International Journal of Information Processing, 7(2), 82-89, 2013
ISSN : 0973-8215
IK International Publishing House Pvt. Ltd., New Delhi, India

Speaker Independent Continuous Speech to Text Converter for Mobile Application

R Sandanalakshmi[a], P Abinaya Viji[a], M Kiruthiga[a], M Manjari[a], A Sharina[a]

[a]Department of Electronics and Communication Engineering, Pondicherry Engineering College, Puducherry, India, Contact: sandanalakshmi@pec.edu.

An efficient speech to text converter for mobile application is presented in this work. The prime motive is to formulate a system which would give optimum performance in terms of complexity, accuracy, delay and memory requirements for mobile environment. The speech to text converter consists of two stages namely front-end analysis and pattern recognition. The front end analysis involves preprocessing and feature extraction. The traditional voice activity detection algorithms which track only energy cannot successfully identify potential speech from input because the unwanted part of the speech also has some energy and appears to be speech. In the proposed system, VAD that calculates energy of high frequency part separately as zero crossing rate to differentiate noise from speech is used. Mel Frequency Cepstral Coefficient (MFCC) is used as feature extraction method and Generalized Regression Neural Network is used as recognizer. MFCC provides low word error rate and better feature extraction. Neural Network improves the accuracy. Thus, a small database containing all possible syllable pronunciation of the user is sufficient to give recognition accuracy closer to 100%. Hence the proposed technique entertains realization of real time speaker independent applications like mobile phones, PDAs *etc.*.

Keywords : Feature Extraction, Neural Network, Speech to Text Converter.

1. INTRODUCTION

Speech Recognition(SR) is the translation of spoken words into text. It is also known as *Automatic Speech Recognition*, *ASR*, *Computer Speech Recognition*, *Speech To Text*, or *STT*. Speech is a natural mode of communication for people and it conveys some information. This information can be used for different purposes like authentication, text conversion or machine control depending upon application. People feel so comfortable to interact with computers *via* speech, rather than resorting to primitive interfaces such as keyboards and pointing devices. The need for speaker independent continuous speech to conversion system lies at the core of many rapidly growing application areas. A speaker independent system is intended for use by any speaker. Continuous speech means naturally spoken sentences, separated by minimum silence which is used for detecting boundaries. Continuous speech recognition is difficult when compared to Isolated words speech recognition. A speech interface would support many valuable applications like telephone directory assistance, spoken database querying for novice users, *handsbusy* applications in medicine or fieldwork, office dictation devices and for controlling electronic devices. Especially, it is useful for embedded systems like smart phones and PDAs having insufficient space for typing or touching and helpful for controlling navigation during car driving. Also, it can be used to build advanced security systems and ATM machines.

At present, there have been a number of successful commercial voice interfaces. The most prominent example is Siri, the voice-activated personal assistant built in the latest iphone. Speech recognition products are also available in Android, the Windows Phone platform and most other mobile systems with considerable limitations. The recognition accuracy and performance of a system would degrade dramatically with small modifications of speech sig-

put the complete word as text. In Transcription system, the duration of an input speech signal is made longer and the corresponding text output is obtained. The recognizer was trained for individual words. This system acts as a basis for real time speech recognition products, where input will never be a single word. Good recognition accuracy has been achieved in both cases as shown in Table 2. These implementations illustrate the potential of optimal configurations of key ASR components.

Theoretically, the accuracy increases with the increase in training data. As a result, memory needed also increases. It is found that carefully forming the database helps a lot in reducing memory requirements and increases recognition accuracy. In training phase, the words are spoken clearly so that it avoids general variations and confusions. In testing phase, the speech signal with minimum pauses should be given as an input. This enables the recognizer to discriminate the words effectively. For future work, the language modeling of HMM can also be utilized in neural network implementation to build an efficient hybrid HMM-NN recognizer including better acoustic modeling accuracy, better context sensitivity, more natural discrimination and a more economical use of parameters.

REFERENCES

1. M Marzinzik and B Kollmeir. Speech Pause Detection For Noise Spectrum Estimation By Tracking Power Envelope Dynamics, *in IEEE Transactions On Speech And Audio Processing, Barcelona*, 10(2):109–117, Feb.2002.
2. M H Moattar and M M Homayounpour. A simple but efficient real-time Voice Activity Detection Algorithm, *in Proceedings of 17th European Signal Processing Conference (EUSIPCO)*, pages 2549–2553, Aug. 2009.
3. Namgook Cho and Eun-Kyoung Kim. Enhanced Voice Activity Detection Using Acoustic Event Detection and Classification, *in IEEE Transactions On Consumer Electronics*, 57(1):196–202, Feb 2011.
4. Wei HAN, Cheong-Fat CHAN, Chiu-Sing CHOY and Kong-Pang PUN. An Efficient MFCC Extraction Method in Speech Recognition, *in IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 145–148, 2006.
5. A N Mishra, Mahesh Chandra, Astik Biswas, S N Sharan. Robust Features for Connected Hindi Digits Recognition, *in International Journal of Signal Processing, Image Processing and Pattern Recognition*, 4(2), June 2011.
6. Chin Luh Tan and Adznan Jantan. Digit Recognition using Neural Networks, *in Malaysian Journal Of Computer Science*, 17(2):40–54, Dec. 2004.
7. R L K Venkateswarlu, R Vasantha Kumari and G Vani Jayasri. Speech Recognition using Radial Basis Function Neural Network, *in Proceedings 3rd International Conference On Electronics Computer Technology (ICECT)*, 3:441–445, 2011.
8. Wouter Gevaert, Georgi Tsenov and Valeri Mladenov. Neural Networks Used For Speech Recognition, *in Journal of Automatic Control, University of Belgrade*, 20:1–7, 2010.
9. L K V Revada, V K Rambatla and K V N Ande. A Novel Approach to Speech Recognition By Using Generalized Regression Neural Networks, *in IJCSI International Journal of Computer Science Issues*, 8(2):484–489, March 2011.
10. Abderrahmane Amrouche and Jean Michel Rouvaen. Efficient System For Speech Recognition Using General Regression Neural Network, *World Academy of Science, Engineering and Technology*, 1(6):271–277, 2006.

Dr. R Sandanalakshmi is currently working as Assistant Professor Department of Electronics and communication Engineering, Pondicherry Engineering College. She has 12 years of teaching experience. Her research interests includes QoS improvement for next generation Wireless Networks, Non-invasive studies on prognosis of dengue using Signal Processing methods, Speech to Text Conversion for Mobile applications.

P Abinaya viji, M Kiruthiga, M Manjari, A Sharina completed Under Graduate B.Tech in the department of Electronics and Communication Engineering, Pondicherry Engineering College in April 2013 and placed in Core companies of specialization.