Indian Streams Research Journal
Vol.1,Issue.V/June; 12pp.1-4
Charansing N. Kayte
ISSN:-2230-7850

Research Papers

HUMAN COMPUTER INTERACTION USING ISOLATED-WORDS SPEECH RECOGNITION SYSTEM

Dr V.P.Pawar
Siddhant Institute of Computer Application, Pune-412109

Chandrashekhar D. Sonawane
Dept. of Computer Science Singhania University, Pacheri Bari, Jhunjhunu(Raj.),India

Charansing N. Kayte
Dept. of Computer Science Singhania University,Pacheri Bari, Jhunjhunu(Raj.), India

ABSTRACT

This research paper aims to develop an isolated-word automatic speech recognition (IWASR) system based on vector quantization (VQ). This system receives, analyzes, searches and matches an input speech signal with the trained set of speech signals which are stored in the database/codebook, and returns matching results to users. IWASR is meant to assist customers calling a university's telephone operator to respond to their enquiries in a convenient way using their natural speech. Callers are assisted to select language, faculty and the staff name they wish to contact.

To extract features from speech signals, Mel-frequency cepstral coefficients (MFCC) algorithm was applied. Subsequently, vector quantization was used for all feature vectors generated from the MFCC. A codebook was resulted from training the VQ initial codebook and experimental results showed that the recognition rate has been improved with the increase of codebook size and showed that the codebook size of 81 feature vectors had a recognition rate exceeded 85%.
KEY-WORDS: Human-computer, interaction, speech recogion, technology, application.

INTRODUCTION

Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them [Hewett,1992]. The Speech is the most prominent and natural form of communication between humans. Work on speech recognition is not new to our times. For many years people have been trying to make our machines hear, understand and also speak our natural language . This arduous task can be classified into three relatively smaller tasks (Bharti W, 2010):

1. Speech recognition to allow the machine to catch words, phrases and sentences that we speak
2. Natural language processing to allow the machine to understand what we speak, and
3. Speech synthesis to allow machines to speak.

The work described in this paper speech inputs from users, analyzes the speech inputs, searches and matches the input speech with the pre-recorded and stored speeches in the trained database/codebook, and returns the matching result to the users. Developing this system is meant to assist customers calling a university's telephone operator to respond to their enquiries in a fast and convenient

Please cite this Article as : Charansing N. Kayte , Dr V.P.Pawar and Chandrashekhar D. Sonawane , HUMAN COMPUTER INTERACTION USING ISOLATED-WORDS SPEECH RECOGNITION SYSTEM : Indian Streams Research Journal (JUNE ; 2012)

way using their natural speech. Callers are assisted using their own speech inputs to select their language preference, faculty in a university and finally select the staff name they wish to contact.

This research applied the speech recognition technology in telephony domains. The focus was recognizing an isolated distinct word from a set of nine distinct words for three language, faculty and staff selections that could be used in operating the Isolated-Word Automatic Speech Recognition (IWASR) system. The nine distinct words are grouped into two main stages. The first stage is the faculty selection, which includes three faculty options namely Art, Commerce and Science faculties. Finally the third stage is the staff selection. Each of the options Science is

considered as a distinct isolated word since they are regarded as one utterance just like other options i.e. English, Science, Marathi …etc. Therefore, the IWASR operates on the nine distinct words that have just been mentioned.

MATERIALS AND METHODS

The Isolated-Word Automatic Speech Recognition (IWASR) system is designed in such a way that would guide and instruct users to accomplish their call destinations successfully. In general, this isolated-word recognizer was developed in three main simulation systems, which differ in the size of the database/codebook, which would show the recognition rate of each codebook size, and how would the decrease or increase of the codebook size affect the overall performance of the IWASR in terms of the recognition rate (Alan V,1999).The IWASR methodology is carried out is different stages as described below (Flanagan,1999),:

1. Collection of speech samples from different (male and female) speakers
2. Extracting distinguished and discriminative features from the collected speech samples and producing a set of feature vectors.
3. Training the feature vectors against the initial databases/codebooks in order to build unique databases/codebooks.
4. Matching/testing unknown feature vectors against the trained databases/codebooks in order to obtain the accuracy/recognition rate.
5. Evaluating the IWASR.

IWASR is divided into two main architectures. The first IWASR architecture is the training architecture, whereas the second IWASR architecture is the matching/testing architecture. Both architectures were implemented using a pipe and filter architecture, because each individual component of the architecture has a set of inputs and outputs and processes the data sequentially. IWASR architectures using pipe and filter style support reusability and they are easy to maintain and enhance. Each filter in the IWASR architectures reads a stream of data on its input and produces a stream of data on its outputs. Figure 1 shows the IWASR context diagram that represents the external look of the system where a caller performs a call request via the IWASR and receives the responds after processing the speech input, whereas both Figures 2 shows the IWASR training and matching/testing architecture.

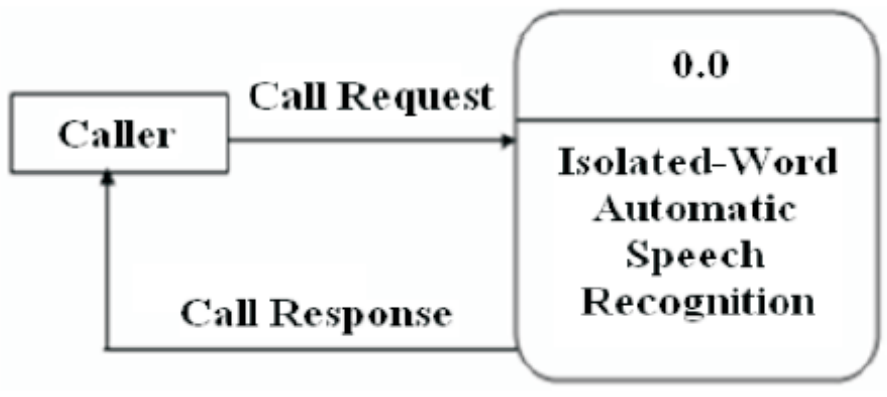

FIGURE 1
IWASR CONTEXT DIAGRAM (LEVEL-0 DIAGRAM)

Please cite this Article as : Charansing N. Kayte , Dr V.P.Pawar and Chandrashekhar D. Sonawane , HUMAN COMPUTER INTERACTION USING ISOLATED-WORDS SPEECH RECOGNITION SYSTEM : Indian Streams Research Journal (JUNE ; 2012)

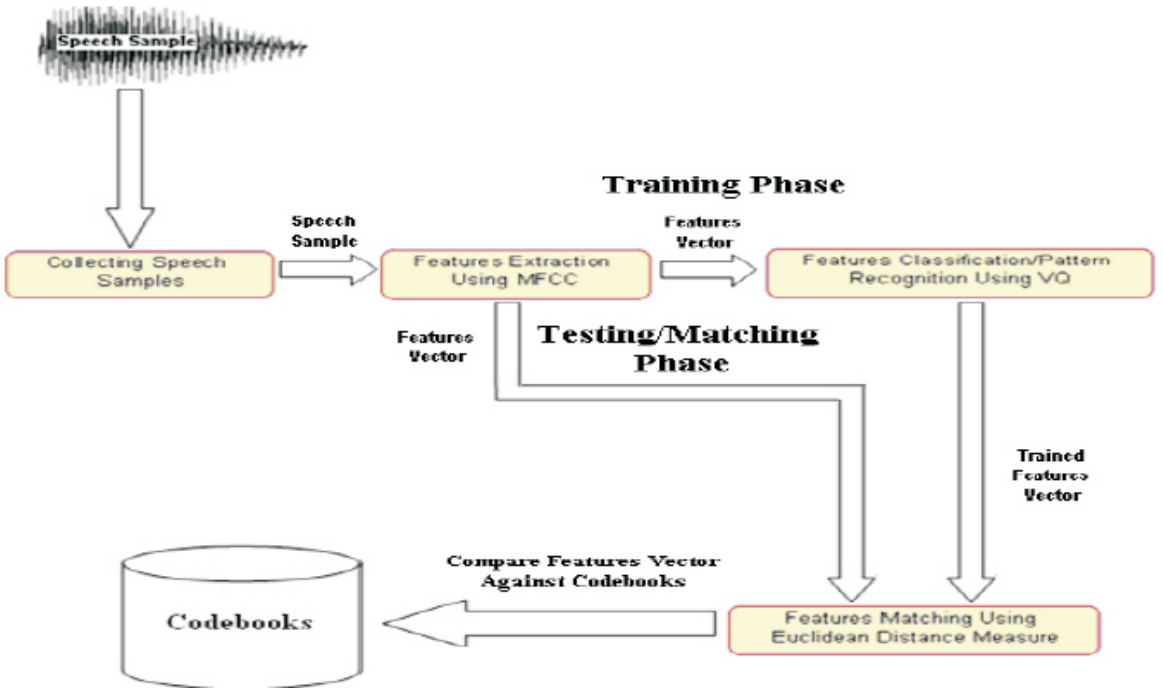

FIGURE 2
IWASR TRAINING AND MATCHING/TESTING ARCHITECTURE

RESULTS AND DISCUSSION

The experimental results of performing the MFCC algorithm for extraction features from unknown speech samples and matching/testing them against the trained VQ codebooks using the Euclidean distance measure. The features matching/testing stage was done for the three different versions of IWASR. It is found that the bigger the codebook size, the higher the recognition rate. This is found through testing some unknown speech samples against the codebooks provided by each IWASR version. Table 1 shows a total of 90 unknown speech samples were tested using the first version of IWASR and yielding an accuracy rate of "31.11%". Out of those 90 unknown speech samples only 28 samples were matched correctly and the other 62 samples were matched wrongly.

TABLE 1
TESTING RESULTS OF THE IWASR VERSION 1

IWASR Version	Total Testing Samples	Correct Matching	Wrong Matching	Accuracy Rate (%)
Version 1	90	28	62	31.11%

Table 2 shows a total of 63 unknown speech samples were tested using the second version of IWASR and yielding an accuracy rate of "85.71%". Out of those 63 unknown speech samples, 54 samples were matched correctly and only 9 samples were matched wrongly.

TABLE 2
TESTING RESULTS OF THE IWASR VERSION 1

IWASR Version	Total Testing Samples	Correct Matching	Wrong Matching	Accuracy Rate (%)
Version 1	36	32	4	88.88%

Please cite this Article as : Charansing N. Kayte , Dr V.P.Pawar and Chandrashekhar D. Sonawane , HUMAN COMPUTER INTERACTION USING ISOLATED-WORDS SPEECH RECOGNITION SYSTEM : Indian Streams Research Journal (JUNE ; 2012)

The rationale behind this slight increase in the recognition rate is mainly because IWASR version 3 has 9 speech samples collected from four different speakers who are males and females, which are trained to produce the VQ codebook. This version has shown better performance than the second IWASR version. Therefore, the bigger the codebook size the higher the recognition rate.

LITERATURE CITED

Alan V. O. and Ronald W. S. (1999). Discrete Time Signal Processing. 2nd Edition,Prentice Hall, New Jersey, USA.

Brad A. Myers. (1998) "A Brief History of Human Computer Interaction Technology". ACM interactions. 5: 44-54.

Gholampour, I. and Nayebi, K. (1999). High Performance Telephony Speech Recognition via CASCADE HMM/ANN HYBRID. Fifth International Symposium on Signal Processing and its Applications, Australia,645-648.

Bharti W Gawali, Santosh Gaikwad, Pravin Yannawar, Suresh C Mehrotra, 2010. Marathi Isolated Word Recognition System using MFCC and DTW Features. Journal of Computer Applications, 10:3.

Hewett, T., Baecher, R., Card, S., Carey, T., Gasen, J.,Mantei, M., Perlman, G., Strong, G., and Verplank, W.(1992). "ACM SIGCHI curricula for human-computer interaction." Report of the ACM SIGCHI Curriculum Development Group.

Martens, J.P. (2000). Continuous Speech Recognition over the Telephone. Electronics and Information Systems, Ghent University, Belgium.

Yuk, D. and Flanagan, J. (1999). Telephone Speech Recognition Using Neural Networks and Hidden Markov Models. IEEE International Conference on Acoustics,Speech, and Signal Processing,157-160.

Please cite this Article as : Charansing N. Kayte , Dr V.P.Pawar and Chandrashekhar D. Sonawane , HUMAN COMPUTER INTERACTION USING ISOLATED-WORDS SPEECH RECOGNITION SYSTEM : Indian Streams Research Journal (JUNE ; 2012)