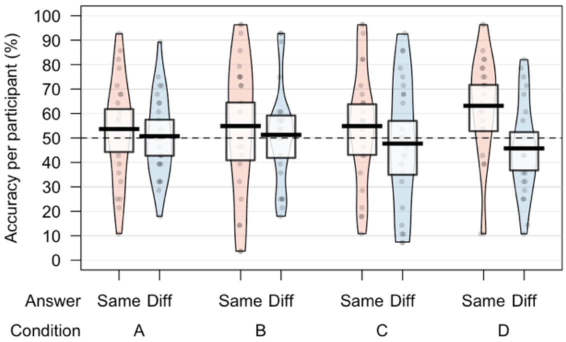
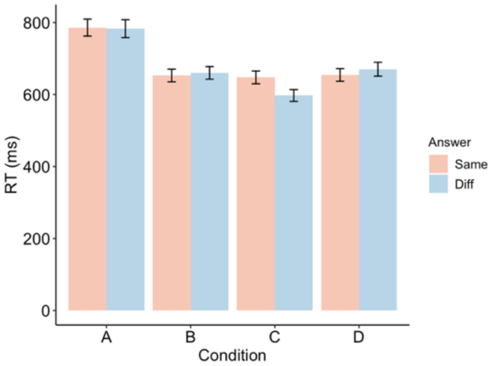
WHETHER FACE-VOICE (A)SYNCHRONY AFFECTS MULTIMODAL IDENTITY RECOGNITION BY NON-NATIVE SPEAKERS OF ENGLISH

Wenxi Fei & Yu-Yin Hsu (The Hong Kong Polytechnic University)
wen-xi.fei@connect.polyu.hk, yyhsu@polyu.edu.hk

This study aimed to investigate the impacts of synchronization of face-voice information on the recognition of speakers' identity, and the perceptual patterns of non-native speakers when matching faces and voices of unfamiliar speakers. Previous research has been shown that native English speakers can match faces and voice of unfamiliar speakers at about 70% accuracy [1][2]. However, the designs of these studies did not reflect real-life scenarios in which speakers' face and voice always overlap and can involve individuals with different language backgrounds, such as participants in a video conference trying to communicate under varying levels of internet connection. To investigate the effects of time synchronization, we implemented four conditions: A. static face and voice, B. time-matched face and voice, C. voice-first and D. face-first asynchronization. To ensure comparability with previous studies [1][2], we adopted the same material design with all the contents in English. Participants were asked to determine whether the face and voice belonged to the "same" or "different" person. We then analyzed accuracy and reaction times by (a)synchronization and answer types.

Our data included 25 participants (M_{age} = 24.2, SD_{age} = 2.83, N_{female} = 16) who were Mandarin native speakers learning English as a second language. The results showed that non-native speakers achieved an overall accuracy rate of 52.74%, indicating that the task posed considerable difficulty for them. The effect of conditions on accuracy was found to be slightly higher in the face-first condition (D) than in other conditions, but the differences were not significant. It suggested that facial information being dynamic or not may not be crucial, but rather, the earlier presentation of facial information holds greater dominance in non-native listeners' identity judgements. Moreover, face-voice pairs that were identified as the "same" also reached significant higher accuracy rates than the "different" ones (χ^2[1] = 16.79, p < .001). A significant interactive effect of condition and answer (χ^2[3] = 10.28, p = .0016) further indicated that participants tended to perform with higher accuracy when responding the trials of "same" answers. Regarding reaction times, non-native speakers spent significantly more time to judge static faces (A) compared to other conditions (all ps < .03), indicating the challenges associated with static facial information for the purpose of identification. More data from both non-native and native speakers will be needed for further investigation.





[1] Smith, et al. (2016). Concordant cues in faces and voices: Testing the backup signal hypothesis. [2] Lavan, et al. (2021). Explaining face-voice matching decisions: The contribution of mouth movements, stimulus effects and response biases.