

Abstract

An Efficient First Pass of a Two-Stage Approach for Automatic Language Identification of Telephone Speech

By
Jonathan K. Law

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Professional Studies
in Computing

June 2002

Automatic language identification, recognizing a speaker's language from a speech signal, is gaining increased importance in the context of economic globalization. The most accurate systems for such identification use multiple, large vocabulary, continuous speech recognizers. But their scalability is limited, as each new language added requires complete recognizers and an enormous amount of training. For increased efficiency, we propose a Two-Stage approach: an efficient clustering algorithm (least cost) first selects the top candidates at an accuracy level of over 80%. Then, speech recognizers (high cost) would be used to narrow the field from approximately three candidates to correctly identify the language. In this research, we describe the first phase of this Two-Stage process in which we are able to narrow the list of possible languages through an extremely scalable and flexible system. We cover how test and reference patterns (acoustic feature vectors) are extracted from speech utterances, how cepstral coefficients are used, and how reference models are generated from the reference patterns using Vector Quantization (VQ) clustering algorithm. Various distance measures are also examined in the selection phase to find the best method. We show using the top-N strategy in this first stage leads to substantial improvements over existing systems in discriminating between different languages, and identifies the top candidates with a combined accuracy of over 80 percent. Further research is underway to improve the entire process, but combining this particular methodology with current best practices has proven an extremely efficient and cost effective way to address the challenges of automatic language identification.