Visual Speech Modifies the Phoneme Restoration Effect

Erin Cvejic, Jeesun Kim & Chris Davis

MARCS Auditory Laboratories, University of Western Sydney, Bankstown, Australia
e.cvejic@uws.edu.au, j.kim@uws.edu.au, chris.davis@uws.edu.au

Abstract

The current study examined how seeing the talker (having visual speech information) affected phoneme restoration. For this purpose, six phonemes (/b/, /p/, /d/, /t/, /g/, /k/) in word contexts were used as target sounds in a 2IFC task in which participants were required to choose the white noise filled interval in which a target phoneme was present. The results showed that choosing the interval in which the phoneme was present was more difficult (i.e., more errors) in the auditory-only presentation condition. The results were discussed in terms of two potential effects of visual speech, i.e., eliciting an illusion of speech or as a distracter to task performance.

Index Terms: phoneme restoration, visual speech

1. Introduction

Phoneme restoration is an auditory phenomenon that occurs when a phoneme contained within a spoken word is replaced with extraneous noise, and the listener hears the word as being intact (the phoneme is restored within the speech signal) [1]. In this paper, we examine what effect the presence of visual speech may have on the phoneme restoration effect. Previous attempts to examine this phenomenon have been inconclusive [2]. We therefore hypothesize that the presence of visual speech will either induce an ‘illusion’ of speech, making it more difficult to know if a real phoneme was presented (as seeing a talker can create the perception of speech sounds in noise, even though they are absent [3]), or have a ‘signal boosting’ effect, making it easier to know if a real phoneme was presented (since providing visual speech has been shown to improve both speech intelligibility and detectability in noise, suggesting that seeing the talker may effectively boost the speech signal [4]).

2. Method

The stimuli items used were the words reprobate, interrogate, syncopate, suffocate, annorate and accommodate with the phoneme preceding the ‘ate’ in each word designated the critical phoneme. These items allowed for a comparison of aperiodicity (voiced/voiceless) and articulation type (bilabial/velar/alveolar).

Participants were presented with two speech intervals. One interval contained the phoneme signal and static white noise at the temporal location of the critical phoneme (‘added’), whilst the other interval contained only static white noise at the temporal location of the critical phoneme (‘replaced’). The task of the participant is to choose which interval was the ‘added’ interval.

Stimuli were presented in noise at an estimated 75% correct threshold level in auditory only (AO) and auditory-visual (AV) conditions in a 2IFC task, with items blocked by presentation condition and within-block presentation order randomisation. Each block was presented ten times (20 blocks total), with two counter-balanced repetitions per item per block (12 items per block), resulting in 240 items per session. Each participant completed four sessions in total.

3. Results

Percentage of errors for each presentation condition were compared in a 2x2x3 (presentation x aperiodicity x articulation) ANOVA, revealing main effects of Presentation Condition, Aperiodicity and Articulation type all achieving significant $F$ values. When visual speech was provided more errors were made than in the auditory only condition, suggesting visual speech information made the task more difficult.

4. Discussion

The finding of more errors with the presentation of visual speech supports the ‘illusion’ hypothesis, with speech sounds perceived despite their absence in the signal. However, the visually distinct phonemes (bilabials) did not produce more errors than the visually non-distinct ones (velars). Perhaps visual speech was distracting participants from the auditory decision aspect of the task, resulting in an increased amount of errors for audiovisual presentations. A follow-up experiment is currently underway using non-matched auditory-visual speech to explore this ‘distraction’ hypothesis.

5. References