D37 Measuring Typicality in Speech Features in American English Dialects: Towards Likelihood Ratios in Speaker Recognition Casework

Reva Schwartz, MA, Forensic Services Division, U.S. Secret Service, 950 H Street, NW, Suite 4200, Washington, DC 20223; Wade Shen, MA*, Joseph Campbell, PhD, and Robert Granville, BA, MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420; Shelley Paget, BA, and Melanie Heighway, BA, Appen Pty. Ltd., North Tower, Level 6, Chatswood, AUSTRALIA; and Stephanie Strassel, BA, Meghan Lammie Glenn, BA, and Chris Cieri, PhD, Linguistic Data Consortium, 3600 Market Street, Suite 810, Philadelphia, PA 19104*

After attending this presentation, attendees will be informed about work in speaker recognition.

This presentation will impact the forensic community by describing new methods being developed to measure the typicality of certain features in American-English dialects.

This presentation will focus on new methods that are being developed to measure the typicality of certain features in American English dialects. The presentation will demonstrate the ongoing development in speaker recognition analysis. These methods have been enhanced by a growing knowledge of what is typical for various dialects in American English. The goal is to eventually build a large annotated corpus sufficient for establishing dialect norms for a variety of linguistic phenomena. It is hoped that such a corpus will assist forensic phoneticians and sociolinguists to quantify variation between and within American English dialects. It is also believe these methods will improve technologies for automatic dialect identification and automatic speaker recognition.

Claims about speaker recognition vary across multiple methods. In the field of forensic phonetics, applied phoneticians routinely identify speakers from phonetic characteristics that are hypothesized to be speaker specific.

Quantifiable norms of language- and dialect-dependent features are necessary for forensic examiners to assess if a given phonological or phonetic feature is speaker specific or commonly found in that speaker's dialect. To obtain these norms, detailed annotation of large sets of speech data must take place. This research utilizes rapid, semiautomatic annotation techniques of detailed phonological and morphological phenomena for large-scale speech corpora. Resulting annotations and corpora will support both large-scale linguistic dialect analysis and automatic dialect identification.

The techniques used in the detailed annotation of large-scale speech corpora will be described in detail. "Regions-of-interest" (ROIs) were used, where an annotator is asked to make a judgment of whether or not a certain transformation of a given feature occurred in orthographically transcribed data. Transformations from General American English to a specific dialect are based on rules of occurrence in the specific dialect. These transformations are currently phonetic only and are among the most commonly occurring in spoken dialects of American English. Morphological rules will be applied at a later date. Phonetic rules were developed by a team of sociolinguists and sociophoneticians. The use of ROIs, together with an annotation tool, allows a large amount of data to be processed in a shorter amount of time. The output of these judgments is a likelihood ratio between speech samples. With enough judgments from the speech corpora a measure of typicality can be used, allowing for likelihood ratios between speech samples and compared against a given population.

The ROI method has been carried into speaker recognition casework for testing purposes. Each case starts with a detailed orthographic transcription. The transcription and original audio are processed to create a file where the audio is aligned with a transcript that contains a word layer and a phone layer. Once the transcription is checked for errors the word and phone layers are fed into a tool to generate ROIs based on the above-described rule-development process. The ROIs are then judged by an expert examiner in a different tool. It was found that the process must be carried out by expert examiners with backgrounds in linguistics and phonetics since judgments must be informed and interpretation of output is highly technical in nature. The presentation will lay out each step in the analysis procedure from raw audio to final conclusion.

Speaker Recognition, Likelihood Ratios, Linguistics/Phonetics

 ** Presenting Author*