

## 5. Large Scale Spoken Document Retrieval

**Pedro J. Moreno and Jean Manuel Van Thong, HP Labs**

[http://www.hpl.hp.com/personal/Pedro\\_Moreno](http://www.hpl.hp.com/personal/Pedro_Moreno)

[http://www.hpl.hp.com/personal/Jean-Manuel\\_Van\\_Thong](http://www.hpl.hp.com/personal/Jean-Manuel_Van_Thong)

Search engines like Google or Yahoo have been extremely successful over the years in facilitating the search and retrieval of text pages and written documents. However, only recently these technologies have been extended to spoken documents. While there are many similarities with standard text search engines, spoken document retrieval is sufficiently different.

In this tutorial we provide an introduction to the field of spoken document retrieval with an special emphasis on large audio collections. We will start with a general introduction to speech recognition, then continue with various approaches to audio indexing and then continue with a global description of the architecture needed for large scale indexing. We will conclude with several demos of existing engines and technologies.

### 5.1. Tutorial Outline

1. Extracting metadata from raw audio.
  1. Fundamentals of speech recognition.
    1. Acoustic modeling.
      1. Words based, phone based.
    2. Language modeling
    3. Search
  2. Speech recognition approaches for audio indexing.
    1. Phonetic search.
    2. Word spotting approaches.
    3. Large vocabulary speech recognition.
    4. Syllable based speech recognition.
    5. Limitations and advantages of all approaches.
    6. The out-of-vocabulary (OOV) problem.
  3. Text audio alignment.
2. Indexing and searching metadata.
  1. Searching versus indexing.
  2. Content segmentation.
  3. Modification to text indexing, long documents vs. short documents.
  4. Index fusion approaches.
  5. Acoustic search versus semantic search.
3. Architecture design for large scale indexing.
  1. The web search model for audio indexing.
    1. Audio (and video) crawling.
    2. Audio to text transcription.
    3. Index construction.
  2. API's for querying and index update.
  3. The user interface design.
  4. Putting everything together.
4. Demos of several systems.
5. Conclusions: Where is audio indexing headed?

### 5.2. Target Audience

This tutorial is designed for information retrieval and computer scientists with no previous knowledge of speech recognition and information retrieval.

Pedro J. Moreno is a senior researcher at the Cambridge Research Lab, which is part of Hewlett-Packard Labs. His main interests are in the practical applications of machine learning techniques in several fields such as audio indexing, image retrieval, text classification and noise robustness. Dr. Moreno has been involved in the design of HP Labs audio indexing engine SpeechBot. Lately his main interests are in the areas of bioinformatics and bio signal interpretation.

JM Van Thong is a senior researcher at the Cambridge Research Lab, which is part of Hewlett-Packard Labs. His current research interests are bioinformatics, media indexing, and information retrieval systems as well as user interfaces. During his 17 years spent in research, JM has been involved in several successful projects including SpeechBot, the first large scale web audio indexing system, RedBot, a web-based tool for automatic red-eye correction, an information retrieval system for hand-helds, a real-time streaming phoneme recognizer for a facial animation package, and planar maps technology for a sketching software.