International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com, Special Issue 42 (AMBALIKA) (March 2017), PP. 30-34

MULTIPLE USER SPEECH RECOGNITION SYSTEM

[1] Diksha Singh, [2]Ayushi Shukla [3]Dinesh Kumar Tiwari [4]Nupa Ram Chauhan
[1,2,3,4]Department of CSE FGIET
Raebareli India
[1] erdikshashingh4@gmail.com, [2] ayushi.fgi@gmail.com [3] dkt10dec@gmail.com [4] nrcua80@gmail.com

***Abstract*— Speech recognition system are become more useful in current scenario .Different interactive speech recognition applications are available in market and they are mainly made in such a way that only single user can use this system at a time. In this paper, our goal is to implement a general purpose speech recognition system for multiple users with thesecure manner. For multi user here we provide user ID and password (ie, user's own voice). Generally for implementing anefficient Speech Recognition System techniques MFCC (Mel Frequency Cepstrum Coefficients), HMM (Hidden Markov Model) are used.**

***Index Terms*—** ASR(Automatic speech recognition), SRS(speech recognition system), SR(Speech Recognition),Mel Frequency Cepstrum Coefficient (MFCC),STT(Speech-To-Text).

I. INTRODUCTION

"Speech recognition is the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words."Some speech recognition systems use "training" where an individual speaker or user reads text or isolatedvocabularyinto the system. The system anatomizes the user's specific voice and uses it to fine-tune the recognition of that speech of person, that's result is improved accuracy.

Voice User Interface are included in Speech Recognition such as voice dialing (e.g. "Call home"), domestic appliance control, call routing (e.g. "I would like to make a collect call") , search, simple data entry (e.g., entering a credit card number), STT processing (e.g., word password or emails), and aircraft (asDirect voice input).

A. PROPOSED WORK:

In one computer system one person/user can use the Speech Recognition System.so if another user want to use that particular system then he has to set up all the Speech Recognition System according their voice that will cause more waste of time and also not easy to set again and again whole settings.[So here I try to implement the LOG IN system by using the VOICE as PASSWORD for different USERS in one system]

B. PROPOSED NAME:

MINERVA (Goddess in GREEK MYTHOLOGY. Her name's real meaning is wisdom intellect and innovative ideas and inventiveness)

II. CLASSIFICATION OF SPEECHRECOGNITION SYSTEM

A. Types of Speech Recognition (based on utterances*)*

1) Isolated Word

Isolated word recognition system which recognizes single utterances like. Single word, isolated word recognition is suitable for situations where the user is required to give only one word response or commands, but it is very unnatural for multiple word inputs. It is simple and easiest for implementation because word are predefined and the words pronunciation have to be clear which is the major advantage of this type.

The drawback of this type is choosing different boundaries affects the results [1].

2) Continuous Speech

• Users are allowed to communicate in basic nature by this system, while computer arbitrate its content

• Its nature is like dictation. In this system words which have closer meaning run parallel without break. The idea of development of continuous speech recognition is a great challenge.

.

3) Spontaneous Speech

• When we speak, the voice/speech comes from our mouth is continuous and natural .So this system assimilate spontaneous speech. An Automatic Speech Recognition system with spontaneous speech is able to handle heterogeneity of natural speech features such as words being run-together. In this system speech can include mispronunciation, falsestarts and blank words.

B. Types of Speech Recognition (based on Speaker Model*)*

Every human being has unique voice because of his unique personal attributes and physical body. This system is classified in 3 main classes and they are as follows:

1) Speaker Dependent Model

Speaker dependent systems aredevelopedfor a particular type of speaker. They are generally more accurate for the particular speaker, but can be less accurate for other type of speakers.

These systems are usually cheaper, easier to develop and more accurate .But these systems are not flexible as speaker independent systems.

2) Speaker Independent Models

Speaker Independent system can recognize a different type of speakers withoutany prior training.

• This is developed to handle for any unique type of speaker. Speech independent system is used in (IVRS) Interactive Voice Response System for large number of different users who can accept inputs.

- Limitation of words is one of drawback of this system.
- Its precision is less than speaker dependent system and it is pricy.

3) Speaker Adaptive Models

• The speaker dependent data is used by speech recognition system which is adaptable to the best suited speaker for recognizing the speech. The error rate can be decreases by an adaption [2].

• They adjust operations according to user's characteristics.

C. Types of Speech Recognition (based on Vocabulary)

Whether size of vocabulary is large or small, In both cases it affects following features as:complexness, operation and the recognition rate of Automatic Speech Recognition system. So that Automatic Speech Recognition system can be classified as in vocabulary formas following:

• The limits 1 to 100 words or sentences are calledSmall Vocabulary.

• The limits 101 to 1000 words or sentences are called Medium Vocabulary.

• The limits 1001 to 10,000 words or sentences are calledLarge Vocabulary.

• The limits more than 10,000 words or sentences are calledVery largevocabulary.

III. SPEECH RECOGNITION SYSTEM'SFUNCTIONING

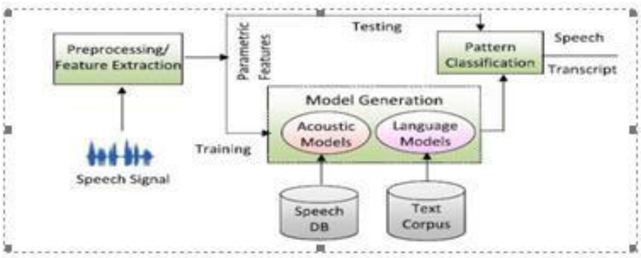


Fig 1: Automatic Speech Recognition System (SYSTEM ARCHITECTURE) [3]

A. Pre-processing/Digital Processing

The recorded acoustic signal is an analog signal. An analog signal can't directly transfer to the ASR systems.

So these speech signals need to transform in the form of digital signals and then only they processed.

These digital signals are move to the first order filters to spectrally level the signals. This procedure increases the energy of signal at higher frequency. This is the preprocessing step.

B. Feature Extraction

• The step of Feature extraction results the set of parameters of speech or stated expressions which have audial correlation with speech signals and above parameters canbe computedby processingof the audial waveform. We can call theseparameters asfeatures.

• Feature extractor's main work is to retain relevant information and disposeof irrelevant one.

• When this operation starts, acoustic signal is divided by feature extractor into 10-25ms.Window function is multiplied by Data acquired in those frames.

• We can divide window functions in many types that can be as, Blackman, Welch or Gaussian, hamming Rectangular etc. In this wayevery framehas given features. Feature extraction can be categories in different methods such as:

>Mel Frequency CepstrumCoefficient: MFCC) [4],
>Linear Predictive Cepstrum Coefficient; LPCC) ,
>Perceptual Linear Prediction ;(PLP), wavelet
>Relative Spectral Transform; RASTA-PLP [5] Processing etc.

C. Acoustic Modeling

• The foundation of Automatic Speech Recognition system is the designing Acoustic modeling [6].

• Inacoustic modeling,there is wellknown demonstration between the acoustic information and thephonetics.

• The important role of an acoustic model is decided by performance of the system . The responsibility of computational load is also given by this model.

• The co-relation between the basic voice units andthe acoustic cognizance is established by Training.

• The requirement of system's training is to create a pattern representation for the features of class in which one or more patterns can be used which can be corresponding to speech sounds of that equalclass.

• There are number of models are available for audial modeling. HiddenMarkov Model (HMM) is used over a broad rangeand accepted [8] because it's efficient algorithmfor trainingandrecognition.

D. Language *Modeling*

• This model hold the structural constraints which are used to generate probability of occurrence ,available in the language .this add up the probability of a word occurrence after a word sequence [9].for every language there are predefined grammar rules and constrains.

• Usually Bi-gram, Tri-gram, N-gram language modelsare used in speech recognition to finding accurate sequence of word by foretelling the probability of the nthword, using the N-1 earlier words.

• Inspeech recognition, the computer system matchessounds with word sequence.

• The words and phrase which have same sounds are differentiating in this model.

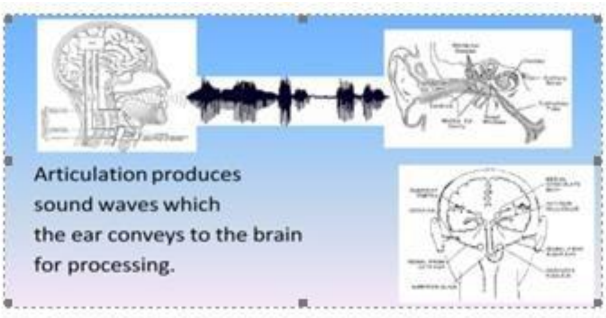
Example, In American English, the phrase like “recognize speech" and "wreck a nice beach"they have similar pronunciation but meanings are distinct.when prove from the language model is assimilate with the pronunciation model and the acoustic model the obscurities are easy to conclude.

E. Pattern Recognition

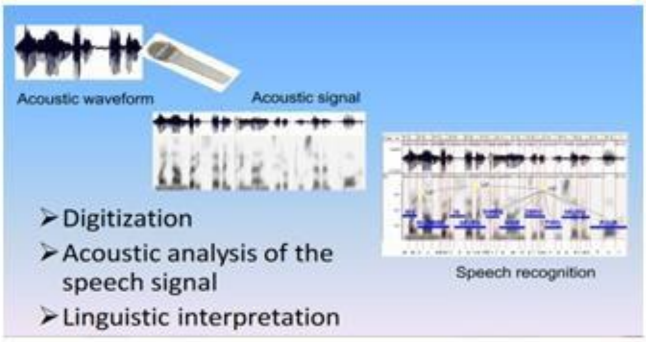
• In this recognition computing thea measure of similarities between each sound class reference pattern and equating the unknown test pattern are processed.
• After completion of training to recognize the speech during testing of patterns are divided

IV. DISCRIPTION

A. HOW HUMAN DO THIS?(fig-2)



B. HOW MIGHT COMPUTER TO DO THIS?(fig-3)



C. DESIRABLE ATTRIBUTES OF A VOICE RECOGNITION SYSTEM:

i Feature should occur naturally and frequently in speech
ii Easily measurable.
iii Doesn’t change over time or be affected by speakers health.
iv Isn’t affected by background noise.
v Not be subject to mimicry.

D. COMMON VOICE RECOGNITION TECHNIQUES:

i. DISCRETE FOURIER TRANSFORM
ii. LINEAR PREDICTIVE CODING
iii. CEPSTRAL ANALYSIS
iv. DYNAMIC TIME WARPING
v. HIDDEN MARKOV MODELS

V. RELATED WORK

• There are different type of speech recognition applications are available. Dragon, IBM and Philips had been developed different software’s.

• Microsoft developed Genie speech recognition that is interactive software. Different speech applications are developed; some are developed by AT&T that permits users tocommand on their computers, aswrite messages by voice.

• Various applications are showing up now a day.
  ➢ The SPHINX speech recognizer of CMU gives languages model in reference with acoustic model
  ➢ This recognizer worked on the Hidden Markov Models (HMM). The Universityof Colorado developed the SONIC recognizer.
  ➢ Let’s take other recognizers suchas XVoicefor Linux, in which input is taken by IBM’s ViaVoice that’s exclusive for Windows.
  ➢

• A speech recognition process has some demerits .One of main disadvantage is Background Noise.It disorients the recognizer which results difficulty in hearing the input speech.

• For reducing the energy consumption with balancing the quality of the application we advanced HP Smart Badge IV embedded system for optimizing speech recognition on it.

• There is such scalablesystem proposed which reducing the computational load and combine it with compression that is also scalable as the requirement of bandwidth in DSR (Distributed Speech recognition)there are different speech recognizers telecommunications field as Directory assistance and voice banking.

VI. PROJECT DESCRIPTION

• Our research refers the wide area of speech recognition system where it emphasizesto a system that provides customized speech recognition data to multiple voices and users.

• In the process of speech recognition computer received Text Words which are get converted by using microphones (Microphone receives Acoustic Signals) these words are used in different software application of computer likewise data entry, presentations (document), control and command etc.

• For example, indifferent medical devices and system, also in controlling and operating of other commands, speech recognition system areused. A surgeon or different user control through their voice commands that direct the functionality of a device controlled by this system. Like, the surgeon givesa voice command to settle patient table.

• To enable speech recognition in an operating room, medical devices and/or other equipment are connected with a component (e.g. a call system) through communication channels (e.g. an Ethernet connection, device bus, etc.). A speech recognition system is also connected providing of the connected devices, to implement control act, system will send right signal.

• Speech recognition systems are generally customizable for a particular user. For example, a user may customize particular settings of the system. A user may further configure or train the system to recognize his or her particular voice and custom commands.

- Customization and configuration is performed at the time of initializing the system or upon restarting the system. As such, prior art systems are not customizable or reconfigurable during use and without interruption. Therefore when a second user desires to operate the system, the system must rebooted and/or restarted to reconfigure the system for the second user. In an operating room setting where time is critical, such delay is highly undesirable.
- It is therefore desired to provide speech recognition system and method including an improved means of user configuration and customization. It is further desired to provide a speech recognition system and method operable by two or more users continuously.

A. Research idea-

- POSSIBILITY OF THIS idea!!!!!!!(yes)_in one computer system one person/user can use the SRS.so if another user want to use that particular system then he has to set up all the SRS according their voice that will cause more waste of time and also not easy to set again and again whole settings.
- So here I try to implement the LOG IN system by using the VOICE as PASSWORD for different USERS.
- INDEPENDENT SPEAKER.
- MULTIPLE USER ACCOUNTS.
- ENHANCE THE SECURITY BY ATTACHING USER ID AND PASSWORD(FOR EVERY SINGLE PERSON/USER)(for windows SRS).
- MORE USABILITY THEN PRESENT.

VII. CONCLUSION AND FUTURE ENHANCEMENTS

A. CHALLENGES

- External noise interfering
- Imprecision of pronunciation
- Large vocabularies
- Different speakers
- Overlapping speech
- Firstly Vocabulary had to contain 4 words that are pronounced likewise to enable the speech recognition system and they are- Play, Pause,Stop and Repeat.For one media file Play and Stop command are used.Here’s one difficulty is that between sound card and software conflictions are possible.

B. .HOW TO SOLVESHORTCOMINGS

- Using High Quality Microphone.
- Use Good Quality of Sound Card.
- System must be trained properly.
- If Possible work in quit environment

C. FUTURE SCOPE

To extract and apply all levels and information from the speech signal conveying speaker identity

▪ Acoustic – use spectral features conveying vocal tract information

▪ Prosodic - use features derived from pitch, energy tracks to classify information

▪ Phonetic – use phone sequences to characterize speaker specific pronunciations

▪ Idiolect – use words to characterize user specific word patterns

▪ Linguistic – use linguistic patterns to characterize speaker specific conversation style.

Mainly the works are implemented on the Acoustic properties. By implementing all in one the developed system will be [Just A Really Very Intelligent System] that is termed as (J.A.R.V.I.S).

1) APPLICATIONS

- Telephony and other domains.
- People with disabilities
- Training air traffic controller
- High performance fighter air craft
- Voice recognition gadgets(TV)
- Subtitle reading(Automatic)
- Aerospace (e.g. space exploration, spacecraft, etc.)
- Interactive Voice Response.
- Real time Speech Writing(In court reporting time)
- Hands-free computing.
- Home automation.
- Automatic Translation.

CONCLUSION

This research paper is based on distinct approaches and basics of speech recognition system. By implementing these different types of approaches, techniques and models the rate of speech recognition system can be get better and to develop improved quality speech recognition.

a. For future development of speech recognition main priority of making large vocabulary in system that will be Independent of speaker and follow the Continuous speech recognition system.

b. To implement these some famous techniques are used at top level and they are-HIDDEN MARKOV MODEL (HMM)and ARTIFICIAL NEURAL NETWORK(ANN)

c. Here we are going to make a system in which more than one speaker can use a computer system by instructing computer using their own voice.

REFERENCE

[1] Sanjivani S. Bhabad, Gajanan K. Kharate, "An Overview of Technical Progress in Speech Recognition"International Journal of Advanced Research in Computer Science and Software Engineering,Volume3,Issue3,March 2013
[2] Pratik K. Kurzekar,Ratndeep R.Deshmukh,Vishal B.Waghmare,Pukhraj P.Shrishrimal,"Continuous Speech Recognition System: A Review", Asian Journal of Computer Science and Information Technology, 2014
[3] R.K.Aggarwal, M.Dave "Integration of Multiple acoustic and language models for improved Hindi speech Recognition system", Springer Science Business Media, LLC 2012.
[4] Bishnu Prasad Das1, Ranjan Parekh,"Recognition of Isolated Words using Features based on LPC,MFCCZCR and STE,with Neural Network Classifiers", International Journal of Modern Engineering Research ,Vol.2, Issue.3, May-June 2012.
[5] S.B. Magre, R.R. Deshmukh, "A Review on Feature Extraction and Noise Reduction Technique",International Journal of Advanced Research in Computer Science and Software Engineering Vol 4, Issue 2, February 2014.
[6] Rajesh Kumar Aggarwal, Ph.D_Thesis, "Improving Hindi Speech Recognition Using Filter Bank Optimization and Acoustic Model Refinement",December 2012.