International Journal of Engineering & Technology, 7 (1.9) (2018) 268-272

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Real time device control over voice recognition in offline

C. Rukkumani [1]*, Dr. Krishna Mohanta. S [2], Govindaraj. S [1]

[1] *Research Scholar, Bharath Institute of Higher Education and Research*
[2] *Professor, Department of CSE, Kakatiya Institute of Technology and Science for women, Nizamabad, Telangana, India*
**Corresponding author E-mail:rukmanic@gmail.com*

Abstract

The consistent increase in the number and ownership population of mobile devices introduces a variety of limitations. A set of this limitations revolve around interactivity. The overly dependent haptic mechanism of interaction has caused device falls, slower time to interaction, health concerns, and limited support for the disabled among other problems. There is need to formulate innovative techniques that facilitate our interaction with these devices for users. In order to achieve this, a Real-time Voice Recognition Algorithm is formulated that lets users of mobile devices acquire freedom to move about and reduce the need for constantly glancing at their screen. This is achieved by allowing users to verbally command their devices to carry out ordinary tasks. An added unique feature is that it also offers offline access as any commands given by a user are processed and executed locally on the device.

Keywords: *Mobile Devices; Real Time; Voice Recognition; Screen.*

1. Introduction

The Smart phone market is a standout among the most focused markets on the planet today with different contenders, for example, Samsung, Apple, Asus, Google, Sony, Microsoft, and Research in Motion among others being secured a tight race to keep up or develop in the top position. At to begin with, it was about the devices" equipment yet now programming has emerged as one of the key determinants of a gadgets disappointment or achievement. One key programming that is making waves is the Mobile Digital Assistant. An application that permits one to associate with their gadgets through discourse. The most surely understood of these applications are Apple's Siri and Google Now. These applications permit one to achieve errands, for example, making telephone calls, setting cautions and checking for information, for example, climate. Despite the fact that these applications do take care of business, there are a couple of disadvantages. First off, a large portion of them frequently require that a Wi-Fi/versatile information association be available with the end goal for them to work. This is on the grounds that they first catch the user's discourse as the client verbally issues an order and after that continue to transfer this information to an online server which then translates the information and unravels the charge which is then sent back to the gadget lastly executed. Without web, the usefulness of such web subordinate applications is incredibly injured. These applications additionally require a high data transfer capacity in order to have the capacity to execute the users' charges in a convenient manner. In the event that this is not accessible, the applications have a tendency to work at a lazy pace and can regularly bother the client. There are likewise extra modules which are not accessible on the shut source arrangement sin the market that should be created. Also, these merchants frequently have an inclination to predominantly focus on the environment that backing their interests. Also, web association is exorbitant particularly in creating nations as well as it is a noteworthy worry as far as battery utilize particularly when imparted to other dynamic applications.

2. Related work

2.1. Automatic speech recognition: b. h. juang & lawrence r. rabiner

One of the problems faced in speech recognition is that the spoken word can be vastly altered by accents, dialects and mannerisms. In South Africa, there is a large variety of languages and dialects. Even the most basic speech recognition systems perform poorly when trying to recognise words spoken by English second language speakers.

Speech is the primary means of communication between people. For reasons ranging from technological curiosity about the mechanisms for mechanical realization of human speech capabilities, to the desire to automate simple tasks inherently requiring human-machine interactions, research in automatic speech recognition (and speech synthesis) by machine has attracted a great deal of attention over the past five decades. The desire for automation of simple tasks is not a modern phenomenon, but one that goes back more than one hundred years in history. By way of example, in 1881 Alexander Graham Bell, his cousin Chichester Bell and Charles Sumner Tainter invented a recording device that used a rotating cylinder with a wax coating on which up-and-down grooves could be cut by a stylus, which responded to incoming sound pressure (in much the same way as a microphone that Bell invented earlier for use with the telephone). Based on this invention, Bell and Tainter formed the Volta Graphophone Co. in 1888 in order to manufacture machines for the recording and reproduction of sound in office environments. The American Graphophone Co., which later became the Columbia Graphophone Co., acquired the patent in 1907 and trademarked the term "Dictaphone." Just about the same time, Thomas Edison invented the phonograph using a tinfoil based cylinder, which was subsequently adapted to

wax, and developed the "Ediphone" to compete directly with Columbia. The purpose of these products was to record dictation of notes and letters for a secretary (likely in a large pool that offered the service as shown in Figure 1) who would later type them out (offline), thereby circumventing the need for costly stenographers. This turn-of-the-century concept of "office mechanization" spawned a range of electric and electronic implements and improvements, including the electric typewriter, which changed the face of office automation in the mid-part of the twentieth century. It does not take much imagination to envision the obvious interest in creating an "automatic typewriter" that could directly respond to and transcribe a human's voice without having to deal with the annoyance of recording and handling the speech on wax cylinders or other recording media.

2.2. An efficient speech recognition system, suma swamy1 and k. v ramakrishnan

This paper describes the development of an efficient speech recognition system using different techniques such as Mel Frequency Cepstrum Coefficients (MFCC), Vector Quantization (VQ) and Hidden Markov Model (HMM). This paper explains how speaker recognition followed by speech recognition is used to recognize the speech faster, efficiently and accurately. MFCC is used to extract the characteristics from the input speech signal with respect to a particular word uttered by a particular speaker. Then HMM is used on Quantized feature vectors to identify the word by evaluating the maximum log likelihood values for the spoken word

The idea of human machine interaction led to research in Speech recognition. Automatic speech recognition uses the process and related technology for converting speech signals into a sequence of words or other linguistic units by means of an algorithm implemented as a computer program. Speech understanding systems presently are capable of understanding speech input for vocabularies of thousands of words in operational environments. Speech signal conveys two important types of information: (a) speech content and (b) The speaker identity. Speech recognisers aim to extract the lexical information from the speech signal independently of the speaker by reducing the inter-speaker variability. Speaker recognition is concerned with extracting the identity of the person. [3] Speaker identification allows the use of uttered speech to verify the speaker's identity and control access to secure services. Speech Recognition offers greater freedom to employ the physically handicapped in several applications like manufacturing processes, medicine and telephone network. Figure 1 (a) shows the speech recognition system without speaker identification. Figure 1 (b) shows how the speaker identification followed by speech recognition improves the efficiency. With this approach, the database will be divided into smaller divisions (SP1 to SPn) with respect to different speakers. Hence the speech recognition rate improves for the corresponding speaker.

2.3. Large vocabulary speech recognition system, yasuhisa fujii, kazumasa yamamoto, seiichi nakagawa

In this paper, we describe large vocabulary Continuous Speech Recognition (LVCSR) system SPOJUS++ which has been developed in our laboratory for over 20 years and recently fully reimplemented from scratch. SPOJUS++ employs a context-dependent Hidden Markov Model (HMM) as an acoustic model and an N-gram model as a language model to decode speech. SPOJUS++ has many novel features including a dynamic expansion of linear dictionary. Also, SPOJUS++ can construct a confusion network which leads to word error rate minimization recognition. Constructed confusion networks can be used in many kinds of post-processing applications which require automatic speech recognition results. We evaluated SPOJUS++ in terms of word accuracy, real time factor and search error. Experimental results showed that SPOJUS++ is comparable to state-of-the-arts.

2.4. Speech recognition based system to control electrical appliances arvinder singh, gagandeep singh

An important pre-processing step in Automatic Speech Recognition systems is to detect the presence of noise. It has been shown that accurate speech endpoint detection improves the isolated word recognition accuracy. Also, proper location of regions of speech reduces the amount of processing.

The objective of this Project is the development of "Speech recognition based system to control electrical appliances" and the analysis techniques that would provide sharply improved speech recognition accuracy in any type of noisy environments. Speech is a natural medium of communication for humans, and in the last decade various speech technologies like automatic speech recognition (ASR), voice response systems etc. have considerably matured. Moreover in robotics the ASR is sharply arranging its place from the last pair of decades as it is the only one medium by which these human-like robots are getting converted in the humanoids. The above systems rely on the clarity of the captured speech but many of the real-world environments include noise and others that mitigate the system performance.

2.5. Template based continuous speech recognition

Mathias De Wachter, Mike Matton, Kris Demuynck, Patrick Wambacq, Member, IEEE, Ronald Cools and Dirk Van Compernolle, Member

Despite their known weaknesses, Hidden Markov Models have been the dominant technique for acoustic modeling in speech recognition for over two decades. Still, the advances in the HMM framework have not solved its key problems: it discards information about time dependencies and is prone to overgeneralization. In this paper, we attempt to overcome these problems by relying on straightforward template matching. The basis for the recognizer is the well-known DTW algorithm. However, classical DTW continuous speech recognition results in an explosion of the search space. The traditional top-down search is therefore complemented with a data driven selection of candidates for DTW alignment. We also extend the DTW framework with a flexible subword unit mechanism and a class sensitive distance measure – two components suggested by state-of-the-art HMM systems. The added flexibility of the unit selection in the template based framework leads to new approaches to speaker and environment adaptation. The template matching system reaches a performance somewhat worse than the best published HMM results for the Resource Management benchmark. But thanks to complementarity of errors between the HMM and DTW systems, the combination of both leads to a decrease in word error rate with 17% compared to the HMM results.

2.6. Sophisticated automated speech recognition wiqas ghai and navdeep singh

Automatic speech recognition, which was considered to be a concept of science fiction and which has been hit by number of performance degrading factors, is now an important part of information and communication technology. Improvements in the fundamental approaches and development of new approaches by researchers have lead to the advancement of ASRs which were just responding to a set of sounds to sophisticated ASRs which responds to fluently spoken natural language. Using artificial neural networks (ANNs), mathematical models of the low-level circuits in the human brain, to improve speech-recognition performance, through a model known as the ANN-Hidden Markov Model (ANN-HMM) have shown promise for large-vocabulary speech recognition systems. Achieving higher Recognition accuracy, low Word error rate, developing speech corpus depending upon the nature of language and addressing the issues of sources of variability through approaches like Missing Data Techniques & Convolutive Non-Negative Matrix Factorization, are the major consider-

ations for developing an efficient ASR. In this paper, an effort has been made to highlight the progress made so far for ASRs of different languages and the technological perspective of automatic speech recognition in countries like China, Russian, Portuguese, Spain, Saudi Arab, Vietnam, Japan, UK, Sri-Lanka, Philippines, Algeria and India.

2.7. The application of real-time voice recognition to control critical mobile device operations

Omyonga Kevin1, Kasamani Bernard Shibwabo

The consistent increase in the number and ownership population of mobile devices introduces a variety of limitations. A set of this limitations revolve around interactivity. The overly dependent haptic mechanism of interaction has caused device falls, slower time to interaction, health concerns, and limited support for the disabled among other problems. There is need to formulate innovative techniques that facilitate our interaction with these devices for users. In order to achieve this, a Real-time Voice Recognition Algorithm is formulated that lets users of mobile devices acquire freedom to move about and reduce the need for constantly glancing at their screen. This is achieved by allowing users to verbally command their devices to carry out ordinary tasks such as setting an alarm, making a call, or even starting any application. An added unique feature is that it also offers offline access as any commands given by a user are processed and executed locally on the device.

3. Proposed work

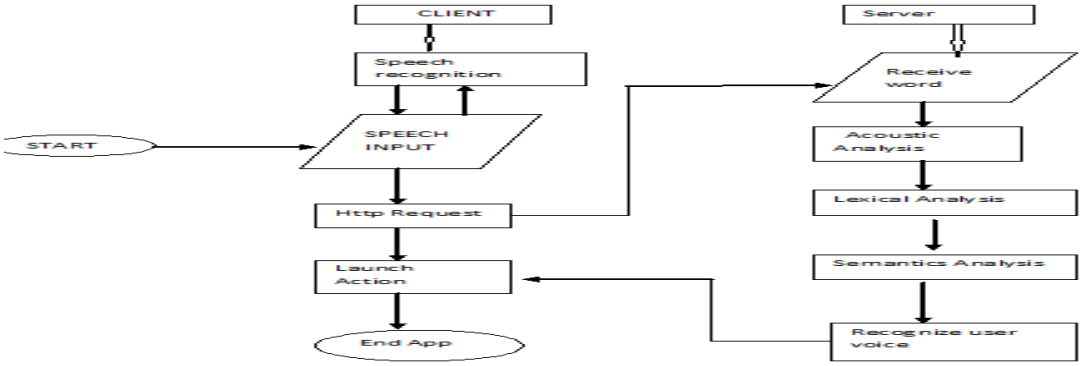


Fig. 1: System Architecture.

3.1. Module description

Modules identified for a speech recognition system

i) Speech Signal acquisition: it is process of fetching the voice in the form of signal and those signal is converted into machine readable format
ii) Feature Extraction: it involves analysis of speech. It is classified as two techniques temporal and spectral analysis. In temporal analysis technique the voice flow itself is used for analysis. In spectral analysis technique the representation of voice signal is used for analysis
iii) Acoustic Modeling: it is used to corresponds the relationship between an audio signal and distinct unit of voice or sound in a particular or specified language to make up the speech or voice
iv) Language modeling: is a prospect distribution over a series of words. This model tries to match sounds or voice with continuous words. This model provides the situation to differentiate between words and phrases that sounds as same.
v) Recognition: it is ability of the program to find words and phrases in voice command and convert it to machine readable format. It has some restricted vocabulary of words and phrases these may be identified if they are spoken correctly.
vi) Lexical modeling: it concerns the vocabulary of the words or a word which carries the meaning of it of a specified language.

Two of these modules Speech acquisition and Feature extraction are common to both the phases of ASR.

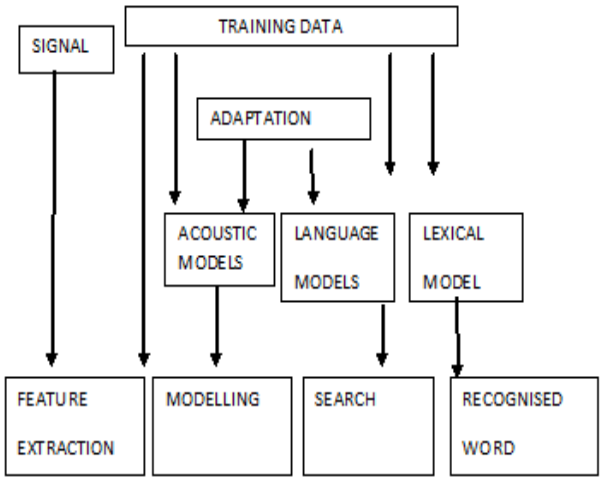


Fig. 2: Work Flow.

Model adaptation is meant for minimizing the dependencies on speakers‟ voice, acoustic environment, microphones and transmission channel, and to improve the generalization capability of the system.

Lexical Models

Lexicon is developed to provide the pronunciation of each word in a given language. Through lexical model, various combinations of phones are defined to give valid words for the recognition. Neural networks have helped to develop lexical model for non-native speech recognition.

Speech Data Collection

Text corpus is used finally to record the words/sentences through a single speaker or number of speakers depending upon the requirements. Precisely it involves the following steps:

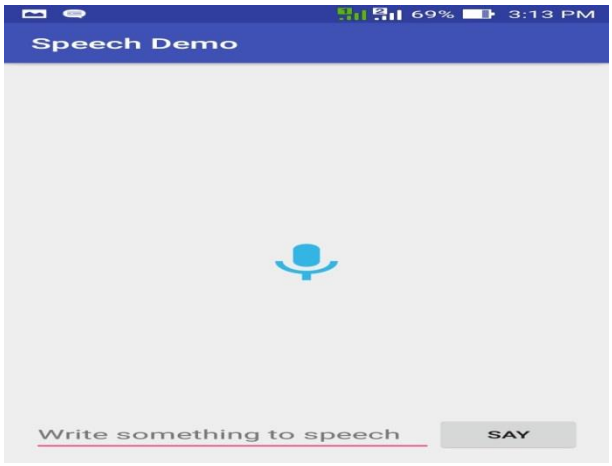
a) Selection of Speaker
b) Data Statistics
c) Transcription Correction

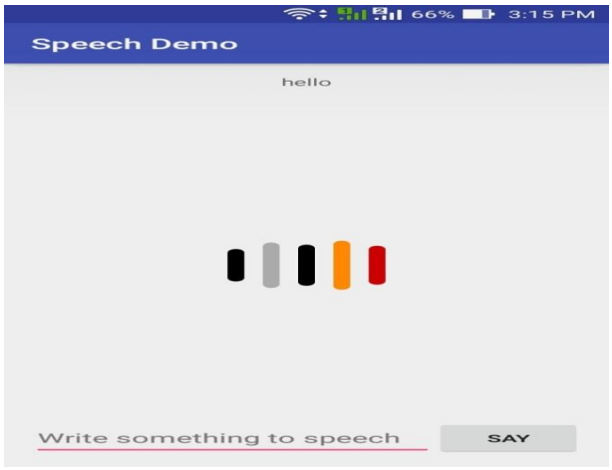
Lot of research has been done on clear speech using written text material which is generally read by a talker in conversational or clear styles. It has been found that read clear speech and spontaneously elicited clear speech is not same acoustic phonetically. Kain et al. [17] have proposed a communicative setting which allows talkers to naturally hear themselves while speaking and allows listening sound pressure levels to be controlled. They compared the words spoken in two different conditions of normal hearing and simulated hearing loss respectively. Acoustic phonetic properties of keywords in the first format were found to correspond to conversational speech and keywords spoken in the second format were found to correspond to moderately clear speech. There are languages which have regional and social variations with regard to pronunciation [19]. As a result, all or majority of the dialect regions are to be considered while going for the design of speech corpus. Modern standard Arabic is one of such languages in Algeria. While designing speech corpus Algerian Arabic speech database by Hamdani [18], six regions from southern and northern Algeria so as to cover all the regional and social variations of MSA. It has been found that these regions are having maximum homogeneous phonetic features and minimum phonological differences and this fact helped the researchers to get higher recognition rate in their speech recognizer.

4. General mobile assistants

Up until recently, the idea of interacting with a mobile device through speech was deemed pure science fiction. Such applications were only seen in movies (Bosker, 2013). That has all changed. Various companies like Google, Apple and Microsoft are spearheading the development of this applications with each attempting to deliver the most efficient user experience. These applications are designed to use a voice-controlled interface. As a result, a growing number of people now talk to their mobile smart phones asking them to carry out various activities such as sending emails, making phone calls and sending text messages.

5. Results





6. Conclusion

Speech recognition has been developed steadily over the last decades and it has been consolidated into numerous new applications. For most applications, the simplicity and understandability of speech recognition have reached the acceptable level. However, in flow, text preprocessing, and pronunciation fields there is still much work and improvements to be done to achieve more natural sounding speech or voice. Natural voice has so many changes that perfect naturalness may be impossible to achieve. However, since the markets of speech recognition related applications are increasing and developing regularly, the interest for giving more efforts and funds into this research area is also increasing. Present speech recognition systems are so difficult that one researcher cannot handle the entire system or app. But it is possible to divide the system or procedure into several individual modules whose developing process can be done separately if the communication between the modules is made correctly.

We are concluded with following future enhancement

1) Offline feature in voice recognition to control a Smart phones.
2) SOS message service through another mobile in case of device loss.
3) Added a fall detection using a accelerometer sensor, when the device accidentally fall from an user pocket It automatically make a Beep Noise, It helps the user to get the device when fall down.
4) Also Flashes a Flash light to find the device in a dark place

References

[1] Davis, K., Biddulph, R., and Balashek, S., "Automatic Recognition of Spoken Digit," J. Acoust. Soc. Am. 24: Nov 1952, p. 637. https://doi.org/10.1121/1.1906946.
[2] Hemdal, J.F. and Hughes, G.W., A feature based computer recognition program for the modeling of vowel perception, in Models for the Perception of Speech and Visual Form, Wathen-Dunn, W. Ed. MIT Press, Cambridge, MA.
[3] Watcher, M. D., Matton, M., Demuynck, K., Wambacq, P., Cools, R., "Template Based Continuous Speech Recognition", IEEE Transaction on Audio, Speech, & Language Processing, 2007.
[4] Samoulian, A., "Knowledge Based Approach to Speech Recognition", 1994.
[5] Tripathy, H. K., Tripathy, B. K., Das, P. K., "A Knowledge based Approach Using Fuzzy Inference Rules for Vowel Recognition", Journal of Convergence Information Technology Vol. 3 No 1, March 2008.
[6] Savage, J., Rivera, C., Aguilar, V., "Isolated word speech recognition using Vector Quantization Techniques and Artificial Neural Networks", 1991.
[7] Debyeche, M., Haton, J.P., Houacine, A., "Improved Vector Quantization Technique for Discrete HMM speech recognition system", International Arab Journal of information Technology, Vol. 4, No. 4, October 2007.

[8] Hatulan, R. J. F., Chan, A. J. L., Hilario, A. D., Lim, J. K. T., and Sybingco, E., "Speech to text converter for Filipino Language using Hybrid Artificial Neural Network and Hidden Markov Model", ECE Student Forum December 1, 2007 De La Salle University.

[9] K.Sathesh Kumar, K.Shankar, M. Ilayaraja and M. Rajesh, "Sensitive Data Security In Cloud Computing Aid Of Different Encryption Techniques, Journal of Advanced Research in Dynamical and Control Systems, vol.18, no.23, 2017.

[10] Sendra, J. P., Iglesias, D. M., Maria, F. D., "Support Vector Machines For Continuous Speech Recognition", 14th European Signal Processing Conference 2006, Florence, Italy, Sept 2006.

[11] Jain, R. And Saxena, S. K., "Advanced Feature Extraction & Its Implementation In Speech Recognition System", IJSTM, Vol. 2 Issue 3, July 2011.

[12] Aggarwal, R.K. and Dave, M., "Acoustic Modelling Problem for Automatic Speech Recognition System: Conventional Methods (Part I)", International Journal of Speech Technology (2011) 14:297–308. https://doi.org/10.1007/s10772-011-9108-2.

[13] Aggarwal, R. K. and Dave, M., "Acoustic modelling problem for automatic speech recognition system: advances and refinements (Part II)", International Journal of Speech Technology (2011) 14:309–320. https://doi.org/10.1007/s10772-011-9106-4.

[14] 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. https://doi.org/10.1109/89.536930.

[15] Yasuhisa Fujii, Y., Yamamoto, K., Nakagawa, S., "Automatic Speech Recognition Using Hidden Conditional Neural Fields", Icassp 2011: P-5036-5039.

[16] 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.

[17] Sorensen, J., and Allauzen, C., "Unary data structures for Language Models", INTERSPEECH 2011.

[18] 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.

[19] Hamdani, G. D., Selouani, S. A., Boudraa, M., "Algerian Arabic Speech Database (Algasd): Corpus Design and Automatic Speech Recognition Application", the Arabian Journal for Science and Engineering, Volume 35, Number 2c, Dec 2010.