An Investigation of Subword Unit Representations for Spoken Document Retrieval

Kenney Ng and Victor W. Zue

Spoken Language Systems Group
MIT Laboratory for Computer Science
545 Technology Square, Cambridge, MA 02139 USA
{kng, zue}@mit.edu

This study investigates the feasibility of using subword unit representations for spoken document retrieval as an alternative to using words generated by either keyword spotting or word recognition. Our investigation is motivated by the observation that word-based retrieval approaches face the problem of either having to know the keywords to search for a priori, or requiring a very large recognition vocabulary in order to cover the contents of growing and diverse message collections. In this study, we examine a range of subword units of varying complexity derived from phonetic transcriptions. The basic underlying unit is the phone; more and less complex units are derived by varying the level of detail and the length of sequences of the phonetic units. We measure the ability of the different subword units to effectively index and retrieve a large collection of recorded speech messages. We also compare their performance when the underlying phonetic transcriptions are perfect and when they contain recognition errors. We find that with the appropriate subword units it is possible to achieve performance comparable to that of text-based word units if the underlying phonetic units are recognized correctly. In the presence of recognition errors, performance degrades but many subword units can still achieve reasonable performance.