Automatic Speaker and Language Recognition

Doug Reynolds and Marc Zissman
MIT Lincoln Laboratory

Tutorial Outline

1. Introduction/Background
    - History
    - Definitions
    - Applications
    - Basics of speech processing
2. Speaker Recognition
    - Approaches using acoustic information; Speech features conveying speaker information
    - Approaches using other levels of information; Prosodics; Idiolect; Phonotactics; Pronunciations;
    - System fusion
    - Computation issues
    - Performance of speaker recognition techniques
3. Language Recognition
    - Approaches using acoustic information; Speech features conveying language information
    - Approaches using phonotactics; Using phone strings and statistical language models
    - System fusion
    - Computation issues
    - Performance of language recognition techniques
4. Conclusions and Future Directions

Abstract

The speech signal conveys several levels of information beyond the words, such as information about the identity of the speaker and the language being spoken. These other levels are very useful for augmenting the word transcripts to produce *rich transcripts* for indexing and searching of audio archives. In this tutorial we will provide an overview of state-of-the-art techniques for extracting, modeling and evaluating speaker and language information from the speech signal. We will provide an overview of the area with some historical context and describe different applications that use speaker and language recognition technology. For both speaker and language recognition technology, we will discuss the theory and the practice of how these systems are designed, trained, and evaluated, from the extraction of features from the speech signal, to corpora used to gage performance. A summary of expected performance in recent NIST evaluations will show expected performance levels. We will finally discuss some of the critical challenging issues and future research directions.