Effect of Native Language Experience on Audio-visual Perception of English Fricatives by Korean and Mandarin Natives

Yue Wang [1], *Dawn Behne* [2], *Haisheng Jiang* [1], *Angela Feehan* [1]

[1] Department of Linguistics, Simon Fraser University, Canada
[2] Department of Psychology, Norwegian University of Science and Technology, Norway
yue_wang@sfu.ca

Abstract

This study examines how native language (L1) experience affects audio-visual (AV) perception of nonnative (L2) speech. Mandarin, Korean, and English perceivers were presented English CV syllables containing fricatives with 3 places of articulation: labiodentals nonexistent in Korean, interdentals nonexistent in Korean and Mandarin, and alveolars occurring in all L1s. The stimuli were presented as: audio-only (A), visual-only (V), congruent AV, and incongruent AV. Results show that all groups performed better in the AV than A or V condition. The English perceivers outperformed the nonnatives. Mandarin perceivers relied more on V input, and greater AV integration with incongruent AV, whereas Koreans benefited more from A. These findings indicate that nonnatives had less effective AV integration than natives with sounds involving visually unfamiliar places of articulation. The nonnatives' AV processing was differentially influenced by L1 experience. Conversely, similarities across groups indicate possible perceptual universals. Together they point to an integrated network in speech processing across modalities and linguistic backgrounds.

Index Terms: audio-visual, speech perception, nonnative.

1. Introduction

1.1. Background

Language experience often involves face-to-face interaction with simultaneous perception of a speaker´s voice and facial movements. Previous research has shown that for native (L1) perceivers, speech perception is enhanced when visual information is available [1,2]. This enhancement is especially effective when auditory distinctiveness decreases [2,3]. The relative contribution of audio and visual information has also been revealed by what is known as the "McGurk effect" [4], where an audio /ba/ dubbed onto a visual /ga/ may produce a /da/ percept, suggesting the tendency to integrate cross-modal inputs. These results suggest that humans have the ability to integrate audio-visual speech information via seeking optimally efficient combination of input channels [5,6].

Extension of this claim may also apply to the perception of nonnative (L2) speech, for which the auditory input is "non-optimal" because of the novel nature of L2 stimuli. However, although nonnatives may be facilitated by visual information from language-specific articulations would not be able to efficiently use L2 visual information as they are not familiar with the visual cues specific to the L2. Subsequent questions arise as to whether and how audio-visual speech perception in an L2 is affected by previous linguistic experience.

1.2. Current study

This study explores the extent to which nonnative speakers make use of visual information in L2 speech perception and the extent to which audio-visual speech perception is affected by L1 background.

Fricatives of three places of articulation (labiodental, interdental, and alveolar) were tested with English, Mandarin, and Korean perceivers. These native groups are included for their differences in L1 phonetic inventories, where English contains fricatives and corresponding visual cues that are unfamiliar to Mandarin and Korean natives (e.g., [17-19]). In particular, while all groups have alveolar fricatives in their L1, neither Mandarin nor Korean contains the interdental fricatives. The labiodental place, on the other hand, occurs in the Mandarin, but not Korean, fricative inventory.

Moreover, previous research has tested nonnative perception of A, V, and AV congruent conditions [12, 13], or mismatched A and V components [20]. The current study extends this research by including all the congruent and incongruent AV modalities (A, V, AVc, and AVi). To tease apart the influence of A and V information in a single stimulus, this study makes use of the McGurk illusion [4], by using responses to an incongruent AV stimulus to determine the contribution of each component modality. Of particular interest was how the Mandarin and Korean groups would respond in an incongruent AV condition, and to what extent they remain anchored to the A and/or V components, versus perceive the non-native AV-fused intermediate English percept?

These differences raise the questions addressed in this study: how do Mandarin and Korean natives differ from the native English group, and from each other, in the audio-visual perception of the unfamiliar fricatives, given their different L1 experience with these fricatives? Most essentially, how do these differences reflect the nature of audio-visual speech processing?

2. Method

2.1.