Audiovisual Integration and Speech Perception in Bilingual Speakers

Rachel Tessmer, Nicole Tsao & Rachel Reetzke

Faculty Advisors: Bharath Chandrasekaran and Li Sheng, Department of Communication Sciences and Disorders

Background

- There is debate over when and how the auditory and visual systems come to work together in development.
- The principle of inverse effectiveness posits that as the quality of independent auditory and visual signals decrease, the strength of multisensory integration increases.
- At the intermediate signal-to-noise ratio (SNR) of -12 dB, there is evidence for a window of maximal multisensory integration beyond the predictions of this principle.
- Bilinguals may have greater difficulty than monolinguals perceiving speech in noisy environments, despite performing similarly in quiet conditions.

Aims & Predictions

- The aim of this project is two-fold: 1) to understand how the bilingual experience modulates audiovisual processing, and 2) to examine the multimodality of speech and language processing in noisy listening environments. Ultimately, we seek to determine whether a more diverse linguistic input across multiple modalities in bilingual speakers generalizes to a greater utilization of visual cues.
- The current study investigates the effect of bilingualism on a speech perception in noise (SPIN) task across age groups. We predict that bilingual speakers, both children and adults, will rely more on visual cues as listening environments become increasingly difficult. This is because bilingual speakers have a more diverse linguistic input and therefore are expected to use a more multistrategy in speech perception.

Methods

Participants

- We recruited 30 children (14 monolingual and 16 bilingual children, ages 6-10) and 37 adults (20 monolingual and 17 bilingual adults, ages 18-27) to participate in this study.
- All bilingual speakers were screened to ensure L2 onset was prior to age 3 and that L2 proficiency and daily use met the study’s requirements.

Stimuli

- Stimuli consisted of 80 sentences with 4 target words each. All sentences were produced by the same native English male speaker.
- Each sentence was masked by one of two types of noise:
  - 1) Informational masking: Two-talker babble (2T)
  - 2) Energetic masking: Pink noise (P)
- Each sentence was presented at one of five SNR levels:
  - 0 dB, -4 dB, -8 dB, -12 dB, & -16 dB

Prior to testing, all participants completed a hearing screening. Next, they were asked to focus on a fixation cross centered on the screen. 80 sentences total were presented as either audio only (AO), where no video of the speaker was played, or audiovisual (AV), where a video of the speaker accompanied the acoustic stimulus. Within these conditions, half of the sentences were masked with Pink noise and the other with Two-talker babble. Sentences were presented at one of five SNR levels. Each of these components was randomized. Participants were instructed to listen for the target sentence, which was presented 500 milliseconds after masker onset. Adult participants were asked to type what they heard and child participants were asked to say what they heard while experimenters typed their responses. After completing the SPIN task, participants or their guardians filled out an in-depth language history questionnaire and completed a nonverbal intelligence test (KBIT-2).

Results

- Bilingual and monolingual children exhibited similar performance across all conditions.
- After analyzing variance via two-way ANOVA, we found a significant age effect for all conditions, with adult performance exceeding child performance.
- Additionally, we found a significant language group effect in the 2T AV condition, with monolingual adults performing significantly better than bilingual adults.
- Adult bilinguals showed a steady increase in visual enhancement in Pink noise conditions as SNR levels decreased.
- Monolingual adults showed peak utilization of visual cues at -12 dB in Pink noise conditions.
- We found an effect of age on visual enhancement for Pink noise conditions, with adults utilizing visual cues significantly more than children.

Discussion

Against predictions, monolingual and bilingual children did not differ significantly on their performance in the SPIN task.

Monolingual adults exhibited an exaggerated peak for visual enhancement at -12 SNR, replicating Ross et al.’s findings of a window of maximal multisensory integration beyond the predictions of the principle of inverse effectiveness. All four groups showed increased visual enhancement at easier SNRs, while adult groups continued to show increased visual enhancement at harder SNRs, particularly in the Pink noise condition. This suggests that adults exhibit greater utilization of visual cues in more difficult listening environments, while children do not. Bilingual children perform more similarly to their monolingual counterparts than bilingual adults. These results may indicate that the divergent use of visual cues in speech perception between bilingual and monolingual speakers occurs later in development.