

Audio Hot Spotting

Qian Hu

781-271-2959 • qian@mitre.org

MITRE Sponsored Research

The logo consists of a cluster of 3D cubes in yellow, orange, and blue, arranged in a stepped pattern. To the right of the cubes, the text "MITRE Technology Program" is written in a bold, sans-serif font.

MITRE
Technology
Program

Problem

Nuggets of information are buried within huge mountains of multimedia data. A system is needed to quickly and automatically identify and retrieve these audio “hot spots.”

The current approach to audio information retrieval of simply combining text-based information retrieval with automatic speech recognition does not meet user needs in real applications.

Background

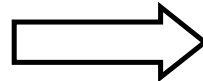
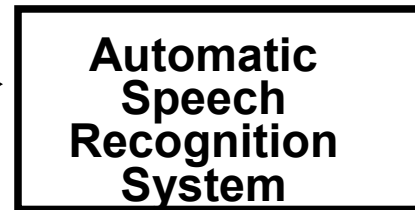
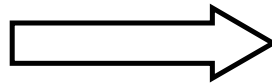
Current Approach: ASR + IR Doesn't Meet the Needs of Critical Applications

“.. cross into **Korea**”



Smuggler ID 719

Background surf noise



“... cross into career”

Nothing Found

[TREC 6-9: Spoken Document Retrieval]

Objective

- **Provide prototype hot-spotting capability to enable efficient audio filtering and retrieval**
 - **Determine the limitations of algorithms in existing component technologies**
 - **Leverage, integrate, and extend the best technologies for the hot-spotting problem**
 - **Research and develop new algorithms when COTS and GOTS systems fail to meet the requirements of audio hot-spotting**
 - **Research and develop audio-specific query algorithms making use of multiple types of audio information**

Activities

- **Speech Recognition:** Evaluated COTS/GOTS with real application data; experimented with language model adaptation
- **Speaker Identification:** Extended GOTS algorithms to improve “speaker-change” detection in domains including several unknown speakers
- Experimented with query expansion techniques to compensate for transcription errors
- Extended Audio Hot Spotting prototype to:
 - support foreign languages
 - search across multiple “audio documents”
 - search documents on reusable, predefined text queries

Highlight

Input

Real Media
WAV Files
AVI
MPG
MPA
MPEG
MP4
MP3

Preprocessing

<i>Speech Recognition / LM experimentation</i>
<i>Speaker ID / Speaker Change</i>
<i>Automatic Keyword Indexing</i>
<i>Data Structuring</i>
<i>Audio Feature Extraction</i>

Output

Time Indexed Text
Time Indexed Speakers & Speaker Statistics
Lexical Index List
Structured Data in Relational Database
Time Indexed Audio Features

Blue = MITRE

Red = GOTS/COTS

ATF agents located the TNT ..

Identified Selection Available as Original Media or Text



<i>Query Language</i>
<i>Hot Spotting GUI</i>
<i>IR for High Error Rate Speech</i>

Demonstration

Built Audio Hot Spotting prototype to query and retrieve key words, speakers, and some audio effects for both English and Spanish

The screenshot shows the 'Audio Hot Spotting' web application in Microsoft Internet Explorer. The browser address bar shows the URL: http://mm106647-2k.mitre.org/pls/audio/audio_rev14.results_single_combined. The page features a logo with the text 'AUDIO HOT SPOTTING' and a magnifying glass over the word 'SPOTTING'. Below the logo, there are search fields for 'Word and Phrase Search' (containing 'terrorism'), 'Speaker Search' (containing 'Marty Faga'), and 'Sound effect search' (containing 'applause'). A vertical yellow bar labeled 'QUERY CUES' is positioned to the right of these search fields, with arrows pointing to each field. A yellow callout box above the search results area states 'Allows Combination of Query Cues'. The search results section displays: 'Search results: Word/Phrase: "terrorism" Speaker: "Marty Faga"'. Below this, it says 'Your search produced 3 hit(s)'. A table with three columns: 'Time', 'Text', and 'Speaker' is shown. The 'Time' column has a speaker icon and the text 'Click to play audio'. The 'Text' column has the text 'Click to display text excerpt'. The table contains three rows of search results. A yellow callout box to the right of the table states 'Provides both multimedia and text retrieval'. A video player window in the top right corner shows a man speaking at a podium. At the bottom of the page, there is a text excerpt: 'Department of Health and Human Services and started a new office of public health preparedness to detect and responsive bio terrorism this is this is the office that has the national responsibility for bio terrorism and'.

QUERY CUES

Media file name: Forum 2002
Date: 22-MAY-02

Word and Phrase Search:
terrorism

Speaker Search:
Marty Faga

Sound effect search:
applause

Search results:
Word/Phrase: "terrorism"
Speaker: "Marty Faga"

Your search produced 3 hit(s).

Time Click to play audio	Text Click to display text excerpt	Speaker
00:22:50	and responsive bio terrorism this	Marty Faga
00:22:55	responsibility for bio terrorism and	Marty Faga
00:56:03	new threat of terrorism inside	Marty Faga

Department of Health and Human Services and started a new office of public health preparedness to detect and responsive bio terrorism this is this is the office that has the national responsibility for bio terrorism and

Impacts

- Presented our work at the 5th International Conference on Text, Speech and Dialogue
- Presented a paper at 2002 Fall IEEE Conference on Technologies for Homeland Security
- Briefed government organizations, including NSA, DARPA, ARDA, FBIS, and AFRL
- Worked with industry leaders to improve component technologies

Future Plans

- Audio-specific query algorithms
- Domain-specific applications
- IR & audio mining research
- Explore multiple ASR engines

- Audio feature extraction
- Speaker change detection
- Foreign language support
- Query expansion research

- Integrate and extend component technologies
- Develop a prototype system to **hotspot** speaker and/or keywords

