Imperial Journal of Interdisciplinary Research (IJIR)
Vol-2, Issue-3 , 2016
ISSN : 2454-1362 , http://www.onlinejournal.in

Android Application For Deaf People

Patil Diksha Y.[1], Jadhav Ashwini A.[2], Chavan Amruta V.[3] & Nikam Shrutika M.[4]
Department of Computer Engineering, Matoshri College of Engineering and Research Centre, Eklahare,Nashik

Abstract : *The advancement of Information and Communication Technology has impacted all aspects of the human life.It has changed the way we work, travel, conduct business, study and communicate. For the Deaf community, the use ICT has improve their quality of life by developing systems that can help them communicate better with the rest of the world and amongst themselves. Sign Language is the primary means of communication in the deaf and dumb community. The problem arises when dumb or deaf people try to express themselves to other people with the help of these sign language grammars and vice a versa. The application provides deaf people a way of getting more closed to advanced technology by using speech to image translation. This helps deaf people to learn new technologies by looking toward images which are being converted to images by using speech recognition system.*

Keywords— *Android platform, SQLite database, Speech recognization, Image processing, clustering.*

1. Introduction

Deaf people normally use sign language in order to communicate with each other. In this communication system, deaf people are not able to represent their ideas or messages to other people which they want to say. In today's world technology has been developed very fast and presents each action in digital form then it may be in images or audio format. In order to make their life more advanced, application is needed to be developed so they can get opportunity to learn new thing and can get a chance to introduce with new technologies.

Dumb people are usually deprived of normal communication with other people in the society. It has been observed that they find it really difficult at times to interact with normal people with their gestures, as only a very few of those are recognized by most people. Sign Language is the primary means of communication in the deaf and dumb community. For the Deaf community, we try to improve their quality of life by developing systems that can help them communicate better with the rest of the world and amongst themselves. Speech-to-sign technology enables audible language into sign language translation on smart phones.

2. Literature Survey

The problem arises when dumb or deaf people try to express themselves to other people with the help of these sign language grammars. This is because normal people are usually unaware of these grammars. As a result it has been seen that communication of a dumb person are only limited within his/her family or the deaf community. At this age of technology, their is the demand for a computer based system for the dumb community. Interesting technologies are being developed for speech recognition but no real commercial product for sign recognition is actually there in the current market.

The Stephen Cox, Michael Lincoln and Judy Tryggvason in 'TESSA, a system to aid communication with deaf people', 2002 proposed the the speech to sign coversion algorithm is being used in this paper to recognize the speech and convert it into a images. The S.M. Halawani and Zaitun A.B. proposed 'An Avatar Based Translation System from Arabic Speech to Arabic Sign Language for Deaf People', in 2008 which describe the importance of web to search multimedia content such as image or video which is classified into two categories such as text based search and content based search. D. Molla & J.L Vicedo discussed that the 'Restricted domain QA' ,in 2007 extension of text based QA (Question Answer) to research based multimedia QA to manage the range of factoid. H. Cui, M.Y. Kan proposed that the 'Definitional QA', in 2008 Queries are classified into two classes namely related query or non-related query. R. C. Wang, W. W. Cohen, E.Nyberg propose that the paper 'List QA' in 2008 to collect image and video data we need to generate queries through engine.

1. Input

The system takes input in the form of speech data. Application mainly focuses on giving output as an image. User is going to create session and allows to fetch images from database related to the speech.

2. Speech recognition system

Speech recognition basically means talking to a computer having it recognized what we are saying, and lastly, doing this in real time. This process fundamentally functions as a pipeline that converts PCM digital audio form from a sound card into recognized speech.

3. Proposed System

The system is being developed mainly in two parts which consists of speech recognition and conversion of speech into image form. The first part receives the input in the form of a speech. The speech is captured through a speech recognition system as an analogue signal. Then it is digitized and translated to sign language. The second part of the system converts the sign language to avatars which are displayed on the computer screen for the user to see. The system will refer to two sets of databases. The first database, which we call the Sign and Speech database will contain all the images of the alphabets and words. While the second database called the Avatar database will contain the equivalent avatar for each alphabet and words.

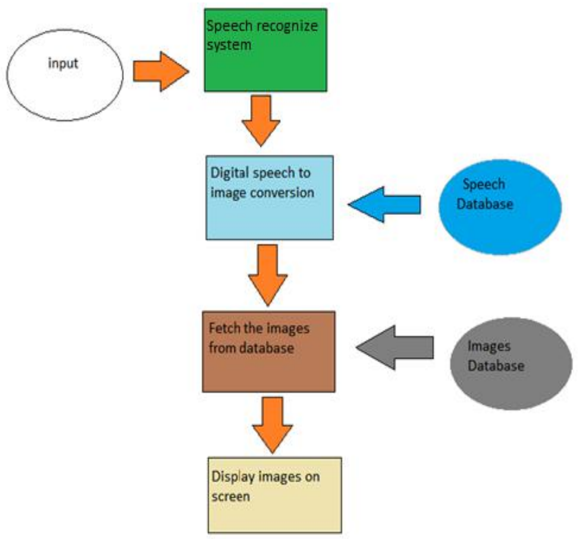

Figure. 1. Overall Architecture of the System

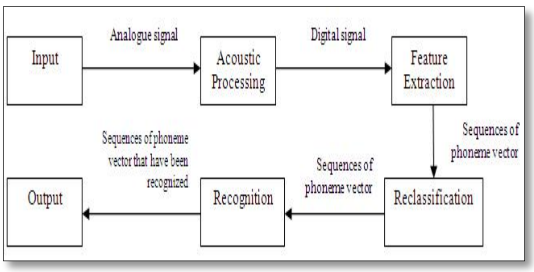

Figure 2: The Speech Recognition Process

Acoustic Processing

Acoustic modeling of speech typically refers to the process of establishing statistical representations for the feature vector sequences computed from the speech waveform. Hidden Markov Model (HMM) is one most common type of acoustic models. Other acoustic models include segmental models, super-segmental models (including hidden dynamic models), neural networks, maximum entropy models, and (hidden) conditional random fields, etc.

Feature Extraction

In speaker independent Automatic Speech Recognition Feature extraction is the process of retaining useful information of the signal while discarding redundant and unwanted information or we can say this process involves analysis of speech signal.

Reclassification

The sequence of phoneme vector is being classified and sent it to the recognition block. It classifies the given input into the category on sign languages.

3. Digital speech to image conversion

The speech is captured through a speech recognition system as an analog signal. Then it is digitized and translated to sign language. Then it fetches the image from the database and displays on the screen.

4. Database

Database is mainly to maintain the data and images related to that data. Whenever user will create a session and will start the lecture then the speech related images is being fetched from the database and will be displayed on the screen. In order to have a lecture without an internet then SQLite database which is inbuilt in android can be used. SQLite is an Open Source Database which is embedded into Android. SQLite supports standard relational database features like SQL syntax, transactions and prepared statements. In addition it requires only little memory at runtime (approx. 250 Kbyte). SQLite is a software library that implements a self-contained, server less, zero- configuration on, transactional SQL database engine. SQLite is the most widely deployed SQL database engine in the world.

5. Displaying images

This is the last stage of process where user can see the images related to the speech and can understand the lecture.

4. Algorithm

A.Speech-Recognition Algorithm

1. Hidden-Markov models (HMMs)

Hidden-Markov models (HMMs) are popular statistical models used to implement speech-recognition technologies.The time variances in the spoken language are modeled as Markov processes with discrete state spaces.Each state produces speech observations according to the probability distribution characteristics of that state. The speech observations can take on a discrete or a continuous value. In either case, the speech observations represent a fixed time duration (i.e., a frame). The states are not directly observable, which is why the model is called the hidden-Markov model.

The original minimal HMM (min_HMM) algorithm was implemented on a floating-point C language program platform running under the UNIX operating system.

2. The Acoustic Front End

The sampled signal is split into frames; each frame represents a finite time slice. Each slice is short enough to allow a speech wave segment to be considered stationary within the frame. A technique called windowing is used to achieve this result. Windowing allows the portion of the sample that is closest to the center of a window to be more heavily weighted than the parts of the sample that are further away from the center of the window. This weighting technique minimizes spectral leakage. Another window function, the hamming window, is an error-trapping routine that is used to detect data errors inside individual windows. By stepping a data window along a sampled speech signal, a sequence of frames that represent the whole speech wave sequence is obtained. Typical window lengths are 30 ms (where the length is a representation of a forward time-slice sequence), but the frames can be stepped at a shorter elapsed time interval, such as 20 ms, so that the frames overlap. Individual windowed speech frames are then processed further to capture the characteristic information in a more compact form. Texas Instruments uses linear predictive coding (LPC) to perform speech spectral analysis.LPC conversions use a pitch-asynchronous automatic-correlation method with a frame rate of 20 ms. The gestures generated by the acoustic front end of the recognizer are obtained by orthogonalization of the 14-mel spaced filter-bank outputs. These values, in addition to the corresponding 14 difference values, the actual speech level, the relative speech level, the transitional characteristics of the speech, and the energy difference values, make up a 32-element vector called the generalized speech parameters (GSPs). This 32-element vector is then multiplied by a linear transform (LT) value [6] to provide a 10-element vector named the generalized speech features (GSFs), which is used in the final acoustic-distance calculation.

5. Conclusion

In this way, we have completed the design part of the project with the requirement specification. Modules of the project are designed and are well studied in order to fulfil the requirements of the project. Moreover, the testing part with all the test cases of the modules of the project is being carried out. Thus, the completion of partial report is being completed with full hard work and complete support and guidance of our guide and project plan is made to ensure the proper planning of the project.

6. Acknowledgements

We take this opportunity to thank our project guide Prof. Praveen .Andhale for their guidence and providing all the necessary facilities for this paper.

7. References

[1]Stephen Cox, Michael Lincoln and JudyTryggvason, "TESSA, a system to aid communication with deaf people". Proceedings on the Automatic speech to text conversion,Automatic translation of English text to sign language and display it as a sequence of signs using computer graphics techniques, 2002.

[2] S.M. Halawani and Zaitun A.B., “An Avatar Based Translation System from Arabic Speech to Arabic Sign Language for Deaf People”,Arabic Speech to Arabic Sign Language for Deaf People, 2012.

[3] Sunguk Lee, " Creating and Using Databases for Android Applications," Open source SQLite database, 2012.

[4] Surabhi Bansal and Ruchi Bahety, "Speech Recognition System," Transform the PCM digital audio into a better acoustic representation and Unit matching system, 2013.

[5] A. Sujith Kumar and Shahabaz Begum,“Sign Mobiles: An Android App for Specially Able People,” Speech to-Text Translation, ASL Dictionaries (TXT2ASL), 2014.