SPEECH TO TEXT AND TEXT TO SPEECH MODELS & METHODOLOGIES

[1]Md. Khaja Kutubuddin, [2]N. Lokeswari
[1]Student, [2]Assistant Professor
Department of Computer Science and Engineering,
Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam, India.

Abstract : Communication plays a major role in human life to share one's ideas or information to others. There is a need to develop technologies for blind and dumb people who cannot talk or see others to communicate with others. So the concept of speech to text and text to speech conversion comes into picture. However, there are many algorithms and methods we can use to process speech signals easily and recognize the speech and convert them to text. Here, we will discuss different techniques and algorithms that are applied to achieve the mentioned functionalities.

Keywords - Speech to text, text to speech recognition, Machine Translation.

1. Introduction:

Speech Synthesis is used for generation of text using the voice input. Speech Synthesis converts the input signal into waveform received from the speaker and then converts the input into phenomes, small words by processing the speech signals natural language techniques. In this process, various applications such as speech coding, speech synthesis, speech recognition and speaker recognition technologies, speech processing is used. In speech recognition, the generated speech wave is used to generate a set of words. Here we use Hidden Markov Model which is a stochastic model and issued to connect various states of transition with each other is majorly used. For converting speech to text, we have various algorithms like HMM model and neural n/w methods for generation of stream of words, recognizing them and speaker adaption.

2. Literature Review:

2.1 Text to Speech: For text to speech recognition, we first recognize the stream of characters one by one and the recognized stream is then used to create the neural network. For recognizing the characters , we collect the images of the characters and then store them in single vector. This data is saved as data file for training in neural network. We use feedforward neural n/w for increasing the accuracy of conversion with some weights for each node(frequency of each word) and biases initialized. The goal is assigned between 0.01 and 0.05. The created neural network has to be trained by adjusting weight and bias of network until the performance reaches to goal. Here we are using inbuilt feature of windows for conversion and firstly we check the condition that if win 32 SAPI is available in the computer or not. If it is not available then error will be generated and win 32 SAPI library should be installed on the computer. Then get the input string and extracts the voice by firstly select the voice which are available in library. Chooses the pace of voice and initializes the wave player to convert the text into speech.

2.2 Speech to Text : Basic speech recognition model follows the following steps.

i)Pre-processing: The analog signal is transformed into digital signals for later processing. This digital signal is moved to the first order filters flatten the signals. This helps in increasing the signals energy at a higher frequency.

ii) Feature Extraction: We generate the parameters of the input signals by computing the sequence of input signals. Commonly used feature extraction feature technique is Dynamic Time Warping.

iii) Accoustic Models : It is the fundamental part of Automated Speech Recognition system where a connection between the acoustic information and phonetics is established. Training reduces the inaccuracy of conversion by reducing the error generated during conversion by adjusting corresponding weights and bias and repeat the process until the error becomes negligible. The error generated may be due to noice in the input signals.

iv) Language Models : This model is used to find the probability of occurrence of each word using the frequency of that word in a sentence.

v) Pattern Classification : This is used to compare the error in expected outcome and actual outcome by comparing our input signal with a known signal and output. Different approaches for pattern matching are Template Based Approach, Knowledge Based approach, Neural network based approach.

3. Speech to Text Conversion Methods :

Speech to Text Conversion works similar to that of speech recognition. STT uses various techniques and some widely used conversion methods are discussed below.

i)Hidden Markov Model (HMM) : HMM is a statistical model used in speech recognition because a speech signal can be viewed as a piece wise stationery signal or a short time stationary signal. HMM, models are most effective for speech to text conversion. It depends on the following parameters.

Accuracy : Accuracy is measured on the basis of the difference between the given input pattern and a known pattern. The given input pattern will be compared with the output of the known test pattern and error produced must be negligible and the output must be independent of the speaker.

Speed of Recognition: The designed system must be efficient and must take less time for recognition as the users lose their interest. For more efficiency, the signals undergoes the pre-processing, HMM training and HMM Recognition in HMM model.

4. TEXT TO SPEECH CONVERSION :

Text to speech involves collection of voice input, analysis, speech processing and conversion of input signal into wave format and then convert into speech.

Text Processing : The input text is analyzed, normalized and transcribed into phonetic or linguistic representation.
There are many speech synthesis techniques like concatenative synthesis, Articulatory synthesis, Formant synthesis, Domain specific synthesis, Unit selection synthesis, Diphone synthesis, HMM bases synthesis etc…

Articulatory Synthesis depends on physical models based on the human speech system like vocal tract, tongue movement and so on. Diphone Synthesis uses a database of prerecorded voices and phrases and the output voice is created based on the pattern of the input signal.

Applications of Speech to Text & Text to Speech :

1. For blind and dumb people for communication.
2. Voice enabled websites which works on speech to text and text to speech engines.
3. When there is a need for human to machine interaction like in warning systems, clocks.
4. Taking speechnotes like Google Docs Voice Typing etc…

5. Conclusion : We covered various techniques and algorithms used for speech to text and text to speech and also their applications. The need of maintaining accuracy in conversion is evolving as there are many applications which depends on these API's. The HMM model is the best model for speech to text and for text to speech, we have many techniques but the formant synthesis is the most effective technique.

6. References :

[1]. Ms. Prachi khilari, Bhope V. , International Journal of Advanced Research in Computer Applications & Technology, Volume 4 Issue 7

[2]. Chaw Su Thu, Theingi Zin, International Journal of Engineering Research & Technology(IJERT) Vol:3 Issue 3, March-2014

[3]. Ayushi Trivedi, Navya Pant, IOSR Journal of Computer Engineering(IOSR-JCE) Volume 20, Issue 2, Ver 1 (Mar-Apr)

[4]. Kaladharan N. , International Journal of Computer Science & Information Security, vol 5, Issue 10, March 20

[5]. M.A. Anasuya, S.K. Katti, International Journal of Computer Science and Information Security Vol 6, No. 3, 2009.

[6]. Suhas R. Mache, Manasi R.Baheti, C. Namrata Mahendar, Review on Text-to-Speech Synthesizer, International Jouranal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 8, August 2015

[7]. Mouiad Fadiel Alawneh, Tengku Mohd Sembok Rule-Based Machine Translation Evaluation Methodologies, International Journal of Computer Applications Vol 121 No. 23, July 2015

[8].G.E. Dahl, D. Yu, L. Deng, A.Acero Large Vocabulary continuous speech recognition with context-dependant DBN-HMMs, In Proceedings of IEEE International Conference on Accoustics, Speech and Signal Processing(ICSSAP), pp.4688-4691,2011.

[9]. Pere Pujol Marsal, Susagna Pol Font, Astrid Hagen, H. Bourland and C. Nadeu, Comparison And Combination of Rasta-Pip and Ff Features in a Hybrid HMM/Mlp Speech Recognition System, Speech and Audio Processing, IEEE Transactions on Vol 13, Issue 1, 20 December 2004.