INTELLIGENT LIP READING SYSTEM FOR HEARING AND VOCAL IMPAIRMENT

R.Nishitha[1], Dr K.Srinivasan[2], Dr V.Rukkumani[3]
[1]Student, [2]Professor and Head, [3]Associate Professor,
Electronics and Instrumentation Engineering, Sri Ramakrishna Engineering College, Coimbatore

Abstract
In this paper, we proposed an idea for developing system for lip reading and speech to text converter for the deaf and mute. It basically uses digital image processing technique for lip reading and conversion into audio and digital signal processing for speech to text conversion.
Keywords: lip reader, speech to text converter, digital image processing, digital signal processing

I. INTRODUCTION

Though there are many devices and techniques that aid auditory and vocal impairment, none of them are followed widely as the same. There are many challenges to those who aren't born with such impairments, but unfortunately get impaired in the middle of their life. All that the deaf needs is to comprehend what is told to them and all that the mute needs is to convey their message.

What if a device is available which can lip read and convert it to a text (for deaf) to comprehend what is being told and that could be converted into a speech (for dumb) to interpret what the person is trying to convey. The idea of lip reading is based on perceptual phenomenon called McGurk effect which is an audio-visual integration. It is the combination of tracking lips, teeth and tongue.

The lip reading system can be developed by lip finding and tracking, visual feature extraction under which shape-based feature extraction and appearance-based feature extraction are employed and finally audio-visual integration is done [1].

With the help of image processing, fuzzy logic and artificial intelligence, this task could be achieved. Image processing will aid in lip finding and tracking. Fuzzy logic will help in processing possible texts or combination of words which match with the lip movements. It would be necessary to have Artificial Intelligence (AI) in the lip reading system because it would have the job of remembering the matching sets of the motions of the mouth to the correct word. It would also remember previous conversations and use those as references for the guessed words, evolving the program to be more accurate in its guesses [2].

Speech Recognition systems operate in two phases which are- a training phase, during which the system learns the reference patterns representing the different phonetics (speech sounds) that constitute the vocabulary of the application. Each reference is learned from spoken examples and stored either in the form of templates obtained by some of the averaging methods or models that characterize the statistical properties of pattern. The other phase is a recognizing phase, during which an unknown input pattern, is identified by considering the set of references.

The system also performs speech analysis using the linear predictive coding (LPC) method. From the LPC coefficients we get the cepstral time derivatives and weighted cepstral coefficients, which form the feature vector for a frame. Then, the system will perform vector quantization using a vector codebook. The resulting vectors form the observation sequence. For each word in the vocabulary, the system builds a Hidden Markov Model or HMM [3] and trains the model during

the training phase is performed using the PC-based C programs.

II. PROPOSED IDEA

The idea of the project is to integrate lip reading system and speech to text conversion system in an efficient manner. This system must prove to be useful for both the deaf and mute.

This idea is basically implemented because of the difficulties faced by the deaf and mute to comprehend and convey.

II.1. LIP READING

There is naturally an individual variation in the ability to lip read, and as with any skill, the competence varies, with level of hearing loss not necessarily the ability to predict.
Those with monaural loss may also need to lip read, as may some without diagnosed hearing loss but with a specific type of processing disorder. In lip reading, we tend to lip-read part and guess or just predict the rest.

Lip reading is more or less 80% guesswork. When lip reading is considered, we must be aware that there is a high chance of visual impairment within the hearing impaired population. Whether someone is following sign language or a spoken communication, their visual attention is important continually.

Factors which make sentences more difficult to lip read are:

- The background noise
- not knowing about the context
- obscurity of the context (the more immediate the topic, the easier for the receiver)
- some abstract concepts
- use of colloquial language
- sentence complexities
- long sentences
- difficulties in segmenting the sentences as receiver (detecting the word boundaries)
- unfamiliar or unusual use of vocabulary
- non-visual consonants, like; /s/ /k/, especially at the beginning of the key words
- visually confusable consonants, like; /t/ /d/ /n/
- speed of delivery- especially if too fast or too slow
- exaggerated speech pattern usage
- absence of normal facial expressions while talking
- non-specific gesture usage
- difficulties at close range focusing on lips and maintaining eye contact- require adequate distance to take in all the available clues
- lack of concentration while observing

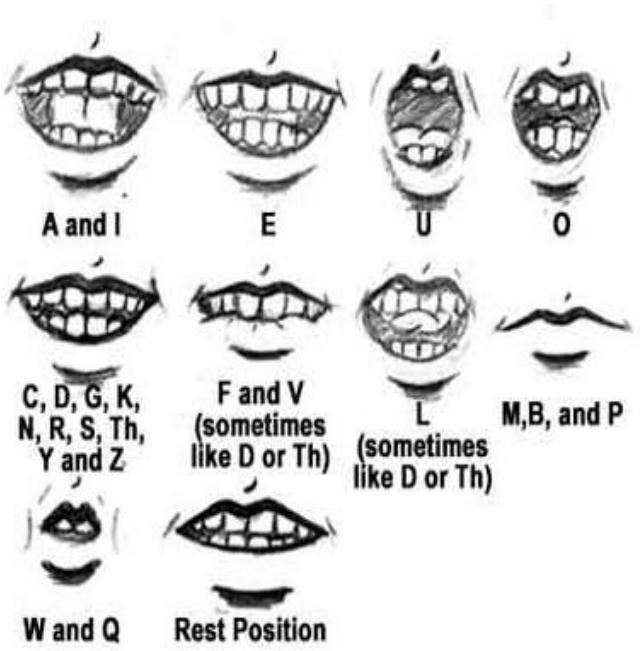

Figure 1. Lip movement based on phonemes

II.2. IMAGE PROCESSING OF LIP

An automatic lip reading technique is a rapidly evolving speech recognition technology that can be used for monitoring a speaker's motions so that, specially designed computer programs can interpret sentences.

Initially, automatic detection and tracking of the mouth region must be done. It is done based on lip tracking and lip finding. Lip finding is applied only when no previous information of the lip position is available. This will happen in the first frame of a sequence or whenever lips have not been correctly located in the previous frame. Lip finding is based on the geometric model of the face [1].

The structures of pixels are evaluated to know if their relative positions match a simplified prior model of face. Lip tracking is done based on the references of the previous data and so lip tracking is much more reliable and requires less resource than lip finding.

In [5], Active Appearance Model (AAM) is employed to extract location of specific points of face from every frame of video sequence. The shape information extracted by the AAM from the face image is used to compute a set of suitable parameters that describe the appearance of facial features. The key points to optimally track speech related movements are selected. These key points are then transformed into a set of representative parameters. It helps in the interference engine with the data that encapsulates the most important aspects of speech.

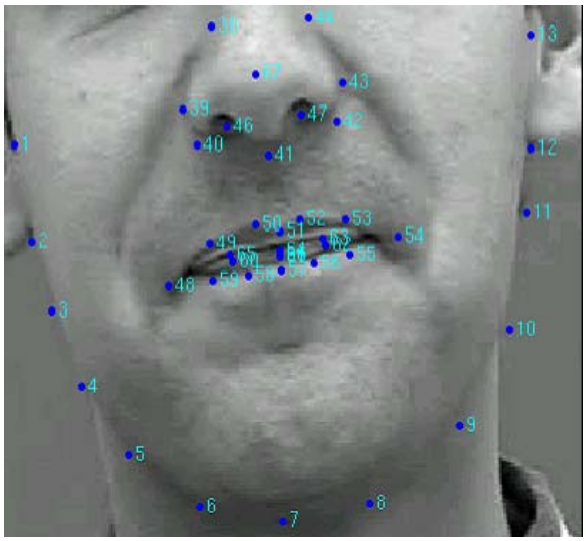

Figure 2. Active appearance model

Fuzzy logic would be utilized in order to adapt the data produced from the image processing system. As Kaehler Steven said, fuzzy logic (FL) provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information. Fuzzy logic's way to control problems mimics the way a person would make decisions, in a much faster manner. For the proposed lip reading system, the fuzzy logic would choose and suggest the most probable sentences available from analyzed data gathered by the image processor [2].

It would be necessary to have an Artificial Intelligence in the lip reading system because it would have to remember the matching sets of the motions of the mouth to the correct word.

It would also remember all the previous conversations and use them as references for the guessed words; evolving the program to be much more accurate in its guesses.

II.3. TEXT TO SPEECH CONVERSION

The acquired sentences in the form of text from the display must be read out. This will help the mute to communicate with others just as any common man. This process will include the synthesis of speech from text.

Speech synthesis is the process of converting the message written in text to equivalent message in spoken form. A Text-To-Speech (TTS) synthesizer is a computer-based system that has to be able to read text. This generally involves two steps, which are, text processing and speech generation. The main function of text-to-speech (TTS) system is to convert an arbitrary text to a spoken waveform [8]. This task generally consists of text analysis, text normalization, text processing, acoustic processing and finally speech generation.

Text analysis part is a preprocessing part which analyzes the input text and organizes into manageable list of words. Text normalization is the transformation of text to pronounceable form.

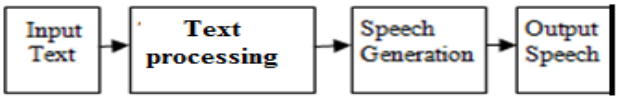

Figure 3. Proposed method

II.4. SPEECH TO TEXT CONVERSION

The other part of the system is to comprehend the speech signals and depict them as texts. This is the part of the system which will facilitate the deaf.

Speech-to-text system will convert speech to text when instantly given a voice. The system will not synthesize the quality of the recorded human speech. There are different technologies which are suitable for different applications.

Basically, there is no simple metric that could be applied to any of the STT [Speech to Text] systems and which would give a clear concept of the overall quality of any system [12].

The main reason is, the STT systems should not be assessed in isolated place, but they should be evaluated for their respective uses. There are many uses for the STT systems, so they should be given to their exact destinations.

Automatic Speech Recognition (ASR) systems [11] operate in two phases. First, a training phase, during which the system learns

the reference patterns which represent the different speech sounds that constitute the vocabulary of the application. Each reference will be learned from spoken examples and stored either in the form of templates obtained by some averaging methods or models that characterize the statistical properties of the pattern. Second, a recognizing phase, during which an unknown input patterns, will be identified by considering the set of references.

The speech recognition system may be viewed as working in four stages- analysis, feature extraction, modeling and finally, testing.

III. BLOCK DIAGRAM

The block diagram in Figure 4 represents the whole idea of this system. It varies according to people with respect to the challenge they face.

The deaf would not need the text to speech converter and audio output, whereas, the mute would not need the speech input and the speech to text converter.

The main advantage of this system is to make available both the facilities so as to benefit the deaf and mute. The inputs of the block are speech signal or lip reading system. The speech input would need to be interpreted by the system to convert into text and display.

The other input is to the lip reading system. The lip reading process is carried out by image processing.

By fuzzy logic and artificial intelligence, the images are converted to text.

The text is both displayed and also converted into speech output with the help of text to speech converter.

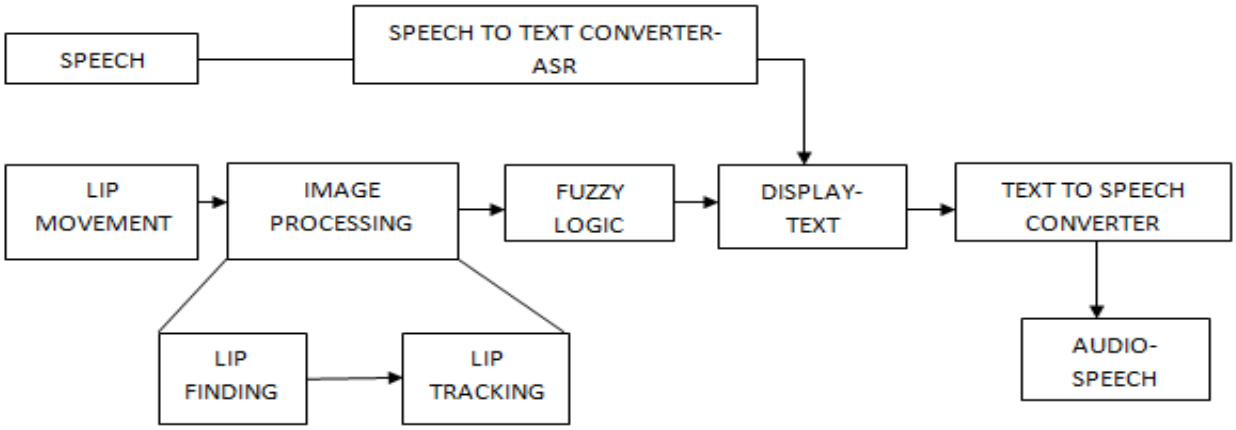

Figure 4. Block diagram of the system

IV.RESULT

In this proposed method, we evaluated the propriety of the suggested visual feature by recognizing an isolated word with utterance unit of speaker having a word. The recognition accuracy of the trained words was found to be better than that of the untrained words.

It was also found that the more definite visual speech element for a word was given, the higher accuracy the result was.

Then, as result of test about the scale of dictionary (database), the larger the dictionary is, the lower accuracy is.

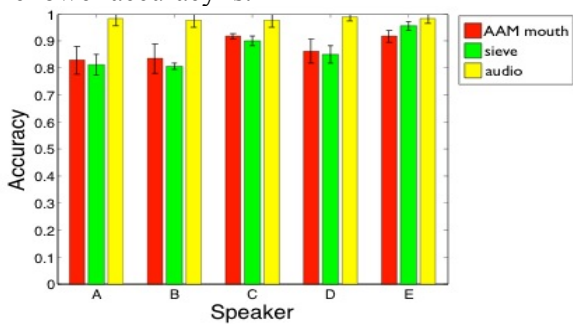

Figure 5. Graph of the result

In addition to lip reading, speech was also taken as an input. From the database created consisting of various words and syllables, speech was converted to text. The performance of the system was found to be good. It recognizes the words with the help of the added database. Only if the words match, the system shows the output. Its accuracy is estimated to be 85%.

With the same database created, text to speech conversion was done using character recognition and speech synthesis techniques. The output was commendable and proved to be reliable.

V. CONCLUSION

The lip reader and speech to text converter is a technique which benefits the deaf and mute at the same time, unlike any other system which can benefit them individually only.

The outcome of the project will be efficient conversation of the deaf and mute, without the need for feeling inferior and this could help in boosting their confidence and their personality.

It aids in ensuring their security in case of any emergencies and this could help them get out of their insecurities.

VI.REFERENCES

[1] Jesus F. Guitarte Perez, Alejandro F. Frangi Eduardo Lleida Solano and Klaus Lukas, "Lip Reading for Robust Speech Recognition on Embedded Devices," ICASSP 2005, pp. I-473-476.

[2] Michelle Kim, "Personalized Lip Reading Program Using Artificial Intelligence System," 7th Intelligence Technology: Personalised Lip Reading Using Intelligence System, pp.1-8.

[3] Ha Jong Won,Li Gwang Chol,Kim Hyok Chol,Li Kum Song, "Definition of Visual Speech Element and Research on a Method of Extracting Feature Vector for Korean Lip-Reading," College of Computer Science, DPR of Korea, pp.1-18.

[4] Suma Swamy and K.V Ramakrishnan, "An efficient speech recognition system," Computer Science & Engineering: An International Journal (CSEIJ), No.4, Vol. 3, 2013, pp. 21-27.

[5] Alin G. Chithu, Leon J.M. Rothkrantz, "Visual speech recognition- Automatic system for lip reading of Dutch," Information technologies and control, plenary paper, 2009, pp. 2-9.

[6] Ziheng Zhou, Guoying Zhao and Matti Pietikainen, "Towards a Practical Lipreading System," In Machine Vision Group, pp. 137-144.

[7] Kaladharan N, "An English Text to Speech Conversion System," International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 10, 2015, pp. 1-5.

[8] Kaveri Kamble, Ramesh Kagalkar, "A Review: Translation of Text to Speech Conversion for Hindi Language," International Journal of Science and Research (IJSR), Volume 3, Issue 11, 2014, pp. 1027- 1031.

[9] Poonam.S.Shetake, S.A.Patil, P. M Jadhav, "Review of text to speech conversion methods," International Journal of Industrial Electronics and Electrical Engineering, Volume 2, Issue 8, 2014, pp. 29- 35.

[10] Deepa V.Jose, Alfateh Mustafa, Sharan R., "A Novel Model for Speech to Text Conversion," International Refereed Journal of Engineering and Science (IRJES), Volume 3, Issue 1,2014, pp.39- 41.

[11] Miss.Prachi Khilari & Prof. Bhope V. P., "A review on speech to text conversion methods," Volume 4, Issue 7, 2015, pp. 3067-3072.

[12] Nuzhat Atiqua Nafis and Md. Safaet Hossain, "Speech to Text Conversion in Real-time," International Journal of Innovation and Scientific Research, Volume 17, No. 2, 2015, pp. 271- 277.