

BANO, S., JITHENDRA, P., NIHARIKA, G.L. and SIKHI, Y. 2020. Speech to text translation enabling multilingualism. In *Proceedings of the 2020 International conference for innovation in technology (INOCON 2020)*, 6-8 November 2020, Bangluru, India. Piscataway: IEEE [online]. Available from: <https://doi.org/10.1109/INOCON50539.2020.9298280>

Speech to text translation enabling multilingualism.

BANO, S., JITHENDRA, P., NIHARIKA, G.L. and SIKHI, Y.

2020

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses.

Speech to Text Translation enabling Multilingualism

Shahana Bano
Department Of CSE
Koneru Lakshmaiah Education
Foundation
Vaddeswaram, India
shahanabano@icloud.com

Pavuluri Jithendra
Department Of CSE
Koneru Lakshmaiah Education
Foundation
Vaddeswaram, India
jithendrapavuluri43@gmail.com

Gorsa Lakshmi Niharika
Department Of CSE
Koneru Lakshmaiah Education
Foundation
Vaddeswaram, India
niharikagorsa2000@gmail.com

Yalavarthi Sikhi
Department Of CSE
Koneru Lakshmaiah Education
Foundation
Vaddeswaram, India
ysikhi@gmail.com

Abstract— Speech acts as a barrier to communication between two individuals and helps them in expressing their feelings, thoughts, emotions, and ideologies among each other. The process of establishing a communicational interaction between the machine and mankind is known as Natural Language processing. Speech recognition aids in translating the spoken language into text. We have come up with a Speech Recognition model that converts the speech data given by the user as an input into the text format in his desired language. This model is developed by adding Multilingual features to the existent Google Speech Recognition model based on some of the natural language processing principles. The goal of this research is to build a speech recognition model that even facilitates an illiterate person to easily communicate with the computer system in his regional language.

Keywords— *SpeechRecognition, Tkinter, Recognizer, Multilingual text, Natural Language.*

I. INTRODUCTION

In many areas of research for processing the Natural Language, we were using Google, Siri, and many other inbuilt tools for conversion of Natural Language to text.

So, in our research, we have performed the conversions based on the references of all the existing speech recognition tools. To start our exploration to understand the working mechanism of speech recognition we have continued the entire research in python. This makes it easy for the conversion of Natural Language to multilingual text.

As we discussed earlier, language acts as a bridge between people considering it in mind, we have built a model that takes the input from the user who is willing to speak. This recognized speech data is recorded in the database and it is translated into the language that they select it to be displayed. This model is developed by adding Multi linguistic features to the existing Google Speech Recognition model based on some of the Natural Language processing principles. When you try to create a model that supports the speech recognition feature then you need to import some packages such as SpeechRecognition, PyAudio, Tkinter. This makes our model easy to recognize the user's voice. SpeechRecognition makes to recognize the speech, it depends on the utterances, the echo of the source, and one must speak more clearly to get recognized by the computer system. Once the speech is recognized then it breaks the sound and translates it into the desired text. So here we used the Tkinter package, to use this Tkinter package one must go to the prompt and try to install the package using the

command `pip install tkintertable`. This package allows us to obtain the pop-up window-based output.

So, when we take the input from the user the model translates all the speech data into the desired text. The Natural Language processing makes the computer system to recognize the speech and the imported packages help to develop the pop-up message box which contains text output.

II. PROCEDURE

A. Importing all the packages:

To make sure to run our model we need to install some speech recognition packages. This SpeechRecognition package has numerous amounts of inbuilt hidden classes. One among those classes is recognizer which makes them the system to recognize what the source is trying to tell. It breaks the sound with the frequency distribution and translates with the help of an instance by using various APIs that are available in the SpeechRecognition package. There are many APIs in the recognizer class, some of them can be used online and one among them is used when the computer system is working in offline mode. Here in our model, we have used google API which does translate into different languages that the user selects to translate.

Later we need to install the PyAudio package where it deals with microphone class. That records our speech using a microphone. When we use the listen method in our program it enables the model to record the audio of the user and it aids in translation in the later stage using the SpeechRecognize package.

For the whole purpose of creating output and the visualization of the entire model, we need to install the Tkinter package. This makes the output window to be more effective like a pop-up window which displays the desired text as an output.

B. Working Principle:

For the conversion of Speech into the desired Text. The input is taken as speech from the user, and then the user selects the stream of language for the conversion of speech into text. Later speech is broken into various vocal sounds and utterances that are made from the source. Then every sound with frequency is converted into text. Here for the recognition of the sound, we use the SpeechRecognition package which contains a class named recognizer. This method of recognizer class enables the model to recognize and helps in the conversion of sound to text with the help of an online tool that is google API which is used for

translations of different languages of speech. This model makes the user understand any spoken language by looking at the text output.

For the output window which displays the converted text, we used the Tkinter package which helps to display the pop-up window.

- Step 1: Start.
- Step 2: Need to select the language in which the speech is to be translated.
- Step 3: Place the microphone near to the person or device in which the speech input is going to be given.
- Step 4: If the source recognizes the speech then the model goes to the translation phase.
Else ask the user to speak clearly.
- Step 5: Speech Recognition package comes into the functionality.
- Step 6: Conversion/Translation of speech to text in the desired language of the user.
- Step 7: Displays the text.
- Step 8: Click done to finish with the translation.
- Step 9: Repeat the steps 1,2,3,4,5,6,7,8.
- Step 10: End.

III. FLOWCHART

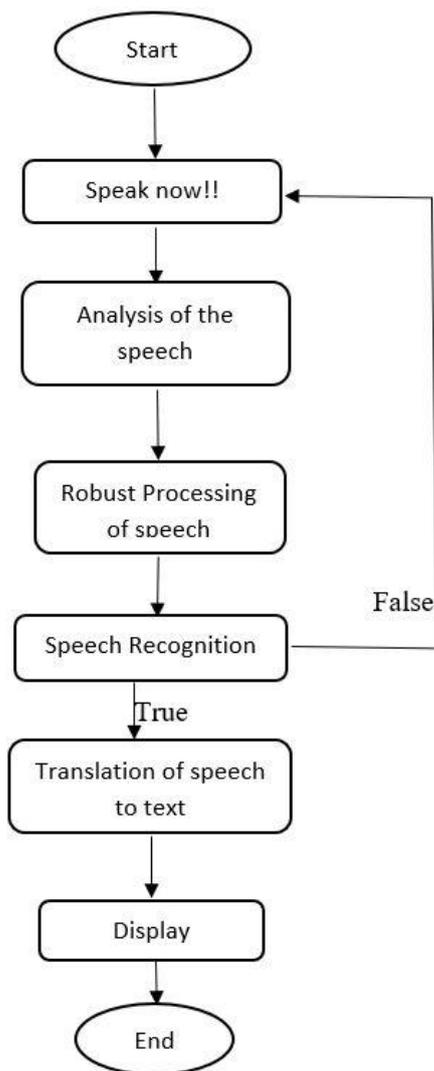


Fig. 1. Overview of the process.

The flow of our model starts with the run command. When you try to execute the command then a pop-up prompt appears with a list of languages. Then the user needs to select his desired language and click on done. Then another prompt appears asking the user to speak now! Then by using PyAudio and SpeechRecognition packages it analyses the speech and records it into the database. The robust processing of speech takes place and it recognizes the speech and goes to the next stage called translation else it leaves a message that the system doesn't recognize the audio. Then it repeats the steps from the starting. The later stage of recognition goes to the translation phase, in which the speech translation is done in the desired language selected by the user. It displays the text as output in the later stage of conversion or translation of the speech. Therefore, the end of the process of Speech to Text Translation. Fig.1.

IV. RESULTS

Step 1: A pop-up window appears on the screen and it allows the user to choose any one of the languages among the available languages. Fig.2.

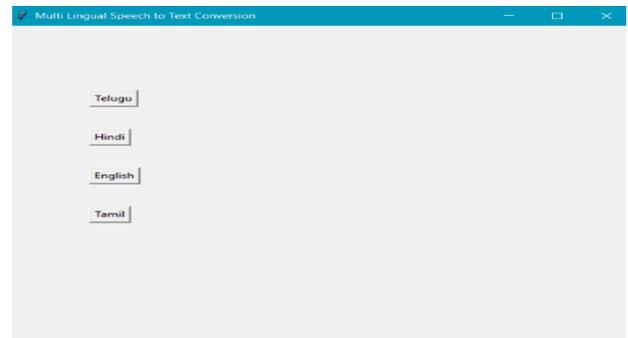


Fig. 2. Pop-up window to select the language.

Step 2: After choosing one of the languages another pop-up window appears on the screen and then the user should click on Ok and then he should start speaking. Fig.3.



Fig. 3. Speech Recognition.

Step 3: The user can go on speaking and when he stops speaking a pop window appears showing a Done message and again the user should click on OK. Fig.4.

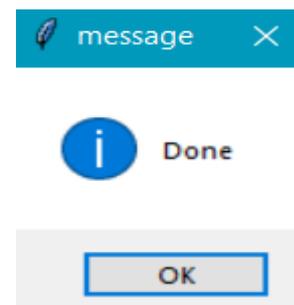


Fig. 4. Speech recognition is done.

Step 4: After clicking OK the user can see the output text message of what he has spoken through the microphone. These kinds of output pop-up windows shown in the figures are displayed only when we use the Tkinter package in the code.

So the SpeechRecognition package analyzes the speech provided to the system and later on it recognizes the speech data and gives us the output in the form of text. Fig. 5,6,7,8.

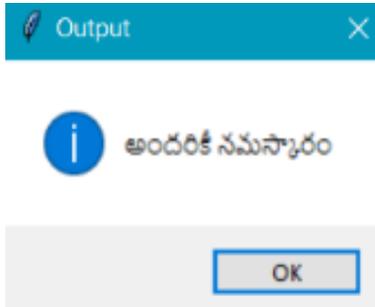


Fig. 5. Output in the Telugu Language.

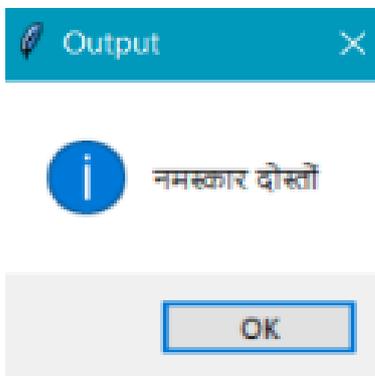


Fig. 6. Output in the Hindi Language.

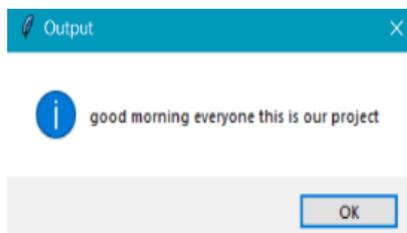


Fig. 7. Output in the English Language.

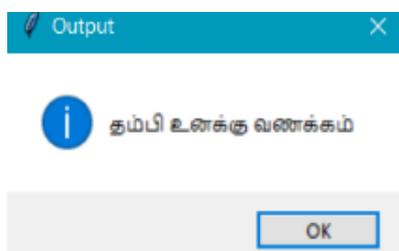


Fig. 8. Output in the Tamil Language.

V. CONCLUSION

By implementing this model we have learned how can we use the SpeechRecognition packages to build a Speech Translation model. The more we use these kinds of packages we get more flexibility in the code and output that is to be displayed. This model can be used in any purpose of translation of speech to text. This model has high advantages,

one among them is you can survive in unknown places where you don't know the language to speak but with help of this model you can translate that regional speech to text and it can also be used in areas like telecommunication and multimedia. Moreover, this model is useful for providing effective communication between man and the machine.

REFERENCES

- [1] Mrinalini Ket al: Hindi-English Speech-to-Speech Translation System for Travel Expressions, 2015 International Conference On Computation Of Power, Energy, Information And Communication.
- [2] Development and Application of Multilingual Speech Translation Satoshi Nakamura', Spoken Language Communication Research Group Project, National Institute of Information and Communications Technology, Japan.
- [3] Speech-to-Speech Translation: A Review, Mahak Dureja Department of CSE The NorthCap University, Gurgaon Sumanlata Gautam Department of CSE The NorthCap University, Gurgaon. International Journal of Computer Applications (0975 - 8887) Volume 129 - No.13, November2015.
- [4] Sequence-to-Sequence Models for Emphasis Speech Translation. Quoc Truong Do,Skriani Sakti; Sakriani Sakti; Satoshi Nakamura, 2018 IEEE/ACM
- [5] Olabe, J. C.; Santos, A.; Martínez, R.; Muñoz, E.; Martínez, M.; Quilis, A.; Bernstein, J., "Real time text-to speech conversion system for spanish," Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84. , vol.9, no., pp.85,87, Mar 1984.
- [6] Kavalier, R. et al., "A Dynamic Time Warp Integrated Circuit for a 1000-Word Recognition System", IEEE Journal of Solid-State Circuits, vol SC-22, NO 1, February 1987, pp 3-14.
- [7] Aggarwal, R. K. and Dave, M., "Acoustic modeling problem for automatic speech recognition system: advances and refinements (Part II)", International Journal of Speech Technology (2011) 14:309–320.
- [8] Ostendorf, M., Digalakis, V., & Kimball, O. A. (1996). "From HMM's to segment models: a unified view of stochastic modeling for speech recognition". IEEE Transactions on Speech and Audio Processing, 4(5), 360– 378.
- [9] Yasuhisa Fujii, Y., Yamamoto, K., Nakagawa, S., "AUTOMATIC SPEECH RECOGNITION USING HIDDEN CONDITIONAL NEURAL FIELDS", ICASSP 2011: P-5036-5039.
- [10] Mohamed, A. R., Dahl, G. E., and Hinton, G., "Acoustic Modelling using Deep Belief Networks", submitted to IEEE TRANS. On audio, speech, and language processing, 2010.
- [11] Sorensen, J., and Allauzen, C., "Unary data structures for Language Models", INTERSPEECH 2011.
- [12] Kain, A., Hosom, J. P., Ferguson, S. H., Bush, B., "Creating a speech corpus with semi-spontaneous, parallel conversational and clear speech", Tech Report: CSLU-11- 003, August 2011
- [13] Speech recognition for English to Indonesian translator using Hidden Markov Model. Hariz Zakka Muhammad, Muhammad Nasrun, Casi Setianingh, Muhammad Ary Murti, 2018 International Conference on Signals and Systems (ICSigSys).
- [14] Automatic Speech-Speech translation form of English language and translate into Tamil language. J Poornakala, A Maheshwari. International Journal of InnovativeResearch in Science.Vol. 5, Issu e3, March 2016 ... DOI:10.15680/IJRSET.2016.0503110. 3926.
- [15] Multilingual speech-to-speech translation system for mobile consumer devices.Seung Yun, Young-Jik Lee, Sang-Hun Kim.2014 IEEE Transaction on Consumer Electronics.
- [16] Speech Recognition using Python, Speech To Text Translation in Python, Python Training, Edureka, "https://www.youtube.com/watch?v=sHeJgKBaiAI&feature=youtu.be".
- [17] Leija, L.Santiago, S.Alvarado, C., "A System of text reading and translation to voice for blind persons," Engineering in Medicine and Biology Society, 1996.
- [18] R.Sproat, J. Hu, H. Chen, "Emu: An e-mail preprocessor for text-to-speech," Proc.IEEE Workshop on Multimedia Signal Proc., pp. 239-244, Dec. 1998.

- [19] Kimmu Parssinen, "Multilingual Text to Speech system for Mobile Device" University of technology April 2007.
- [20] M.T.Bala Murugan and M.Balaji , "SOPC-Based Speech-to-Text Conversion" National Institute of Technology, Trichy.
- [21] H.Ward, Jr.,"Hierarchical grouping to optimize an objective function," J.Amer. Statist. Assoc., vol.58, pp.236-244, 1963.