

IMPLEMENTATION OF VOICE TRANSLATION SYSTEM

Pratik Tamakuwala, Prof. Jayashree Prasad, Siddhant Jain, Rahul Waghresha, Amey Wadodkar

Abstract—: 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. This application is aimed at desktop users who can then communicate with other users, irrespective of the other user's ability to understand the speaker's language. It can find varied applications in businesses, teaching and voice response systems.

Index Terms—Free TTS, sphinx4, voice recognizer, voice synthesizer.

1. INTRODUCTION

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 [10] 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 work, authors propose language translation on personal computers. The application is aimed at desktop users who can then communicate with other users, irrespective of the other user's ability to understand the speaker's language. The system is mainly [9] divided into 3 major modules:

1. Sphinx
2. Free TTS
3. Google API.

The work executes in proper stepwise manner. As the system architecture is shown in figure-1. The execution starts from client side. First of all the user must need to connect with internet as we are using Google API for translation purpose.

1. Recognition of Human Voice (Speech recognizer):

As execution starts [8] first of the all microphone is being checked by the system. If it is present then the recognition processes is carried out. Recognition step is done with help of Sphinx4 technology.

2. Google API:

This part is divided into NLU parsing model, information extractor, natural language generation, semantics and word translator. The text-to-text translation

Process is done in this step. The string to be translated is supplied to the googleapi which in response provides with a corresponding translated string in the desired language.

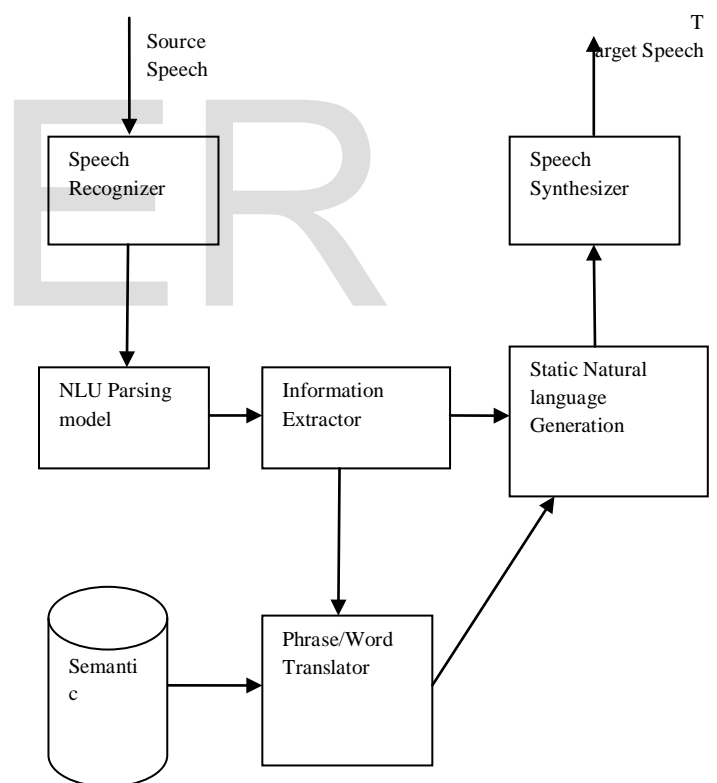


Fig. 1: System Architecture [9]

3. Speech synthesizer:

In speech synthesizing technic authors are going to use FreeTTS technology. The translated string or sentence obtained as an output from the translator is narrated by the computer in its corresponding language. The narrator can be configured in male as well as female voices as per the desire.

Algorithm:

Voice Messaging:-

Step1: Start

Step2: First import packages[5]

- package edu.cmu.sphinx.demo.helloworld;
- javax.swing.DefaultListModel;
- edu.cmu.sphinx.frontend.util.Microphone;
- edu.cmu.sphinx.recognizer.Recognizer;
- edu.cmu.sphinx.result.Result;
- edu.cmu.sphinx.util.props.ConfigurationManager;

Step3: Create class as EnglishRecognizer, Define cm object for ConfigurationManager;

Step4: if philosopher is equal to NULL then Call Englishhelloworld.config.xml and initialize it to cm object; Else Call Englishhelloworld.config.xml and initialize it to cm object;

Step5: if microphone does not start recording then Print message as "cannot start microphone";

Step6: while TRUE then Print message as "start speaking..."

Step7: Create object as result for storing recognized sentence.

If result is not equal to NULL then print recognized sentence.

Else print error message as I can't hear what you said.

Step8: Translate the generated English sentence to Hindi sentence.

Step9: At last synthesize the translated Hindi sentence.

Step10: Stop

2. RESULTS

Now, it is necessary to check the application for various conditions and on different platforms. Based on the behavior and results of the application authors have to create the performance table and its graphical representation, here they are using two examples for testing on different conditions;

Example1: Run Application on system having its inbuilt microphone.

Analysis of application:-

- 1) Install application on system.
- 2) Run application
- 3) Give voice input from inbuilt microphone of system
- 4) When users are passing voice input as "My name is Peter" there is possibility of error occurrence because of external noise and respected issues as a result it will not generate the expected outcome.
- 5) So passing voice input through inbuilt microphone of system (desktop) gives less accuracy.
- 6) According to the correctness of application authors create following graph shown in figure.

Table 1: Performance Test [3]

	Noisy Environment	Correctness
D1	0	4
D2	1	3
D3	2	2
D4	3	1

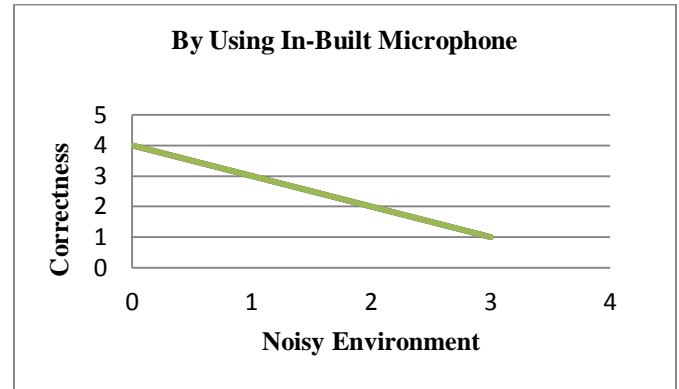


Fig. 2: Correctness vs. Noisy Environment.

Example2: Run Application on system by using external microphone.

Analysis of application:-

- 1) Install application on system.
- 2) Run application
- 3) Give voice input through external microphone which is connected to the system.
- 4) Now users have to pass voice input as "My name is Peter" then there is more possibility of correct output. Here authors conclude that[6] instead of using inbuilt microphone users can increase the accuracy of application by using external microphone.
- 5) So passing voice input through external microphone of system gives chances of more accuracy.
- 6) If users use this application in noise-free environment then it will give high accuracy.
- 7) According to correctness of application authors create following graph shown in figure.

Table 2: Performance Test based on Environment [7]

	Normal Environment (In-Built Microphone)	Normal Environment (External Microphone)	Noisy Environment (In-Built Microphone)	Noisy Environment (External Microphone)
D1	0.89	0.95	0.50	0.75
D2	0.60	0.75	0.40	0.55
D3	0.28	0.45	0.19	0.37
D4	0.20	0.30	0.10	0.15

Performance table [2] is drawn for Normal Environment and Noisy Environment, for two cases i.e. for in-built Microphone and External Microphone.

D1: Simple Sentence. E.g. "I ate".

D2: Medium sentence. E.g. "I am going to school by bus".

D3: Complex sentence. E.g. "Every morning, I go for walk in garden and do exercise".

D4: More complex sentence. E.g. "I pretend, I am ignoring you but inside I am dying for you".

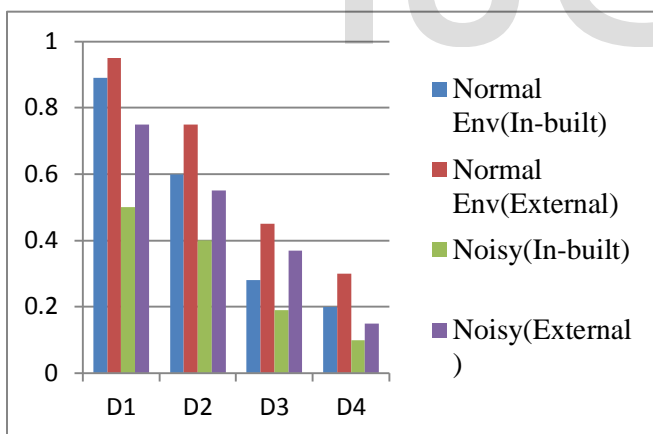


Fig. 3: Accuracy test for different sentences. [3]

This graph shows, there [3] are four different speakers with their unique speaking style. Let D1, D2, D3, D4 be four different speakers. D1, D2 speaks with less voice frequency so correctness is less as compare to D3, D4 which speaks with good voice frequency. Therefore the output which is required will change according to voice frequencies of different users.

Table 3: Performance Test [3]

	Voice Frequency	Correctness
D1	0.2	0.2
D2	0.4	0.4
D3	0.6	0.6
D4	0.8	0.8

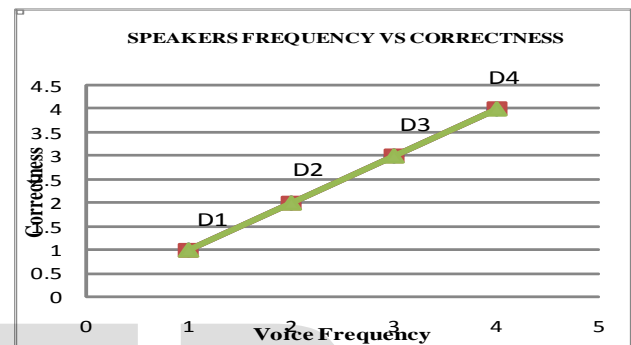


Fig.4: Correctness vs. Voice frequency [7]

3. APPLICATIONS

The translation of speech from one language to other will remove the bane of [6] language barrier for communication. The speech translation may prove beneficial even while online calling or video conferencing. A person speaking in English can be interpreted on the other side of the call in any language by speech translation. A person giving a conference or lecture or speech in some language can be understood by people in different languages simultaneously with the help of speech translation.

The speech translation feature is considered to be a boon for the tourism industry. Tourists visiting the country certainly don't know the local language [Hindi in most cases]. However they usually face problems in communicating [1] with local people. Speech translation feature can help convert English speech to Hindi speech and vice versa thus a fruitful two way communication can be established without any hindrance benefitting both the host and visitors. Speech translation can be used during video calling like: on Skype, where people of different regions are communicating and the language of communication is different, than the system will translate it

into their respected local language. It can also be used for learning different language just by translating any language into a particular language you want to learn.

4. CONCLUSION

Language has always been a barrier to effective communication. As businesses expand and technology engulfs the entire globe, reliable and real-time translation becomes imperative. While considerable progress has been done in this direction, more efforts need to be taken in order to reduce the enormous processing time involved with it. With this paper, authors propose a new system model to ensure effective real-time communication between two users who do not speak a common language while ensuring minimal computing time.

Authors studied different modules as well as performance of each module and tests for implementation of actual module. Finally, the optimized speech translation method is achieved.

Authors conclude that if users use this application in noisy Environment then conflict of words will occur and due to which there will not be a proper communication through voice translation. But if they use this application in Noise free environment then Voice translation is more accurate as compared to noisy environment.

If there are different users speaking with different voice frequency then it may affect the actual process of translation. If there are complex sentences for translation purpose it will require some time for the processing of sentence recognition and it will give some delay in output. Author's main purpose of this application is to overcome the language barrier of regional languages.

5. REFERENCES

1. Schlenker -Schulte, 1991; Perfetti et al. 2000 with respect to reading skills among deaf readers (Stinson et al.1999: Accuracy). Leitch et al.2002.
2. ZaidiRazak, Noor Jamaliah Ibrahim, EmranMohd Tamil, Mohd Yamani IdnaIdris, "Quarnic Verse Recitation Feature Extraction Using Mel-Frequency Cepstral Coefficient (Mfcc)" Department Of Al-Quran & Al-Hadith, Academy of Islamic Studies, University of Malaya.
- 3 D. R. Reddy, "An Approach to Computer speech Recognition by direct analysis of the speech wave", Tech. Report No.C549, Computer Science Department, Stanford University, sept.1996.
4. Cremer, Inge (1996): "Prüfungstexteverstehbargestalten", Hörgeschädigtenpädagogik 4, 50 Jahrgang, Sonderdruck.
5. Willie Walker, Paul Lamere, Philip Kwok, Bhiksha Raj, Rita Singh, EvandroGouvea, Peter Wolf, Joe Woelfel" Sphinx-4: A Flexible Open Source Framework for Speech recognition" SMLI TR2004-0811 c2004 SUN MICROSYSTEMS INC.

6. B Zhou, Y Gao, J Sorensen, et al. , "A hand held speech to speech translation system", Automatic Speech Recognition and Understanding, 2003.

7. IBM (2010) online IBM Research Source: -

<http://www.research.ibm.com/Viewed> 12 Jan 2010.

8. Y. Yan and E. Bernard, "An approach to automatic language identification based on language dependant phone recognition", ICASSP'95, vol.5, May.1995 p.3511.

9. L. R. Rabiner and B. H. jaung, Fundamentals of Speech Recognition Prentice-Hall, Englewood Cliff, New Jersey 1993.

10. AakashNayak, SantoshKhule, Anand More, AvinashYalgonde, Dr. Rajesh S. Prasad, "Study of various issues in voice translation", Vol2, No 2(2013): IJARCET February-2013.