Danica Reid
Simon Fraser University

# An Analysis of the Acoustic Properties of Similar Voices

**Background**: Forensic Phonetics work is often time sensitive and having a method of acoustic analysis that can be condu4(onr wor)4(k is o)-2(fte,n6t)oQEMs 7 792 reW*nBT/F1 1aliemsicu4(onr wor)4(k is ET

compared the voices of a perpetrator and a suspect to establish whether the suspect had made a series of obscene phone calls. The three measures used were: speaking fundamental frequency (F0), long-term average spectrum (LTS), and long-term formant frequency (LTF). 7 K H V H S U R Y L G H D Q R Y H U D O O S L F W X U H R I D V S H D by any individual sounds. These measurements, especially in the case of LTF, are also relatively easy to conduct as it is not necessary to separate out certain sounds to do fine-grained analyses (Gold, French, & Harrison, 2013). In addition to these benefits, long-term measures can also be conducted on languages that are not known by a given researcher as no specific knowledge of the phonetic inventory of a language is needed (Jessen, 2008). While these measures have been used to provide evidence for speaker identity (Nolan & Grigoras, 2005), it is unclear how good this method is at distinguishing highly similar voices. The goal of this project was to determine the usefulness of long-term acoustic measures in the comparison of two remarkably similar voices.

**Analysis**: Two hosts from the National Public Radio (NPR) podcast *Invisibilia* were used as the speakers for this analysis. Both the audience as well as the hosts themselves have made note of the striking similarity of their voices. This is despite the speakers being raised in different parts of the United States and having an age difference of approximately ten years. Once separating the useable speech from two episodes of the podcast, I was left with 4 minutes of speech for one host and 9 minutes of speech for the second host which was split into two parts to test reliability. The F0 measurements are based on pitch tracks extracted from Praat. This is also how the formant measures were taken for the LTF analysis. The LTS analysis also used Praat to draw the distribution. This project focused on an observation of the shape of each distribution.

**Results**: An informal analysis by ear confirmed that the two speakers have clearly similar voices. Both speakers make use of creaky voice and uptalk, though to a slightly different extent, and have similar ranges in pitch. However, the differences that are detectable through this informal analysis are often not sufficient to provide conclusive evidence as to speaker identity which is why the additional measures are necessary. Previous work has suggested that the shape of the F0 distribution stays relatively constant despite any intra-speaker pitch variation (Kinoshita, Ishihara, & Rose, 2002). Based on the resulting distributions for the F0 of the two speakers, there were not substantial differences in the shape of the distributions despite a 10 to 15 Hz difference in 792W*nBT/F1 12 Tfds.u9*nJTJ-3(a)4by-sub speaker. This suggests that F0 might not be particularly informative in the comparison of similar voices. The LTS analysis provided similar results. Throughout the distribution curve, the two speakers only showed a

two speakers were in the first and third formants. The differences in the third formant are particularly telling as the higher formants are generally thought to be relatively stable for any given speaker (Gold et al, 2013; Reetz & Jongman, 2009). As the field of Forensic Phonetics is still growing, results such as these are important as they highlight the need for more research into the analyses used in speaker identification and voice comparison

Danica Reid
Simon Fraser University

References