Forensic Comparison of Voices, Speech and Speakers

Tools and Methods in Forensic Phonetics

Jonas Lindh

Department of Philosophy, Linguistics and Theory of Science

Abstract

This thesis has three main objectives. The first objective (A) includes Study I, which investigates the parameter fundamental frequency (F0) and its robustness in different acoustic contexts by using different measures. The outcome concludes that using the alternative baseline as a measure will diminish the effect of low-quality recordings or varying speaking liveliness. However, both creaky voice and raised vocal effort induce intra-variation problems that are yet to be solved.

The second objective (B) includes study II, III and IV. Study II investigates the differences between the results from an ear witness line-up experiment and the pairwise perceptual judgments of voice similarity performed by a large group of listeners. The study shows that humans seem to be much more focused on similarities of speech style than features connected to voice quality, even when recordings are played backwards. Study III investigates the differences between an automatic voice comparison system and humans' perceptual judgments of voice similarity. The experiments' results show that it is possible to see a correlation between how speakers were judged as more or less different using multidimensional scaling of similarity ranks compared to both the automatic system and the listeners. However, there are also differences due to the fact that human listeners include information about speech style and have difficulties weighting the parameters, i.e. ignoring them when they are contradictory. Study IV successfully investigates a new functional method for how to convert the perceptual similarity judgments made by humans and then compare those to the automatic system results within the likelihood ratio framework. It was discovered that the automatic system outperformed the naïve human listeners in this task (using a very small dataset).

The third objective (C) includes study V. Study V investigates several statistical modelling techniques to calculate relevant likelihood ratios using simulations based on existing reference data in an authentic forensic case of a disputed utterance. The study presents several problems with modelling small datasets and develops methods to take into account the lack of data within the likelihood ratio framework.

In summary, the thesis contains a larger historical background to forensic speaker comparison to guide the reader into the current research situation within forensic phonetics. The work further seeks to build a bridge between forensic phonetics and automatic voice recognition. Practical casework implications have been considered throughout the work on the basis of own experience as a forensic caseworker and through collaborative interaction with other parties working in the field, both in research and in forensic practice and law enforcement. Since 2005, the author has been involved in over 400 forensic cases and given testimony in several countries.

Keywords: forensic phonetics, automatic voice recognition, disputed utterance, speech, language technology