ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013

Voice Translation on Mobile Phones

Parul Mann, Chaitrali Morde, Madhuri Sankpal
Information Technology Department, Mumbai University

***Abstract*— current voice translation tools and services use natural language understanding and natural language processing to convert words. However, these parsing methods concentrate more on capturing keywords and translating them, completely neglecting the considerable amount of processing time involved. In this paper, we are suggesting techniques that can optimize the processing time thereby increasing the throughput of voice translation services.**

***Index Terms*— Session based Cache, Sphinx, Template Matching, TTS,**

I. INTRODUCTION

This project is aimed at mobile phone users who can communicate with other users using offline software model that ensures effective real time communication between two users who do not speak a common language while ensuring minimal computing time.

II. MOTIVATION

Language has always been the basis of any form of written or speech communication. However, the presence of multiple language and dialects has been a hindrance to effective communication. Especially in a nation like India where the language and dialect changes with region, the requirement of a middle translation layer that can eliminate the linguistic barriers becomes essential. Speakers from different regional identities should be able to interact with one another without the need to understand individual languages.

III. LITERATURE SURVEY

A. Existing Systems

Voice translation combines technologies in the areas of automatic speech recognition, understanding and text-to-speech synthesis. The tight coupling of speech recognition and understanding effectively mitigates the effects of speech recognition errors and non-grammatical inputs common in conversational colloquial speech (as opposed to well-formed written text or read speech in dictation or broadcast news) on the quality of the translated output, resulting in a robust system for limited domains.

B. Proposed Systems

In the proposed system the user voice is taken as the input using microphone in the mobiles, it is analyzed by the speech recognizer in which acoustic signals captured by the micro phone are converted to a set of meaningful words. When speech is produced in a sequence of words, language models or artificial grammars are used to restrict the combination of words. After that words obtained are parsed in Natural Language Parsing model. Then information is extracted from the open source library. Then the input text is matched with database text. Then the output text is converted to voice and sent to speech synthesizer and output voice is heard by the user.

IV. IMPLEMENTATION

The voice translation model consists of four main components namely: speech recognition, natural language interpretation and analysis, sentence generation and text-to-speech synthesis. The optimization is provided by the natural language interpretation and analysis module is which further divided into four parts namely: template matching, indexing frequently used words, session-based cache and translation to target language. Figure depicts the overall system model for optimization of voice translation on mobile phones.

A. Speech Recognition Component

The initiator dials a number on his mobile phone. This number is connected to a centralized server. The first phase is the speech recognition component. Speech recognition is the process of converting an acoustic signal captured by the mobile's microphone to a set of meaningful words. An isolated word speech recognition system requires that the initiator on the mobile phone pause briefly between words, whereas a continuous speech recognition system does not. Speech recognition modules also take into consideration the speaking accents of the caller. Recognition is more difficult when vocabularies are large or have many similar-sounding words.

B. Language Interpretation and Analysis

The different techniques used for language interpretation can be listed as follows:-

- Template Matching :-
  Template matching checks the source input for commonly used phrases or sentences. Every language consists of a set of commonly spoken words or sentences. Converting such sentences to the target language and replacing when required drastically reduces the processing time needed for translating the sentences.
- Indexing Frequently Used Words :-
  The large lexical database of words will add to the time complexity of the process. The words are indexed based on the number of times a particular word has been used. The probability search algorithm is used to index the words in the database. In probability search the most probable element is brought at the beginning. When a key is found, it is swapped with the previous key. Thus if the key is accessed very often, it is brought at the beginning. Thus the most probable key is brought at the beginning. The efficiency of probability search increases as more and more words are being translated and indexed.
- Session Based Cache :-
  The system maintains a session-based cache for each user requesting for the service. This works on the lines of a web cache which caches web pages. This is done to reduce to reduce bandwidth usage, server load, and perceived lag. It is assumed that when a user engages in a conversation, there are bound to be multiple repetitions of certain words. Based on this assumption, we cache such words along with their translated text so that server processing time is saved.
- Translation to Target Language:-
  After the sentence passes through the first three phases of language interpretation and analysis, the final phase is translation to target language.

C. Sentence Generation

The collective set of translated word is converted to a meaningful sentence generation for the target language. Sentence generation is a natural language processing task of generating natural language from a logical form (set of translated words).

D. Text to Speech Synthesis

Text-to-speech synthesizer converts the sentence obtained from the sentence generation module into human speech form in the target language. This is done by using free TTS software. This process can be further explained as follows:-

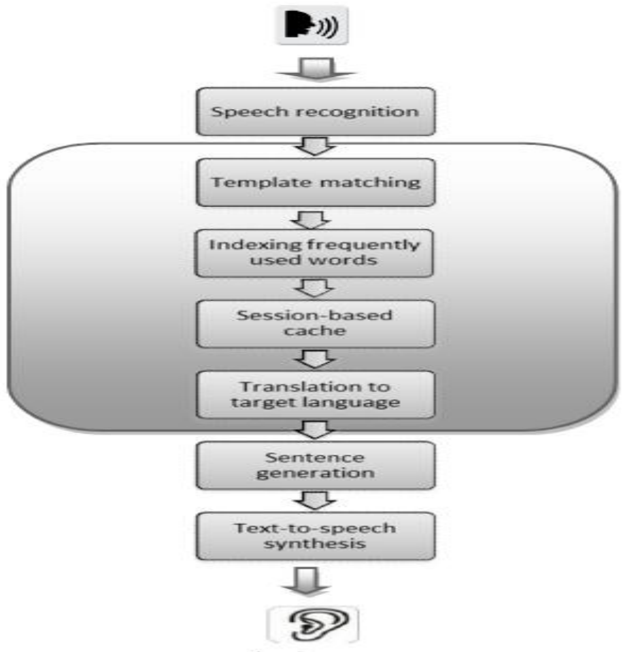

Fig 1. System Model

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013

V. KEY TECHNOLOGY

A. Sphinx

To implement this speech recognition module, Sphinz, a speech recognition system is used. Sphinx-4 is a state-of-the-art speech recognition system written entirely in the JavaTM programming language. It was created via a joint collaboration between the Sphinx group at Carnegie Mellon University, Sun Microsystems Laboratories, Mitsubishi Electric Research Labs (MERL), and Hewlett Packard (HP), with contributions from the University of California at Santa Cruz (UCSC) and the Massachusetts Institute of Technology (MIT).Sphinx-4 started out as a port of Sphinx-3 to the Java programming language, but evolved into a recognizer designed to be much more flexible than Sphinx-3, thus becoming an excellent platform for speech research.

B. Free TTS

Text-to-speech synthesizer converts the sentence obtained from the sentence generation module into human speech form in the target language. This is done by using free TTS software. It is an open source speech synthesis system written entirely in the Java programming language. It is based upon Flite. Free TTS is an implementation of Sun's Java Speech API. Free TTS supports end-of-speech markers. Gnopernicus uses these in a number of places: to know when text should and should not be interrupted, to better concatenate speech, and to sequence speech in different voices.

VI. SCOPE

Text-based translation services mainly focus around capturing words and converting them to target language. However, voice based translation services have remained few and slow. This is because most of these models concentrate mainly on language interpretation and language generation. They fail to take into consideration the large amount of back-end processing that takes place while translation. Most translation methods make use of customized dictionaries to find the translated words. However, searching for relevant words and synonyms from such large dictionaries is slow and time-consuming. More so it also depends on the content of the sentence being translated. In this project, we propose language translation on mobile phones. This project is aimed at mobile phone users who can then communicate with other users, irrespective of the other user’s ability to understand the speaker’s language.

VII. CONCLUSION

Speech technology has now advanced to the stage where it offers great promise for human-computer interaction in a variety of applications. We are moving from a world of the visual paradigm to the voice paradigm. Thus Speech has the potential for future incorporation into a multi-modal interface thus leading to a new era. Techniques like template matching, indexing frequently used words using probability search and session-based cache can considerably enhance processing times. More so, these factors become all the more important when we need to achieve real-time translation on mobile phones.

VIII. ACKNOWLGMENT

We are grateful to our institute Padmabhushan Vasantdada Patil Pratishthan’s College of Engineering, to channelize our skills and energy. We are indebted to our inspiring **Madam Mrs. Vidya Kawtikwar**, who has extended all valuable guidance, help and constant encouragement through the various difficult stages for the development of the paper.

REFERENCES

[1] Willie Walker, Paul Lamere, Philip Kwok“Free TTS- A performance case study”, August 2002.

[2] Yen Chun Lin“An optimized approach to voice translation on mobile phones”, 2010.

[3] Titus Flex FORTUNA”Dynamix Programming Algorithms in Speech recognition”, 2008.

[4] Waleed H. Abdulla and Nikola K. Kasabov, Department of Knowledge Engineering Lab, Information Science Department, University of OtagoNew Zealand.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013
[5] Geraldo Lino De Campos and Evandro Bacci Gouvea Escola Polytechnic University of Sao Paolo, Department of Computer Engineering.

[6] Dr. Joseph Picone Institute for Signal and Information Processing, Department of Electrical and Computer Engineering, Mississippi State University.

[7] Professor Matti Karjalainen and Sami Lemmetty, Helsinki University of Technology.