International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

# Extraction of Lip features for the Identification of Vowels Utterances using MFCC and Geometrical Aspects

**Srikanth G N, M. K Venkatesha**

***Abstract - Identification of a person's speech by his lip movement is a challenging task. Even though many software tools available for recognition of speech to text and vice versa, some of the words uttered may not be accurate as spoken and may vary from person to person because of their pronunciation. In addition, in the noisy environment speech uttered may not perceive effectively hence there lip movement for a given speech varies. Lip reading has added advantages when it augmented with speech recognition, thus increasing the perceived information. In this paper, the video file of a individual person are converted to frames and extraction of only the lip contour for vowels is done by calculating its area and other geometrical aspects. Once this is done as a part of testing it is compared with three to four people's lip contour for vowels for first 20 frames. The parameters such as mean, centroid will remain approximately same for all people irrespective of their lip movement but there is change in major and minor axis and hence area changes considerably. In audio domain vowel detection is carried out by extracting unique features of English vowel utterance using Mel Frequency Cepstrum Coefficients (MFCC) and the feature vectors that are orthonormalized to compare the normalized vectors with standard database and results are obtained with approximation.***

***Keywords - MFCC, Orthonormal Vectors, Vowels.***

## I. INTRODUCTION

Speech contains information more than the text for example prosodic features gives information about the inner feelings of the human being therefore speech signal analysis itself is a different area of research in the field of engineering, but speech under noisy environment suffers from drawback being not perceived properly by humans in such situation.

when it is supplemented with lip movement gives an added strength in proper understanding of the information to be conveyed even under noisy environment. In the system of HMI communication, it is quite significant to realize normal and user friendly interfacing. Recognition of speech or lip reading systems using computers have been studied and improved as a means of HMI communication. This is importance to supply similar interface at the output.

Manuscript received on January 02, 2020.
Revised Manuscript received on January 15, 2020.
Manuscript published on January 30, 2020.
* Correspondence Author
**Srikanth G N*,** Research Scholar, Dept. of Electronics & Inst. Engg, R N S Institute of Technology, Bengaluru, India. srikanthgn@yahoo.com
**M. K Venkatesha,** Principal, R N S Institute of Technology, Bengaluru, India. mkvenkatesha@gmail.com



The movement of lip synthesis play a vital role in the system of human-machine communication. It has been studied that intelligibility of the speech in noisy environments is enhanced by augmenting the face information i.e. lip movements to the speech signals aggravated by the acoustic noise.

Moreover, for the hearing impaired men this system can compensate the lack of auditory information. Further it can also be used for reduction of bandwidth required in transmitting video signal where in bandwidth required for transmitting entire video, we can use the speech and derive the lip movements from speech and only morphing the lip movements on the speaker keeping background and other aspects of the video unchanged. Its application can be in news transmission on television and even in mobile communication.

The literature pertaining to lip reading has been developed over the years. There are two main spheres of research on lip reading given by data parameterization and inference phenomenon. The techniques used for deriving or mapping of lip movements from the speech signal requires time synchronization and this mapping the process requires intensive computations. It is known that speech to lip movement mapping is one to one and also for some vocabulary it is many to one i.e. same lip movement for many speech utterances. Due to rapid progress in speech recognition methods, hitherto the most widely used method for morphing of lip movements from speech signal which gives promising results is by using HMM (Hidden Markov Model and its derivatives) [1], [2].

Quite recently in this area some work is also attempted in implementation of deriving using Deep Learning techniques. The applications that can be influenced by this work can be in the area of human-machine communication, Robotics, Advanced speech recognition engines etc.

## II. RELATED WORK

Mel Frequency Cepstral Coefficients (MFCC) are quiet well applied features in most of the speech identification applications as well as speakers. In 1980s, considerable attempts has been done to develop MFCC [3], [4]. While employing MFCC in the development of applications, many problems such as use of frequency estimation algorithms, design of the efficient filter banks and the total selected features in extracting the speech and its dynamics all play a significant role in the performance of speech recognition systems. But the main challenges encountered in MFCC extraction are the increased computational complexity, demand for robust features, and the degraded performance in the presence of noise.

So, there must be sufficient compatibility between the number of robust features limit on the one hand and the required performance of the complete system on the other hand in addition to having the knowledge of the type of the environment and its noise. The mentioned factors greatly affect the performance and robustness of the speech recognition systems.

The present methods of visual feature extraction and its illustration can be categorized into two types by considering spatio-temporal relativity. They are stationary based techniques and the movement related ones. In the previous one, a video of lip developments are split into images of sequentially arranged upon which the features such as texture and shape of lip in every frame can be utilized to represent visual features.

The effective extraction of significant features can be done by applying pre-processing procedure on the extracted images from the video. Segregating video into an image sequence is followed by extracting the mouth portion from each image of the subject. Operations are performed on the extracted images so as to extract significant features related to lip portion to be used for morphing of lip movements corresponding to the speech utterance. There are various algorithms developed for implementing synthesis of lip movements from speech namely Vector Quantization (VQ) methods and Hidden Markov Models (HMM).

Vector quantization method is a well known quantization technique [1] that lets the modeling of probability density functions by the distribution of prototype vectors. This is mainly applied for data compression application. It functions by categorizing large set of points into batches having roughly similar quantity of points which are nearest to them. Every group is indicated by its centroid like K-means and clustering techniques. In vector quantization, density matching property is effective, particularly for the recognition of density for large and multi-dimensional data represented with many dimensions. Since data points are indicated in terms of indices of their closest centroid point, frequently appearing data have less error and rarely occurring data have larger error. Hence, VQ is fit for lossy compression. Vector quantization is very close to the sparse coding as well as self-organizing map model.

The basic principle of Hidden Markov Model (HMM) is to categorize speech signal into their stochastic models where the several phonemes which construct the word shows states of the HMM while their probabilities transition may be the probability of next phoneme being uttered. Words models in vocabulary are generated during training phase. During the recognition phase when the speaker utters a word it is cut into phonemes as done during the training phase and it's HMM states are generated. After the utterance of a specific phoneme, the phoneme to follow with highest probability of occurrence is found from the probabilistic models by evaluating and comparing the probabilities using each probabilistic model. The sequence of phoneme to phoneme extends and eventually at certain stage we get the most likely word out of the existing words which the user might have uttered and hence identification is obtained in a limited vocabulary system. Thus the HMM based stochastic model for the speech signal is used in conjunction with image feature vectors to develop hybrid model for the combination of speech and lip movement images.

## III. PROPOSED METHODOLOGY

Most of the speech signals processing systems contain significant feature extraction as the main module in them. In the process of extraction of features, a small quantity of data from the voice signal is extracted and this is utilized for representing a voice signal. In the proposed methodology, speech feature is extracted using MFCC model. The objective of the system is to transform waveform of speech into parametric form for further investigation and also for signal processing. It is in general considered as front end speech processing unit. When investigated over a sufficiently small duration of time i.e. 20 - 40 m sec., the properties of a speech signal are reasonably stationary. But, for a longer duration, signal property change to indicate different speech sounds being spoken. Hence, STFT (Short Time Fourier Transform) is the most preferred technique for the analysis of speech signals. Various methods exist for parametrically representing the speech signal for the speaker identification application, they are as follows.

- LPC (Linear Prediction Coding)
- MFCC (Mel Frequency Cepstrum Coefficients).

### A. MFCC for Speech Utterance

The following block diagram of MFCC system for speech utterance is depicted in fig. 1. The input of a speech signal is acquired at a typical sampling rate equal to or above 8 KHz. When sampled at 8 KHz, these signals can record frequencies up to 4 kHz, this includes most of the energy of sounds which are produced by humans. The importance of MFCC system is to mimic the functional characteristics of the human ears. Instead of speech waveforms itself, MFFC's were considered to be less vulnerable to any changes.

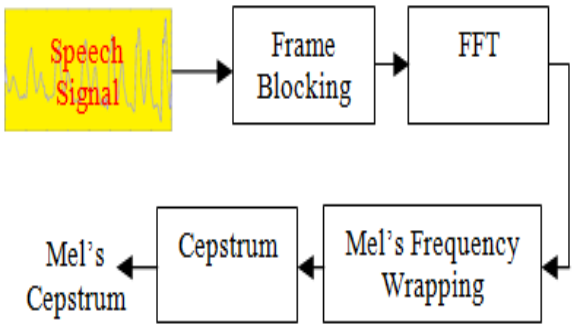


**Fig. 1. MFCC for Speech Utterance.**

For a continuously changing audio signal, in order to solve these, it is assumed that on short duration segments of an audio signal the characteristics of the signal are statistically stationary. The typical speech segment (Frame) will be 20 to 30 msec and the frame interval will be 10 msec. If the speech segment is in very small duration, there are sufficient samples to get a reliable spectral data. If the duration is wider, the changes are fast in the entire frame. Fig. 2 shows signal pattern for Mel's frequency.

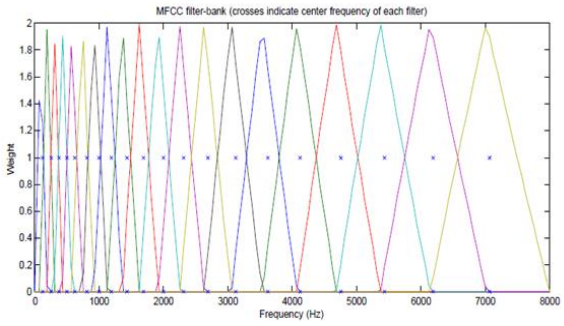


**Fig. 2. Mel's Frequencies Scale Signal**

The Mel scale maps frequency which perceived or single tone pitch to its actual calculated frequency. For discriminating small changes in pitch at low frequencies than at high frequencies humans are most suitable. Facilitating Mel scale makes the extracted features match more aptly than humans hear. The equation for mapping frequency from linear scale to Mel scale which is given by,

$$M(f) = 1125 \ln\left(1 + f/700\right) \quad (1)$$

To obtain the original frequency, the inverse equation is given by

$$M^{-1}(m) = 700\left(\exp\left(m/1125\right) - 1\right) \quad (2)$$

Here, analog speech signal segmented to blocks of *N* samples with nearby frames being separated by *M* (*M* < *N*). In this case, 1st frame contains first *N* samples and the 2nd frame starts with *M* samples after 1st frame of the 1st sample and *N* - *M* samples overlap. In the same manner, 3rd frame. This continues till the entire speech is accounted. The *N* and *M values a*re *N* = 256 that equivalent to approximately 30ms of window length and facilitate the radix-2 FFT with *M* = 100.

The further processing step is to windowing each speech segment individually so as to eliminate discontinuities in the speech signal at the starting and also at the end of every frame. The technique used here is minimization of the spectral distortion by applying the window to taper the signal to zero at the starting and end of each segment. If the window is expressed as $w(n)$, $0 \le n \le N-1$, where *N* is the number of samples in each speech segment, then the result of windowing process is

$$y_l(n) = x_l(n).w(n), \quad 0 \le n \le N\text{-}1 \quad (3)$$

In general, Hamming window is applied for windowing the speech signal. In the next processing, the FFT converts every segment of *N* samples to frequency domain from time domain. The result after this process is often referred to as spectrum or periodogram of signal.

The Psychophysical studies reveals that, perception of human for the spectral content of acoustic signals may not possess linear characteristics. Hence, for every single tone signal with an actual frequency *f* a subjective pitch or frequency is calculated using 'Mel' scale. The Mel scale is linear frequency scale below 1 kHz and a logarithmic scale above 1 kHz. The equation to calculate the Mel frequency for a given frequency is,

$$Mel(f) = 2595 Log_{10}\left(1 + f/700\right) \quad (4)$$

One method for simulating the Mel spectrum is to use a filter bank, with their centre frequencies spaced uniformly on the Mel scale. That filter bank has a triangular shaped band pass frequency response for each filter in the filter bank, and the spacing between the centre frequencies as well as the bandwidth of each filter is determined by a constant Mel frequency interval.

The feature vector of MFCC provides only single frame for power spectral envelope but speech would also have information in the dynamics i.e. trajectories of the MFCC coefficients over time. To compute delta coefficients, the following equation can be used.

$$d_t = \frac{\sum_{n=1}^{N} n.\left(c_{t+2} - c_{t-2}\right)}{2\sum_{n=1}^{N} n^2} \quad (5)$$

here, $d_t$ = delta coefficient, for the frame 't' calculated as static coefficients $c_{t+N}$ to $c_{t-N}$. Typically, the N value is 2. Delta i.e. coefficients of acceleration are computed in the same manner, but they are computed from the deltas. By applying procedure explained above for each speech segment of around 30msec with overlap, a set of MFCC is computed. This set of coefficients is referred as a *vector of acoustic feature*. Hence, each input utterance frame is transformed into an acoustic vector.

### B. Audio MFCC and Lip Tracking

Using silence detection or voice activity detection methods the time stamps corresponding to the presence of speech utterances are identified and using the obtained time stamps the image sequences corresponding to the speech utterances are extracted. Obtaining speech feature vectors (MFCC) and image Feature vectors (Geometric parameter or Spectral parameters) from the entire speech signal and extracted image sequences respectively. Fig.3 shows the general steps adopted in audio and image signal analysis.

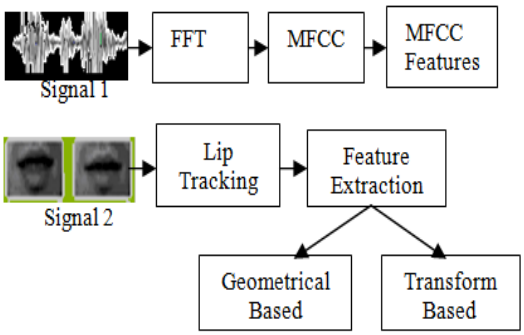


**Fig. 3. Frame work of Audio MFCC and Visual Lip Tracking.**

Video database created for Kannada language words, and created database repository. Video is segregated to audio stream and corresponding images from the dataset. Speech signal is divided into frames by windowing and MFCC features are extracted for each frame and obtained 51 features representing the speech feature and dynamics of speech that is 17 Cepstral coefficients and 17 delta and 17 delta-delta coefficients per frame, which is supposed to be correlated with lip features.

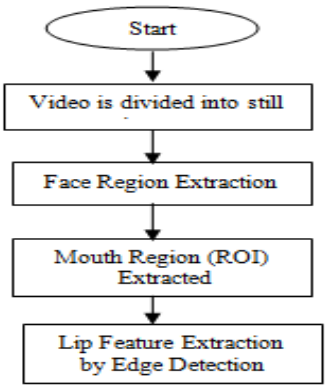


**Fig. 4. Flow chart for Mouth and Lip Feature Extraction**

Extraction of the lip features related to speech is a challenging task. Initially lip contour which was extracted identifying the mouth portion by using cascade object detector class was used. Further by using thresholding operation, the lip portion is extracted in every image and the significant geometrical aspects related to lip portion is populated in a vector. The following figures illustrate the lip contour extraction.

**(a) (b) (c)**

**Fig. 5. Results of Lip Contour Extraction: (a) Mouth Region, (b) Colour to Gray image and (c) Contour Extraction by Edge detection.**

Geometrical properties like area, centroid, major axis length, minor axis length, and perimeter are obtained now these are populated as tuples of vector per frame and used as feature vector representing lip parameters.

- Detection of vowels in speech using principle of orthogonality. The reason being it allows you to easily decouple a vector since vowels are independent has separate identity.
- Lip contour extraction from still images to obtain geometrical features.

## IV. EXPERIMENTAL RESULTS AND DISCUSSION

**Experimental Results to Determine the Utterance of Vowels:** Speech signal database related to the utterance of vowels (A, E, I, O, U) are prepared, MFCC vectors are obtained for the entire database. In the testing phase given a vowel utterance by person it can be identified by comparing Mel's frequency coefficients of speech samples with standard database. Since vowels are distinct they obtained MFCC are orthonormalized, properties of orthonormal vectors are applied to compare the normalized vectors with standard database and results are obtained with approximation. Further this method can also be applied to identify phonemes such as fricatives, nasals and stop consonants. Thus it is possible to identify the uttered words by extracting the features of speech signal. The following table shows the results for utterance of vowels.

**Table I. Features obtained using MFCC, Vowel A**

| Standard Data | Test Data | Maximum Absolute Difference | Accuracy |
|---|---|---|---|
| 0.0562 | 0.0726 | 0.0164 | 77.50% |
| 0.0516 | 0.0701 | 0.0185 | 73.70% |
| 0.0497 | 0.0669 | 0.0172 | 74.30% |
| 0.0483 | 0.0627 | 0.0144 | 77.10% |
| 0.0457 | 0.0625 | 0.0168 | 73.20% |
| **Average accuracy = 75.16 %** | | | |

**Table II. Features obtained using MFCC. Vowel E**

| Standard Data | Test Data | Maximum Absolute Difference | Accuracy |
|---|---|---|---|
| 0.0627 | 0.0814 | 0.0187 | 77.30% |
| 0.0523 | 0.0695 | 0.0172 | 75.30% |
| 0.0512 | 0.0625 | 0.0113 | 82.00% |
| 0.0481 | 0.0595 | 0.0114 | 80.90% |
| 0.046 | 0.0573 | 0.0113 | 80.30% |
| **Average accuracy = 79.16 %** | | | |

**Table III. Features obtained using MFCC. Vowel I**

| Standard Data | Test Data | Maximum Absolute Difference | Accuracy |
|---|---|---|---|
| 0.0432 | 0.0543 | 0.0111 | 79.60% |
| 0.0428 | 0.0527 | 0.0099 | 81.30% |
| 0.0457 | 0.0625 | 0.0168 | 73.20% |
| 0.0431 | 0.059 | 0.0159 | 73.10% |
| 0.046 | 0.0662 | 0.0102 | 84.60% |
| **Average accuracy = 78.36 %** | | | |

**Table IV. Features obtained using MFCC. Vowel O**

| Standard Data | Test Data | Maximum Absolute Difference | Accuracy |
|---|---|---|---|
| 0.0539 | 0.0618 | 0.0079 | 87.30% |
| 0.0538 | 0.06 | 0.0062 | 89.70% |
| 0.0459 | 0.0581 | 0.0086 | 85.20% |
| 0.046 | 0.0562 | 0.0102 | 81.90% |
| 0.0446 | 0.0552 | 0.0106 | 80.80% |
| **Average accuracy = 84.98 %** | | | |

**Table V. Features obtained using MFCC. Vowel U**

| Standard Data | Test Data | Maximum Absolute Difference | Accuracy |
|---|---|---|---|
| 0.0454 | 0.0609 | 0.0155 | 74.60% |
| 0.0443 | 0.0581 | 0.0138 | 76.30% |
| 0.0426 | 0.0575 | 0.0149 | 74.10% |
| 0.0409 | 0.0574 | 0.0165 | 71.30% |
| 0.0406 | 0.0551 | 0.0145 | 73.70% |
| **Average accuracy = 74 %** | | | |

Equation for obtaining orthogonal and orthonormal vectors is presented for reference here

$$U_i = V_i - \sum_{j-1}^{i-1} \text{Proj}_{u_i} V_i \qquad q_i = \frac{u_i}{\| u_i \|}$$

The following figures shows the Spectrum and MFCC of Utterance of vowels: 'A', 'E', 'I', 'O', 'U'.

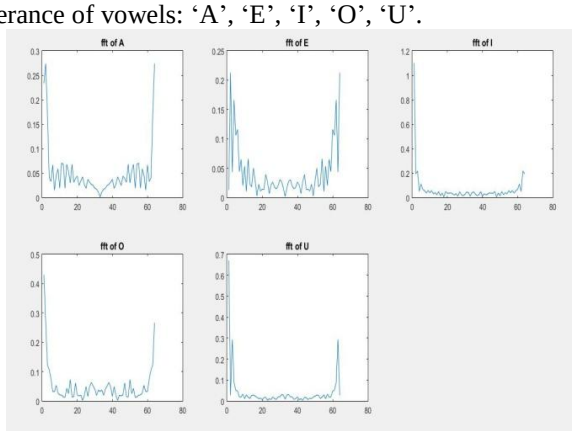


**Fig. 6. Spectrum of speech Utterance for vowels 'A', 'E', 'I', 'O', 'U'.**

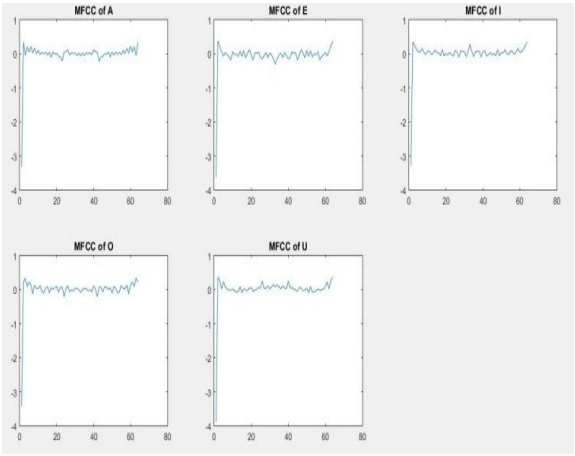


**Fig. 7. MFCC of speech Utterance of vowels 'A', 'E', 'I', 'O', 'U'.**

MFCC coefficients and their orthonormal vectors in the form of matrix

**Table VI. Cross Product to identify Vowels.**

A =

| | | | | |
|---|---|---|---|---|
| 0.0885 | 0.0649 | 0.0604 | 0.0570 | 0.0567 |
| 0.1053 | 0.0818 | 0.0697 | 0.0651 | 0.0629 |
| 0.1099 | 0.0716 | 0.0601 | 0.0523 | 0.0504 |
| 0.0961 | 0.0747 | 0.0743 | 0.0558 | 0.0539 |
| 0.1158 | 0.0682 | 0.0516 | 0.0464 | 0.0460 |
| 0.1270 | 0.0871 | 0.0746 | 0.0637 | 0.0609 |

B =

| | | | | |
|---|---|---|---|---|
| 0.0885 | 0.0649 | 0.0604 | 0.0570 | 0.0567 |
| 0.0010 | 0.0053 | -0.0015 | -0.0021 | -0.0039 |
| 0.0129 | -0.0077 | -0.0022 | -0.0053 | -0.0036 |
| -0.0014 | -0.0010 | 0.0102 | -0.0039 | -0.0036 |
| -0.0000 | 0.0002 | -0.0000 | -0.0009 | 0.0007 |
| 0.0000 | -0.0000 | 0.0000 | 0.0000 | -0.0000 |

The tables I to V illustrates that the maximum absolute difference between the standard and test vectors and hence the accuracy attained. Table VI express the correlation (similarity) between vector related to vowel A standard and test vector whereas not matching with other vowel hence and principle of orthogonally is proved. In this case it is vowel A. The calculation and results are follows.

A.A= 0.1389 A.E= $-2.583e^{-18}$ A.I= $8.619e^{-1}$
A.O= $7.589e^{-18}$ A.U=$1.722e^{-17}$

**Experimental Results for Lip Feature Extraction to Identify Uttered Vowels:**

In the noisy environment speech can be perceived understood properly when speech signal is along with the lip movements the various advantages are as discussed before. In this work, separating the images related with the vowel utterance for the detection of mouth contour.

This function creates an object model for particular region of the face and has different value for each region fixed by the software. The lip portion is extracted using Hough transformation.

The geometrical features of the lips considered are area, centroid, major and minor axis, energy and standard deviation using cascade object detector and canny edge

detection. The connected component of the lip contour of each vowel is analyzed and corresponding parameters were calculated. The values related to geometrical parameters are populated in the vector and hence determining the deviations from the standard data recorded. The following figure demonstrates the mouth portion and lip contour extraction.

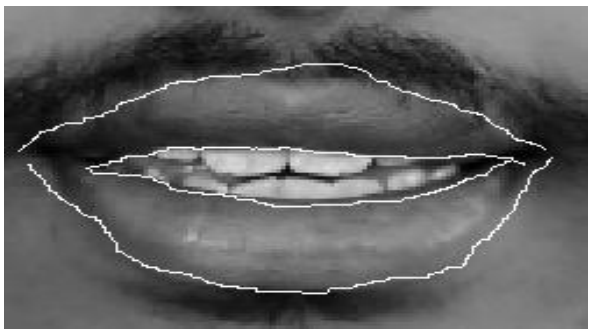

**Fig. 8. Mouth portion and lip contour extraction**

The sample results are obtained in matrix form of dimension m×1 for vowels are as shown below.

| | | |
|---|---|---|
| Area | | 0.77 |
| Centroid | | 0.64 |
| Major Axis Length | | 357.70 |
| Minor Axis Length | = | 180 |
| Perimeter | | 283.15 |
| SD | | 67.91 |
| Energy | | 0.5 |
| Mean | | 0.75 |

**Fig.9. Construction of Matrix for Vowels Sample Results**

The following Table VII shows the error calculation for the different set of frames.

**Table VII. Accuracy obtained for the given set of frames.**

| No. Frames (Set) | Std. Frame Data (Area) | Persons Frame Data (Area) | Accuracy |
|---|---|---|---|
| 1 | 0.9658 | 0.5169 | 53.52 % |
| 2 | 0.7632 | 0.7039 | 92.23 % |
| 3 | 0.7319 | 0.6531 | 92.12 % |

## V. CONCLUSION

The Experimental results obtained justify the identifying utterance of vowels, and its associated lip parameter extraction. However next step will be to improve the accuracy of speech recognition and lip tracking by incorporating MFCC feature vectors augmented delta (velocity) and delta-delta (acceleration ) coefficients which takes care of dynamics of the speech and recognition. With respect to lip parameter extraction transformation based approach like DCT /2d-DWT to be used for integrating and implementing lip movement mapping from speech signal.

In the future work will be on transform based approach for feature extraction of speech and its corresponding lip movements will be produced. We would like to express our gratitude to RNS Institute of Technology and Management for extending the support and providing R&D facilities.

## REFERENCES


1. Yamamaoto E, Nakamura S and Shikano K, "Lip movement synthesis from speech based on hidden Markov models "In proc. IEEE International Conference on Automatic Face and Gesture Recognition, pp.154-159 (1998).
2. Yamamaoto E, Nakamura S and Shikano K, " Speech to lip movements Synthesis maximizing audio-visual joint probability based on EM algorithm", .IEEE Second workshop on Multimedia Signal Processing (Cat.No.98EX175). pp. 53-58 (1998).
3. Lazaretto, F. "Converting Speech into Lip Movements "A Multimedia Telephone for Hard of Hearing People, *IEEE Trans. on Rehabilitation Engineering*, Vol. 3, NO.I, pp. 90-92 (1995).
4. Mahesh Goyani et. al "Performance Enhancement in Lip Synchronization Using MFCC Parameters", International Journal of Engineering Science and Technology, Vol. 2(6), 2010, 2364-2369.
5. Eli Yamamoto, Satoshi Nakamura, Kiyohiro Shikano, "Subjective evaluation for HMM-Based speech to lip movement synthesis", Graduate School of Information Science, Nara Institute of Science & Technology 8916-5 Takayama, Ikoma, Nara 630-01, JAPAN.
6. Sayf a. majeed,Hhafizah husain, Salina abdul samad, Tariq f. idbeaa " Mel frequency cepstral coefficients (mfcc) feature extraction enhancement in the application of speech recognition a comparison study" Journal of theoretical and Applied Information Technology 10th September 2015. Vol.79. No.1 © 2005 – 2015, ISSN:1992-8645.
7. T. Nitta 'Feature extraction for speech recognition based on orthogonal acoustic-feature planes and LDA'1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings, ICASSP99 (Cat. No.99CH36258).
8. Chung J S, Senior A,Vinyals,O& Zisserman A." Lip reading sentences in the wild".*arXiv preprint arXiv: 1611.05358*. (2016).
9. Y. Huang, X. Ding, B. Guo, H.Y. Shum, "Real-time face synthesis driven by voice", Proc. of Computer-Aided Design and Computer Graphics, Kunming, PRC, Aug., 2001.
10. T. Chen, "Audiovisual Speech Processing: Lip Reading and Lip Synchronization", IEEE Signal Processing Magazine, 2001.
11. T.A. Faruquie, A. Kapoor, R. Kate, N. Rajput, L.V. Subramaniam, "Audio driven facial animation for audio-visual reality",Proc. of ICME'01, Japan, 2001.
12. Exposito, R. Gutierrez-Osuna, P. Kakumanu, O.N. Garcia, "Optimal Data-Encoding for Speech-Driven Facial Animation", WSU Report, WSU-CS-02-04.


## AUTHORS PROFILE

**Srikanth G N,** Associate professor in department of Electronics and Instrumentation Engineering, has 21 years of experience out of which 3 and half years of R & D experience wherein designed and developed blood parameter analyzers, has17 and half years of teaching experience, published papers in the field of signal processing, speech and image processing. Instrumentation

**Dr. M K Venkatesha**, Principal (since 2008), RNSIT is a "JC Bose Gold Medalist" and has more than 35 years of teaching experience, has completed Post Graduation from the University of Manitoba (CANADA) on a fellowship. The area of interest is in signal processing and has published a number of research papers with many citations. Worked as a Registrar (Evaluation) of VTU, Belagavi (2001-04), Adviser of AICTE, New Delhi (2004-05) & Principal of BMSCE, Bengaluru (2005-08).