

Title: The identification and modification of consonant perceptual cues in natural speech

This tutorial will cover the psychoacoustic and signal processing techniques for the identification and manipulation of consonant cues in natural speech. Humans are able to recognize naturally-spoken consonants with a remarkable accuracy despite the huge amount of variability that is present. Understanding which cues humans use to decode these variable consonants is fundamental to any research involving natural speech.

In this tutorial, we will first review the history of speech cue research, including the work at Haskins Labs using samples of synthetic speech, and corresponding testing of early perceptual cue hypotheses using naturally-spoken speech samples. This will lead into a discussion of the development of modern speech cue concepts and current approaches to analysis. Specifically, the use of classification models to assess the importance and interactions of existing consonant cues, as well as how one can use cue enhancement and modification for verification.

The HSR research group at UIUC has developed a psychoacoustic approach to identifying and determining the roles of consonant cues in naturally-spoken speech. We will show how data from three basic psychoacoustic experiments can be used to isolate the necessary and sufficient consonant cue region in natural, variable speech. Using sound demonstrations, we will show how removal of non-necessary energy does not affect correct recognition. The demonstrations will also show how naturally-occurring conflicting cues can completely change the perception of the consonant if the necessary cue region is masked or attenuated.

Novel concepts that are covered by the tutorial include the use of the 3DDS psychoacoustic method to isolate the necessary consonant cue region in individual speech samples (Li et al. 2010, 2012). We will discuss how manipulation of this small necessary cue region directly affects the noise-robustness of a speech signal (Kapoor and Allen 2012). As well as the binary nature of consonant perception in masking noise (Singh and Allen 2012) and the effects of naturally-occurring conflicting consonant cues (Li et al. 2010, 2012).

There is a wide target audience for this presentation, in general terms the audience is researchers who are interested in a basic understanding of what acoustic elements humans use to decode natural speech. This includes researchers working on de-noising, automatic speech recognition, and improving hearing aid algorithms.

Presentation Outline:

[Presenter Jont Allen]

- Introduction to the work of the Human Speech Recognition group at UIUC
- Primary Research Objectives
 - Identify the acoustic cues in naturally-spoken consonants (over 100 tokens used)
 - Isolate the necessary cues
 - Understand the role of naturally-occurring conflicting cues
- Background/History of Consonant Perception
 - Rayleigh (1910) to Shannon (1948)
 - Findings and drawbacks of synthetic speech approaches

- Findings with natural speech
- Harvey Fletcher and the articulation index (AI)
- Recent trends and achievements in human speech perception (last 15 years)
 - Acoustic characteristics of vowels and stop consonants
 - Acoustic characteristics of fricatives (focus on work of Jongman et al.)
 - Effects of cue enhancement and modification
 - Classification models to determine discriminating cues (McMurray et al., Alwan et al.)

[Switch to Presenter Feipeng Li]

- HSR research group at UIUC
 - Methods
 - The AIgram and SNR_e
 - Confusion patterns (including morphs, primes)
 - 3-Dimensional Deep Search (3DDS) methodology

[Switch to Presenter Andrea Trevino]

- Results
 - Stop consonants, necessary cues in natural speech
 - Stop consonants, conflicting cues
 - Stop consonants, demo with sound
 - Stop consonant verification experiments
 - Importance of the burst vs. formant transition (Cvengros et al.)
 - Effects of amplification/attenuation of necessary cue region (Kapoor et al.)
 - Stop consonant cue variability, relation to the AI (Singh et al.)
 - Fricative consonants, necessary cues in natural speech (Li, Trevino, et al.)
 - Fricative consonants, conflicting cues
 - Fricative consonants, demo with sound
 - Summary cue results
- Relating these results to the state of the art
- Summary and Conclusions

Presenter Bios:

Jont Allen received an MS/PhD from the University of Pennsylvania in 1970. Upon graduation Allen joined Bell Laboratories, Murray Hill NJ, with the Acoustics Research Department. In 2003, Allen retired from AT&T and joined the Dept. of Electrical Engineering at UIUC, as an Associate Professor, with Tenure (2006).

Andrea Trevino received her BS in Electrical Engineering from Rice University in 2007 and MS in ECE from the University of Illinois UC in 2009. She is currently pursuing her PhD in ECE at the University of Illinois UC with Prof. Allen.

Feipeng Li received his BS/MS, in EE, from Wuhan University in China. In 2003, he became a PhD student in the ECE Department at UIUC. Following graduation in 2009, Li joined the research group of X. Wang at Johns Hopkins University (JHU), and later transferred to the Hermanski lab at JHU.