**PREDICTING SEARCH TERM RELIABILITY  
FOR SPOKEN TERM DETECTION SYSTEMS**

Amir Hossein Harati Nejad Torbati and Joseph Picone

*Abstract*— Spoken term detection is an extension of text-based searching that allows users to type keywords and search audio files containing recordings of spoken language. Performance is dependent on many external factors such as the acoustic channel, language, pronunciation variations and acoustic confusability of the search term. Unlike text-based searches, the likelihoods of false alarms and misses for specific search terms, which we refer to as reliability, play a significant role in the overall perception of the usability of the system. In this paper, we present a system that predicts the reliability of a search term based on its inherent confusability. Our approach integrates predictors of the reliability that are based on both acoustic and phonetic features. These predictors are trained using an analysis of recognition errors produced from a state of the art spoken term detection system operating on the Fisher Corpus. This work represents the first large-scale attempt to predict the success of a keyword search term from only its spelling. We explore the complex relationship between phonetic and acoustic properties of search terms. We show that a 76% correlation between the predicted error rate and the actual measured error rate can be achieved, and that the remaining confusability is due to other acoustic modeling issues that cannot be derived from a search term’s spelling.

*Keywords*— spoken term detection, voice keyword search, information retrieval

Manuscript submitted December 30, 2012. Manuscript revised and resubmitted on May 8, 2013.

A. Harati and J. Picone are with the Department of Electrical and Computer Engineering at Temple University, 1947 North 12th Street, Philadelphia, Pennsylvania 19027 USA (phone: 215-204-4841; fax: 215-204-5960; email: joseph.picone@isip.piconepress.com).