

## **The development of audiovisual speech perception is subject to cultural background: Evidence from Cantonese-speaking children**

Yi WENG, Yicheng RONG, and Gang PENG  
The Hong Kong Polytechnic University

While speech perception has been intensively studied from the unimodal auditory perspective, it has been proven to be a multisensory process that entails the cooperation and interaction between both auditory modality and other modalities, in particular visual modality. The term audiovisual speech perception emphasizes combining auditory and visual signals to perceive and understand speech. Throughout the development of audiovisual speech perception, children appear to experience a developmental shift in sensory dominance: from relying more on isolated auditory information to taking audiovisual bimodal cues into consideration. Previous studies have suggested that the developmental shift is universal, while the timepoint of its occurrence is subject to language and cultural backgrounds (Weng et al., under review). For instance, the developmental shift has been found to occur at the age of ten in English-speaking children while at the age of five in Mandarin-speaking ones (Hirst et al., 2018, Weng et al., under review). To examine this hypothesis, the current study recruited Hong Kong Cantonese-speaking children and adults to track down their developmental trajectory of audiovisual speech perception. In addition, the current study also investigated the role of auditory noise, as there was little knowledge about how it impacted the strategy adopted by children in audiovisual speech perception along development.

A total of 38 typically developing Hong Kong Cantonese-speaking children aged 4–9 years (23 females), together with an adult control group ( $n = 20$ , 10 females) were recruited in Hong Kong. The classic McGurk paradigm was adopted, with auditory noise level manipulated, giving rise to clean, 10dB SNR, -10dB SNR. A Cantonese-speaking young female was invited to record the articulation of three monosyllabic stimuli with a high-level tone: “Ba” (/ba/), “Da” (/da/) and “Ga” (/ga/). There were three congruent stimuli whose auditory components were matched with the visual ones (i.e., “Ba”, “Da” and “Ga”) and one incongruent stimulus with conflicting audiovisual information (auditory “Ba” dubbed on visual “Ga”, “AbVg”). Each block, consisting of congruent and incongruent stimuli varying in auditory noise levels, was repeated seven times. Participants were required to make responses to the stimuli by pointing at one of the three options: “Ba”, “Da” and “Ga”, and the experimenter would document their responses by pressing on a response recorder.

Based on the results from the identification of congruent stimuli, children aged 6–9 were found to show comparable performance relative to adults only when the auditory condition was free of noise, but to achieve a significantly lower accuracy in both noisy conditions. For the perception of incongruent stimuli, a developmental shift was uncovered in which child participants exhibited significantly more auditory-dominant (“Ba”) responses while fewer audiovisual-integrated (“Da”) responses compared to the adult group under clean and 10dB SNR conditions.

For the ability to identify congruent stimuli, it appeared adult-like in Cantonese-speaking children at around 6 years of age when the information from both modalities was free of noise, while it seemed to be immature even in children aged 9 years

under auditorily noisy circumstances, which was in accordance with the finding that children were more susceptible to auditory noise relative to adults (Johnson, 2000). For the perception of incongruent stimuli, since children aged 8–9 years still demonstrated a stronger auditory reliance compared to adults, it could be inferred that the occurrence of the developmental shift in audiovisual speech perception considerably lagged behind their Mandarin-speaking counterparts who exhibited the adult-like audiovisual-integrated strategy at around five (Weng et al., under review). The later shift might be attributed to the complex phonology in Cantonese (Bauer & Benedict, 1997, Peng et al., 2006) and to the strict regulations on mask-wearing implemented during the COVID-19 pandemic in Hong Kong (Former Food and Health Bureau of the Government of Hong Kong SAR, 2020; The Government of Hong Kong SAR, 2023).

**Keywords:** audiovisual speech integration, development, Cantonese-speaking children