ISSN:0975-9646
Naveenkumar T. Rudrappa et al / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 14 (2) , 2023, 16-19

HiTEK-Hybrid Software Engineering Approach for Noisy and Noiseless Speech: NLP

Naveenkumar T. Rudrappa[#1], Dr.Mallamma V. Reddy[*2], Dr.M Hanumanthappa[#3]

[#1,]*2Department of Computer Science, #3 Department of Computer Science and Applications,
Rani Channamma University Belagavi, Karnataka, India.
Jnanabharati Campus, Bangalore University, Bangalore, Karnataka, India.
[1]naveenkumartr@rcub.ac.in
[2]mvreddy@rcub.ac.in
[3]hanu6572@bub.ernet.in

***Abstract*— Natural Language Processing is a field of linguistics that consists of computer science, information engineering and artificial intelligence concerned with interactions between computers and human communication languages to process and analyze large amounts of natural language speech data. Acoustics is the study of sound its production, transmission and effects. Acoustic feature extraction is to analyze a speech signal by a predetermined number of components. Acoustic feature enhancement is to improve the quality of identified components of a speech signal. In this paper we use hybrid software engineering methods to facilitate extracting the acoustic features, converting noisy to a noiseless recorded wave file and enhance its quality for further pre-processing by developing a graphical user interface which has microphone as a recording aid and speaker as an output device. This interface involves a mixture of waterfall model and incremental development which will be used for cross language information retrieval of Hindi, Telugu, English and Kannada languages.**

***Keywords*— Acoustic, Artificial Intelligence, Cross Language Information Retrieval, Natural Language Processing.**

I. Introduction

The World Wide Web (WWW) is a rich source of information, growing at an enormous rate with an estimate of more than 29.7 billion pages as on February 2007. English is still the dominant language. However, statistics on global internet usage reveal that the number of non-English internet users is steadily on the rise which gives rise to a huge repository of information that can render data in regional languages. Web access to non-English internet users is a major challenge and hence an effort is made to solve this issue using Multi-Lingual Cross Language Information Retrieval (CLIR) [1] where voice pronunciation, storage, interpretation and retrieval play a major role. Speech is a natural means of communication among human beings and gives a very good platform for man-machine-man interaction. Translation is very important for human computer interactions as it allows environmental barriers to be removed for people who migrate from one state to another state for job, research work, transfer across states and for marketing of goods and services.

The present era is of human machine interaction, educationally backward and rural communities of India are being deprived of technologies that spread awareness of interconnected computers and communications. A good solution to this problem would be computers talking to a common man in a regional language of his choice as India owns language diversity, as per 2001 census India has 1599 languages, 122 major languages and 22 official languages in which some of them are Assamese, Bengali, English, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Tamil, Telugu, and Urdu [2, 3] as per 8th Schedule. These are the naturally spoken languages in India. NLP has recently gained much attention for representing and analyzing human language computationally. It has spread its applications in various fields such as acoustic, machine translation, information extraction, summarization and medical question and answering etc.

Natural Language Processing is a field of linguistics, computer science, information engineering and artificial intelligence concerned with interactions between computers and human communication languages, how to program computers to effectively process and analyze large amounts of natural language corpora.

Software engineering and natural language processing are connected to each other. They both are branches of computer science and engineering. [4] Natural language processing is the process carried out on computers for a variety of naturally used languages whereas software engineering is a branch of knowledge to develop software in a systematic form for a purpose. It is of a firm opinion by using SE tools and technique of one research area in the context of another, better software will be developed. NLP can be applied in the every phase of Software Development Life Cycle. Speech converted plain text can be used as input for various NLP tasks.

II. Terminologies Related To Speech and Text

The various terms and phrases related to spoken speech and text with reference to the consideration of Hindi, Telugu, English and Kannada (HiTEK) languages are listed and illustrated as follows:

1. Phonemes are basic building blocks of a language, the smallest part of a word that may cause a change of meaning.
2. Phone is a unit sound utterance.
3. Phonetics is a science that involves detailed analysis of

human speech and its perception.
4. Acoustic phonetics is learning of the physical transmission of speech sounds from the speaker to the listener.
5. Articulatory phonetics is the study of movement of speech organs.
6. Auditory phonetics is the study of speech reception and recognition of speech sounds by the listener.
7. Phonology is a branch of phonetics that studies phonemes and its pronunciation in a particular language.
8. Phonotactics is a branch of linguistics that govern the rules for possible phoneme sequences in a language.
9. Character Unigram is a unique single letter.
10. Character bigram is a unique two-letter chain sequence
11. Character trigram is a unique three-letter chain.
12. Character n-gram is a unique n-character long sequence of letters.
13. N-gram frequency: How frequently a n-gram chain repeats in some sample text. [5]

III. PROPOSED METHODOLOGY

Human being is a social animal. He has a unique capability to speak which other animals do not possess. He has to communicate to others either through speech or through hand gestures. The proposed methodology focuses on acoustics which is a study of human sound, its production, transmission and effects. Speech has seven characteristics namely articulation, pronunciation, disfluency, pause, pitch, speech rate [6] and rhythm. These characteristics differentiate human speech from other animal sounds. Storing and processing the human speech is a challenging because of the combination of hardware that is microphone for input, and pre-processing of a recorded wave file which involves multiple speech synthesis techniques.

The proposed methodology consists of both waterfall model and incremental development with a concatenative technique and framed a biresearch model called Software Engineering Methodology in Naturally Used Languages (SEMiNUL) which is mixture of Software engineering techniques and speech processing techniques. The various steps involved in the processing of speech file are as shown in figure 1:

1. Speech Recording is the process of generating spoken language by a human and capturing the same with the microphone in real time. It uses concatenative technique and its descendants for further pre-processing. [7, 8]

2. Speech Feature Extraction is to represent a speech signal by a predetermined number of components of a signal. Features are extracted using RASTA filtering technique which analyses the speech to a maximum precision. [9, 10]

3. Noise Removal and Feature Enhancement [11] is to remove or suppress the background noise which may be fan rotation, wind, vehicle movement sounds etc to improve the quality of identified components of a speech. The filtered speech signal is analysed for stationary background noise using Spectral Subtraction enhancement technique [12, 13, 14] so that the resulting speech is more natural and pleasant, stored in a wave file with .wav [15] file format.

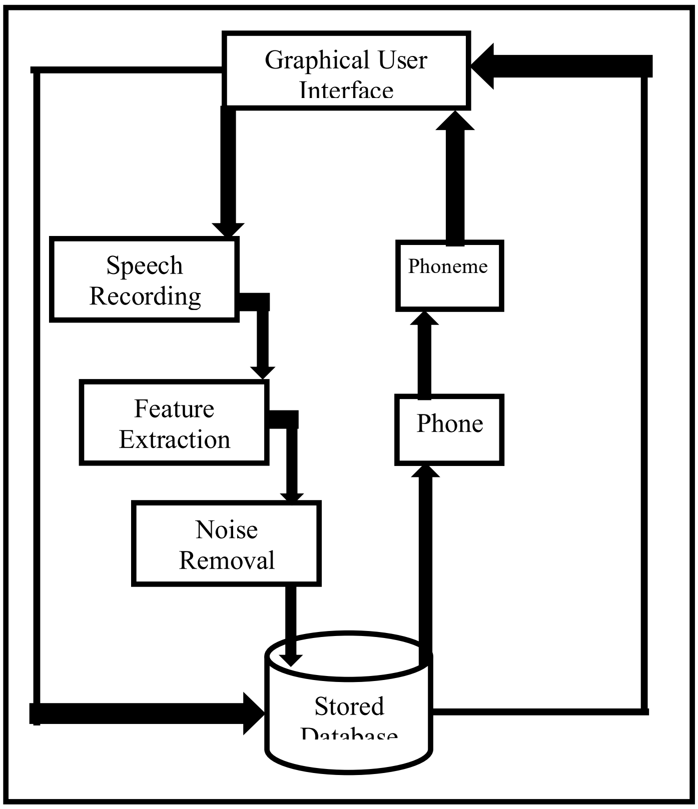

Fig.1: System Architecture for Speech Processing using SEMiNUL

The outcome of the proposed model is noise free speech signal which is further divided into appropriate phones by applying a stop word. The phones are compared with a stored database and if a match is found the phones are further divided into phonemes, unigram, bigram and trigram till we obtain individual phonemes iteratively. If the word is not present in the dictionary for example name of a person then the split up phones are concatenated and pronounced as it is. Once we obtain individual phonemes we concatenate them to produce appropriate phones which are pronounced out through the speaker.

The working of proposed model is depicted algorithmically as follows:

Algorithm:
Step 1: Record the voice and store in database
Step 2: Apply concatenative Synthesis
Step 3: Extract features of concatenated speech
Step 4: Apply enhancement to speech signal
Step 5: Check whether enhanced speech is found in the dictionary
Step 6: If match found, present the phone stop: goto step 7
Else: goto step 10
Step 7: Split up the speech into phones
Step 8: Split up the phones into phonemes

Step 9: Repeat step 7 and 8 until we get individual phonemes
Step 10: Output speech phones on speaker and store in database
Step 11: Exit

IV. EXPERIMENTAL SETUP AND RESULTS

To develop this application, the Software Requirements are Windows 10 Operating System, tools used are Anaconda 3, Jupyter with Python Programming. It uses front end as Jupyter IDE with Python programming language and in the back end a file system database to store, raw audio and processed audio file into the local drive with .wav file format for further processing. This application uses pyaudio library from python to activate microphone for input and speaker for output.

The process of implementation starts with creating a graphical user interface for recording the voice and to store it in a database. The stored file is used as an input for synthesizing, extracting features and enhancing speech by removing noise. The clean speech is compared with the dictionary until we obtain phones. These phones are repeatedly fragmented into phonemes, unigram, trigrams until individual phonemes are segregated. The phonemes so obtained are concatenated to obtain phones and are output on the speaker as well as stored as a processed new file.

The results are displayed using pyplot and matplotlib to plot a graph of input wav file and processed wav file, and compared. Results are as shown below:

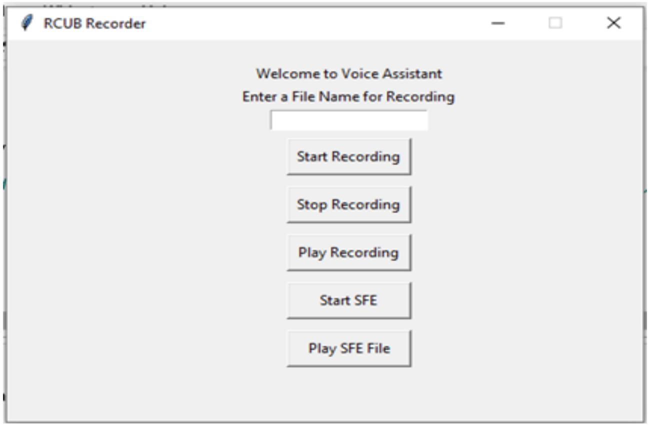

Fig. 2: Screen shot of Graphical User Interface

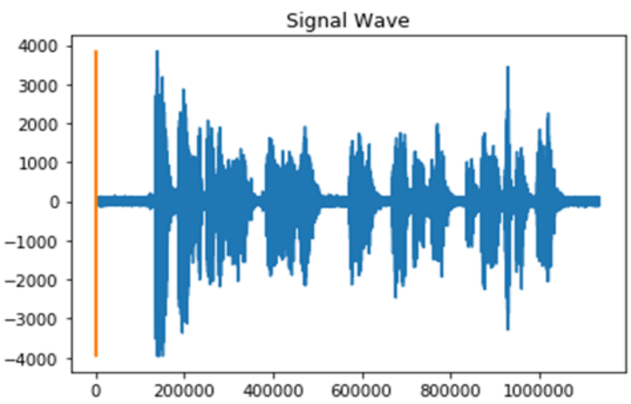

Fig. 3: Screen shot of Graph plot of recorded sound

V. CONCLUSION

This paper includes the basic information about the speech production, devices used for capturing human speech, transmission of wave file, storage and effects. The dictionary is useful to learn spoken languages which is a media for communication. The dictionary serves as a prototype to learn and spell Indian languages. The developed prototype will be used for cross language information retrieval of Hindi, Telugu, English and Kannada languages. The future work is associated with dictionary based phonetics, processing of phones and phone phrase identification.

ACKNOWLEDGMENT

I thank Rani Channamma University, Belagavi, Karnataka for providing a separate lab for Research Scholars and the necessary infrastructure to carry on with the research work. I would like to thank the NFST wing, Ministry of Tribal Affairs for selecting me as a Research Fellow and providing the necessary fellowship grant. I thank Dr. Mallamma V. Reddy for her constant support, mentoring and guidance.

REFERENCES

[1] Retrieval Jitendra Kumar, Sudha Shanker Prasad and Dinesh Kumar Prabhakar, Hindi Language based Cross Lingual Information Department of Computer Science and Engineering, Indian Institute of Technology(ISM), Dhanbad, Jharkhand, India Email- {jitendrakumar, sudha.shanker} @cse.ism.ac.in, dinesh.nitr@gmail.com Richa Gupta Department of Computer Science and Engineering, Indraprastha Institute of Information Technology, Delhi, India Email-guptaricha239@gmail.com, International Journal of Computer Science and Information Security (IJCSIS), Vol. 14, No. 10, October 2016
[2] Mallamma V. Reddy*, Hanumanthappa M. **, Jyothi N.M***, Rashmi S, Phonetic Dictionary for Natural Language Processing: Kannada, *(Department of Computer Science, Rani Channamma University, Vidyasangam, Belgaum-591156, India) ** (Department of Computer Science, Bangalore University, Jnanabharathi Campus, Bangalore-561156, India) ***(Department of Master of Computer Applications, Bapuji Institute of Engineering and Technology, Davangere-577004, India) Mallamma V. Reddy et al Int. Journal of Engineering Research and Applications www.ijera.com ISSN : 2248-9622, Vol. 4, Issue 7(Version 3), July 2014, pp.01-04
[3] https://en.wikipedia.org/wiki/Languages_of_India
[4] Ashwin CH , Akhila SP & Annapoorna Shetty Natural Language Processing In Software Engineering by International Journal of Latest Trends in Engineering and Technology Special Issue SACAIM 2017, pp. 496-498 e-ISSN:2278-621X
[5] Mallamma V. Reddy *, Hanumanthappa M **, Kannada Phonemes to Speech Dictionary: Statistical Approach, * Department of Computer Science, Rani Channamma University, Vidyasangam, Belagavi-591156, India ** Department of Computer Science, Bangalore University, Jnanabharathi Campus, Bangalore-560056, India Mallamma V. Reddy. Int. Journal of Engineering Research and Application www.ijera.com ISSN : 2248-9622, Vol. 7, Issue 1, (Part -4) January 2017, pp.77-80
[6] Judit B_ona, Temporal characteristics of speech: The effect of age and speech style, Department of Phonetics, E€otv€os Lor_and University, H-1088 Budapest, M_uzeum krt 4/A, Hungary bona.judit@btk.elte.hu, [http://dx.doi.org/10.1121/1.4885482]
[7] S. P. Panda, A. K. Nayak, and S. Patnaik, A Survey on Speech Synthesis Techniques, "Text to Speech Synthesis with an Indian Language Perspective", International Journal of Grid and Utility Computing, Inderscience, Vol. 6, No. 3/4, pp. 170-178, 2015
[8] Santosh K.Gaikwad, W.Gawali, Review on Speech Recognition Technique, Research Student Department of CS& IT Dr.Babasaheb

Ambedkar Marathwada University Aurangabad Bharti, Associate Professor Department of CS& IT, Dr.Babasaheb Ambedkar Marathwada University Aurangabad Pravin Yannawar Assistant Professor, Department of CS& IT Dr.Babasaheb Ambedkar Marathwada University Aurangabad, International Journal of Computer Applications (0975 – 8887) Volume 10– No.3, November 2010

[9] Shreya Narang[1], Ms. Divya Gupta[2], Speech Feature Extraction Techniques: A Review, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.3, March- 2015, pg. 107-114

[10] Manish P. Kesarkar, Feature Extraction For Speech Recogniton : (Roll No: 03307003), M.Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay, Submitted November2003

[11] David Ferris, Techniques and Challenges in Speech Synthesis, Thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Engineering in Electrical Engineering at The University of Newcastle, Australia.

[12] Ramesh Nuthakki, Speech Enhancement Techniques, 5th International Conference on Emerging Trends in Engineering,Technology, Science and Management, Institution of Electronics and Telecommunication Engineers, Ganganagar, Bengaluru Karnataka, India(ICETETSM-17) ISBN:978-93-86171-60-3

[13] Hynek Hermansky, Eric A. Wan and Carlos Evandano,Speech Enhancement based on temporal processing: Oregon Graduate Institute of Science and Technology, Department of Electrical Engineering and Applied Physics

[14] Speech Feature Smoothing For Robust Asr: Chia-Ping Chen Jeff Bilmes, Department of Electrical Engineering, University of Washington, Seattle, WA 98195-2500 {chiaping,bilmes}@ee.washington.edu, Daniel P. W. Ellis Department of Electrical Engineering, Columbia University New York 10027 dpwe@ee.columbia.edu

[15] A Technique to Encrypt-decrypt Stereo Wave File Jihad Nadir, Ashraf Abu Ein and Ziad Alqadi* Computer Engineering Department Albalqa Applied University Amman - Jordan *Email: natalia_maw [AT] yahoo.com, International Journal of Computer and Information Technology (ISSN: 2279 – 0764) Volume 05 – Issue 05, September 2016.