# *Spoken Term Detection*

## A New NIST Evaluation Initiative

Information retrieval is a major economic activity around the world, and digitized speech from many sources is growing rapidly in volume. Responding to the opportunity and need, NIST is designing a new evaluation initiative called Spoken Term Detection (STD). The goal is to facilitate research and development of technology for finding short word sequences rapidly and accurately in large heterogeneous audio archives — an important precursor to more sophisticated retrieval technologies. If you are working on or interested in this type of technology, please send email to STD-info@nist.gov to be added to the STD interest group mailing list. A brief description of the current plan follows.

**The Task** – The STD task is to index a speech corpus and then to find all occurrences (if any) of specified terms (word strings) in that corpus.

Systems will index the corpus without knowing the terms and will search for terms individually. A wide range of search terms and a mixture of data types will be used to probe the technology. Speed, accuracy, and robustness will be key evaluation measures.

A search term will consist of one or more contiguous words specified orthographically. Thus, "cat", "cattle", "grasshopper", "New York", "Albert Einstein", and "in terms of" are all potential search terms.

For each search term, a system will output a list of hypothesized starting and ending times and associated confidence scores plus a confidence score cutoff threshold for making hard decisions.

**The Test Corpus** – The test corpus will contain multiple source data types — broadcast news, telephone conversations, and meetings — and will be in three languages — English, Chinese, and Arabic. There will be up to 10 hours of speech per language and up to 1000 search terms per language (including single-word and multi-word terms, common and rare terms, terms that are in the test corpus and those that are not). Systems will process all of the data and all of the terms for at least one language.

**The Development Test Corpus** – To help participants debug their systems, NIST will provide a development test corpus plus sample search terms and scoring software.

**The Evaluation** – *Accuracy* will be measured in terms of misses and false alarms. A hypothesized term location will be considered correct if it is within ~ 0.5 seconds of the ground truth reference and if the orthography is an exact match (disregarding case). Thus, for example, "grasshoppers" will not match "grasshopper", and "rite" will not match "right". On the other hand, "wind" (moving air) will match "wind" (twist) despite distinctly different pronunciations.[1]

*Robustness* will be assessed in terms of the accuracy on various subsets of search terms (to determine the effect of parameters such as acoustic complexity and average duration) and underlying source data types. Systems may be strong in some areas and not in others.

*Speed* is of great importance in potential applications. Systems will preprocess (index) the source data (knowing the language and source data type but not knowing the search terms) then search the index (rather than the audio data) for the designated terms, reporting index size, indexing speed and search speed.

**The Workshop** – Although the evaluation is open to all interested parties, attending the follow-on technical workshop is a privilege limited to evaluation participants. During the workshop, principal investigators will be expected to discuss the results, describe their systems, exchange ideas, and help define future evaluations.

**The Schedule** – NIST will provide a detailed evaluation plan in late May 2006, development data in July 2006, and evaluation data in November 2006. The evaluation will span two weeks in November. The two-day workshop will occur in December.

[1] This is a practical expedient.