12th International Seminar on Speech Production
Providence, RI, USA, June 1-4, 2020

Setting the stage for speech production: Infants prefer listening to speech sounds with infant vocal properties

Linda Polka[1], Matthew Masapollo[2], & Lucie Ménard[3]
McGill University[1], Boston University[2], University of Quebec at Montreal[3]

During the development of speech production, infants must learn to control articulatory movements by effectively monitoring time-varying patterns of auditory and somatosensory sensation (Guenther, 2016). Yet, surprisingly little is known about the mechanisms by which infants establish linkages between self-generated vocal tract movements and their kinematic and acoustic consequences. According to the Native Language Magnet Theory (Kuhl *et al*., 2008), exposure to ambient language establishes auditory patterns in memory that are specific to that language, and these representations subsequently guide infants' successive motor approximations of speech until a match is achieved via a vocal imitation mechanism. This account is consistent with studies showing that by 10 months, infants exposed to different languages produce babbling with some language-specific vowel characteristics when measured acoustically (de Boysson-Bardies, Hallé, Sagart, & Durand, 1989; Rvachew *et al*., 2006). In addition, infants have been reported to rapidly modify their vocalizations in response to audio-visual recordings of vowels produced (non-interactively) by an adult on a television (Kuhl & Meltzoff, 1996). Such findings suggest that infants have some ability to process their own vocal output and can link auditory and optical patterns with sensory-motor patterns that they are attempting to emulate.

But what factors facilitate such rapid speech motor learning during infancy? One possibility is that infants are perceptually biased toward speech elements that align with their own emerging vocal production patterns (cf. Vihman, 2014). Evidence supporting this hypothesis is provided by recent research demonstrating that infants (at 4-6 months) preferentially attend to infant vowel sounds over adult vowel sounds (Masapollo, Polka & Ménard, 2016). However, this "infant vowel" bias could be given an alternative explanation; namely, infants' preference for listening to infant vowels could just as well derive from their higher voice pitch (or fundamental frequency, F_0). A preference for the higher F_0 of infant vowels would not be surprising given that infants prefer listening to infant-directed speech with this same vocal property (Fernald & Kuhl, 1987). Masapollo *et al*. (2016) point out, however, that if infants prefer both the high F_0 and vocal resonances of infant vowels, then that would provide support for the former account, since both of those vocal properties jointly form infant speech signals.

Indeed, Masapollo *et al*. (2016) found that infant vocal resonances were sufficient to elicit a vowel preference. Specifically, 4- to 6-month-olds preferred to listen to /i/ vowels synthesized with infant formant values over /i/ vowels synthesized with adult formant values when both vowel types had F_0 values typical of adult female speech (210 or 240 Hz). However, a non-significant trend (favoring the infant formants) was observed in another experiment in which both vowel types had higher F_0 values (315 or 360 Hz) that fall within the range observed in infant-directed speech as produced by adult female talkers. This latter result suggests that infants' listening preferences may also be influenced by F_0.

In the present research, we synthesized a set of vowel stimuli, using the Variable Linear Articulatory Model (Ménard, Schwartz, & Boe, 2004), to further investigate infants' preference for infant over adult vocal resonances. Across three experiments, infants were tested in a sequential preferential listening procedure. In this procedure, infants were seated on their caregiver's lap and shown a static checkerboard pattern to look at while listening to the vowel stimuli. Infants were

presented with the infant vowels and adult vowels on alternating trials. Trial length was infant-controlled, i.e., trials were initiated by infant fixation on the checkerboard pattern and ended when infants looked away for more than 2 seconds. In each experiment, we collected 12 trials, 6 for each vowel type, and then calculated whether infants choose to listen longer to one type of vowel over another. The type of vowel that infants heard first was counterbalanced in each experiment.

In Experiment 1, we assessed preference for adult versus infant formants when the F_0 of both vowel types were the same (315 or 360 Hz). We tested 5- to 7-month-old infants (mean age = 7:11); this group was slightly older than the 4- to 6-month-olds (mean age = 5:19) tested in Masapollo *et al*. (2016), who failed to show a significant listening preference when presented these same vowels. In this older sample, we found a robust preference for vowels with infant resonances [$t(23)=3.143$, $p=.005$, $r^2=.542$], raising the possibility that infants' preference for infant vocal resonances may increase with age during the developmental period when canonical babbling emerges.

The next two experiments were designed to further assess infant listening preferences in the older age range (5-7 months) when F_0 values are modulated. If infants are more attentive to vowels with infant vocal resonance properties when F_0 values are varied, this would support Masapollo *et al*.'s hypothesis by showing that infants are listening selectively to vowels that resemble the output of their own vocal apparatus. In Experiment 2, infants' looking times were measured while they heard /i/ vowels with infant formants and relatively high F_0 (400 or 450 Hz) versus vowels with adult female formants and relatively low F_0 (315 or 360 Hz). In Experiment 3, looking times from another group of infants were measured while they listened to /i/ vowels with infant formants and relatively low F_0 (316 or 360 Hz) versus vowels with adult female formants and relatively high F_0 (400 or 450 Hz). An analysis of variance – vocal resonance type (infant formants vs. adult formants) × experiment (2 vs. 3), performed on the mean looking times – revealed a significant main effect of resonance type [$F(1,34) = 4.756$, $p=.036$, $\eta^2_p=.123$], such that infants listened longer to the vowels with infant formants, compared to the vowels with adult formants. There was no significant main effect of experiment [$F(1,34)=.023$, $p=.879$], or two-way interaction [$F(1,34)=.906$, $p=.348$], indicating that infants' listening preferences do not derive solely from higher F_0.

Although the present findings do not show how infants perceive and monitor their own self-generated speech-like phonatory and articulatory movements, the perceptual bias identified by Masapollo and colleagues may support them in doing so. We speculate that an attraction to infant vocal properties could help infants focus attention on their own early productions and motivate them to engage the speech articulators, which in turn may forge connections between auditory patterns and articulation. Future research exploring the mechanisms underlying this bias will advance our understanding of the development of sensorimotor integration for speech.

Selected References: Guenther, F.H. (2016). Neural Control of Speech, MIT Press, Cambridge, MA. / Kuhl, P.K., & Meltzoff, A.N. (1996). Infant vocalizations in response to speech: vocal imitation and developmental change. *Journal of the Acoustical Society of America*, 100, 425–438. / Masapollo, M., Polka, L., & Ménard, L. (2016). When infants talk, infants listen: pre-babbling infants prefer listening to speech with infant vocal properties. *Developmental Science*, DOI: 10.1111/desc.12298. / Ménard, L., Schwartz, J-L. & Boe, L.-J. (2004). The role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for French synthesized vowels from birth to adulthood. *Journal of Speech, Language, and Hearing Research*, 47, 1059-1080.