

Infants' use of asynchronous auditory and visual information in the categorization of speech sounds

Speech perception is a multisensory endeavour, and increasing evidence indicates that it is so from early in life. Infants match heard and seen speech,¹⁻⁴ and the addition of mismatching visual information affects auditory perception for infants, as it does for adults.⁵⁻⁶ Recent work suggests that infants' detection of the correspondence between heard and seen speech is present even without specific listening experience, as infants match auditory and visual speech sounds in unfamiliar language,⁷ and change their looking patterns in response to matched versus mismatched audiovisual speech in non-native languages.⁸ What has remained relatively unexplored, however, is whether and how infants with little linguistic experience weigh information from the auditory and visual domains differently, and whether this pattern is similar to that observed in experienced adults.

To explore this question, we conduct two studies, one with English-learning 6-month-old infants, and one with adult speakers of Hindi. For both groups, we use a Hindi speech distinction (voiced retroflex stop vs. voiced dental stop). This phonemic distinction, non-native for the infants, was chosen to enable us to determine how infants process heard and seen audiovisual speech syllables when experiencing them for the first time, and whether this pattern is similar to that exhibited by adults. All stimuli consist of auditory information from one phone and visual information from the other, such that the speech segments are incongruent in content. Audiovisual stimuli were also manipulated temporally, resulting in three types of segment: one in which auditory information precedes visual information, one in which visual information precedes auditory information, and one in which the two modalities are presented synchronously. Adults are presented with the stimuli in a forced-choice task, requiring them to categorize the sounds as either dental tokens or retroflex tokens. Infants, in separate groups corresponding to the three synchronicity conditions, are familiarized to sequences of the speech segments and then tested with auditory-only tokens of the two sound types. We hypothesize that 6-month-old English-learning infants, prior to perceptual narrowing, will categorize the speech sounds using a modality weighting similar to that of the Hindi-speaking adults.

Preliminary results from the infants' auditory-first condition reveal an increase in looking time to the category presented auditorily during familiarization ($M_{audio} = 13900.6$ ms; $M_{video} = 13019.6$ ms). This result indicates that infants may have categorized the speech segments during familiarization using the information from the auditory modality, and then exhibited a familiarity preference during test. However, it is as yet unclear whether or not a preference for categorization based on auditory information is in fact due to greater weighting of the auditory over the visual information, or if infants are simply categorizing speech based on whichever source of information occurs first. The completion of the three conditions of this study, along with the data gleaned from the adults, will more clearly tease apart these two possibilities and thus deepen our understanding of the nature of audiovisual speech perception from the earliest stages of life.

Works Cited

1. Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, 218(4577), 1138-1141.
2. Kuhl, P.K. & Meltzoff, A.N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development*, 7(3), 361-381.
3. Patterson, M.L., & Werker, J.F. (1999). Matching phonetic information in face and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, 22(2), 237-47.
4. Patterson, M.L., & Werker, J.F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, 6(2), 191-6.
5. Burnham, D. & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, 45(4), 204-220.
6. Desjardins, R. & Werker, J.F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, 45(4), 187-203.
7. Pons, F., Lewkowicz, D.J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 10598-10602.
8. Danielson, D.K., Greuel, A.J., Kandhadai, P., & Werker, J.F. (2015). Experience-independent changes in infants' face-scanning patterns while watching incongruent non-native audiovisual speech. 2015 Society for Research in Child Development Biennial Meeting. Philadelphia, PA. 19-21 March.