ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Isolated Word Recognition System for Autistic Speech

Akshata M Sonnad[*1], Anjana Gopan[2], Divya S[3], Sailakshmi N R[4], Ambika R[5]
[*1,2,3,4,5] Department of Electronics and Communication Engg., B.M.S. Institute of Technology, Bangalore, Karnataka, India
akshatasonnad@gmail.com

Abstract

This paper aims to discuss the implementation of an Isolated Word Recognition system for Autistic speech using Hidden Markov Model (HMM). This system will be able to recognize the spoken words of Autistic children. Autism is a neurological disability which resists the functions of neurons in human brain. The most common problem with autistic children is communication. Focusing on their speech, it is observed that there is a high variability compared to normal speakers' speech and it is unclear and not always understandable. Considering this, we aim to build a system to recognize the autistic speech and give the output as text and convert it back to clear speech using speech synthesizers.

Keywords: Isolated word, automatic speech recognition, Hidden Markov Model.

Introduction

Autism is a complex developmental disability that typically appears during the first three years of life. This disorder is four more times prevalent in boys than girls. Symptoms of autism can vary in different ways, taking the category of social interaction and relationships, they include: non-verbal communication development problems such as eye to eye gazing, body language, and facial expressions, failure to make friends with people their own age, lack of interest and lack of empathy. Generally, children with autism experience language problems, restricted interests and activities as well as sensory and intellectual problems.

Hence in this paper, we aim to develop an Isolated Word Recognition system based on Hidden Markov Model (HMM) using an open source tool kit called HTK[1] (Hidden Markov Model Tool Kit) to recognise words spoken by autistic children. Isolated Word" basically refers to the presence of silence between two words. The major difficulties in the implementation of an ASR system are due to different speaking styles and environmental disturbances. So the main aim of an ASR system is to transform a speech signal into text message independent of the device, speaker or the surroundings in an accurate and efficient manner. As an extension, the output text format is converted into standardized speech outputs using a text to speech converter.

Existing Methods

An Automatic Speech Recognition system can be developed using various methods. The four main approaches include[2]:

ACOUSTIC-PHONETIC APPROACH:

Acoustic-phonetic approach assumes that the phonetic units are broadly characterized by a set of features such as format frequency, voiced/unvoiced and pitch. These features are extracted from the speech signal and are used to segment and level the speech. It uses knowledge of phonetics and linguistics to guide the search process. Usually certain rules are defined for easy decoding and it is based on "blackboard" architecture.

KNOWLEDGE BASED APPROACH:

Knowledge based approach attempts to mechanize the recognition procedure according to the way a person applies its intelligence in visualizing, analysing and finally making a decision on the measured acoustic features. Expert system is used widely in this approach.

STATISTICAL BASED APPROACH:

The recognition procedure uses mathematical and statistical tools. It is sometimes viewed as anti-linguistic approach. The statistical processes are used to search through the space of all possible solutions to pick the statistically most likely one.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

PATTERN RECOGNITION APPROACH:

Pattern recognition approach does not require any explicit knowledge of speech. This approach involves two major steps i.e, pattern training and pattern comparison. The essential feature of this approach is that it uses a well established mathematical framework and gives consistent speech pattern representations, for reliable pattern comparison, from a set of labelled training samples through a formal training algorithm. The popular pattern recognition techniques include template matching and Hidden Markov Model (HMM).

Proposed Method

It has been proved that in most of the speech recognition experiments, Hidden Markov Models give highly accurate results. The technique of HMM has been broadly accepted in today's state-of-the art ASR systems, mainly for two reasons: its potential to model the non-linear dependencies of each speech element on the adjacent units and a powerful set of analytical approaches provided for estimating model parameters.

The isolated word recognition system based on HMM is developed and designed in two stages:

- Training the system with input speech samples.
- Testing the system with test samples.

Training the system with input speech samples:

In this stage, speech samples from different people (normal and autistic) are collected. The training corpus so formed trains the system. A limited vocabulary is chosen for forming speech models. The words are converted into acoustic models statistically characterized as HMM. Two major steps involved in the training stage are:

- *Data preparation:* This involves fixing vocabulary. In this project we aim to use the words from one to ten as speech samples.
- *Recording data:* This involves creating transcription files, coding data, creating word models, re-estimate models and fixing silence models.

Testing the system with test samples:

In this stage, the words already stored to be tested are considered. Utterance for each word is recorded as test data, which is decoded to recognize spoken words. Depending on the type of speech recognition unit, automatic speech recognition can be divided into two groups[3]; *isolated word recognition* and *continuous speech recognition*. The isolated word recognition assumes that a word is uttered in a discrete manner so that there are silences at the beginning and the end of each word. The continuous speech recognition is more difficult and complicated compared to the isolated word recognition because word boundaries are not known and are often ambiguous. In this paper we concentrate on the training and testing of Isolated word recognition. The testing and decoding is done using Viterbi decoding algorithm.

HIDDEN MARKOV MODEL[4]

The hidden Markov model (HMM) is one of widely used statistical models to model sequences of speech parameters by well-defined algorithms. It is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (*hidden*) states and has successfully been applied to speech recognition systems. At each time unit (i.e., frame), the HMM changes states according to state transition probability distribution, and then generates an observation ***o(t)*** at time *t* according to the output probability distribution of the current state. Hence, the HMM is a doubly stochastic random process model.

BLOCK DIAGRAM

In this section, a brief discussion about the proposed block diagram and modular description of each block is carried out.

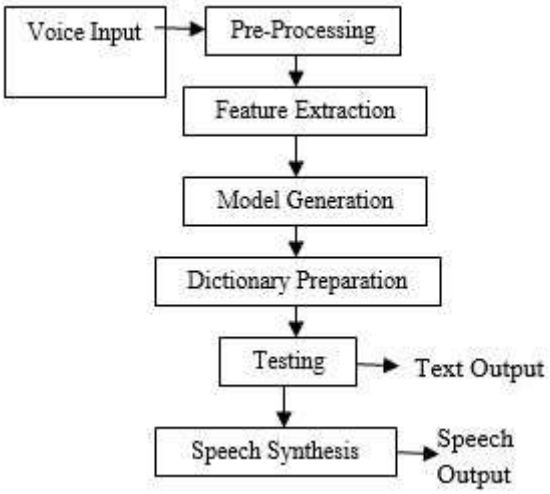

Fig.1 : Block diagram

In our project, the words: ONE, TWO, THREE, FOUR, FIVE, SIX, SEVEN,EIGHT, NINE and TEN are used as samples for training and testing These samples are collected from Autistic children and normal persons. The system performance is first

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

evaluated for normal speech samples and then tested for autistic speech samples.

The steps to be followed are:

1. PRE-PROCESSING[5]:

   Preprocessing makes it convenient for subsequent processing of the speech signal. First, noise is removed from the recorded sample. The signal is then amplified if required. Next, silence is added between each sample, thereby segregating two consecutive words.

2. FEATURE EXTRACTION:

   Feature extraction is the first stage of Speech Recognition, wherein the acoustic signal is converted into a sequence of acoustic feature vectors. It is the most important stage in the entire process, since it is responsible for extracting relevant information from the speech frames, as feature parameters or acoustic vectors. Feature extraction contains three major steps:

   1. *Pre-emphasis*: In order to flatten speech spectrum, a pre-emphasis filter is used before spectral analysis. Its aim is to compensate the high-frequency part of the speech signal that was suppressed during the human sound production mechanism. The most used filter is a high-pass FIR filter.
   2. *Frame blocking and windowing*: The speech signal is divided into a sequence of frames where each frame can be analysed independently and represented by a single feature vector. Since each frame is supposed to have stationary behaviour, a compromise, in order to make the frame blocking, is to use a 20-25 ms window applied at 10 ms intervals (frame rate of 100 frames/s and overlap between adjacent windows of about 50%).
   3. *Mel frequency cepstral co-efficients extraction*[6] : The Mel Frequency Cepstral Coefficient (MFCC) is a representation of the speech signal defined as the real cepstrum of a windowed short-time signal derived from the FFT of that signal. It is first subjected to a log-based transform of the frequency axis (Mel-frequency scale), and then de-correlated using a modified Discrete Cosine Transform (DCT-II). The figure below illustrates the complete process to extract the MFFC vectors from the speech signal. It is to be noted that the MFCC extraction process is applied over each frame of the speech signal independently as shown below:

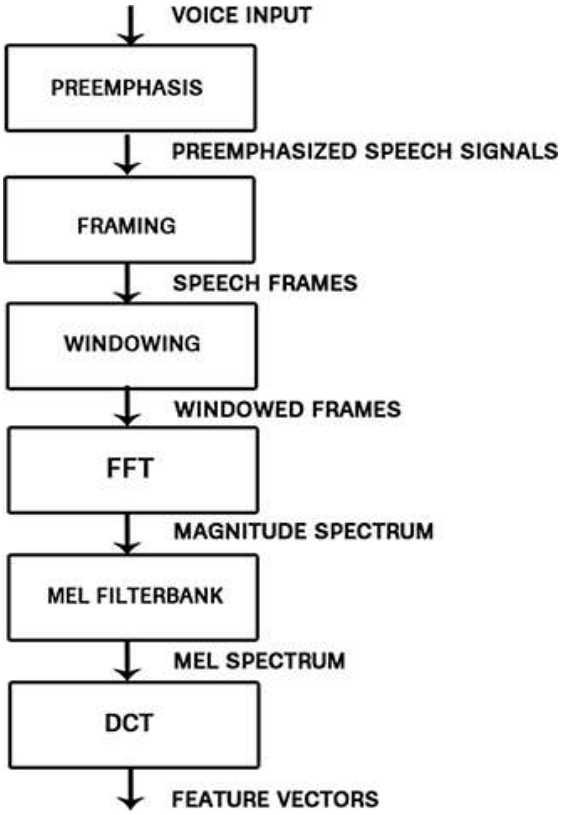

Fig.2 MFCC Extraction Process

Model Generation

After the Mel Frequency Cepstral Coefficients (MFCCs) are obtained, acoustical models must be defined to train and test the system.

Dictionary Preparation

Dictionary is generated by first extracting the word transcriptions directly from database. The input words are then compacted through a compression procedure to retain only representative transcriptions. The compressed dictionary then is used for re-segmenting the training sentences and for decoding the test sentences. A word-pair grammar has been used in continuous speech recognition systems. The word-pair grammar is modified such that context-sensitive grammatical parts are defined to smooth the transitions between words instead of using the grammatical parts directly.

Testing

The system is tested with various test samples for both normal and autistic samples. The accuracy is calculated for both the modules and the

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

performance of the system is evaluated. The output so obtained is in text format.

Speech Synthesis

The output in text format is converted into speech using speech synthesizer. This ultimately gives an understandable speech output of the unclear speech inputs.

Software Tools

The software tools proposed for the implementation of the automatic speech recognition system for autistic speech are:

WAVE SURFER:

Wave Surfer is an open source audio editing tool for sound visualisation and manipulation. It interactively displays sound pressure waveforms. It is widely used for studies of acoustic phonetics, spectrograms, pitch tracks and transcriptions.

HIDDEN MARKOV MODEL TOOLKIT (HTK) [7]:

The HTK toolkit based on Hidden Markov Model (HMM), a statistical approach, can be used to develop an Isolated Word Automatic Recognition System (ARS). **HTK** (Hidden Markov Model Toolkit) is software toolkit for handling HMMs. HTK is an open source tool kit that works in Linux environment. HTK consists of a set of library modules and tools available in C source form. The tool provides sophisticated facilities for speech analysis, HMM training, testing and results analysis. Labelled speech samples are given as input and the HTK extracts Mel Frequency Cepstral Coefficients (MFCC), generates model for each word and gives the output in the form of text.

FESTIVAL:

Festival offers a general framework for building speech synthesis systems as well as including examples of various modules. It offers full text to speech through a number Application Program Interfaces (APIs): from shell level, though a Scheme command interpreter, as a C++ library, from Java, and an Emacs interface. Festival is free software. Festival and the speech tools are distributed under an X11-type licence allowing unrestricted commercial and non-commercial use alike.

Conclusion

In conclusion, this paper gives a detail discussion about the steps involved in automatic speech recognition of isolated words. As an application of ASR technique, the system is mainly designed for the purpose of recognizing a select set of words spoken by autistic children. The various advantages in this approach include expected higher accuracy since HMM approach and MFCC extraction algorithm are used; transformation of unclear speech into clear speech output with the help of speech synthesizers, along with the display of the recognized words; and the advantage of using open source software tools for the entire process.

References

[1] *HTK "Hidden Markov Model Toolkit", available at http://htk.eng.cam.ac.uk,2012.*

[2] *Santosh K Gaikwad, Bharti W Gawali, Pravin Yannawar." A Review on Speech Recognition Technique" International Journal of Computer Applications (0975 – 8887) Volume 10– No.3, November 2010*

[3] *Dongsuk Yook, "Introduction to Automatic Speech Recognition", pp. 12-13*

[4] *R. Rabiner, and B. H. Huang, "An introduction to hidden markov models," IEEE Acoust. Speech Signal Processing Mag., pp. 4-16, 1986.*

[5] *Mikael Nilsson, Marcus Ejnarsson. "Speech Recohnition using Hidden Markov Model, performance evaluation in noisy environment", pp.17-18*

[6] *Zaidi Razak, Noor Jamaliah Ibrahim, Emran Mohd Tamil, Mohd Yamani Idna Idris "Quarnic Verse recitation feature extraction using Mel-Frequency Cepstral Coefficient(MFCC)" Department of Al-Quran & Al-Hadith, AcademyOf Islamic Studies, University of Malaya.*

[7] *Mohit Dua, R.K.Aggarwal, Virender Kadyan,Shelza Dua. "Punjabi Automatic Speech Recognition Using HTK" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012, ISSN (Online): 1694-0814*