

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/326722992>

Instant Bangla Speech to Text Conversion

Article in International Journal of Science and Research (IJSR) · June 2014

CITATIONS

0

READS

1,244

3 authors, including:



Nazia Hossain

Stamford University Bangladesh

6 PUBLICATIONS 2 CITATIONS

SEE PROFILE



Tarin Kazi

Stamford University Bangladesh

2 PUBLICATIONS 0 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Speaker Based Identificaion [View project](#)

Instant Bangla Speech to Text Conversion

Nazia Hossain¹, Dilruba Shareen², Tarin Kazi³

¹Khulna University of Engineering and Technology, Department of Computer Science & Engineering,
51, Shiddeswari road, Dhaka-1217, Bangladesh

²Khulna University of Engineering and Technology, Department of Computer Science & Engineering,
51, Shiddeswari road, Dhaka-1217, Bangladesh

³Ahsanullah University of Science & Technology, Department of Computer Science & Engineering
51, Shiddeswari road, Dhaka-1217, Bangladesh

Abstract: *Speech is natural mode of communication system for people. Among numerous languages, Bangla is the one which is spoken by about 250 million people around the world. However most of the computerized systems are user friendly only for the English language users. Hence, using computer is sometimes problematic for Bangla language users, especially when it comes to the part of writing. For this sake, a framework is introduced for writing Bangla words using voice command. An approach is presented for the activity that is used to improve the Bangla writing procedure. First, the speaker speaks over the microphone. Then the voice is processed instantly, that is, the spoken words or phrases are converted into text-form using speech engine. And then, the text formatted spoken words are matched into the phonemes or phrases which are stored in a XML file. Then using another XML file of corresponding letters for converting English text letters into Bangla, the spoken words are converted into Bangla text words. Thus, anyone can easily write Bangla by using voice. The main target of this paper is to build a framework that can write Bangla word in an editor using voice command.*

Keywords: Phonemes, Speech Recognition (SR), XML, Grammar generation, SAPI.

1. Introduction

Speech is the easiest communication system. So, if a system could be developed that could be operated using voice command, it would be more user friendly than today's traditional system. Likewise, if we could operate the computer for writing Bangla, it would be better for the people who suffer in writing Bangla in any text editor. Our research concerns this idea too.

People who intend to write Bangla in text editors provided by the computers such as Microsoft Word or Notepad often suffer from various problems. For examples, they have to install some software and maintain some rules and regulations which turn out to be an overhead. And, there are also some people who are not able to use their hands for their writing. Our research has also considered those people.

The objective of the research was to develop a software which would be used to write Bangla texts in various editors. Our target was – first, to recognize Bangla speech and to write Bangla words in any text editor using voice command; second, to use punctuation marks in sentences and finally to make an editor which would offer flexibility to user.

Speech recognition is the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words. The recognized words can be the final results or they can also serve as the input to further linguistic processing. To carry out these tasks we have used microphone, speech engine - Microsoft SDK (software development kit), SAPI (speech application programming interface) 5.1, Microsoft Visual Studio2008, and Bijoy2000 software.

2. Related Works

Bangla speech to text or text to speech research is not much satisfactory till date. A word separation algorithm was developed for Continuous speech Recognition. It compared noise energy and zero crossing with speech for limited words. The problem is that, it requires huge memory [1]

Some researchers detected speech pattern by comparing speech energy to noise energy, and by using stress information. [2]

Some have research works about speech segmentation which is the signal processing front-end. [3]

Another group of researchers suggested that speech is segmented into identifiable and meaningful phonemes, there are two types of segmentation: phonetic and syllable like. After segmentation, clustering has been used. [4]

Other researchers detected the word boundary of Bangla speech by detecting speech differences at the syllable boundary, amplitude, power and by the duration of the nucleus vowels. [5]

Most of the speech recognition researchers use HMM (Hidden Markov Model) for Isolated and Continuous Bangla speech recognition. [6]

3. Methodology

3.1 Voice Recording

Our first task for speech to text Bangla conversion is connecting microphone perfectly to provide voice command.

User speaks through the microphone. The analog sound wave coming from the microphone is changed into digital sound signal using sound card of the computer consequently. Secondly, to convert the sound signal into text phonemes, the user should have a speech engine, Microsoft SDK 5.1 likewise. The speech engine is for converting digital sound signal to text format. This text forms are written into text editor using keyboard handler function.

3.2 Grammar Generation

Grammar generation is the most challenging part in the process of speech recognition. This process involves deciding the fact that which words or phrases have actually been spoken. So, Grammar rules are used by speech recognition (SR) process to analyze human speech input. In this process, this attempts to understand what a person is saying. In this research, the grammars are represented in a XML file in text form for the upcoming voice command. In the case of dictation, the grammars are used to indicate some words that are likely to be spoken.

3.3 Speech Recognition Process

At first microphone receives the speech of the speaker, as a sound wave. This sound wave is converted into digital form by the sound card, and the digital signal is converted into phonemes by speech engine. On the other hand, the possibly spoken words are broken into phonemes and stored it into our XML files that is called grammar file. Consequently, the converted text form of the spoken words is matched to the phonemes or words of grammar file.

So, Grammar is the text form of the voice command. The final element of the speech engine is the grammar rules. The XML file where the list of possible spoken words or phrases is listed is like the following:

```
<?XML version="1.0" encoding="utf-8"? >
<GRAMMAR LANGID="445">
<RULE ID="WordRec"
  TOPLEVEL="ACTIVE">
<L>
<P>aa</P>
<P>aam</P>
<P>aami</P>
<P>aamra</P>
<L>
<RULE>
<GRAMMAR>
```

From this XML file, the matched words can be retrieved which are being spoken through the microphone. Here, "Bijoy" is the software which is used for recognizing Bangla letters from the English phonemes in the system. After getting the spoken words in English using speech engine, the program will try to find the corresponding Bangla letters for the individual word which are used to write Bangla, in the Bangla supported text editor from another xml file. The search is held here in letter by letter. When we need to write

‘আ’, in Bangla, it requires the two English letters ‘Av’ in “Bijoy” software, and this would correspond to the phoneme of ‘aa’ That is, when the search found ‘Av’ for ‘আ’, then the control goes to “keyboard event” function. The keyboard function would find the corresponding ASCII and HEX values of ‘Av’. Then it would give the expected Bangla text form in the text editor. The file is like the following,

```
<?XML version="1.0" encoding="utf-8"? >
<GRAMMAR>
<L>
< aa > Av </ aa >
< aam > Avg </ aam >
< aami > Avwg </ aami >
< aamra > Avgiv </ aamra >
<L>
<GRAMMAR>
```

This process is repeated and continuously finds and writes Bangla text until it finds sounds which are the words that are spoken. [8] & [9]

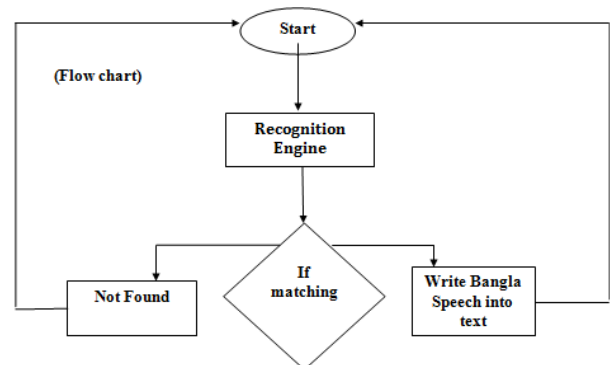


Figure 1: Flow Chart of Converting Bangla Speech into Text

4. Experimental Result

4.1 Sample Output

Here, we have made an environment which is illustrated in figure 2.

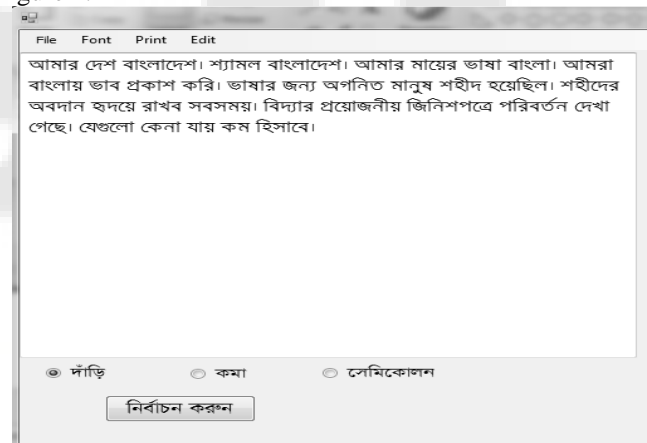


Figure 2: Writing Bangla speech into text

Therefore, when the user speaks through microphone, the speech is converted into text form instantly. Also, here are some punctuation marks such as 'full stop', 'comma' and 'semicolon' which could be used in the written sentences. Wherever the speaker intends to use these marks, they need to press the submit button which is 'Select' button; then the punctuation would be placed in the current position of the cursor.

5. Discussion

This research would help those people who are unable to write Bangla in any editor. By using this application anybody can easily write Bangla words within a short time. Therefore, it would save our time. This application would also reduce the overhead to write Bangla, and ease our life. Moreover, by using this editor, a user can write a document in Bangla as well as save it and open a new file for writing. One can also change the font, font size as well as the font color. In this research, the accuracy of Bangla text conversion should be more concise with respect to time. The next aim of this research is to introduce it as commercial software and build a system that would be helpful for the disabled or semi-disabled people. Especially, for the people, who could not write using hands. Moreover, this software would assist people in writing Bangla easily in computer system. In future, this research could be more developed and improved on different language conversion from English.

6. Acknowledgement

First of all we acknowledge to Almighty for completing this research successfully. Then we are grateful to Ms. Shaheena Sultana, assistant Professor, Department of Computer Science and Engineering, Khulna University of Engineering and Technology whose intelligent direction has made the task easier to accomplish. We wish to express our hearty and sincere gratitude to Nasrin Akhter, Assistant Professor, Stamford University Bangladesh, for her invaluable guidance and wholehearted cooperation. We would like to thank our parents, friends for their invaluable suggestions and critical review of our thesis.

References

- [1] Md. Abul Hasnat, Jabir Mowla, Mumit Khan "Isolated And Continuous Bangla Speech Recognition: Implementation, Performance And Application Perspective" Department Of Computer Science And Engineering, Conference Papers (Centre For Research On Bangla Language Processing), Brac University, 2007
- [2] Nipa Chowdhury(Duet), Md. Abdus Sattar(Buet), Arup Kanti Bishwas (Erl) "Separating Words From Continuous Bangla Speech", Global Journal of Computer Science & Technology, Vol 9, No 4, 2009
- [3] Basudev Kumar Sadhu, Rintu Roy "Bangla Writing Made Easy: A Voice Controlled Method" Department Of Computer Science & Engineering, Khulna University Of Engineering And Technology, Thesis/Project No: Cser-10-21, 2010
- [4] Md. Mijanur Rahman, Md. Farukuzzaman Khan and Mohammad Ali Moni (Pust), "Speech Recognition Front-End For Segmenting and Clustering Continuous Bangla Speech", Daffodil International University Journal of Science and Technology, Vol 5, Issue 1, 2010.
- [5] Firoj Alam, S.M. Murtoza Habib, Mumit Khan "Bangla Text To Speech Using Festival", Conference on Human Language Technology for Development, Alexandria, Egypt, May 2011
- [6] Shammur Absar Chowdhury "Implementation Of Speech Recognition System In Bangla", School of Engineering and Computer Science BRAC University, 2010
- [7] Szu-Chen Stan Jou "Automatic Speech Recognition On Vibrocervigraphic And Electromyographic Signals ", October 2008
- [8] Firoj Alam (Brac) "The First Text To Speech Synthesis for Bangla Language", Department of Computer Science & Technology, 2006.

Author Profile



Nazia Hossain has been received her B.Sc Engineering Degree from Khulna University of Engineering & Technology, Bangladesh. She is now full-time lecturer in Stamford University Bangladesh. And, continuing her graduate study in M.Sc. Engineering in Bangladesh University of Engineering and Technology



Dilruba Shareen has been received her B.Sc Engineering Degree from Khulna University of Engineering & Technology, Bangladesh. She is now full-time Software Engineer in M&H Informatics in Bangladesh.



Tarin Kazi has been received her B.Sc Engineering Degree from Ahsanullah University of Science & Technology, Bangladesh. She is now full-time lecturer in Stamford University Bangladesh.