RESEARCH ARTICLE OPEN ACCESS

Towards understanding speech to text conversion for Ahirani (अहिराणी) language.

Pooja Shirude [1], **Mohit Chaudhari** [2], **Gaurav Baviskar** [3], **Mahesh Kanhere** [4]
Department Of Computer Engineering SSBT'S COET, BambhoriJalgaon, India

ABSTRACT

This paper represents an overview of Automatic Speech Recognition (ASR) for local ethnic language Ahirani, which is a mostly spoken language in the Khandesh region of Maharashtra State. Speech is the most powerful way of communication, with which people express their thoughts and feelings through different languages. With change in languages speech feature are also get change. However, even while communicating in the same language, the dialect varies with each person. This creates difficulty in acknowledge the carried message for some people. Sometimes lengthy messages or speeches are also quite difficult to follow due to reasons such as different accent, pace and so on. Speech recognition which is an integrative field of computational linguistics aids in improving technologies that empower the recognition and restating of speech into text. The research work presented in this report describes an easy and effective method of speech recognition. Extensive experimentation is performed to validate the productivity of the proposed method.

Keywords: - Ahirani, Automatic Speech Recognition (ASR), HMM, HTK, Isolated Word ASR, MFCC, Speaker Independent.

I. INTRODUCTION

Speech is the most important part of communication be express our thoughts and emotions, speech is considered as the main medium for communication. Speech recognition is the procedure of building a machine recognize the speech of different people based on certain words or phrases. It is a technology that allows a computer to recognize the words that a person speaks into a microphone. Speech recognition can be defined as the process of converting an acoustic signal, captured by a microphone, to a set of words [1][2]. Automatic speech recognition (ASR) is one of the fastest growing areas of engineering and technology. already many Automatic speech recognition systems are developed for English, Hindi, Marathi and other major languages which are majorly spoken in developed countries.

Automatic speech recognition systems are under development for Indian languages such as Hindi, Marathi, Tamil, Telugu, Bengali and Assamese. Rural languages like Ahirani are not been entirely explored to date. This work is aim to begin the work on designing and developing a speech recognition system for Ahirani. It is one of the most habitual languages spoken in Khandesh region mainly constitutes Dhulia, Jalgaon and Nandurbar districts. Automatic speech recognition systems have been perform using various toolkits and software.

Hidden Markov Model ToolKit, Sphinx toolkit, HMM Toolbox for Matlab etc. are one of the majorly used tools Among all these HTK toolkit is the most widely used tool to design Automatic Speech Recognition systems, Since it is used in building hidden Markov Models it has applications in other study areas as well. Hidden Markov Model ToolKit is well documented and provides guided tutorials for its use.

II. LITERATURE SURVEY

The pre-existing work which is similar to our work is represented in the literature survey section Speech to text conversion finds applications in various scenarios Ajay S. Patil implemented an experimental, speaker independent isolated word speech for the language Ahirani, using the Hidden Markov Model Toolkit (HTK) The system is trained on 20 Ahirani words by collecting data from 10 speakers and is tested using data collected from another 10 speakers in room environment.

The overall accuracy of their system is 94% and is speaker independent [1].

"A Model for Translation of English Words to Ahirani (Khandeshi) Words" Dr. Rajeev B. Kharat

The proposed model includes database of 1800 English words and their separate original words in Ahirani language the development of the model it was tested and feedback of 130 MCA scholars were taken where maturity of the scholars
i.e. 88 editorialized the model

Translation of English to Ahirani Language Uday 1) Chandrakant Patkar, 2) Dr. Suhas Haribhau Patil, 3) Dr. Prasadu Peddi.

In this paper system proposed a technique to translate the English to Ahirani language. The system use tokenization for dividing a string into several corridor. also by using POS tagging reads text in several languages and assigns part of speech to each word. After relating for each word, its exact transliteration and a proper translation in English to Ahirani language is done.

Kumar and Aggarwal (2011) built a speech recognition system for Hindi using HTK to recognize the isolated words using acoustic word model.

The system is trained for 30 Hindi words collected from eight speakers. Overall accuracy of their system is 94.63%[9].

Dua et al. implemented an isolated word Automatic Speech Recognition system (ASR) for Punjabi using HTK.

The system is trained for 115 Punjabi words collected from eight speakers and is tested using samples from six speakers.

The overall system performance is 95.63% and 94.08%[12].Saini et al. (2013) also built an ASR for Hindi using HTK that recognizes isolated words

and the system is trained for 113 Hindi words collected from nine speakers. The systems overall accuracy is 96.61% [13].

Gawali (2010) et al. developed isolated word recognition system using MFCC and DTW features for Marathi [8].

Our paper aims to discuss design and implementation

of a Ahirani isolated sentence recognizer consisting of 5 sentences and developed to work in speaker independent real time environments.[14]

III. DATASET

Our corpus contains samples from day-to-day life sentences including philosophy, literature, commentary on poetry and grammar. so that collecting a wide range of Ahirani vocabulary. The recordings were primarily collected with the help of volunteers, recording their speeches by using the Recording app on Android phones. Each of these speakers is native Ahirani speaker or bilingual speaker. The speakers in splits, train, validation, test, and out-of-domain test sets are disjoint. The domain of the training data-set primarily is a speech collection of day-to-day life sentences which any people can easily understand. We do not include verses in our current dataset, as modelling ASR systems for verses would require additional resources in terms of both the acoustic model and the language model fronts. The dataset we used is consists of Ahirani short sentences. The dataset contains 400 sound files of 5 different classes. Our task is to extract various features from these files and classify the corresponding audio files into respective categories.

IV. FEATURE EXTRACTION

The very basic step for a speech recognition system is Feature Extraction of the speech signal. where the audio signal is filtered in terms of various parameters known as the feature extraction technique. The goal of feature extraction is to extract a set of properties or features of an utterance that have acoustic correlations to the speech signal, that is parameters that can somehow be computed or estimated through the processing of the signal waveform. Such parameters are termed features. It includes the process of measuring some important feature or properties of the signal such as energy, amplitude or frequency response (i.e. signal measurement), amplify these measurements with some meaningful derived measurements of signal (i.e. parameterize the signals), and statically conditioning and mapping these numbers to form observation vectors. For the audio signal analysis, we have used feature extraction technique known as Mel Frequency Cepstral Coefficient (MFCC).

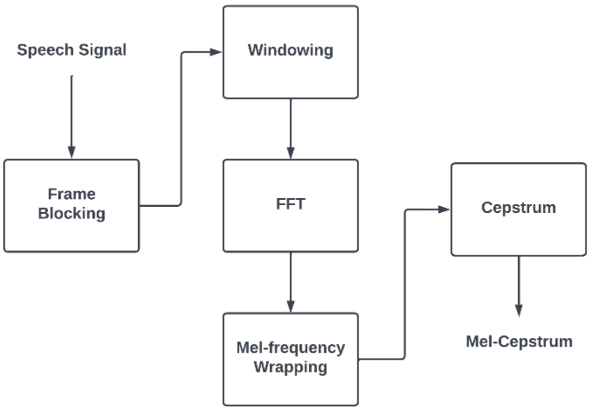

Figure : MFCC Extraction Process

To reduce noise in audio signal along with audio classification and identification of the speaker MFCC technique is used. The important parameter of speech signal in feature extraction method is Cepstral coefficients and pitch frequency, listed parameters are used for speech recognition, synthesis and verification of the speaker, etc. It summarizes the frequency distribution across the window size, to analyze both the frequency and time characteristics of the sound. The audio representation will allow to identify features for classification. In this extraction technique first, the speech is analyzed over short frame window. When the output of the FFT is passed through a Mel-filter then the Mel spectrum is obtained which performs the Cepstral coefficients.

V. EXPERIMENTAL RESULTS

Speech to text is basically a process in which audio signals are converted into textual format based on the dataset provided. We have used the audio classification method with the help of a Hidden Markov Model to classify to classify the sentences into the appropriate category. The audio sample provided to the model is first filtered and then features from that audio are extracted with the help of a MFCC technique which extracts features from the audio and based on these features extracted audio is classified in its appropriate category. Once classified label assigned to it is provided as a output.

Sample rate – Sample rates is basically number of audio samples recorded per unit second. 22050 is the default sample rate for reading the files using librosa. Depending on the libraries sample rate can vary..

2-D Array – Recorded samples of amplitude are shown by the 1st axis and number of channels are represented by the second axis respectively. The channels have two types– Monophonic and stereo. Monophonic has only one channel while the stereo has two channels.

Batch Size	Epoch	Accuracy %
110	38	46.8%
120	40	51.2%
140	50	64 %
160	60	79.8%
180	**80**	**82.8%**

Table: Accuracy Benchmark table

As shown in the table accuracy of the model varies as we change the batch size and number of epochs. We can see gradual increase in accuracy of model as batch size and number of epochs increases, Highest accuracy achieved till now is **82.8%** at the batch size of 180 and number of epochs is 80.

VI. CONCLUSION

We presented a new Speech Corpus for Ahirani Language along with the large vocabulary of it along with that developed the speech recognition system for Ahirani language. where the system will recognize the short sentences using acoustic word model. The training of the system will be done using the Ahirani sentences. During the development of the system, the training data will be collected from the four disjoint speakers, and accuracy achieved till date is 82.8%. The use of such an ASR solves the problem of technology acceptance in India, majorly in rural and remote parts by bringing human interaction closer to human-Machine interaction. In future we have plan to create more robust speech corpus for the Ahirani language which will cover the wide variety of sentences and words.

REFERENCES

[1] Ajay S. Patil, Automatic Speech Recognition for Ahirani Language Using Hidden Markov Model Toolkit (HTK) ,International Journal of Computer Science Trends and Technology (IJCST) – Volume 2 Issue 3, May-Jun 2014.

[2] Uday Chandrakant Patkar, Dr. Suhas Haribhau Patil , Dr. Prasadu Peddi , Translation of English to Ahirani Language, International Research Journal of Engineering and Technology (IRJET) Volume 2 Issue 6,— June 2020.

[3] Vinnarasu A., Deepa V. Jose, Speech to text conversion and summarization for effective understanding and documentation, International Journal of Electrical and Computer Engineering (IJECE) Vol. 9, No. 5, October 2019, pp. 3642 3648 ISSN: 2088-8708, DOI: 10.11591/ijece.v9i5.pp3642-3648.

[4] Sunil S. Nimbhore, Ghanshyam D. Ramteke, Rakesh J. Ramteke, Implementation of English-Text to Marathi- Speech (ETMS) Synthesizer, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278- 0661,p-ISSN: 2278-8727, Volume 17, Issue 1, Ver. VI (Jan – Feb. 2015), PP 34-43 www.iosrjournals.org.

[5] P. P. Shrishrimal, R. R. Deshmukh, Vishal B. Waghmare, DEVELOP MENT OF ISOLATED WORDS SPEECH DATABASE OF MARATHI WORDS FOR AGRICULTURE PURPOSE, Asian Journal of Computer Science And Information Technology volume 2 Issue 7 July 2012

[6] Dr. Rajeev B. Kharat, "A Model for Translation of English Words to Ahirani (Khandeshi) Words", International Journal of Latest Research in Humanities and Social Science (IJLRHSS) Volume 03 - Issue 02, 2020 www.ijlrhss.com —— PP. 46-48

[7] Devaraja Adiga1, Rishabh Kumar, Amrith Krishna, Preethi Jyothi, Ganesh Ramakrishnan, Pawan Goyal, Automatic Speech Recognition in Sanskrit: A New Speech Corpus and Modelling Insights, India.

[8] Gawali, Bharti W., Gaikwad, S., Yannawar P., Mehrotra Suresh C., Marathi Isolated Word Recognition System using MFCC and DTW Features (2010), Int. Conf. on Advances in Computer Science 2010, DOI: 02.ACS.2010. 01.73.

[9] Kuldeep Kumar, R. K. Aggarwal, Hindi Speech Recognition System using HTK, International Journal of Computing and Business Research, Volume 2 Issue 2 May 2011.

[10] Dr. Rajeev B. Kharat, A Model for Translation of English Words to Ahirani (Khandeshi) Words, International Journal of Latest Research in Humanities and Social Science (IJLRHSS), Volume 03 -Issue 02, 2020