Intelligent Personal Assistant with the comprehension of two languages simultaneously

Nahid Hossain   
Department of Computer Science  
Georgia State UniversityAtlanta, Georgia  
[nhossain5@student.gsu.edu](mailto:nhossain5@student.gsu.edu)

# Introduction

An Intelligent Personal Assistant (IPA) is a software, usually found on smartphones, that offers human assistance by answering questions and performing actions using natural language processing (NLP) and is supported by artificial intelligence (AI).[2] Some examples are Siri, Alexa, Cortana, and Echo. Most of these virtual assistants only understand English, if not, they have the capability of understanding Spanish. However, Google Assistant has the capability of understanding a couple more languages.

Bengali, also known as Bangla, is one of the most spoken languages in the world. It is ranked 7th among most spoken and it has about 265 million speakers, both native and non-native.[3] Bangladesh and India, where most of the native and non-native speakers are, have two of the world’s largest diaspora populations. They have 7.8 million and 17.5 million, respectively.[4] Most of these diasporas speak English as their second or even first language. Personally speaking, a combination of Bengali and English has seen a relevance within the diaspora communities. The combination of Bengali and English has been called Banglish. This word has seen some cases even in television shows back in South Asia.

The main goal of this research project is to make an IPA that can understand Banglish, which is the speaking of Bengali and English simultaneously. Not only is Banglish a combination of two languages, but it also takes words from one or the other and adjusts it with an ending of some sort. Here are some examples. Figure 1

|  |  |  |
| --- | --- | --- |
| English | Shower | Pray |
| Bengali | Ghusl | Namaz |
| With -ing | Showering/Ghusling | Praying/Namazing |

1. Examples of Banglish words

As you can see, the examples of the Banglish words tend to be Bengali words with the English ending -ing. There are many more words and of different endings (especially -ed), and it is planned to be added to the IPA’s dictionary along with the English and Bengali words.

# Proposed method

There has already been a team that has made an IPA that can understand some Bengali and English words.[1] I want to take it a step further by adding more words from Bengali, English, and Banglish. I also want the IPA to be able to become a Voice User Interface (VUI). This makes it such that the IPA can have voice output. This aspect is not seen in every IPA.[2] The team who have previously made the Bengali and English IPA, said that it is possible with some tweaking to make the IPA into a VUI. That is my goal along with adding Banglish words, which many do not know of unless they are from a diaspora community. By using methods, they have introduced and finding better ways to implement a voice output, a new generation of VUI may be made. This can be further developed by incorporating combinations of other languages as well as slang. Theoretically speaking, a VUI can comprehend three languages at the same time with much coding, but highly unlikely unless someone can speak three languages simultaneously.

# Research methodology

When I began creating the bilingual IPA, I needed to find the perfect speech recognition software. My main key factors were that it had to be open-source, available on Windows, and programming in Java was highly preferred. I finally came to the conclusion by picking CMU(Carnegie Mellon University) Sphinx.

CMU Sphinx has the ability to create models for languages that are currently not using the software. That is perfect for my project, because it’s main goal is to understand English, Bangla, and Banglish.

I used CMU’s we based tool, known as lmtool, which allows users to create two important text-based components which we are going to use for this project. One of the components is the lexical model and the other component is the language model.

The lexical model is basically the dictionary that the program will understand and the language model is like a list of combinations of words from the lexical model.

You can see the lexical model here:

##### Table i.

|  |  |
| --- | --- |
| **Word** | **Pronunciation** |
| BAJE | B EY JH |
| BANGLISH | B AE NG G AH L IH SH |
| BHONDHO | B HH AA N D HH OW |
| BYE | B AY |
| CHROME | K R OW M |
| CLOSE | K L OW S |
| CLOSE(2) | K L OW Z |
| DATE | D EY T |
| EDGE | EH JH |
| EXCEL | IH K S EH L |
| EXPLORER | IH K S P L AO R ER |
| INTERNET | IH N T ER N EH T |
| IS | IH Z |
| ITUNES | AY T UW N Z |
| KI | K IY |
| KOITA | K OY T AH |
| KORO | K AO R OW |
| KOTO | K OW T OW |
| KULO | K Y UW L OW |
| MEDIA | M IY D IY AH |
| NOTEPAD | N OW T P AE D |
| OPEN | OW P AH N |
| PAINT | P EY N T |
| PLAYER | P L EY ER |
| POWERPOINT | P AW ER P OY N T |
| SNIPPING | S N IH P IH NG |
| TA | T AA |
| TARIK | T AE R AH K |
| THE | DH AH |
| THE(2) | DH IY |
| TIME | T AY M |
| TOOL | T UW L |
| WHAT | W AH T |
| WHAT(2) | HH W AH T |
| WINDOWS | W IH N D OW Z |
| WORD | W ER D |
| WORDPAD | W ER D P AE D |

When I was coding the project, I had to configure the lexical model and language model. There is also something known as an acoustic model, but I used the one provided by CMU Sphinx. An acoustic model is essentially a huge dictionary.

I made 28 methods in the project. Some examples of methods include: Open “Program Name”, Close “Program Name”, What is the time, and What is the date. These commands can be said in all three languages that are being focused on. One command is not listed below and is called “Bye Banglish”, this closes the program with it outputting “See you later!” I also made it ask for the user to repeat their statement if the speech was not recognized. The program shows the users input in every case.

You can see the list of full commands here:

##### Table ii.

|  |  |  |  |
| --- | --- | --- | --- |
| **Command** | **English** | **Bangla** | **Banglish** |
| Opening Chrome | Open Chrome | Chrome Kulo | Chrome Open Koro |
| Closing Chrome | Close Chrome | Chrome Bhondho Koro | Chrome Close Koro |
| Opening Edge | Open Edge | Edge Kulo | Edge Open Koro |
| Closing Edge | Close Edge | Edge Bhondho Koro | Edge Close Koro |
| Opening iTunes | Open iTunes | iTunes Kulo | iTunes Open Koro |
| Closing iTunes | Close iTunes | iTunes Bhondho Koro | iTunes Close Koro |
| Opening Excel | Open Excel | Excel Kulo | Excel Open Koro |
| Closing Excel | Close Excel | Excel Bhondho Koro | Excel Close Koro |
| Opening PowerPoint | Open PowerPoint | PowerPoint Kulo | PowerPoint Open Koro |
| Closing PowerPoint | Close PowerPoint | PowerPoint Bhondho Koro | PowerPoint Close Koro |
| Opening Snipping Tool | Open Snipping Tool | Snipping Tool Kulo | Snipping Tool Open Koro |
| Closing Snipping Tool | Close Snipping Tool | Snipping Tool Bhondho Koro | Snipping Tool Close Koro |
| Opening Word | Open Word | Word Kulo | Word Open Koro |
| Closing Word | Close Word | Word Bhondho Koro | Word Close Koro |
| Opening Internet Explorer | Open Internet Explorer | Internet Explorer Kulo | Internet Explorer Open Koro |
| Closing Internet Explorer | Close Internet Explorer | Internet Explorer Bhondho Koro | Internet Explorer Close Koro |
| Opening Notepad | Open Notepad | Notepad Kulo | Notepad Open Koro |
| Closing Notepad | Close Notepad | Notepad Bhondho Koro | Notepad Close Koro |
| Opening Paint | Open Paint | Paint Kulo | Paint Open Koro |
| Closing Paint | Close Paint | Paint Bhondho Koro | Paint Close Koro |
| Opening Windows Media Player | Open Windows Media Player | Windows Media Player Kulo | Windows Media Player Open Koro |
| Closing Windows Media Player | Close Windows Media Player | Windows Media Player Bhondho Koro | Windows Media Player Close Koro |
| Opening Wordpad | Open Wordpad | Wordpad Kulo | Wordpad Open Koro |
| Closing Wordpad | Close Wordpad | Wordpad Bhondho Koro | Wordpad Close Koro |
| The time is: | What is the time | Koita Baje | Time ta ki |
| The date is: | What is the date | Koto Tarik | Date ta ki |

When the program runs each command, it reads out the left-most column. I used a speech synthesizer to make that possible. After adding this, the IPA has become a VUI. A speech synthesizer is the artificial version of a human voice. The speech synthesizer I used is called FreeTTS Speech Synthesis System. It is also handled by Carnegie Mellon University. If a user’s input is not understood the speech synthesizer does not say anything, because that would be redundant and very repetitive.

Overall, I made a speech assistant for personal computers that operate with the Windows Operating System. There were some errors while making some methods. For example, I could run open or close certain programs. Most programs that I could open though were created by Microsoft. Another issue I tackled was that it takes more time for the program to understand Bangla or Banglish compared to English. That is probably because my English transliteration for Bangla might have been slightly incorrect.

# Future Work

I recommend adding more methods that can be ran using Java, Windows, or the user’s preferences. I also recommend making a more appealing speech synthesizer or variants of such. One more important feature can be a pause or idle function. I noticed that the program is always going to pick up the microphone’s input unless the program is closed, so this is why a pause/idle feature would be important. Last but not least, adding more slang or acronyms of words can be helpful too. If an individual would like an even better program, they can make it such that the VUI asks questions back to the user for further clarification.

##### References

1. S. S. Chowdhury, A. Talukdar, A. Mahmud and T. Rahman, "Domain Specific Intelligent Personal Assistant with Bilingual Voice Command Processing," *TENCON 2018 - 2018 IEEE Region 10 Conference*, Jeju, Korea (South), 2018, pp. 0731-0734, doi: 10.1109/TENCON.2018.8650203.J. Clerk Maxwell, A Treatise on
2. J. Kurpansky, "What Is an Intelligent Digital Assistant?", Medium, 2017.
3. M. Hasan, “Bangla ranked at 7th among 100 most spoken languages worldwide,” *Dhaka Tribune*, 17-Feb-2020. [Online]. Available: https://www.dhakatribune.com/world/2020/02/17/bengali-ranked-at-7th-among-100-most-spoken-languages-worldwide. [Accessed: 29-Jan-2021].
4. A. Papadopoulos, “Countries with the largest diaspora population in the world, 2019,” *CEOWORLD magazine*, 18-Sep-2019. [Online]. Available: https://ceoworld.biz/2019/09/18/countries-with-the-largest-diaspora-population-in-the-world-2019/. [Accessed: 29-Jan-2021].