor cortex during relevant tasks (d'Ausilio, Bufalari, Salmas, & Fadiga, 2009; Möttönen, & Watkins, 2009).

We take the findings summarized in this and the preceding few subsections as converging evidence that infants perceive amodal articulatory information in speech. The attested effects of self-produced speech vocalizations or speech organ motions, along with the perceptual findings summarized earlier, are consistent with the premise that articulatory information provides the common metric between perception and production that is needed for vocal imitation to occur and develop and ultimately for learning to produce native words recognizably.

Linking perception and production

Infant vocal imitation

Research in this area has focused more on vowels (Kuhl & Meltzoff 1982, 1996) and prosodic properties than on consonants. However, consonantal imitations would offer more insights about the articulatory basis of perception–production relations. Whereas vowels involve only relatively wide within-organ constrictions of the tongue dorsum and root that differ

primarily in location, some with coordinated privative lip rounding or velum lowering gestures, consonants use the full range of speech articulators and constriction degrees as well as more complex constellations of articulator constrictions. As a result, consonantal imitation would be expected to show a more protracted developmental trajectory than for vowels. Indeed, infants imitate vowels notably more often and earlier than consonants or consonant–vowel combinations (e.g., Kokkinaki & Kugiumutzakis, 2000; Moran, Krupka, Tutton, & Symons, 1987; Papoušek & Papoušek, 1989).

A small handful of studies have investigated infants' imitative responses to consonantal articulations presented by laboratory models and/or by caregivers in more naturalistic contexts. In one, a model presented newborns with live AV productions of the consonant /m/ versus the vowel /a/ (held for 4 s each). The babies responded with significantly more lip-closure/tightening ("mouth-clutching") gestures to /m/ than /a/ and conversely with more mouth-opening gestures to /a/ than /m/ (X. Chen, Striano, & Rakoczy, 2004). In a more naturalistic study, mother–infant pairs were recorded in the laboratory at 2, 3, and 5 months while interacting "as they normally do at home" (Papoušek & Papoušek, 1989). In that study, infants displayed some consonant imitation but notably more vowel imitation. Focusing on the consonant imitations (Papoušek & Papoušek, 1989, Figure 5, p. 147), proportionally more glottal and velar than labial or coronal consonants were imitated at 3 and 5 months, and no labials were imitated at 2 months. Fricative and trill imitation were also higher at 3–5 than at 2 months, when only posterior stops and nasals and a few *glottal* fricatives (/h/) were imitated. We interpret these patterns to indicate that (a) consonantal imitation does occur and fairly early, (b) posterior consonantal constrictions are imitated earlier and more often than anterior ones, and (c) constriction degrees (manners) requiring articulatory precision are not imitated before 3–5 months. Regarding the third point, though 2-month-olds did imitate a small proportion of glottal fricatives, we note that [h] is the most posterior of consonants and that it requires only simple vocal fold abduction rather than the precise degree of articulatory constriction that must be achieved for supraglottal fricatives.

Clearly, more research is needed on infant speech imitation and its developmental trajectory past the 1st half-year, particularly with respect to consonantal imitation. However, several findings provide complementary important insights about infants' spontaneous nonimitative consonant production in interactions with their caregivers. One study found that 6- to 9-month-olds produce more frequent and more varied supraglottal consonants (i.e., made with lips or tongue) during play interactions with their primary caregiver, if they vocalize while mouthing an object or their fingers/hand (contact with mouth, lips, or tongue) than if they are not mouthing an object. The authors concluded that active oral engagement with objects introduces and enhances variations in vocal tract closures (i.e., consonantal gestures) during social play vocalizations due to the increased multimodal information provided by mouthing and vocalizing simultaneously, which encourages exploration of consonant production (Fagan & Iverson, 2007). This interpretation may extend to understanding the bias toward posterior consonant imitation by younger infants. The posterior articulators develop and become active earlier than anterior ones in prenatal development, for example, in swallowing and hand sucking, and are highly used on a daily basis in suckling. They get earlier and more usage in early infancy.

Conversely, mothers respond contingently to their 8-month-olds' vocalizations, in particular responding vocally more often and with more varied response types to consonant–vowel than to vowel-only productions by the infant (Gros-Louis, West, Goldstein, & King,

2006). Moreover, maternal responses that are contingent on their 6- to 10-month-olds' vocalizations increases the proportion of mature syllabic forms in their infants' immediately following vocalizations (i.e., consonant–vowel combinations, tighter consonant–vowel timing) relative to noncontingent maternal responses (M. H. Goldstein, King, & West, 2003). The suggestion that this pattern reflects a higher level type of "imitation" of mature structural organization is supported by a study that found infants selectively increase consonant–vowel productions if mothers produce consonant–vowel syllables contingent on their infants' vocalizations but instead produce more vocalic vocalizations if their mothers produced only vocalic elements contingently (M. H. Goldstein & Schwade, 2008). It is important to note that in naturalistic interactive contexts with their mothers, 12-month-olds produce more consonant–vowel combinations than simple vowel sounds during book reading, and mothers produce more imitations or expansions of infants' consonant–vowel, but not vowel-only, utterances during book reading (Gros-Louis, West, & King, 2016). This research indicates that both self-initiated oral exploratory behavior and caregivers' contingent vocalizations in social interactions systematically increase the complexity and maturity of consonantal elements in infants' prelinguistic vocalizations.

Early production and recognition of native words

Toddlers' immediate and delayed imitations of words (from memory without an immediate adult target) offer further insights into the role articulatory gestures play in perception and production of lexical forms in the 2nd year during the early word-learning period (∼11–17 months). Young learners often produce the same word in variable ways and conversely may produce several different words in seemingly the same way, with few or none of their forms fully and correctly matching the sequence of consonants and vowels in the adult targets. Nonetheless, articulatory gestural analyses of such erroneous productions have revealed that they often contain (most of) the correct articulators and gestures, but they appear in the wrong order and/or with the wrong constriction parameter settings or intergestural coordinations. For example, in one study a 2-year-old girl (Studdert-Kennedy & Goodell, 1995) saw a picture of a hippopotamus and spontaneously named it ['ɑpɪnz]. This corresponds to the final three syllables of the target word produced as just two syllables, consistent with her general production limit to just two syllables of multisyllabic targets. Her utterance includes all component articulatory gestures of the consonants in the target's final two syllables, /-təməs/, but with sequencing and intercoordination errors: the tongue tip and glottis-abduction gestures of /t/ are split between her [n] and [p], the bilabial closure and velum-lowering gestures of /m/ between her [p] and [n], and the critical-narrow tongue tip and glottis-abduction gestures of /s/ between her [z] and [p] (i.e., she managed only one of the two glottis-abductions in the target bisyllable). When mom replied, "Oh, hippopotamus!" the child responded with four repetitions of the form she had produced earlier as immediate imitations of adult productions, ['hɪpɑs], which instead includes the correct gestures *and* intergestural coordinations for the target's first and final syllables. This example suggests closer articulatory matching for immediate imitation than delayed (spontaneous retrieval), as would be expected. However, even immediate imitations of new, phonologically complex words can include erroneous sequencing and intercoordination of (most) of the component articulatory gestures, sometimes spread across multiple imitative attempts. The latter was seen in the same child's heroic six attempts to repeat her mother's presentation of <apricot>. The imitative attempts were all trisyllabic forms, that is, the correct syllable

number, but included different subsets of the target articulatory gestures in varying incorrect sequences and combinations. In contrast to the articulatory gesture account, alternative hypotheses that children instead reproduce acoustic or phonetic features of target words do not account well at all for either these cases or other imitations and spontaneous productions of newly learned words by this child and other children (see reviews by Studdert-Kennedy, 2002; Studdert-Kennedy & Goodell, 1995). Conversely, additional findings compatible with the articulatory gestures analyses come from a study on the Articulatory Organ Hypothesis, which examined the articulatory components of the initial consonant of 13- to 19-month-olds' spontaneous productions of adult target words (i.e., delayed imitation) in a separate corpus, as rated by adult listeners. The children matched the target onset's oral tract articulator significantly more often than chance and more often than they matched the constriction degree (stop, fricative, etc.), voicing (glottic gesture), or nasalization (velum gesture), none of which fell above chance (L. M. Goldstein, 2003). Together, these findings support the premise that young children perceive and attempt to reproduce the articulatory gestures of target words rather than their acoustic properties per se. However, again more research in this vein is needed for deeper understanding the impact of contexts and constraints on articulatory variations in early word imitations.

Requisite to imitation is children's perception of the articulatory composition of adult targets and their recognition of how those articulatory patterns relate to their own production preferences/limitations, which in turn may bias the words they choose to imitate as well as how they structure their imitative attempts. A number of studies have found that reliable, persisting preferences in the consonants that individual infants use in their prelinguistic babbling during the final quarter of the 1st year (termed articulatory or vocal motor routines, or word templates), preferences that vary from infant to infant, are carried forward into each child's consonantal preferences in early word productions (Blake & de Boysson-Bardies, 1992; Keren-Portnoy et al., 2009; Velleman & Vihman, 2007; Vihman & Croft, 2007; Vihman et al., 1985). That these consonantal preferences in babbling also sharpen the children's perception of consonants in native speech is supported by two recent passage-listening preference studies with 9- to 12-month-olds. In both, infants' consonantal babbling preferences were systematically related to their listening preference between passages with nonsense words containing those same consonants and otherwise identical passages in which the nonsense words contained consonants not found in their preferred babbling routines. The authors interpreted this as evidence that prelinguistic babbling provides a lens that focuses the child's attention to specific articulatory patterns in heard speech (DePaolis, Vihman, & Keren-Portnoy, 2011; DePaolis, Vihman & Nakai, 2013).

These early imitation findings offer converging evidence that speech perception and production are deeply interdigitated in development from prelinguistic infancy through at least the early phase of lexical development in the 2nd year. Moreover, they implicate the child's reliance on the same articulatory gesture information on both sides of the equation.

Conclusions and future research directions

Our aims in this article have been twofold. First, we provided an integrative, critical review of research into the effects of language experience on developmental changes in infants' speech perception and speech-like production skills. The two domains certainly must be integrally related to one another in the service of developing spoken language, yet they have almost

exclusively been examined and considered separately. Second, supported by both logical analysis and reinterpretation of the patterns of findings observed in those literatures, we argued that the informational primitives for both perception and production of speech by infants are the amodal articulatory gestures that create the correlated multimodal dynamic properties of spoken utterances (across acoustic, visual, haptic, proprioceptive/kinesthetic, and aerotactile modalities). This premise offers a more parsimonious alternative to mainstream assumptions that acoustic properties are the absolute primitives for speech perception and that motoric representations are the primitives for speech production. As we argued, those assumptions leave the young infant with a need for additional specialized mechanisms (e.g., computational modules) to "translate" the two incommensurable types of information back and forth in order to accomplish interrelated development across the two domains. A common articulatory metric across perception and production has the advantage of providing a simpler, more straightforward foundation for their obvious interdependence and mutual influences in early language development. We put forth PAM-AOH (Best & McRoberts, 2003; L. M. Goldstein, 2003; which combines the principles of the Perceptual Assimilation Model (PAM) and the Articulatory Organ Hypothesis (AOH) as a parsimonious theoretical framework that accounts for nearly the full range of existing findings on native language attunement in infant speech perception and production and have updated it to include privative gesture contrasts.

The updated PAM-AOH offers a fruitful framework for further examination of the perception–production relationship across infancy, particularly on interdependencies in the coattunement of infant speech perception and production toward the spoken language environment. First, the framework is useful for designing perceptual studies that could tease out the basis for puzzles we identified in native/nonnative infant speech perception findings to date. For example, as noted earlier, both acoustic-only (AO) and AV /b/-/v/ is discriminated during the 1st half-year whether or not it occurs contrastively in the infants' language environment (MAINTENANCE) and is still discriminated at 11 months if /b/-/v/ is used by the native language but is no longer detected if the native language lacks this contrast (DECLINE). Conversely, AO findings suggest failure to discriminate /d/-/ð/ from early infancy through to as late as 4 years even if the native language uses it (INDUCTION) as well as /b/-/v/ contrastively (English); discrimination is lacking throughout the life span if the native language lacks both /d/-/ð/ (FLAT) and /b/-/v/ (DECLINE) as phonemic contrasts but presents both as position-dependent allophonic variations (e.g., as in Spanish). Both contrasts are within-organ (lips and tongue tip, respectively), involve critical differences in both location and degree of constriction (refer to Figure 2), and present similar acoustic and frequency-of-occurrence patterns. However, PAM-AOH and its amodal articulatory premises raise other relevant differences between the contrasts that may contribute to divergent perceptual development: (a) the labial (lips) distinction is more fully visible than the tongue tip distinction and would be more easily tracked in the multimodal situations involved in infant language learning and (b) young infants produce more and earlier functional, controlled lip and tongue dorsum-specific rather than tongue tip-specific motions both in speech-like and nonspeech oral actions (facial emotion expressions such as smiling, lip pursing, and mouth opening and tongue protrusion, which is accomplished by tongue dorsum and root actions that extend the tongue tip as a relatively "passive rider," and in suckling, which also involves coordinated lip and tongue dorsum and root motions rather than active tongue tip motions, e.g., Geddes, Kent, Mitoulas, & Hartmann, 2008; Hayashi, Hoashi, & Nara, 1997; Iwayama & Eishima, 1997).

The current PAM-AOH framework may also offer insights on other developmental findings, such as English-learning 6- and 11-month infants' discrimination of both within- and between-organ place of articulation contrasts, not only for English anterior fricatives (native) but also for Nuu Chah Nulth (NCN) posterior fricatives (Tyler et al., 2014). The proposed Articulatory Geometry (Figure 2) combined with our reasoning about detection of articulatory information in multimodal speech perception suggests why discrimination is maintained across the 1st year for both within- and between-organ contrasts, though for different reasons with the native English and the nonnative NCN contrasts. Detection of the two English contrasts is supported by experience with both audible and visible articulation differences (upper teeth contacted by lips [f] vs. tongue tip [θ] vs. partially visible tongue tip contact with the alveolar ridge behind the upper teeth [s]). Conversely, although the two NCN contrasts are not only nonnative to English but also invisible on the face, detection of these differences may be maintained through infancy due to infants' substantial proprioceptive/ kinesthetic experience with coordinated tongue dorsum, root, and pharynx motions involved in suckling and swallowing from the fetal period throughout infancy (Geddes et al., 2008; Hayashi et al., 1997; Iwayama & Eishima, 1997). Interestingly, by comparison, adult English listeners like 10- to 12-month-olds discriminate these NCN contrasts moderately well (~80%correct), which they perceptually assimilate to English /h/ but with notable differences in goodness ratings as /h/. Notably, however, the adults had the greatest difficulty with the NCN pharyngeal-glottal voiceless fricatives /ħ/-/h/, a between-organ distinction (pharyngeal vs. laryngeal/glottis) that was not tested in infants and which the adults assimilated as equally good /h/ (Kencalo, Best, Tyler, & Goldstein, 2007; Tyler, Best, Avesani, Bohn, & Vayra, 2016). This may reflect the fact that pharyngeal/ aryepiglottic gestures are not used contrastively in English, possibly worth investigating in further perceptual tests with infants and adults.

It would also be useful to test PAM-AOH assumptions for discrimination of click consonant contrasts by infants learning a nonclick language such as English, given that adults fail to assimilate these consonants to their native phonological system, instead hearing them as nonspeech sounds. Will the within- versus between- versus privative-organ hypotheses of AOH apply to developmental change in these nonnative but nonassimilable nonspeech contrasts? Would the trajectories differ for infants learning other click languages, that is, for whom the clicks would presumably be perceived as speech (see Best, in press)?

Further research is needed, especially regarding the developmental interrelationship between speech perception and production in infants. We believe the current PAM-AOH framework offers a rich range of possible issues to examine, including the relationship between lexical acquisition and phonological development as well as phonetic and phonological aspects of developmental language disorders (e.g., dysphasia, dyslexia, specific language impairment). In addition, we suggest that this approach is likely to have broader relevance to advancing understanding of speech perception and production in late L2 learners and hence help to optimize instruction to include effective L2 listening and accent training. It could also help provide insights into how fluent bi/multilinguals negotiate the phonetic and phonological systems of their two languages, especially in code-switching situations. In any case, we hope this critical review and beginning steps toward integrating the literatures on early speech perception and production will encourage additional investigation of questions that might otherwise not have arisen.

Funding

This work was supported by the Australian Research Council [Grant numbers DP0772441 and DP130104237] and the National Institute on Deafness and Other Communication Disorders [Grant number DC000403].

References

Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech*, *46*, 155–182.

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. 2, pp. 67–96). New York, NY: Academic Press.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*, 321–324.

Auban, J., & Ménard, L. (2006). Compensation for a labial perturbation: An acoustic and articulatory study of child and adult French speakers. *Proceedings of International Symposium on Speech Production (ISSP)*. Ubatuba, Brazil: CEFALA.

Aziz-Zadeh, L., Iacoboni, M., Zaidel, E., Wilson, S., & Mazziotta, J. (2004). Left hemisphere motor facilitation in response to manual action sounds. *European Journal of Neuroscience*, *19*, 2609–2612.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The Celex lexical database*. Philadelphia: University of Pennsylvania. Linguistic Data Consortium.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech contrasts: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 289–304). Dordrecht, The Netherlands: Kluwer Academic.

Best, C. T. (1994). Learning to perceive the sound pattern of English. In C. Rovee-Collier, & L. Lipsitt (Eds.), *Advances in infancy research* (Vol. 9, pp. 217–304). Norwood, NJ: Ablex.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 167–200). Timonium, MD: York Press.

Best, C. T. (in press). Perceiving non-native click consonant contrasts. In B. Sands (Ed.), *The handbook of click languages*. Leiden, The Netherlands: Brill.

Best, C. T., Kroos, C. H., & Irwin, J. (2010). Now I see what you said: Infant sensitivity to place congruency between audio-only and silent-video presentations of native and nonnative consonants. *Proceedings of AVSP (AudioVisual Speech Perception)*. Hakone, Japan: AVISA.

Best, C. T., Kroos, C. H., & Irwin, J. (2011). Do infants detect A→V articulator congruency for nonnative click consonants? *Proceedings of AVSP (AudioVisual Speech Perception)*. Volterra, Italy: AVISA.

Best, C. T., Kroos, C. H., & Irwin, J. (2014, July). *Baby steps in perceiving articulatory foundations of phonological contrasts: Infants detect audio→video congruency in native and nonnative consonants*. Paper presented at Laboratory Phonology, Tokyo, Japan.

Best, C. T., & McRoberts, G. W. (2003). Infant perception of nonnative consonant contrasts that adults assimilate in different ways. *Language & Speech*, *46*, 183–216.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology. *Journal of the Acoustical Society of America*, *1097*, 775–794.

Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two non-native consonant contrasts. *Infant Behavior and Development*, *18*, 339–350.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 45–60.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. Munro, & O.-S. Bohn (Eds.), *Second language speech learning* (pp. 13–34). Amsterdam, The Netherlands: John Benjamins.

Blake, J., & de Boysson-Bardies, B. (1992). Patterns in babbling: A cross-linguistic study. *Journal of Child Language*, *19*, 51–74.

Bosch, L., & Ramon-Casas, M. (2011). Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics*, *39*, 514–526.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English/r/and/l: IV: Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*, 2299–2310.

Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*, 201–251.

Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180.

Browman, C., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP: Bulletin de la Communication Parlée*, *5*, 25–34.

Brown, R. (1958). *Words and things*. Glencoe, IL: Free Press.

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*, *112*, 13531–13536.

Burnham, D., & Dodd, B. (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*, 204–220.

Chen, L. M., & Kent, R. (2010). Segmental production in Mandarin-learning infants. *Journal of Child Language*, *37*, 341–371.

Chen, X., Striano, T., & Rakoczy, H. (2004). Auditory–oral matching behavior in newborns. *Developmental Science*, *7*, 42–47.

Cruttenden, A. (1970). A phonetic study of babbling. *International Journal of Language & Communication Disorders*, *5*, 110–117.

Curtin, S., Byers-Heinlein, C., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, *39*, 492–504.

Cutler, A. (2008). The abstract representations in speech processing. *Quarterly Journal of Experimental Psychology*, *61*, 1601–1619.

Davis, B. L., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research*, *38*, 1199–211.

d'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, *19*, 381–385.

de Boysson-Bardies, B., Sagart, L., & Bacri, N. (1981). Phonetic analysis of late babbling: A case study of a French child. *Journal of Child Language*, *8*, 511–524.

de Boysson-Bardies, B., Sagart, L., & Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language*, *11*, 1–15.

de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, *67*, 297–319.

de Boysson-Bardies, B., Vihman, M. M., Roug-Hellichius, L., Durand, C., Landberg, I., & Arao, F. (1992). Material evidence of infant selection from the target language: A cross-linguistic phonetic study. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 369–391). Timonium, MD: York Press.

DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development*, *34*, 590–601.

DePaolis, R. A., Vihman, M. M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development*, *36*, 642–649.

Derrick, D., & Gick, B. (2013). Aerotactile integration from distal skin stimuli. *Multisensory Research*, *26*, 405–416.

Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, *45*, 187–203.
Eilers, R. E. (1977). Context-sensitive perception of naturally produced stop and fricative consonants by infants. *The Journal of the Acoustical Society of America*, *61*, 1321–1336.
Eilers, R. E., Gavin, W., & Wilson, W. R. (1979). Linguistic experience and phonemic perception in infancy: A cross-linguistic study. *Child Development*, *50*, 14–18.
Eilers, R. E., & Minifie, F. D. (1975). Fricative discrimination in early infancy. *Journal of Speech, Language and Hearing Research*, *18*, 158–167.
Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech, Language, and Hearing Research*, *20*, 766–780.
Eimas, P. D., & Miller, J. L. (1980). Discrimination of information for manner of articulation. *Infant Behavior and Development*, *3*, 367–375.
Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303–306.
Engstrand, O., Williams, K., & Lacerda, F. (2003). Does babbling sound native? Listener responses to vocalizations produced by Swedish and American 12- and 18-month-olds. *Phonetica*, *60*, 17–44.
Esling, J. H. (1996). Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetic Association*, *26*, 65–88.
Fadiga, L., Craighero, L., & Olivier, E. (2005). Human motor cortex excitability during the perception of others' action. *Current Opinion in Neurobiology*, *15*, 213–218.
Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, *73*, 2608–2611.
Fagan, M. K., & Iverson, J. M. (2007). The influence of mouthing on infant vocalization. *Infancy*, *11*, 191–202.
Fenwick, S. E., Davis, C., Best, C. T. & Tyler, M. (2015). The effect of modality and speaking style on the discrimination of non-native phonological and phonetic contrasts in noise. *FAAVSP-2015*, 67–72.
Flege, J. E. (1984). The effect of linguistic experience on Arabs' perception of the English /s/ vs. /z/ contrast. *Folia Linguistica*, *18*, 117–138.
Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*, 1–34.
Fowler, C. A. (2004). Speech as a supramodal or amodal phenomenon. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 189–201). Cambridge, MA: MIT Press.
Fowler, C. A., & Dekle, D. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 816–828.
Fowler, C. A., & Galantucci, B. (2008). The relation of speech perception and speech production. In D. B. Pisoni, & R. Remez (Eds.), *The handbook of speech perception* (pp. 632–652). New York, NY: Wiley.
Fowler, C. A., Rubin, P., Remez, R., & Turvey, M. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production, 1: Speech and talk* (pp. 373–420). London, UK: Academic Press.
Friederici, A. D., & Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, *54*, 287–295.
Geddes, D. T., Kent, J. C., Mitoulas, L. R., & Hartmann, P. E. (2008). Tongue movement and intra-oral vacuum in breastfeeding infants. *Early Human Development*, *84*, 471–477.
Ghosh, P. K., Goldstein, L. M., & Narayanan, S. S. (2011). Processing speech signals using auditory-like filterbank provides least uncertainty about articulatory gestures. *The Journal of the Acoustical Society of America*, *129*, 4014–4022.
Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, *462*, 502–504.
Gick, B., Ikegami, Y., & Derrick, D. (2010). The temporal window of audio-tactile integration in speech perception. *The Journal of the Acoustical Society of America*, *128*, EL342–EL346.
Goldstein, L. M. (2003). Emergence of discrete gestures. *Proceedings of the International Congress of Phonetic Sciences*, *15*, 85–88.

Goldstein, L. M., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 215–249). Cambridge, UK: Cambridge University Press.

Goldstein, L. M., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller, & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159–207). Berlin, Germany: de Gruyter.

Goldstein, L. M., Nam, H., Kulthreshtha, M., Root, L., & Best, C. (2008, June). *Distribution of tongue tip articulations in Hindi versus English and the acquisition of stop place categories*. Paper presented at Laboratory Phonology 11, Wellington, New Zealand.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *100*, 8030–8035.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*, 515–523.

Gottlieb, G. (1976). The roles of experience in the development of behavior and the nervous system. In G. Gottlieb (Ed.), *Development of neural and behavioral specificity* (pp. 1–35). New York, NY: Academic Press.

Gottlieb, G. (1981). The roles of experience in species-specific perceptual development. In R. N. Aslin, J. R. Alberts, & M. R. Peterson (Eds.), *Development of perception: Psychobiological perspectives: Vol. 1. Audition, somatic perception and the chemical senses* (pp. 5–44). New York, NY: Academic Press.

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, *30*, 509–516.

Gros-Louis, J., West, M. J., & King, A. P. (2016). The influence of interactive context on prelinguistic vocalizations and maternal responses. *Language Learning & Development*, *12*, 280–294.

Guion, S., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, *107*, 2711–2724.

Hayashi, Y., Hoashi, E., & Nara, T. (1997). Ultrasonographic analysis of sucking behavior of newborn infants: The driving force of sucking pressure. *Early Human Development*, *49*(1), 33–38.

Holmberg, T. L., Morgan, K. A., & Kuhl, P. K. (1977). Speech perception in early infancy: Discrimination of fricative consonants. *The Journal of the Acoustical Society of America*, *62*, S99.

Irwin, O. C. (1947). Development of speech during infancy: Curve of phonemic frequencies. *Journal of Experimental Psychology*, *37*, 187–193.

Irwin, O. C. (1948). Infant speech: Development of vowel sounds. *Journal of Speech & Hearing Disorders*, *13*, 31–34.

Iwayama, K., & Eishima, M. (1997). Neonatal sucking behaviour and its development until 14 months. *Early Human Development*, *47*(1), 1–9.

Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Science*, *106*, 1245–1248.

Jakobson, R. (1968). *Child language, aphasia, and phonological universals*. The Hague, The Netherlands: Mouton. (English translation of R. Jakobson, 1941, *Kindersprache, aphasie und allgemeine lautgesetze*. Uppsala, Sweden: Almqvist & Wiksell).

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*, 402–420.

Jusczyk, P. W., & Luce, P. A. (1994). Infants′ sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*, 630–645.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 812–832.

Kencalo, T. A., Best, C. T., Tyler, M. D., & Goldstein, L. M. (2007). Perception of native and non-native fricative contrasts by Ukrainian-Australian English bilinguals and Australian English monolinguals. *Australian Journal of Psychology*, *59*, 40.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America*, *72*, 353–365.

Keren-Portnoy, T., Majorano, M., & Vihman, M. M. (2009). From phonetics to phonology: The emergence of first words in Italian. *Journal of Child Language*, *36*, 235–267.

Kern, S., Davis, B. L., & Zink, I. (2009). From babbling to first words in four languages: Common trends, cross language and individual differences. In F. d'Errico & J.-M. Hombert (Eds.), *Becoming eloquent: Advances in the emergence of language, human cognition and modern culture* (pp. 205–232). Amsterdam, The Netherlands: John Benjamins.

Kohler, E., Keysers, C. M., Umilta, A., Fogassi, L., Galles, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, *297*, 846–848.

Kokkinaki, T., & Kugiumutzakis, G. (2000). Basic aspects of vocal imitation in infant-parent interaction during the first 6 months. *Journal of Reproductive and Infant Psychology*, *18*, 173–187.

Konopczinski, G. (1990). *Le langage émergent: Caractéristiques rythmiques* [Emerging language: Rhythmic characteristics]. Hamburg, Germany: Buske Verlag.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259–274). Dordrecht, The Netherlands: Kluwer Academic.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews: Neuroscience*, *5*, 831–843.

Kuhl, P. K., Conboy, B. T, Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B*, *363*, 979–1000.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period." *Language Learning and Development*, *1*, 237–264.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138–1141.

Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development*, *7*, 361–381.

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*, 2425–2438.

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*, F13–F21.

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences*, *105*, 11442–11445.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, *20*, 215–225.

Lee, S. A. S., Davis, B., & MacNeilage, P. (2010). Universal production patterns and ambient language influences in babbling: A cross-linguistic study of Korean- and English-learning infants. *Journal of Child Language*, *37*, 293–318.

Lee-Kim, S. I., Kawahara, S., & Lee, S. J. (2014). The 'whistled' fricative in Xitsonga: Its articulation and acoustics. *Phonetica*, *71*, 50–81.

Levitt, A. G., & Aydelott Utman, J. G. (1992). From babbling towards the sound systems of English and French: A longitudinal case study. *Journal of Child Language*, *19*, 19–49.

Levitt, A., Jusczyk, P. W., Murray, J., & Carden, G. (1988). Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 361.

Levitt, A. G., & Wang, Q. (1991). Evidence for language-specific rhythmic influences in the reduplicative babbling of French- and English-learning infants. *Language and Speech*, *34*, 235–249.
Liberman, A. M. (1996). *Speech: A special code*. Cambridge MA: MIT Press.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–461.
Liberman, A. M., & Mattingly, I. (1985). The motor theory revised. *Cognition*, *21*, 1–36.
Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, *13*, 110–114.
MacKain, K. (1982). On assessing the role of experience on infants' speech discrimination. *Journal of Child Language*, *9*, 527–542.
MacKain, K., Best, C. T., & Strange, W. (1981). Categorical perception of /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*, 369–390.
MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. *Science*, *219*, 1347–1349.
MacNeilage,, P. F., Davis, B. L., Kinney, A., & Matyear, C. L. (2000). The motor core of speech: A comparison of serial organization patterns in infants and languages. *Child Development*, *71*, 153–163.
Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, *19*, 1994–1997.
Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 576.
Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Development*, *55*, 1777–1788.
Matyear, C. L., MacNeilage, P. F., & Davis, B. L. (1998). Nasalization of vowels in nasal environments in babbling: evidence for frame dominance. *Phonetica*, *55*, 1–17.
Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*, 122–134.
Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.
Mehler, J., Peña, M., Nespor, M., & Bonatti, L. (2006). The "soul" of language does not use statistics: Reflections on vowels and consonants. *Cortex*, *42*, 846–854.
Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*, 75–78.
Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development and Parenting*, *6*, 179.
Ménard, L., Davis, B. L., Boë, L. J., & Roy, J. P. (2009). Producing American English vowels during vocal tract growth: A perceptual categorization study of synthesized vowels. *Journal of Speech, Language, and Hearing Research*, *52*, 1268–1285.
Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*, 1006–1010.
Moffitt, A. R. (1971). Consonant cue perception by twenty-to twenty-four-week-old infants. *Child Development*, 717–731.
Moisik, S. R., & Esling, J. H. (2011). *The 'whole larynx' approach to laryngeal features*. Proceedings of ICPhS. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Moisik/Moisik.pdf
Moran, G., Krupka, A., Tutton, A., & Symons, D. (1987). Patterns of maternal and infant imitation during play. *Infant Behavior and Development*, *10*, 477–491.
Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology*, *14*, 477–492.
Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience*, *29*, 9819–9825.
Mugitani, R., Kobayashi, T., & Hiraki, K. (2008). Audiovisual matching of lips and non-canonical sounds in 8-month-old infants. *Infant Behavior and Development*, *31*, 307–310.

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupé (Eds.), *Approaches to phonological complexity* (pp. 299–328). Berlin, Germany: Mouton de Gruyter.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. 1, pp. 93–112). New York, NY: Academic Press.

Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Erlbaum.

Oller, D. K., & Eilers, R. E. (1982). Similarity of babbling in Spanish- and English-learning babies. *Journal of Child Language*, *9*, 565–77.

Opie, I. (1951). Humpty-Dumpty. In I. Opie, & P. Opie (Eds.), *The Oxford English dictionary of nursery rhymes* (pp. 213–215). Oxford, UK: Clarendon Press.

Oudeyer, P. Y. (2006). *Self-organization in the evolution of speech*. Oxford, UK: Oxford University Press.

Papoušek, M., & Papoušek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language*, *9*, 137–158.

Pegg, J. E., & Werker, J. F. (1997). Adult and infant perception of two English phones. *The Journal of the Acoustical Society of America*, *102*, 3742–3753.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*, 115–154.

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/-/ð/ perception: Evidence for a new developmental pattern. *The Journal of the Acoustical Society of America*, *109*, 2190–2201.

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*, *106*, 10598–10602.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*, 188–194.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.

Rogalsky, C., Love, T., Driscoll, D., Anderson, S. W., & Hickok, G. (2011). Are mirror neurons the basis of speech perception? Evidence from five cases with damage to the purported human mirror system. *Neurocase*, *17*, 178–187.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, *59*, 347–357.

Roug, L., Landberg, I., & Lundberg, L. J. (1989). Phonetic development in early infancy: A study of four Swedish children during the first eighteen months of life. *Journal of Child Language*, *16*, 19–40.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Brain Research: Cognitive Brain Research*, *23*, 429–435.

Sato, M., Troille, E., Ménard, L., Cathiard, M. A., & Gracco, V. (2013). Silent articulation modulates auditory and audiovisual speech perception. *Experimental Brain Research*, *227*, 275–288.

Saxe, J. G. (1873). The blind men and the elephant. In J. G. Saxe, *The poems of John Godfrey Saxe* (pp. 135–136). Boston MA: James R. Osgood.

Shafer, V. L., Yan, H. Y., & Datta, H. (2011). The development of English vowel perception in monolingual and bilingual infants: Neurophysiological correlates. *Journal of Phonetics*, *39*, 527–545.

Stark, R. (1980). Prespeech segmental feature development. In G. Yeni-Komshian, J. Kavanaugh, & C. Ferguson (Eds.), *Child phonology* (Vol. 1, pp. 73–92). New York, NY: Academic Press.

Stockman, I. J., Woods, D. R., & Tishman, A. (1981). Listener agreement on phonetic segments in early infant vocalizations. *Journal of Psycholinguistic Research*, *10*, 593–617.

Strafella, A. P., & Paus, T. (2000). Modulation of cortical excitability during action observation: A transcranial magnetic stimulation study. *Neuroreport*, *11*, 2289–2292.

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

Studdert-Kennedy, M. (1998). The particulate origins of language generativity: From syllable to gesture. In J. R. Hurford, J. R. M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: Social and cognitive bases* (pp. 202–221). Cambridge, UK: Cambridge University Press.

Studdert-Kennedy, M. (2002). Mirror neurons, vocal imitation and the evolution of particulate speech. In M. I. Stamenov, & V. Gallese (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 207–227). Amsterdam, The Netherlands: John Benjamins.

Studdert-Kennedy, M., & Goldstein, L. (2003). Launching language: The gestural origin of discrete infinity. In M. Christiansen, & S. Kirby (Eds.), *Language evolution* (pp. 235–254). Oxford, UK: Oxford University Press.

Studdert-Kennedy, M., & Goodell, E. W. (1995). Gestures, features and segments in early child speech. In B. de Gelder, & J. Morais (Eds.), *Speech and reading: A comparative approach* (pp. 65–88). East Sussex, UK: Erlbaum/Taylor & Francis.

Sundara, M., Namasivayam, A. K., & Chen, R. (2001). Observation-execution matching system for speech: A magnetic stimulation study. *Neuroreport*, *12*, 1341–1344.

Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-/ in monolingual and bilingual acquisition of English. *Cognition*, *100*, 369–388.

Sundara, M., & Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics*, *39*, 505–513.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*, 86–132.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *38*, 579[COMP]: Tees & Werker 1984 not in text. Add to text or delete from references.

Thevenin, D. M., Eilers, R. E., Oller, D. K., & Lavoie, L. (1985). Where's the drift in babbling drift? A cross-linguistic study. *Applied Psycholinguistics*, *6*, 3–15.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, *47*, 466–472.

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America*, *120*, 2285–2294.

Tyler, M. D., Best, C. T., Avesani, C., Bohn, O.-S., & Vayra, M. (2016). *Perception of non-native guttural fricative place contrasts as perceived by native listeners of languages that differ in use of gutturals: English, Danish and Italian*. Manuscript in preparation.

Tyler, M. D., Best, C. T., Goldstein, L. M., & Antoniou, M. (2014). Contributions of native-language tuning and articulatory organs to infants' discrimination of native and nonnative consonant contrasts. *Developmental Psychobiology*, *56*, 210–227.

Velleman, S. L., & Vihman, M. M. (2007). Phonology in infancy and early childhood: Implications for theories of language learning. In M. Pennington (Ed.), *Phonology in context* (pp. 25–50). London, UK: Palgrave Macmillan.

Vihman, M. M. (2002). Getting started without a system: From phonetics to phonology in bilingual development. *International Journal of Bilingualism*, *6*, 239–254.

Vihman, M., & Croft, W. (2007). Phonological development: Toward a "radical" templatic phonology. *Linguistics*, *45*, 683–725.

Vihman, M., DePaolis, R. A., & Keren-Portnoy, T. (2009). A dynamic systems approach to babbling and words. In E. L. Bavin (Ed.), *The Cambridge handbook of child language* (pp. 163–182). Cambridge, UK: Cambridge University Press.

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language*, *61*, 397–445.

Vitevitch, M. S., Armbrüster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 514.

Watkins, K. E., & Paus, T. (2004). Modulation of motor excitability during speech perception: The role of Broca's area. *Journal of Cognitive Neuroscience*, *16*, 978–987.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*, 989–994.
Wedel, A. (2004). Category competition drives contrast maintenance within an exemplar-based production/perception loop. *Proceedings of the Association for Computational Linguistics: Current themes in computational phonology and morphology* (pp. 1–10). Stroudsburg, PA: Association for Computational Linguistics.
Weikum, W. M., Oberlander, T. F., Hensch, T. K., & Werker, J. F. (2012). Prenatal exposure to antidepressants and depressed maternal mood alter trajectory of infant speech perception. *Proceedings of the National Academy of Sciences*, *109*(Suppl. 2), 17221–17227.
Werker, J. F. (1989). On becoming a native listener. *American Scientist*, *77*, 54–59.
Werker, J. F. (1991). The ontogeny of speech perception. In I. G. Mattingly, & M. Studdert-Kennedy (Eds.), *Modularity and the Motor Theory of speech perception* (pp. 91–110). Hillsdale, NJ: Erlbaum.
Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*, 197–234.
Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, *24*, 672–683.
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*, 49–63.
Werker, J. F., & Tees, R. C. (1992). The organization and reorganization of human speech perception. *Annual Review of Neuroscience*, *15*, 377–402.
Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, *50*, 509–535.
Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, *46*(3), 233–251.
Whalen, D. H., Levitt, A. G., & Goldstein, L. (2007). VOT in the babbling of French- and English-learning infants. *Journal of Phonetics*, *35*, 341–352.
Whalen, D. H., Levitt, A. G., & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language*, *18*, 501–516.
Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*, *24*, 603–612.
Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, *15*, 420–433.