ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 2, Issue 9, September 2013*

# An Analytical Approach for Mining Audio Signals

Manpreet Kaur Mand[1], Diana Nagpal[2], Gunjan[3]

Assistant Professor, Computer Science & Engineering Department, Guru Nanak Dev Engineering College, Ludhiana, Punjab, India[1,2,3]

**Abstract-** Audio mining or audio indexing is a speech recognition technique that is used to search audio signals for occurrences of spoken words or phrases. Speech technology is used to recognize the words or phonemes that are spoken in an audio or video file and audio mining searches can then be carried out to locate specific words and phrases within the audio. These audio mining searches run at speeds that are typically many thousands of times faster than real time, so large quantities of audio or speech can be searched in a short time. This paper poses a light on various audio mining techniques like LVCSR audio mining and Phonetic audio mining along with their comparison. With the voluminous increase in the amount of audio content on the internet and other sources, it is clear that audio mining is a growing technology. Audio mining has been used for searching television captions and other media content because the audio mining search is able to locate the speech content associated with the text for each caption. The purpose of this paper is to provide a systematic overview of audio mining.

**Keywords-** Audio Indexing, Audio Mining, LVCSR Audio Mining, Phonetic Audio Mining, Speech Recognition Technique.

## I. INTRODUCTION

With the advent of inexpensive storage space and faster processing speed over the past decade, the research has started to penetrate new roots in areas of speech recognition and audio processing. Audio Mining has gained interest due to availability of voluminous audio content [1]. For example newscasts, sporting events, telephone conversations, recordings of meetings, Web casts, documentary archives etc. Users want to make the most of this material by searching and indexing the digitized audio content. Traditionally, companies had to create and manually analyze written transcripts of audio content because using computers to recognize, interpret, and analyze digitized speech was difficult enough. But, with the development of faster microprocessors, larger storage capacities, and better Speech-recognition algorithms have made audio mining easier [2].Audio mining also called Audio Searching is a technique that is used to search audio files for occurrences of spoken words or phrases. Speech technology is used to recognize the words or phonemes that are spoken in an audio or video file and audio mining searches can then be carried out to locate specific words and phrases within the audio. These audio mining searches run at speeds that are typically thousands of times faster than traditional systems, so large quantities of audio or speech can be searched in a short time [3]. Audio Mining is a technique by which the content of an audio signal can be automatically analyzed and searched. It is most commonly used in the field of automatic speech recognition, where the analysis tries to identify any speech within the audio. The audio will typically be processed by a speech recognition system in order to identify determined set of audio characteristics that are likely to occur in the spoken content. This information may either be used immediately in pre-defined searches for keywords or phrases or the output of the speech recognizer may be stored in an index file [4]. Stored audio signals can be transcribed and analyzed to determine or predict what is most likely to happen next. The structured information (Text information) represents what the customer “looks” like. The audio transcriptions represent what they “think” [5].

## II. TERMINOLOGY USED FOR AUDIO MINING

A number of different terms are used in relation with audio mining. These include: audio mining, audio indexing, phonetic searching, phonetic indexing, speech indexing, audio analytics, speech analytics, word spotting, and information retrieval. The terms "audio analytics" and "speech analytics" are often used to cover both audio mining and other speech analysis technologies [3].

## III. AUDIO MINING TECHNIQUES

Audio mining is a speaker-independent, speech recognition technique that is used to search or recognize audio or video

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 2, Issue 9, September 2013*

files for occurrences of spoken words or phrases. The speech recognition search engine identifies words or phonemes that are spoken within the file and generates a searchable index that includes a time stamp for each important word or phoneme and its locations within the file. Essentially, it creates indices of words or speech sounds. Thus, it can enable users to search audio and video files similar to the way they use search engines to search the text content [6]. There are two main techniques of audio mining. One uses large vocabulary continuous speech recognition (LVCSR), and the other uses phonetic recognition to carry out phonetic audio mining. An overview of these two approaches to audio mining is given below.

*A. LVCSR Audio Mining*

Also called text based indexing, is a two step process. In the first step (pre-processing or indexing stage), the speech content of the audio is processed by a large vocabulary recognizer to generate a searchable index file. The index file contains information about the sequences of words spoken in the audio or video data. In the second step (search stage), a search term is defined (e.g. a word or phrase), and one or more index files are searched for all occurrences that match the specified search term. The results of the search can be displayed graphically as "search hits" in the audio file, or the relevant portions of the audio or

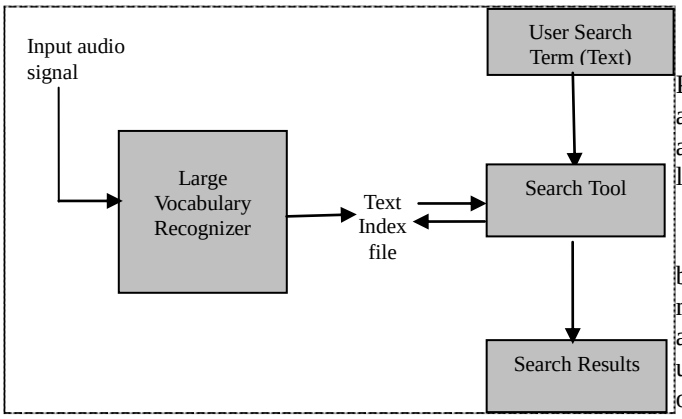


Fig 1: LVCSR Audio Mining

video file can be played to the user [3]. It converts speech to text and then identifies words in a dictionary that can contain several hundred thousand entries. If a word or name is not in the dictionary, the system will choose the most similar word it can find [2].

*B. Phonetic Audio Mining*

Also called phoneme based indexing, Phoneme based indexing doesn't convert speech to text but instead works only with sounds. Phonetic audio mining is also a two-step process. In the first step, audio is processed (indexed) with a phonetic recognizer to generate a phonetic index file. The index file produced by this phonetic approach to audio mining stores the phonetic content of the speech, in contrast to the index files generated by LVCSR methods, which contain information about words [3]. In other words, the system first analyzes and identifies sounds in a piece of audio content to create a phonetic-based index. In second step, system uses a dictionary of several dozen phonemes to convert a user's search term to the correct phonetic string.

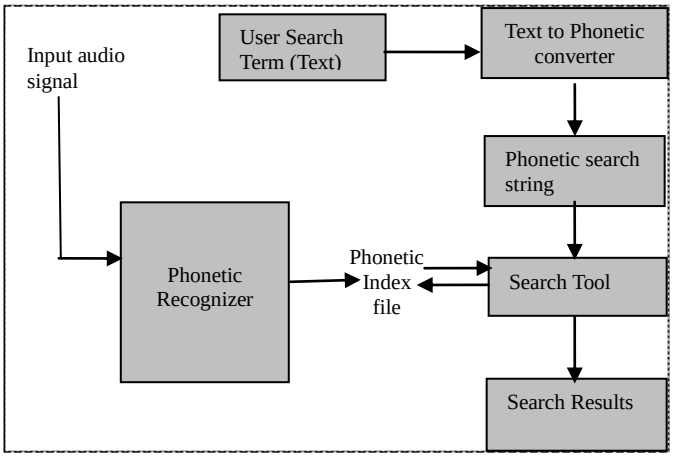


Fig 2: Phonetic Audio Mining

Phonemes are the smallest unit of speech in a language, such as the long "a" sound that distinguishes one utterance from another. All words are sets of phonemes. Finally, the system looks for the search terms in the phonetic index file.

A phonetic system requires a more proprietary search tool because it must phoneticize the query term, and then try to match it with the existing phonetic string output [2]. It is also possible to enter a phonetic search term directly, if the user has sufficient phonetic expertise to enter the sequence of phones that correspond to the pronunciation of the word or phrase they want to search for [3].

## IV.COMAPARISON OF PHONETIC AND LVCSR TECHNIQUE

The differences between the phonetic and large vocabulary techniques to audio mining lead to contrasting advantages and disadvantages in the two techniques. The major differences between above two techniques are listed in Table I [3].

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 2, Issue 9, September 2013*

TABLE 1: Phonetic V/S LVCSR Technique

| **Phonetic Audio Mining** | **LVCSR Audio Mining** |
|---|---|
| Phonetic System works at phoneme level. | LVCSR System works at words level. |
| With phonetic audio mining, the rate at which the audio content can be indexed is many times (about 100 times) faster than with LVCSR techniques. | With LVCSR Audio Mining, The rate at which the audio content is indexed is slow. |
| During the search stage, the computational burden is larger for phonetic search systems than for LVCSR technique. | In this case search stage is typically simple and less time consuming. |
| Phonetic recognition does not require the use of complex language models. | LVCSR approaches must use sophisticated language models, which leads to a much greater computation load at the indexing stage for LVCSR approaches and results in significantly slower indexing speeds. |
| In phonetic approach an open vocabulary is maintained which means that searches for personal or company names can be performed without the need to reprocess the audio. | With LVCSR systems, any word that was not known by the system at the time the speech was indexed can never be found. The LVCSR system has to be updated with a new dictionary that includes all the new words and all the audio has to be pre-processed again, which is a time-consuming task. |

## V.APPLICATIONS OF AUDIO MINING

Audio mining techniques are used in telephony applications, which help to automate quality control aspects of the business where it is important to check that what telephone agents actually said and what they were supposed to say. Audio mining searches on the recorded calls can be made to locate words or phrases that must always be said. This can offer significant advantages in terms of the number of calls that can be checked as the speed at which relevant matches can be found using audio mining is much greater than can be achieved by traditional means (a human listening to the recorded calls).Audio mining has also been used for captioning (subtitling) of TV and other video/media content, as the speech content associated with the text for each caption can be located by running a suitable audio mining search. However, a more effective and efficient way to obtain the start and end times of each word in the caption text is to use speech recognition to automatically align the known text with the speech [3]. This could also be used, for example, to retrieve relevant clips for a news story from a large video archive. Audio mining searches can typically be carried out many thousands of times faster than real time, which makes it possible to search large amounts of speech data when previously this was impossible, due to the time it would take for humans to listen to the material.

## VI.CONCLUSION

This paper describes well known techniques of audio mining. Audio mining is an extremely exciting technology that could add tremendous value to knowledge-sharing, intelligence, and productivity applications. Accuracy and Data Quality are significant issues in this area. Low accuracy rate of audio mining are mainly because of factors like background noise and cross talk. Therefore, a major challenge in speech recognition tools has been recognizing the speech of different users in different environments. Similar to any emerging technology, Audio Mining shows great opportunities but it won't achieve its full capabilities and potential until accuracy improves. In addition, Hybrid approaches must have been used where some phonetic information is retained by large vocabulary systems. At present, a number of different companies produce audio mining and speech analytics software, applications and Software Development Kits (SDKs).

## REFERENCES

[1] S.Shetty, and K.K. Achary, "Audio Data Mining Using Multi-perceptron Artificial Neural Network," *International Journal of Computer Science and Network Security*, vol.8, pp.224-229, Oct. 2008.
[2] N.Leavitt, "Let's Hear It for Audio Mining," IEEE Computer, vol.35, pp. 23-25, Oct.2002.
[3] H.Wright. (2011) Howard Wright's Home Page. [Online]. Available: http://www.hakwright.co.uk/audio-mining.html
[4] Anonymous, (2013) Wikipedia: The Free Encyclopedia. [Online]. Available: http://en.wikipedia.org/wiki/Audio_mining
[5] SAS Business Analytics and Business Intelligence Software. (Oct.2008). Tune into the Voice of Your Customer with Voice Mining [White paper]. Retrieved from http://support.sas.com/resources/papers/sgf2008/voicemining.pdf
[6] A.B.Mack(2009) Audio Mining: Where We are and What's Next.[Online]. Available: http://www.depo.com/E-letters/TheDiscoveryUpdate/2009/January/Articles/audio_mining.html