Special Issue of Engineering and Scientific International Journal (ESIJ)
Technical Seminar & Report Writing - Master of Computer Applications - S. A. Engineering College (TSRW-MCA-SAEC) – May 2015
ISSN 2394-187(Online)
ISSN 2394-7179 (Print)

Human voice Interaction with ATM machine

Rajesh.K[#1], Arun.R[*2], Sivanesan.V[*3]
Master of Computer Applications, S.A Engineering college, Chennai
rajeshkrish12@gmail.com,arunvarun1994@gmail.com,sivavenkat2294@gmail.com

***Abstract*—**The Artificial Intelligence focuses on the mechanisms that generate intelligence and cognition. This research paper deals with how to apply the artificial intelligence in ATM machines to help the visually and physically challenged people. The proposal of Human voice Interaction with ATM machine is for the easy accessibility of the automatic teller machine by visually and physically challenged people by giving voice commands.

Keywords*—ATM, MFCC, CDSR, DTW*

1. Introduction

Artificial intelligence can be described as branch of computer science dealing with the simulation of machine exhibiting intelligent behaviour. A machine can be said as intelligent when it competes or interacts with the human. The interaction of a machine with the human voice commands may carry many advantages such that it could be used by any sort of persons in a Very affective and very easy manner. That too the human voice interaction with a commonly used operating device such as ATM would be more useful for the people with visual impairments. The user interface is designed to provide a convenient means of two-way communication between the user and the inference engine. An end user who tries to find a solution to a problem can describe the context of his problem to the system by means of the user interface. The knowledge base is a file that contains the facts and heuristic that makes up an expert's knowledge. A knowledge base is different from a typical data file or database. In a database, knowledge about the problem domain may be implicitly represented by the structure of the database. The actual contents of a database are the facts, data or information rather than knowledge. On the other hand, in the ESs, knowledge about the problem is explicitly represented in the knowledge base. A knowledge acquisition mechanism is used to acquire human expertise and transform into the knowledge base. This module processes the data entered by the expert and transforms it into a data presentation understood by the system. The inference engine is the knowledge processor that looks at the problem description and tries to find a solution with the help of factual and meta-knowledge. It can be considered as a program that applies domain knowledge to known facts to draw conclusions. The explainer is used to find out how a solution was obtained from an expert system and which individual steps were taken. The user can communicate with the explainer to obtain a report about the operation of the expert system.

The ATM machine stands for Automatic Teller Machine it was invented by Sheppard Baren on June 1967 at Barclays bank in Enfield, United Kingdom. It came to Existence in India in 1987. Despite of the existence and usage of the ATM for more than two decades still the aged people and the people with physical and visual challenges are unable to use the ATM machines. Even though the Braille keypad helps the blinds to enter their authentication information's but they couldn't able to continue with the further transactions.

In order to change this, the paper proposes for the Human voice interaction with ATM machine through Artificial Intelligence.

2. Existing and Proposal

The existing researches were only made for the speech or voice recognition but it was not used in the commonly used operating device such as ATM machines. My paper would carry a research and a proposal of Human voice interaction with ATM machines .Thus results in the proposal of speech or voice recognition of the commonly used operating machine.

The existing researched papers proposed the commonly used algorithms such as the HMM (Hidden Markov Model) and MFCC (Mel Frequency Cepstral Coefficient) for the speech recognition. My proposal would include the existing and the latest algorithms used by the Google servers for speech recognition.

3. Methodology

The research of the Human voice interactions may be accomplished only with the protocols that may help for the request and response of the ATM machine. In general the ATM machines are only a client machine which would only works as a front end to the user or the customer the request of the customer can only be completed and responded by the server which acts as the back end. The protocols are the only guidelines that could connect the client and server and helps for carrying the voice signals to the server and give response for the requested voice. The protocols work in the structure of the given flow chart below in figure 1.

The essential protocols for speech recognition must be studied. This paper describes the development of an efficient speech recognition system using different techniques such as Mel Frequency Cestrum Coefficients (MFCC), Dynamic Time Wrapping (DTW) and Configurable Speech Recognition System(C-DSR).

ISSN 2394-187(Online)
ISSN 2394-7179 (Print)

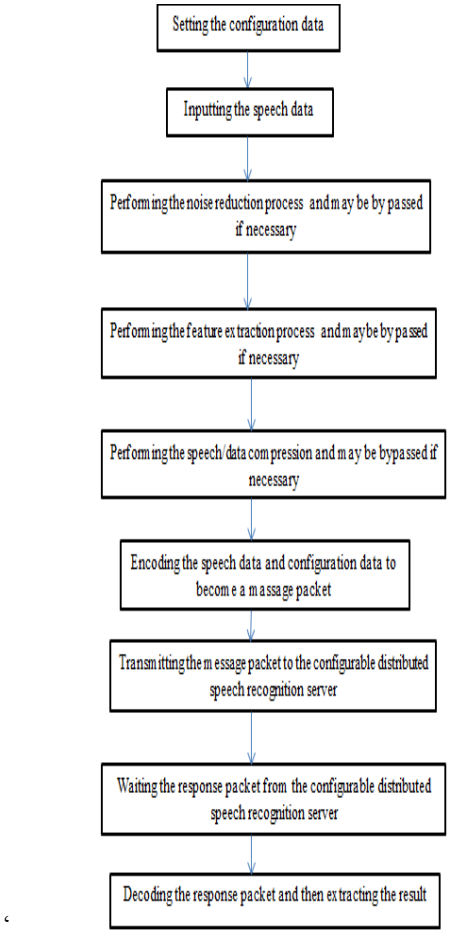

Fig.1: Flowchart for the process of the protocols

3.1 MFCC

In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency.

Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip (a nonlinear "spectrum-of-a-spectrum"). The difference between the cepstrum and the mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum. This frequency warping can allow for better representation of sound, for example, in audio compression.

MFCCs are commonly derived as follows:

- Take the Fourier transform of (a windowed excerpt of) a signal.
- Map the powers of the spectrum obtained above onto the mel scale, using triangular overlapping windows.
- Take the logs of the powers at each of the mel frequencies.
- Take the discrete cosine transform of the list of mel log powers, as if it were a signal.
- The MFCCs are the amplitudes of the resulting spectrum.

There can be variations on this process, for example: differences in the shape or spacing of the windows used to map the scale or addition of dynamics features such as "delta" and "delta-delta" (first- and second-order frame-to-frame difference) coefficients. The European Telecommunications Standards Institute in the early 2000s defined a standardised MFCC algorithm to be used in mobile phones.

3.2 C-DSR PROTOCOL

A C-DSR system of the present invention comprises a configurable distributed speech recognition protocol, and a configurable distributed speech recognition server. Herein, the configurable distributed speech recognition protocol is used to establish data transmitting format, for a client speech mobile device to pack the speech data and configuration data and become a message packet. The configurable distributed speech recognition system receives the message packet from the client speech mobile device, and adjusts speech recognition parameters according to the configuration data, and then returns a result to the client speech mobile device after completing the speech recognition task. Herein, the C-DSR server comprises of a parser, a configuration controller, a configurable distributed speech recognition engine, a history log, a diagnostic tool set, and configurable dialog system. The parser is used to parse and extract the configuration data and speech data in a packet. The configuration controller is used to generate a recognition adjustment parameter according to the configuration data. The configurable distributed speech recognition engine is used to recognize the speech data passed from the parser, and is configurable to the configuration controller. The history log is used to record the result or data generated from the server. The diagnostic tool set generates a diagnostic report according to data in the history log, for tuning the C-DSR engine. The configurable dialog system according to the recognition result to analyze possible lexicon may appearing in dialog, it's for raising the recognition rate and speed of the recognition engine next time.

3.2.1 C-DSR client

A configurable distributed speech recognition protocol, for specifying the speech and configuration data transmission format of a client device, to become a message packet; and a configurable distributed speech recognition server, for receiving said message packet from said client device, said configurable distributed speech recognition server performs speech recognition parameter adjustment according to said configuration data, and returns a speech recognition result to said client device, wherein said configurable distributed speech recognition

ISSN 2394-187(Online)
ISSN 2394-7179 (Print)

server comprises a history log for recoding history data generated by said configurable distributed speech recognition server.

3.2.2 C-DSR server

A parser, for receiving and parsing a message packet, then extracting a configuration data and a speech data included in said message packet;a configuration controller, for processing said configuration data, and according to said configuration data to generate a recognition adjustment parameter, said recognition adjustment parameter is used to configure the resources of said C-DSR server based on the computation, memory, communication, and bandwidth allocation of said client device; a C-DSR engine, for recognizing said speech data sent by said parser, and said C-DSR engine is configured by said configuration controller; and a history log to record history data generated by said C-DSR server.

3.3 DTW

In time series analysis, dynamic time warping (DTW) is an algorithm for measuring similarity between two temporal sequences which may vary in time or speed. For instance, similarities in walking patterns could be detected using DTW, even if one person was walking faster than the other, or if there were accelerations and decelerations during the course of an observation. DTW has been applied to temporal sequences of video, audio and graphics data — indeed, any data which can be turned into a linear sequence can be analyzed with DTW. A well known application has been automatic speech recognition, to cope with different speaking speeds. Other applications include speaker recognition and online signature recognition. Also it is seen that it can be used in partial shape matching application.

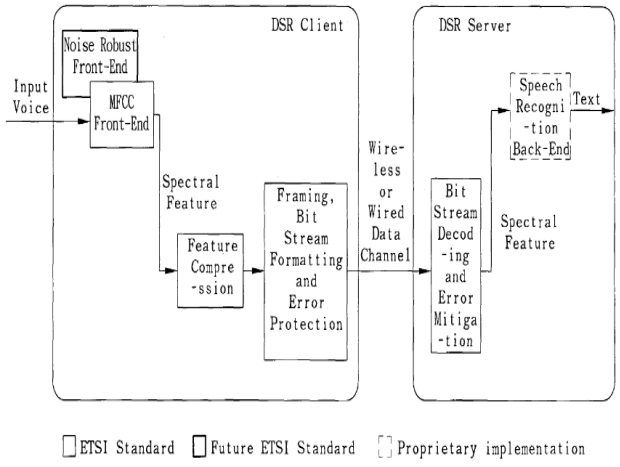

Fig.2: Architecture for the speech recognition

In general, DTW is a method that calculates an optimal match between two given sequences (e.g. time series) with certain restrictions. The sequences are "warped" non-linearly in the time dimension to determine a measure of their similarity independent of certain non-linear variations in the time dimension. This sequence alignment method is often used in time series classification. Although DTW measures a distance-like quantity between two given sequences, it doesn't guarantee the triangle inequality to hold.

The fig represents the way which the DTW works the reference R is the reference words in the database of the server and the test T is the testing words that are uttered by the customer or user of the ATM machine. The line in the graph spots the accuracy of the word uttered by the user and the word in the database of the server.

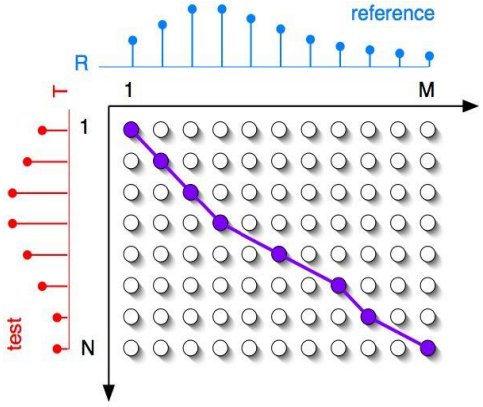

Fig. 3:Graph of the Data Time Wrapping

4. Future enhancement

The paper consists of the proposal of human voice interaction with ATM machine and the algorithm for it are explained this can be enhanced in the future as the machine could also conversant with the user for the transactional purpose.

5. Conclusion

The human voice interaction with the ATM machine could be done with the use of the protocol named configurable distributed speech recognition protocol the architecture is also provided in this paper for the clear vision of the working procedure of the configurable distributed speech recognition protocol.

References

[1] Ibrahim Patel And Dr. Y. Srinivas Rao .“Speech Recognition Using Hmm With Fcc - An Analysis Using Frequency Specral Decomposion Technique “
[2] Raghvendra Priyam And Rashmi Kumari ”ArtificialIntelligence Applications For Speech Recognition”
[3] R.D.Salagar Akshata Patil . “Voice Enabled Atm Machine with IrisRecognition For Authentication”
[4] Anjali Bala And Abhijeet Kumar And Nidhika Birla.“Voice Command Recognition System Based On Mfcc And Dtw”
[5] Suma Swamy And K.V Ramakrishnan.“An Efficient Speech Recognition System”

ISSN 2394-187(Online)
ISSN 2394-7179 (Print)

[6] Christopher Hale, Cam Quynh Nguyen, "Voice Command Recognition Using Fuzzy Logic" ,Motorola, Austin, Texas 78735,
[7] Hubert Wassner and Gerard Chollet, "New Time Frequency Derived Cepstral Coefficients For Automatic Speech Recognition", 8th European Signal Processing Conference (Eusipco'96).

K.Rajesh , holds a under graduate degree in B.Sc. Computer Science from Jaya college of arts and science and pursuing post graduation in Master of computer applications from S.A.Engineering college. This paper is a part of curriculum covered under in (MC7413) Technical Seminar and Report Writing.

R.Arun , holds a under graduate degree in B.Sc. Computer Science from Jaya college of arts and science and pursuing post graduation in Master of computer applications from S.A.Engineering college. This paper is a part of curriculum covered under in (MC7413) Technical Seminar and Report Writing.

V.Sivanesan , holds a under graduate degree in B.Sc Computer Science from Jaya college of arts and science and pursuing post graduation in Master of computer applications from S.A.Engineering college. This paper is a part of curriculum covered under in (MC7413) Technical Seminar and Report Writing.