Phonetic Searching: An Advantage in Search over Speech to Text and Keyword Spotting

Rima A. Patel
Department of Computer Engineering
Vadodara Institute of Engineering, Kotambi Vadodara India

Abstract— This paper represents the concept of phonetic matching and its advantages over speech to text conversion and keyword spotting technique. Phonetic matching is string matching technique that matches the string based on the pronunciation of words. There are several algorithms that have been developed for phonetic matching. Our aim is to present the advantages of phonetic matching algorithms over speech to text conversion and Keyword Spotting technique.
Key words: Phonetic Searching, Keyword Spotting

I. INTRODUCTION

The Speech recognition and string comparison is vast area for research now a days. The searching from any kind of large database is a big issue. Several techniques and algorithms have been developed for searching from large database. In this paper we have discussed the different ways of searching and presented the pros n cons of all of them. In a full-text search, a search engine examines all of the words in every stored document as it tries to match search criteria. Search engine can also recognize the single keyword and provides the result for the search. The another way of searching is by speech recognition. The following section presents the all three techniques of searching.

II. SEARCHING TECHNIQUES INTRODUCTION

A. *Keyword Spotting*

Keyword Spotting (KWS) aims at detecting predefined keywords in an audio stream, and it is a potential technique to provide the desired hands-free interface. [1] There is an extensive literature in KWS, although most of the proposed methods are not suitable for low-latency applications in computationally constrained environments.[1] Most KWS procedures are carried out using one of the three following methods:[4]

- LVCSR-based KWS: where an LVCSR engine produces a transcription of the entire speech database, and the KWS-based application searches the resulting text for the designated keywords.
- Acoustic KWS: where the KWS engine operates on the speech signal and the recognition vocabulary consists only of the designated keywords, represented as sequences of phonemes.
- Phonetic Search (PS) KWS: where a phoneme recognition engine produces a phonetic representation of the entire speech database, and a phonetic search engine searches the resulting phoneme sequence or lattice for the designated keywords.

Keyword spotting is a technique to identify the keywords in utterances. There can be two types.

- Keyword spotting in unconstrained speech
- Keyword spotting in isolated word recognition

The meaning of utterance is spoken word, a statement or a vocal sound. If we want To Determine if a Keyword out of a Predefined Keyword Set was spoken in an Utterance or not, three possibilities are there,

- No need transcribe all the words in the utterance
- Utterances under more unconstrained conditions
- Applications in speech understanding, spoken dialogues, human-network interaction

1) Key Phrase: one or a few keywords connected, or connected with some function words

- e.g. on Tuesday, from China to Hong Kong, Eight Thirty Five a.m.

2) Spotting/Detection of Longer Phrase is More Reliable

- a single keyword may be triggered by local noise or confusing sounds
- Similar verification performed with longer phrase (on frame level, phone level, etc.)
- use of a phrase as the spotting unit

3) Key Phrase Network

- Every arc represents a group of possible key words
- Connected words can be searches easily
- key phrases are easier mapped to semantic concepts for further understanding

4) Example of Phrase Match:

Ads may show on searches that match a phrase, or are close variations of that phrase, with additional words before or after. Ads won't show, however, if a word is added to the middle of the phrase, or if words in the phrase are reordered in any way.

- Symbol: "keyword"
- Example keyword: "men's shirt"
- Example search: buy men's shirt

The Example for keyword phrase is given in below fig.1.

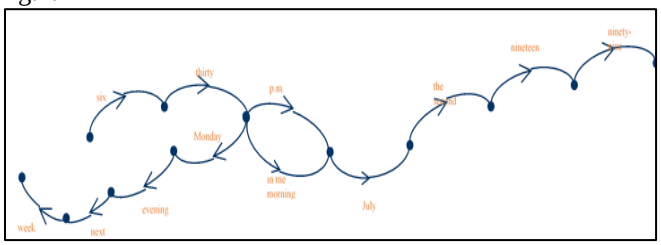

Fig. 1: Keyword Phrase

B. *Speech to Text*

To convert speech to on-screen text or a computer command, a computer has to go through several difficult steps. When we speak, we create vibrations in the air. The analog-to-digital converter (ADC) translates this analog wave into digital data that the computer can understand. [5] To do this, it samples, or digitizes, the sound by taking precise measurements of the wave at frequent intervals. The system rectifies the digitized sound to remove unwanted noise, and sometimes to separate it into different bands of frequency. It also normalizes the sound, or adjusts it to a constant volume level. It may also have to be temporally aligned. People don't always speak at

the same speed, so the sound must be adjusted to match the speed of the template sound samples already stored in the system's memory.

Next the signal is divided into small segments as short as a few hundredths of a second, or even thousandths in the case of plosive consonant sounds -- consonant stops produced by obstructing airflow in the vocal tract -- like "p" or "t." The program then matches these segments to known phonemes in the appropriate language. A phoneme is the smallest element of a language -- a representation of the sounds we make and put together to form meaningful expressions. There are roughly 40 phonemes in the English language (different linguists have different opinions on the exact number), while other languages have more or fewer phonemes.

The model for the same is given below.

Fig. 2: Conversion from Speech To Text

C. Phonetic Search

The phonetic matching is a string matching technique in which searching of particular keyword is done based on pronunciation. As the pronunciation of each and every person is different, it is somewhat difficult to apply this kind of technique. Also the languages of different persons are different. But there are many algorithms available for searching the keyword from the large database based on phonetic search.

Phonetic comparison algorithms are precisely defined methods for quantifying the similarity between speech forms or segments, words, or even entire languages on the basis of their sounds [6].

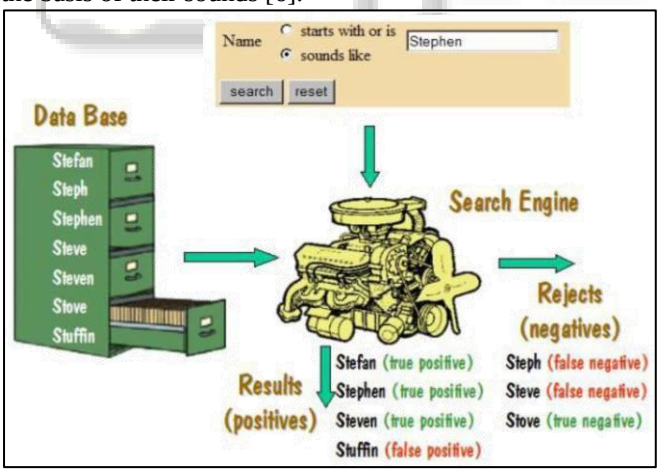

Fig. 3: True and False Negative [7].

To understand the working of matching operation we will discuss the example of large database that consists of the names Stefan, Steph, Stephen, Steve, Steven, Stove, and Stuffin. Suppose that we want to search for the name Stephen. The matches that the search finds are called the positives, and those names that it rejects are called the negatives. Those positives that are relevant are called truepositives, and the others are false positives [7].As an example, let us assume that the matches found when searching for Stephen in the above database are Stefan, Stephen, Steven, and Stuffin. The first three are probably relevant, and are names that we would have wanted to see. So these are the true positives. Stuffin, however, is probably not relevant – it is a false positive.

The names that were rejected are Steph, Steve, and Stove. Of those, Stove is probably not one that we would have wanted. So it is a true negative. But Steph and Steve are ones that we would probably be interested in. They are false negatives [7]. In above example the phonetic matching is applied on the database that contains the names of persons. Same concept can be applied for the keywords, phrases and connected keywords.

D. Advantages of phonetic searching over speech-to-text and keyword spotting

1) Speed, accuracy, scalability. The one-off indexing phase allows high accuracy so that the
2) ly evolving). Phonetic indexing is unconcerned about such linguistic issues, maintaining completely open vocabulary (or, perhaps more accurately, no vocabulary at all).
3) Low penalty for new words. Speech recognition lexicons can be updated with new terminology, names, and other words. However, this exacts a serious penalty in terms of cost of ownership because the entire media archive must then be reprocessed. The dictionary within the phonetic searching architecture, on the other hand, is consulted only during the searching phase, which is relatively fast compared to indexing. Adding new words incurs only another search, and it is often unnecessary to add words, since the spelling-to-sound engine can handle most cases automatically, or users can simply enter sound-it-out versions of words.
4) Phonetic and inexact spelling. Proper names are particularly useful query terms—but also particularly difficult for speech recognition systems, not only because they may not occur in the lexicon as described above, but also because they often have multiple spellings. With phonetic searching, exact spelling is not required.

III. Conclusion

In this paper we have shown the comparison of speech to Text and Keyword Spotting. The phonetic matching technique, which has been used in name search only, can also be used in keyword search from large databse. Also the advantages of keyword spotting over a Speech to Text has been discussed. In future we can combine these two methodologies and can use it in Phonetic search for different languages.

Acknowledgment

I would like to take this opportunity to express my profound gratitude and deep regard to our Institute, for the guidance, valuable feedback and constant encouragement. I would also like to give my sincere gratitude to our management, all the friends and colleagues who filled in the survey, without which this research would be incomplete. I am also very thankful to my husband Mr. Anant B Patel for all his support and help.

REFERENCES

[1] Guoguo Chen, Carolina Parada,Georg Heigold "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions, Small-Footprint Keyword Spotting Using Deep Neural Networks" Center for Language and Speech Processing, Johns Hopkins University, Baltimore, MD2 Google Inc., Mountain View, CA
[2] G. Chen, C. Parada, and G. Heigold, "Small-footprint Keyword Spotting using Deep Neural Networks," in Proc. ICASSP, 2014.
[3] Tara N. Sainath, Carolina Parada "Convolutional Neural Networks for Small-footprint Keyword SpottingGoogle, Inc. New York, NY, U.S.A"
[4] Szöke I, Schwarz P, Matejka P, et al. Comparison of keyword spotting approaches for informal continuous speech. Proceedings of the 9th European Conference on Speech Communication and Technology (Eurospeech); 2005 4-8 Sept; Lisbon, Portugal..
[5] http://electronics.howstuffworks.com
[6] Brett Kessler, Phonetic Comparison Algorithms, Transaction of Philological Society Volume 103:2 243-260, 2005.
[7] Beider, A, Stephen P. Morse, Phonetic Matching: A Better Soundex, March, 2010.

