

Analysis of the Error Pattern of HMM based Bangla ASR

Shourin R. Aura, Md. J. Rahimi, Oli L. Baroi

Ahsanullah University of Science and Technology, Dhaka, Bangladesh
Email: shourinrahman@gmail.com, mjrahimi@gmail.com, olilowna@gmail.com

Received: 04 November 2019; Accepted: 19 November 2019; Published: 08 February 2020

Abstract—Speech Recognition research has been ongoing for more than 80 years. Various attempts have been made to develop and improve speech recognition process around the world. Research on ASR by machine has attracted much attention over the last few decades. Bengali is largely spoken all over the world. There are lots of scopes yet to explore in the research regarding offline automatic Bangla speech recognition system. In our work, a moderate size speech corpus and a HMM based speech recognizer have been built to analyze the error pattern. Audio recordings have been collected from different persons in both quiet and noisy area. Live test has been carried out also to check the performance of the model individually. The percentage of the error and the percentage of correction with the created models are presented in this paper along with the results obtained during the live test. Finally, the results are analyzed to get the error pattern needed for future development.

Index Terms—ASR, HMM, HTK, MFCC, Speech Recognition, Dictionary

I. INTRODUCTION

Speech is the most natural form of communication. Scientists have long been trying to develop the system that can synthesize and recognize human speech. Speech recognition is the process of conversion of speech to text [1]. The idea of human machine interaction led to research in speech recognition. Automatic speech recognition uses the process and related technology for converting speech signals into a sequence of words or other linguistic units by means of algorithm implemented as a computer program. Speech recognition systems presently are capable of understanding speech input for vocabularies of thousands of words in operational environments [2]. Several research papers worked on performance evaluation using different feature extraction techniques like LPC, MFCC [11], [13] using different language corpus. Some documents even help to comprehend how different types of feature extraction techniques work [3], [12], [15] for speech recognition and how to mitigate the effect of noise [14].

The main purpose of this research work was to build an offline Bengali speech recognizer using HTK toolkit. Like other speech to text converter available online, we

tried to develop a completely standalone ASR using HTK toolkit which is very popular and reliable software in this field. The basic principle is identical with ASR of other languages but the set of phonemes used are unique. There still a lot to explore in the field of research regarding the development of a robust offline Bangla speech recognizer. In our work, after developing an offline Bangla speech recognizer we have mainly concentrated to check the performance of it through live test. The error pattern is also closely monitored to find out the way to improve the performance of the offline Bangla speech recognizer. The “Background” gives us a clear idea about the development of ASR technology and “current ASR technology and implications” gives information about recent ASR technology. In “Basic steps of speech recognition process” different steps of speech recognition technique have been discussed. “Methodology” discusses about the Bengali ASR technique. Finally, “Result” gives us the idea of the accuracy of this system.

II. BACKGROUND

During 1950’s, first significant attempts were made to develop techniques in ASR. It was based on direct conversion of speech signal into sequence. During late 1960’s Leonard Baum developed the mathematics of Markov chains at the institute of defense analysis. And then at CMU, Raj Reddy’s student James Baker and his wife Janet Baker began using HMM model for speech recognition for the first time. The use of HMM helped researchers to combine various source of knowledge such as: acoustics, language and syntax in a unified model which is based on probabilities. In the 1970’s, first positive outcome of spoken word recognition came into existence. In this meantime, general pattern matching techniques were discovered and applied [3].

In the present time, ASR can recognize thousands of words. Though Bangla is spoken by about 189 million people [5] and one of the largely spoken languages in the world, only a few works on Bangla ASR can be found in the literature, especially on Bangladeshi accented Bangla. Speech processing approach in Bengali language started only in 21st century. Many attempts have been made to develop various types of recognition software for Bengali languages. For example, there are available works on

Bengali digit recognizer, Bengali ASR for e-mail applications, employing recognizer in three layers back propagation Neural Network as the classifier. Recently Bengali speech recognition system is using HMM models and MFCC. We have delivered our efforts in improving the current Bengali ASR technology.

III. CURRENT ASR TECHNOLOGY

ASR in dialogue applications has some similarities with the other convention recognizers and also some key differences. The dialogue driven systems focus to design for the task the people will be performing when the system is being used. By treating the ASR as stable system designers are able to create toolkit level systems for the constructions of dialogue driven system. But if the designers opt for following the use of more diverse set of response then they need to add flexibilities to the system to support complex grammars, word spotting etc. which requires more complex design strategy.

A speech recognition system can operate in many different conditions. Interruptions are very common in human conversations. So, designers have developed barge in systems where the users can interrupt but the system can still process the speech. But it comes with a tradeoff sacrificing the system efficiency. We can overcome these problems by developing further researches.

IV. BASIC ELEMENTS OF SPEECH RECOGNITION PROCESS

Humans convert words to speech with their speech production mechanism. An ASR system converts speech from a recorded audio signal to text [6]. The basic steps in any speech recognition system includes corpora, HMM, MFCC, LPC and HTK toolkit etc.

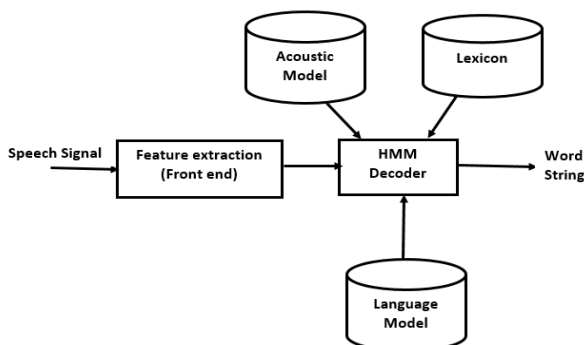


Fig. 1. Diagram of automatic speech recognition system.

Fig. 1 [7] shows the block diagram of a typical ASR system. It is composed of two major components: the front end and the HMM decoder. The front end block extracts spectrum representation of the speech waveform. The decoder block searches the best match of word sequences for the input acoustic features based on acoustic model, lexicon, and language model.

A. Corpus

Corpus is the building block of speech recognition or synthesis engine. It is a body of written or spoken material upon which a linguistic analysis is based. The plural form of Corpus is Corpora [8]. Corpora are only collection of text or speech that is used as basic statistical processing of natural language. The corpus may be open or closed, composed of written language, spoken language or both, mono or multilingual, tagged or untagged. British national corpora are the largest and the best-known corpora. Our corpus is developed using isolated Bengali words or Bengali speech synthesis and it is an over evolving corpus. We used the same corpus as Farzana Noshin Choudhury *et al* [10].

B. HMM

A hidden Markov model is a statistical Markov model. In this model the system being modeled is assumed to be a Markov process with unobserved or hidden states. Most of the speech recognizers are based on HMM. The reasons behind using it are- to model the short time stationary speech signal as Markov model, it can be trained automatically; they are simple and computationally feasible to use.

C. MFCC

The most important part of speech recognition system is feature extraction. The most used feature extraction methods in ASR technology are- MFCC, LPC, PLP and LDA. In our research we used MFCC. It is the most popular feature of extraction in speech recognition [9].

D. HTK toolkit

HTK refers to a portable software toolkit in order to build and manipulate systems that use continuous density Hidden Markov Models. Speech groups at Cambridge University Engineering Development Developed it. HTK is primarily designed to build HMM based speech processing tools. It performs a wide range of task in this domain which includes isolated or connected speech recognition. HTK includes 19 tools that perform task like manipulating transcripts, coding data, HMM training, Viterbi decoding and results analysis etc.

V. METHODOLOGY

This work is done by building recognizer with HTK toolkit using “MATLAB” for ASR. Different GUI like “wav to vec file converter”, “create Monophonic model”, “create Tri-phone model” are main steps to complete the work properly. A speech recognition technique requires a microphone for the person to speak into, speech recognition software, a computer to take and interpret the speech, a good quality sound card for input and output, a proper and good pronunciation, noisy environment etc.

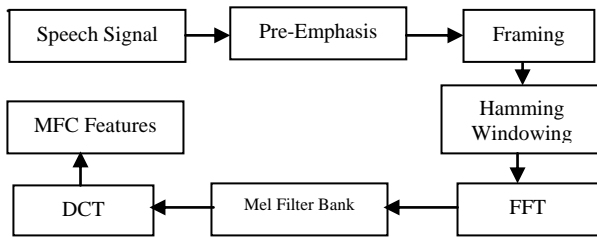


Fig. 2. Steps involved in MFCC feature extraction

In pre-emphasis, we use a high pass filter to give importance to high-frequency components. For exploring the speech signal over a short period of time, we frame the signals and then multiplied them by a Hamming window to minimize the discontinuities of the signal. Then Fast Fourier Transform (FFT) is used to convert speech signal from the time domain to the frequency domain. For more appropriate human hearing the signal further converted to Mel- frequency scale. This is done by using a set of triangular filters called Mel Filter Bank. In the next step the log Mel scale spectrum is converted to time domain using Discrete Cosine Transform (DCT). At the last step we get the MFCC Feature.

A. Building Bengali dictionary

Phonology is a branch of linguistics concerned with the systematic organization of sounds in languages. Each word of a language is basically made from some phonemes. The phonology of Bengali language is like that of its neighboring Eastern Indo-Aryan Language, characterized by a wide variety of diphthongs and inherent vowels. Bengali consists of 7 vowels and 29 consonants. In this research work we made 100 sentences using Bengali dictionary. There are 40 sentences consisting 2 words, 40 sentences consisting 3 words and finally 20 sentences consisting 4 words. These sentences are recorded by 5 different persons using “audacity” software and then they are converted into “WAV” files for training and testing.

B. Encoding data

Encoding data or acoustic analysis is a process in which the speech waveforms are transformed into a more compact and efficient form. Since the recognition tool can't handle the speech waveforms directly, so this step is of great importance in the recognition process. The analysis requires a stationary process to apply a short time DFT or “stDFT”. The speech signal consists both slow and fast variations which are separated by the Cepstral Co-efficient. Then the samples in the same frame (typically 20ms to 40ms duration) are multiplied by Hamming Function. Then using the HCopy HTK tool the original waveform is converted to a series of acoustical vectors.

C. Training the Model

For a fast and precise convergence of the training algorithm the HMM parameters must be initialized with the training data corpus. We did this using the tool HInit. The following command line initializes the HMM by

time-alignment of the training data with a Viterbi algorithm: HInit -A -D -T 1 -S trainlist.txt -M model/hmm0 -H model/proto/protohmm -l label -L label_dirhmm_name Hmm_name is the name of the HMM to initialize. “Prothom” is a description file containing the prototype. Trainlist.txt gives the complete list of the mfcc files forming the training corpus. label_dir is the dictionary where the label files (.lab) corresponding to the training corpus. label indicates which label segment must be used. model/hmm0 is the name of the dictionary where the resulting initialized HMM description will be output. Training procedure is done with the tool HRest. The following command line performs one re-estimation iteration with HTK tool HRest:

```
HRest -A -D -T 1 -S trainlist.txt -M model/hmmi -H model/hmmi-1/hmmfile -l label -L label_dirhmm_name
```

hmm-name is the name of the HMM to train. hmmfile is the description file of the HMM called hmm_name. Trainlist.txt gives the complete list of the mfcc files forming the training corpus. label_dir is the dictionary where the label files (.lab) corresponding to the training corpus. label indicates the label to use within the training data. model/hmmi indicates the index of the current iteration i.

D. Recognition

The recognizer is now complete and its performance can be evaluated. The recognition network and dictionary have already been constructed and test data has been recorded. Thus, all that is necessary is to run the recognizer.

1) *Recognition for noisy data:* At first the noisy database has been prepared. Recording was done in the cafeteria of a university which is an overcrowded place. Then the same procedure is followed for the noisy data.

2) *Running the Recognizer Live:* The built recognizer was tested with live input. Once the configuration file had been set-up for direct audio input the HTK tool Hvite was again used to recognize the live input using a microphone. Proper pronunciation and quiet & calm surrounding are very important for the live test. The speakers have to speak loudly & clearly at normal speed. The speakers have to try more than one time in order to get correct result.

VI. RESULTS

HTK tool was used to analyze the overall performances of the ASR speech recognition technique. An input speech signal was first transformed into a series of “acoustical vectors” using the HTK tool HCopy. The result was stored in a file known as test.scp. The input observation was then processed by a Viterbi algorithm, which matches it against the recognizer's Markov models using the HTK tool HVite. The output was stored in a file which contains the transcription of the input. All the results found are shown below:

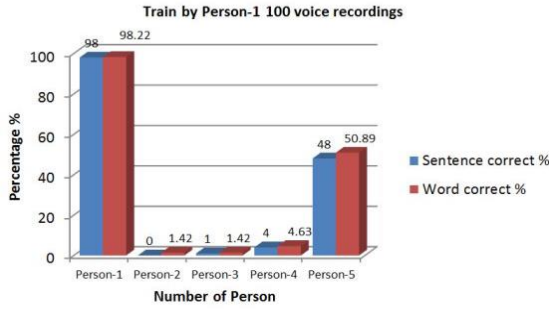


Fig. 3. Percentage of correction training by person-1

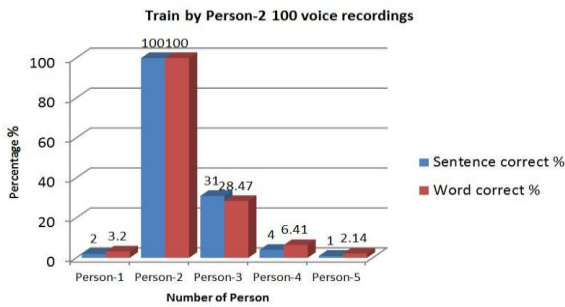


Fig. 4. Percentage of correction training by person-2

Here, we can see that testing with the own model gives more accuracy. Otherwise it gives a poor accuracy. The model which was trained by Person-4 gives better performance than the other three. Accuracy of word is higher than the accuracy of sentences in every case. Person's having similar voice pattern shows more accuracy. Output accuracy is very poor when female voices are tested for male voice model and vice versa. These models were trained by 500 voice recordings and tested by individual person. These give much more accuracy than the 100 voice recordings models. It gives almost 100% accuracy for every user. The accuracy of word is still higher than the accuracy of sentences. Speech Sample were also taken at noisy environment. No convincing results were found. It shows lower accuracy level than the non-noisy voice recordings.

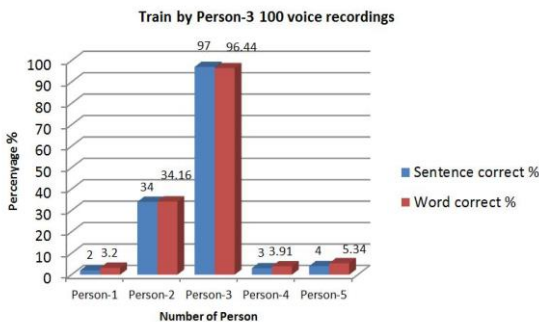


Fig. 5. Percentage of correction training by person-3

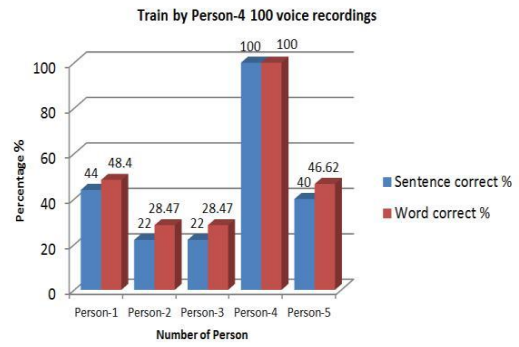


Fig. 6. Percentage of correction training by person-4

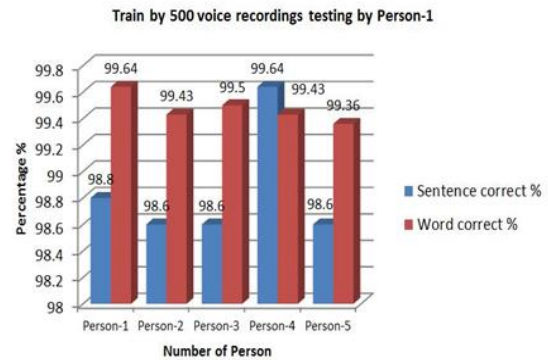


Fig. 7. Percentage of correction training by 500 voices testing by person-1

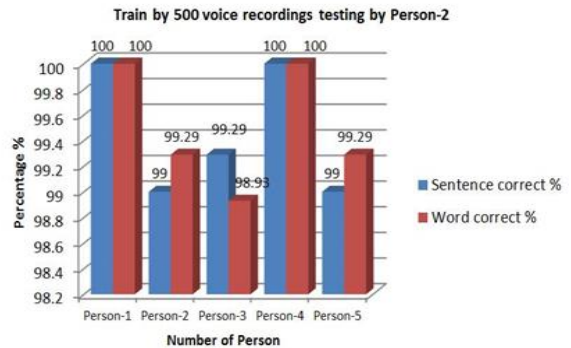


Fig. 8. Percentage of correction training by 500 voices testing by person-2

It seems noise affects the accuracy level as it can't detect the exact word or sentences because of unwanted signals.

Finally, we tested it live. In this case accuracy level is very low. Only a few percentages were correct. If it were done in a quiet environment, then we got less error. Very poor error occurs in large sentences. The error patterns are varied from person to person voice. Frequency, energy or loudness of the voice, also the pronunciation style can affect the error pattern. We found two types of error.

- Wrong: this type of output is always wrong.
- Erroneous: this type of output is not correct for the first time. We get the correct output within second or third attempt.

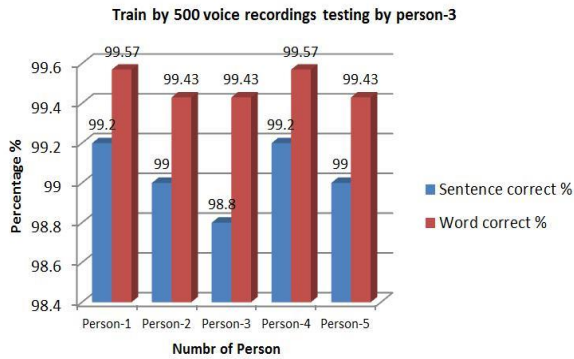


Fig. 9. Percentage of correction training by 500 voices testing by person-3

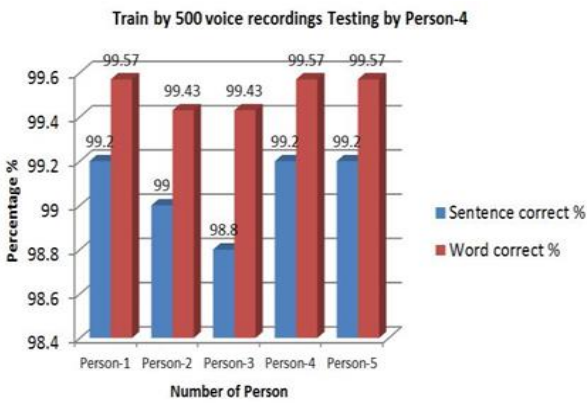


Fig. 10. Percentage of correction training by 500 voices testing by person-4

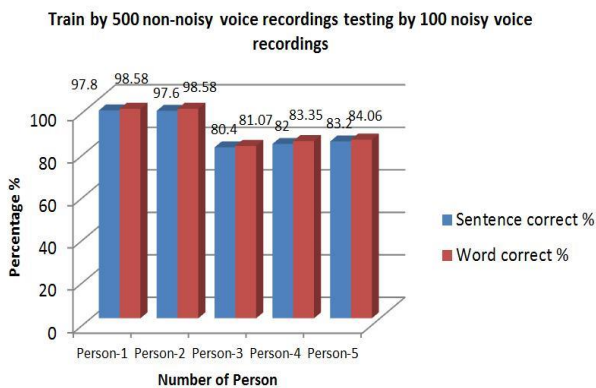


Fig. 11. Percentage of correction training by 500 voices testing by 100 noisy voice recordings

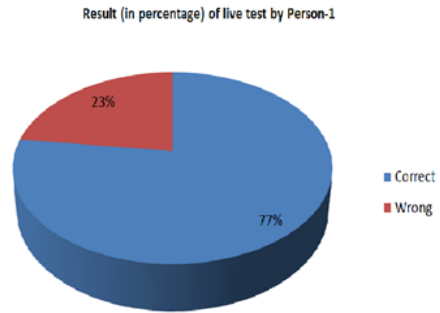


Fig. 12. Result of live test by person-1

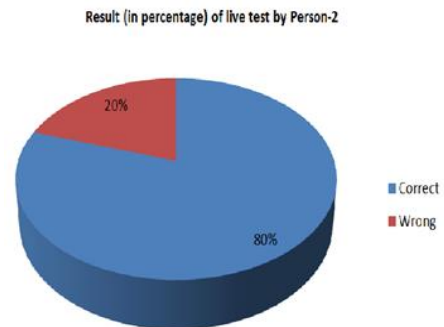


Fig. 13. Result of live test by person-2

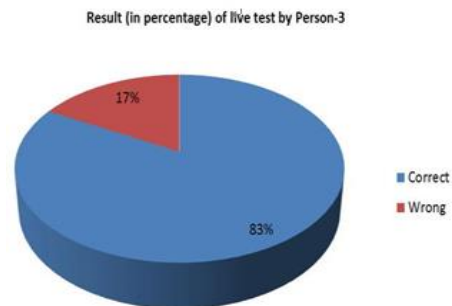


Fig. 14. Result of live test by person-3

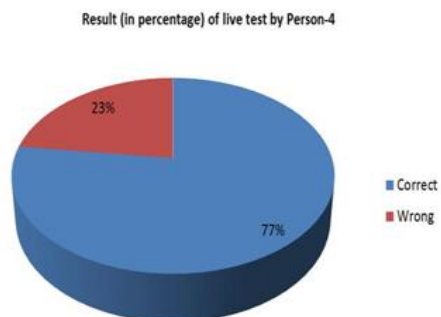


Fig. 15. Result of live test by person-4

Table 1. Error Found During Live Test Of Person-1 & Person-4

Original Sentences	Error Found	Error Type
NARI SHIKKHA NITI	NARI SHIKHKHA NITI	Spelling Error
AUSTRALIAY SHANTI AACHHE	AUSTRELIAY SHANTI AACHHE	Spelling Error
SHABEK BIRODHIDER SOBHA	SHABEK BIDROHIDER SOBHA	Word Error
BOSONTER UTSAB SURU	BOSONTER UTSAB SHURU	Spelling Error
SHIKHKHA BABOWSTHAY EGIYE	SHIKHKHA BAABOWSTHAY EGIYE	Spelling Error
PORIKKHAR JONNO SHOMORTON	PORIKKHAR JONNO SHOMORTON	Spelling Error
BOSHONTER LAL RONGE	BOSONTER LAL RONGE	Spelling Error
CHOTTOGRAM SOHOR KOTO BORO	CHOTTOGRAM SHOHOR KOTO BORO	Spelling Error
MILI AMAR CHELER MAA	MILI AMAR CHELER MAA	Spelling Error
BIDROHIDER DABI NIROBICHCHHINO BIDDUT	BIDROHIDER DABI NIROBICHCHHINNO BIDDUT	Spelling Error
NIRAPOD SHETU BANGLADESHER DABI	NIRAPOD SETU BANGLADESHER DABI	Spelling Error
NIRBHORJOGGO NIRBACHON MOORICHIKAR MOTO	NIRBHORJOGGO NIRBACHON MORICHIKAR MOTO	Spelling Error
TARA EKSATHE CHOTTOGRAM GECHHE	TARA EKSHATHE CHOTTOGRAM GECHHE	Spelling Error

Table 2. Error Found During Live Test By Person-2

ORIGINAL SENTENCES	ERROR FOUND	ERROR TYPE
AALOCHONA	#!MLF#! AALOCHONA	Spelling

AACHHE	AACHHE	error
MACHH UTPADON	JELA WOPOJELAI	Sentence error
NARI SHIKKHA NITI	NARI SHIKHKHA NITI	Spelling error
AUSTRALIAY SHANTI AACHHE	AUSTRELIAY SHANTI AACHHE	Spelling error
SHABEK BIRODHIDER SOBHA	SHABEK BIDROHIDER SOBHA	Word Error
BOSONTER UTSAB SURU	BOSONTER UTSAB SHURU	Spelling error
SHIKHKHA BABOWSTHAY EGIYE	SHIKHKHA BAABOWSTHAY EGIYE	Spelling error
PORIKKHAR JONNO SHOMORTON	PORIKKHAR JONNO SHOMORTON	Spelling error
BOSHONTER LAL RONGE	BOSONTER LAL RONGE	Spelling error
CHOTTOGRAM SOHOR KOTO BORO	CHOTTOGRAM SHOHOR KOTO BORO	Spelling error
MILI AMAR CHELER MA	MILI AMAR CHELER MAA	Spelling error
BIDROHIDER DABI NIROBICHCHHINO BIDDUT	BIDROHIDER DABI NIROBICHCHHINNO BIDDUT	Spelling error
NIRAPOD SHETU BANGLADESHER DABI	NIRAPOD SETU BANGLADESHER DABI	Spelling error
NIRBHORJOGGO NIRBACHON MOORICHIKAR MOTO	NIRBHORJOGGO NIRBACHON MORICHIKAR MOTO	Spelling error
TARA EKSATHE CHOTTOGRAM GHCHHE	TARA EKSHATHE CHOTTOGRAM GHCHHE	Spelling error
AUSTRALIAY SHANTI AACHHE	AUSTRELIAY SHANTI AACHHE	Spelling Error
SHABEK BIRODHIDER SOBHA	SHABEK BIDROHIDER SOBHA	Word Error
BOSONTER UTSAB SURU	BOSONTER UTSAB SHURU	Spelling Error
SHIKHKHA BABOWSTHAY EGIYE	SHIKHKHA BAABOWSTHAY EGIYE	Spelling Error

PORIKHHAR JONNO SHOMORTHON	PORIKKHAR JONNO SHOMORTHON	Spelling Error
BOSHONTER LAL RONGE	BOSONTER LAL RONGE	Spelling Error
CHOTTOGRAM SOHOR KOTO BORO	CHOTTOGRAM SHOHOR KOTO BORO	Spelling Error
MILI AMAR CHELER MA	MILI AMAR CHHELER MAA	Spelling Error
BIDROHIDER DABI NIROBICHCHHINO BIDDUT	BIDROHIDER DABI NIROBICHCHHINNO BIDDUT	Spelling Error
NIRAPOD SHETU BANGLADESHER DABI	NIRAPOD SETU BANGLADESHER DABI	Spelling Error
NIRBHORJOGGO NIRBACHON MOORICHIKAR MOTO	NIRBHORJOGGO NIRBACHON MORICHIKAR MOTO	Spelling Error
TARA EKSHATHE CHOTTOGRAM GECHHE	TARA EKSHATHE CHOTTOGRAM GECHHE	Spelling Error

Table 3. Error Found During Live Test Of Person-3

Original sentences	Error found	Error type
EKSHATHE GAIBE	DHAKAY AACHHE	Sentence Error
BANGLA VASHA	BANGLA AMAR VASHA	Sentence Error
NARI SHIKKHA NITI	NARI SHIKHKHA NITI	Spelling error
AUSTRALIAY SHANTI AACHHE	AUSTRELIAY SHANTI AACHHE	Spelling error
SHABEK BIRODHIDER SOBHA	SHABEK BIDROHIDER SOBHA	Word error
BOSONTER UTSAB SURU	BOSONTER UTSAB SHURU	Spelling error
SHIKHKHA BABOWSTHAY EGIYE	SHIKHKHA BAABOWSTHAY EGIYE	Spelling error
PORIKHHAR JONNO SHOMORTHON	PORIKKHAR JONNO SHOMORTHON	Spelling error
BOSHONTER LAL RONGE	BOSONTER LAL RONGE	Spelling error
CHOTTOGRAM SOHOR KOTO BORO	CHOTTOGRAM SHOHOR KOTO BORO	Spelling error
MILI AMAR CHELER MAA	MILI AMAR CHHELER MAA	Spelling error

BIDROHIDER DABI NIROBICHCHHINO BIDDUT	BIDROHIDER DABI NIROBICHCHHINNO BIDDUT	Spelling error
NIRAPOD SHETU BANGLADESHER DABI	NIRAPOD SETU BANGLADESHER DABI	Spelling error
NIRBHORJOGGO NIRBACHON MOORICHIKAR MOTO	NIRBHORJOGGO NIRBACHON MORICHIKAR MOTO	Spelling error
TARA EKSHATHE CHOTTOGRAM GECHHE	TARA EKSHATHE CHOTTOGRAM GECHHE	Spelling error

These are the errors found during live test. We can see that, errors found during the live test of Person-1 & Person-4 are exactly same. Person-5 also found the same errors except one sentence error. Most of the errors here are spelling errors. Very few are sentence or word error. The spelling errors are common for each person because it doesn't match with the dictionary we built.

VII. CONCLUSION

So, from the experiment, we can come to this conclusion that some error occurs due to the pronunciation differences, some for the dictionary we built before. We got some errors due to spelling mistakes. Otherwise the sentence was correct. Some errors were only for a word but not the whole sentence was incorrect. Error occurred in the large sentences more than the small sentences. We considered both the noisy and noiseless environment for live testing. In noisy environment, the rate of errors has increased.

Therefore, in future, we can try to observe the effects of the filter numbers as well as sampling frequencies and also can enrich our database for getting more practical output. If we can make a model with enrich dataset it will help to get more realistic result. Further by fixing the spelling error in the dictionary, errors can be minimized as we can see the spelling errors are almost same for every user. Finally, an adaptive filter can be used to minimize the errors due to noise. It will filter the unexpected sound if we set the frequency range of human voices. Moreover, we can also introduce more sophisticated method like Deep Learning Method for future development.

Speech recognition can be used to send instant messages, to annotate and to comment, to keep real-time transcripts during conversations. We can also use it in our real life such as health care for example the cripple can command to use the wheel-chair, hand free computing, home automation, voice control home security system. The work implemented in the paper is a step toward the development of Bangla language accessing on web very

fast by sending voice and getting the result in the text format. So, work further with ASR technology will open the door for wider applications in our real life.

ACKNOWLEDGEMENTS

The authors would like to thank Ahsanullah University of Science and Technology for supporting this work.

REFERENCES

- [1] Pelton, Gordon E. Voice Processing, *McGraw-Hill International Edition*, 1993.
- [2] Ganesh Tiwari, "Text Prompted Remote Speaker Authentication: Joint Speech and Speaker Recognition/Verification System".
- [3] Athiramenon.G, Anjusha.V.K, "Analysis of Feature Extraction Methods for Speech Recognition", In: IJSET International Journal of Innovative Science, Engineering and Technology, Vol. 4 Issue 4, April 2017.
- [4] Melanie Pinola, "Speech Recognition Through the Decades: How we Ended Up With Siri", PCWorld.
- [5] Website.[Online]Available: <http://aboutworldlanguages.com/bengali>
- [6] Gruhn R.E., Minker W., Nakamura S., "Statistical Pronunciation Modeling for Non-Native Speech Processing", Chapter -2 Available: <http://www.springer.com/978-3-642-19585-3>
- [7] Website [Online] Available: <http://what-when-how.com/video-search-engines/speech-recognition-audio-processing-video-search-engines/>
- [8] Natural Language Processing Website. [Online] Available: <http://language.worldofcomputing.net/about>
- [9] Eslam Mansour mohammed, Mohammed Sharafsayed, Abdalla Mohammed Moselhy and Abdelaziz Alsayed Abdelnaiem, "LPC and MFCC Performance Evaluation with Artificial Neural Network for Spoken Language Identification". In: International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 6, No. 3, pp. 55-56, June, 2013
- [10] Choudhury, Farzana & Maksud Shamma, Tasneem & Rafiq, Umana & Rahman Shuvo, Hasan & Alam, Shahnewaz (2016), "Development of Bengali Automatic Speech Recognizer and Analysis of Error Pattern". In: International Journal of Scientific and Engineering Research.7.58.
- [11] Arpit Aggarwal, Tanvi Sahay, Mahesh Chandra, "Performance Evaluation of Artificial Neural Networks for Isolated Hindi Digit Recognition with LPC and MFCC". In: 2015 International Conference on Advanced Computing and Communication Systems (ICACCS-2015), Jan. 05 – 07, 2015, Coimbatore, INDIA
- [12] Prabhakar V. Mhadse and Amol C. Wani, "Automation System using Speech Recognition for Speaker Dependency using MFCC," Proc. of the Intl. Conf. on Advances in Computing and Communication (ICACC-2013), pp. 75-79, 27-28 April 2013, Mumbai, Maharashtra, India.
- [13] Eslam Mansour Mohammed, Mohammed Sharaf Sayed, Abdallah Mohammed Moselhy and Abdelaziz Alsayed Abdelnaiem, "LPC and MFCC Performance Evaluation with Artificial Neural Network for Spoken Language Identification," IJSPIPPR, vol. 6, no. 3, pp. 55-66, June 2013.
- [14] Abdul Syafiq B Abdull Sukor, "SPEAKER IDENTIFICATION SYSTEM USING MFCC PROCEDURE AND NOISE REDUCTION METHOD". Available: <https://pdfs.semanticscholar.org/8312/c229a7ed6a8b8456490e1f831b53727f36f8.pdf>
- [15] Abdelmajid H. Mansour, Gafar Zen Alabdeen Salh, Khalid A. Mohammed, "Voice Recognition using Dynamic Time Warping and Mel-Frequency Cepstral Coefficients Algorithms". In: International Journal of Computer Applications (0975–8887) Vol. 116 –No. 2, April 2015

Authors' Profiles



Shourin Rahman Aura received her B.Sc (Hons) majoring in Electrical and Electronic from Ahsanullah University of Science and Technology (AUST) in 2016. She is currently working as a Lecturer in the Faculty of Electrical and Electronic Engineering at the same University. She has been pursuing her M.Sc degree since 2017 from AUST. Her research interests include Speech processing, communications, VLSI and Electronics.



Md. Jakaria Rahimi received his Electrical and Electronic from Bangladesh B.Sc (Hons) and M.Sc majoring in University of Engineering and Technology (BUET). He is currently working as an Assistant Professor at the Faculty of Electrical and Electronic Engineering at Ahsanullah University of Science and Technology. His research interests include Speech processing and Digital signal processing. He has been pursuing his Ph.D. degree since 2016 from BUET



Oli Lowna Baroi received her B.Sc (Hons) majoring in Electrical and Electronic from Ahsanullah University of Science and Technology in 2018. She is currently working as a Lecturer in the Faculty of Electrical and Electronic Engineering at the same University. Her research interests include Speech processing, communications, and Electronics.

How to cite this paper: Shourin R. Aura, Md. J. Rahimi, Oli L. Baroi, " Analysis of the Error Pattern of HMM based Bangla ASR", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.12, No.1, pp. 1-9, 2020.DOI: 10.5815/ijigsp.2020.01.01