The Role of the F2 transition in the Perception of Plosives

Mike Xiang,1 Robert M Cvengros,1 and Jont B Allen1

Dept. of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, 60056, USA

(Dated: 23 July 2019)

A study is done on perceptual features for consonant sounds by surgically removing select acoustic features across time-frequency slices and scoring the consonant reported. Subject responses suggest that the second formant is not relevant towards natural CV sound recognition and that it is the burst that carries the primary perceptual cue.
I. INTRODUCTION

For the entirety of human speech perception research, the nature and composition of the different acoustic features has remained in contention. Among the first acoustic features to be identified and studied were the consonantal burst and the temporal trajectory of the second formant, but the usefulness of the features in recognition remains uncertain. In this study, an experiment has been done to examine the perceptual role of the second formant transition and the consonantal burst, as used to identifying plosive consonants. Historical studies conclusions are all over the place in this regard, with various works contributing towards our understanding of speech perception. In the following section, a brief history of speech perception research is presented, with emphasis on the role of the second formant (F2) trajectory.

1900-1920: After the invention of the telephone, it became possible to perform speech perception experiments. The first research was by ?, who tested the recognition of simple sounds using a commercial public address system, having a microphone and loudspeaker. These first experimental results were soon followed by ?, where the first confusion matrix analysis was carried out. Conclusions were sparse, and the results are mainly of historical interest.

1920-1950: Harvey Fletcher and his team of researchers at the Bell Telephone Laboratory followed Campbell’s preliminary research with extensive perceptual measures under many different conditions of filtering and noise. The main goal was to quantify the !!blank!! These results ended in the development the articulation index, a famous method of modeling speech error (???).

1950-1970: In the late 1940s, Scientists at Haskins Laboratories created a machine called the Pattern Playback which could convert spectrographic pictures into sound. Their machine would work both with photograph copies of actual spectrograms, or with synthetic patterns painted on cellulose acetate. By preparing these hand-painted test patterns, using the recorded spectrograph as a template and reference, the Pattern Playback machine
allowed “systematic study of interactions between the individual sounds of speech (?)” for the first time, leading to the pioneering speech-science research of Alvin Lieberman, Frank Cooper, Pierre Delattre, John M.Borst, and Louis Gerstman (????). After their initial experiments with the Pattern Playback, their focus shifted away from examining the effects of progressively simplifying spectrograms of recorded speech to the study of isolated consonant-vowel elements and their role on speech perception.

The first such isolated acoustic element they analyzed was the burst, which “constitutes the acoustic counterpart of the articulatory explosion” (?). Listeners were presented with hand-created stimulus in which bursts of noise at each one of twelve different frequency positions were combined with each of seven different vowels. Bursts were constant in size and shape and the vowels were composed of two formants. In their own words, the patterns were “very highly simplified” but they were able to find some relationships and concluded that the frequency of the bursts could serve as a cue, but not a completely adequate one, for distinguishing among the plosives /p,t,k/. Bursts at high frequencies were heard as /t/ for all vowels, but at lower frequencies, were heard as /k/ when at or slightly above the second formant of the vowel; otherwise they were heard as /p/.

As a second possible cue, the Haskins group identified the transition between consonant and vowel, seen in the spectrogram as “a curvature of the formants during the vowel onset” (?). To test the effect of the formant transition on listener perception, they created stimuli with second formant transitions which began from a point four harmonics below to six harmonics above the center frequency of the vowel’s second formant. The results, from these two classic studies (??), established the concept of *co-articulation*, a *consonant-vowel context effect* where one acoustic feature (either a burst or a formant transition) can cue different consonants, depending on the following vowel. Interestingly, the earlier AI research of Fletcher assumed there was no coarticulation effect. Syllables (CV, VC, and CVC) were all purposely modeled using independent (C,V) groups.