Native and Nonnative Audio-visual Perception of English Fricatives in Quiet and Cafe-noise Backgrounds

INTRODUCTION

Background
- Native (L1) speech perception is enhanced when visual information is available, especially when a listening environment is poor [1,2,3].
- Nonnative (L2) perceivers may be impeded in correct use of L2 visual cues non-existent in their L1 [4,5].

The current study
- Examined audio-visual (AV) perception of English inter-dental fricatives in quiet and café-noise backgrounds.
- Tested both L1 and L2 perceivers: English natives, and Chinese natives whose L1 does not contain inter-dentals.

METHOD

Stimuli
- /f, θ, s, v, ʃ, z/ combined with vowels /i, a, u/.
- CV syllables in (1) in a quiet background, and (2) embedded in café-noise (S/N=0dB).
- A-only, V-only, AV-congruent, AV-incongruent.

Procedure
- Identification task, 2 repetitions, ISI=4s.
- Stimuli were blocked by background (quiet, café-noise) and modality (A, V, AVc/AVi).
- English Q, Chinese Q, English N, Chinese N.

RESULTS

Fig. 2. Mean % correct responses for A, V, and AV input in quiet and noise by English and Chinese participants.

Fig. 3. Mean % correct responses for incongruent AV stimuli in quiet and noise by English and Chinese participants.

For both groups:
• better performance in A and AV than V conditions.
• better performance in quiet than noise conditions.

Compared to English Ss, Chinese participants showed:
• a lower % correct with the L2 inter-dental AV congruent stimuli.
• a higher percentage of inter-dental responses with the incongruent stimuli.
• a greater degree of reliance on visual information.

CONCLUSIONS

• Nonnatives may attend to visual information to perceive new L2 sounds, but failed to adopt the correct visual cues containing linguistically contrastive information, which suggests language specific processing.
• Similarities between the two groups suggest possible perceptual universals (e.g., A better than V).
• Together, results point to an integrated cross-modal network in speech processing.