ISSN No. 0976-5697

Volume 8, No. 3, March – April 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Audio Visual Technique for Enhancing the Isolated Word Speech Recognition System

N.S Sreekanth
C-DAC Bangalore
#68 Electronics City
Bangalore 560100, India

N.K Narayanan
Dept. of Information Technology
Kannur University
Kerala, India

Abstract: This paper reports the enhancing of isolated word speech recognition system using audio visual features which are extracted from the context speech. Hand gestures captured which are co-expressive with speech are extracted from the context of speech. Both static and dynamic gestures are experimented as part of this study. The extracted gestural features are considered as augmenting features for improving the accuracy of isolated word speech recognition system. The performance of the system are measured under different noise conditions. The experimental results are tabulated under different signal to noise ratio conditions.

Keywords: Speech Recognition, Multimodal, Gesture recognition, Audio Visual Techniques, Non-acoustic features

I. INTRODUCTION

Various methods and techniques are implemented for improving the accuracy of Automatic Speech Recognition (ASR) system. Usage of language based knowledge is most popular among these studies for improving the accuracy of ASR systems. Stochastic based N-gram models are built to compute the next probable word which user might speak. The feature vector comparison will be done against the most probable words, and if it matches system return the result. If the variance is not within the acceptable limit, system will go for next word with higher probability returned by language model. Automatic speech recognition system build by combining acoustic and language models work well in the lab-conditions or indoor environment, i.e either noise free or less noisy environment. The performance of these systems is not acceptable in the noisy environment (noisy for acoustic signals), so deploying the practical solutions at public places may not be reliable. To ensure the reliability of speech recognition system in the noisy environment we may have to ensure an efficient noise-cancellation techniques as part of signal acquisition module i.e., FrontEnd module of the ASR systems [1] [2], [3], [4]. Another way is to extract acoustic independent features from the context of speech and model such parameters as part of feature vector. Context dependent visual features are extracted from the source and model it suitably for recognition application. Understanding and modeling the lips shapes and its movement corresponding to speech sound has been widely investigated by researches and fair results are reported for recognizing the speech in noisy environment and also recognizing the distorted speech [5], [6]. In this paper we introduce a novel method for using visual features - hand gestures that can be used for improving the recognition accuracy of isolated word speech recognition system. The visual features extracted from gestures are used as an augmenting or supporting modality for recognizing speech. Hand based gestures are recognized and fused with speech recognition system for improving the accuracy. The hand based gestures are most frequently co-existing visual feature while human communicate. Two types of gestures are experimented and integrated with speech recognition system for improving the accuracy of speech recognition. They are static gestures, dynamic gestures generated by hands. The details of gesture feature extraction are not discussed in this paper. We encourage the reader to refer the papers cited for understanding the gesture feature extraction process. The integration of multiple modalities for interacting with system not only improves the speech recognition accuracy, but also ensures the effective and natural interaction with machine (like how human beings communicate with each other).

II. REVIEW OF PREVIOUS WORK

Lips modeling is being a popular acoustic independent parameter used for speech recognition and even for speaker recognition [7]. The work done by Petajan (1984) and Mase and Pentland (1991) introduced visual information by the use of lip movement information as an important aid for speech recognition [8], [9]. Kittler et al. (1997) presented a study using geometric features of the lip shapes from model based lip boundary tracking confirming the importance of lip information in identity recognition [10]. Yamamoto et al. (1998) proposed visual information semi automatically mapped to lip movements through the aid of sensors put around the mouth to highlight the lips. The experimental results showed that significant performance could be achieved even by using only visual information [11]. Another interesting study done by Heracleous, P., et al., developed HMM based model for cued speech recognition. Lip and mouth movement based geometrical feature extracted through image processing techniques provided an effective method for recognizing speech in noisy environment. The extracted visual information will act as an augmenting feature for recognizing speech [12]. If we look at human-human way of communication, we use gestures or other forms of visual artifacts along with speech, as an augmented or support system. Especially in noisy environment e.g. inside factory, we extensively use the gestures in addition to speech for communicating with others. In such context it is found that gestures play a very important role for interpolating the missing part of speech. Providing gestures via hand, head, eyes or with other external object is very popular in human-human interaction. The visual parameters from these gestures can be extracted through computer vision (image processing) techniques or sensor based techniques and this can be used as a reliable method for designing the robust human computer interaction systems [13]. Work done by Neti, C et al., at IBM labs reports with a effective audio visual fusion techniques for

building continuous speech recognition system. The system reports performance improvements of 7% over Audio-only recognizer, by introducing the visual modality in clean speech, and 27% at an 8.5 dB SNR audio condition [14]. The work done by Vikramjit, et al., reports the extraction of vocal tract construction gestures fused with acoustic parameters experimented and reliable results were reported [16]. Another interesting work report by Ze Lei, et al., uses hand gesture combined with speech based interaction techniques in the field of control system and robotic environment for controlling machines [16] . The study done by Wu-chun, reports, the use of online-handwriting based techniques combined with speech recognition module to improve man-machine interaction techniques [17]. Detection and modeling of head, hand and arm gestures of a speaker have been studied extensively and these gestures were shown to carry linguistic information. A typical example is the head gesture while saying "yes/no". The detection of gestures is based on discrete pre-designated symbol sets, which are manually labeled during the training phase. The gesture-speech correlation is modeled by examining the co-occurrence of speech and gesture patterns [18]. The work reported by Lei Chen, et al., uses the visual patterns for repairing the distorted speech signal, from video information [19]. In this paper hand gestures extracted from the context of speech are used for recognizing the isolated words in noisy environment.

III. VISUAL FEATURES EXTRACTED FROM THE CONTEXT OF SPEECH

The approach for integrating the visual features for improving the accuracy of speech recognition system depends on the type of application or task that user would like to perform. The recognized gestures are generally substituted or interpolate a missing word or it can be redundant information in the communicated speech. Some time it can even be a deictic reference corresponding to a recognized word in continuous speech. For example if we say "*Please take that*", where the instruction will be complete only if the user point to the object to be taken. So here the pointing gesture is expected to replace the word that in the speech based communication. The approaches for integration of visual features with ASR system for isolated word recognition differ from the continuous speech recognition systems. In both the context the gestures are equivalent to a word or a phrase in speech. In isolated word recognition system it can have a bijective relationship between speech and gestures; it means we expect one-to-one correspondence for each word in speech and gestures. Digit recognition systems, system used for controlling devices, robotic control environment are some examples for this. Whereas for the continuous speech recognition system a bijective relationship between the individual words in the communicated speech with gesture is not guaranteed always. For example, if we call someone to come near to you, you may call "*Could you please come here*" and also show the gesture, i.e, wave your hand with an action of calling. In this case the gesture shown by the person is only equivalent to the phrase "*come here*" but not corresponding to "*can you please*". In this paper we discuss about the inclusion of gestural features which are co-expressed with speech for recognizing the isolated word - speech recognition system.

As discussed earlier, the gestures are co-expressive with speech when human communicate each other. We have used the hand gestures as part of this experiment. The gestures produced by hand during the communication may be either static or dynamic. Static gestures are the gestures where the movement of body part does not involve. For example when we say “three” and also show three fingers while speaking is treated as static gesture. For dynamic gestures, the body part also moves and the orientation and direction of movement of the object take part in the gesture production determine the gesture produced. User says “*go to next page*” and shows gesture of moving the hand from left to right is a good example for dynamic gesture. Similarly user can say a number or a letter and can write in the air with finger. The movement of finger in the air can be tracked and feature vector can be generated. These types of gestures also considered as dynamic gesture. Static gestures are recognized using the orientation and geometrical properties of hand and fingers. Skin colour based segmentation is used to segment the hand form the complex image [20]. The convex hull algorithm returns the enclosed polygon of the hand and along with the fingers orientation. Various geometrical features like area, perimeter, centroid, solidity along with histogram distribution is used for recognizing the static gestures [21]. For recognizing dynamic hand gestures we have used the skin color based segmentation algorithm for segmenting the hand form the complex image frames. The finger tips will be identified form the segmented frames using the convex hull algorithm. The movement of finger will be tracked and the feature vector will be generated using Freeman’s eight directional codes (popular algorithm used for online handwriting recognition) [22]. A dynamic time wrapping based Leventian Minimum Edit distance algorithm is used for classification of incoming gestural features.

IV. ISOLATED WORD SPEECH RECOGNITION SYSTEM USING DYNAMIC TIME WARPING ALGORITHM - WITHOUT USING THE GESTURES

The acoustic features are extracted from speech signals for recognizing the speech. In this experiment the frequency domain based Mel Frequency Cepstral Coefficients (MFCC) features and the time domain based Phase Space Point Distribution (PSPD) parameters are used for speech recognition. The Thirteen MFCC coefficient (12 coefficients and one energy parameter) and it's first and second derivatives generate 39 size frequency based features. These are appended with 20 sized PSPD feature vector extracted from a fixed frame or window. Based on the studies [23], [24] it is reported that the optimum frame length or window size for the extraction of features are found to be 256 samples per frame. It is reported that increasing the window size will reduces the recognition accuracy of speech recognition systems [25]. The extracted features from various trails are saved and this will be used for recognizing the incoming signals during testing phase. Dynamic Time Wrapping (DTW) algorithm is used for recognizing the speech i.e., here isolated words. Dynamic time warping algorithm, which is based on dynamic programming, is a technique that calculates the level of similarity between two time series data in which any of them may be warped in a non-linear fashion by shrinking and stretching the time axis [26]. Dynamic time warping is an algorithm for measuring similarity between two sequences which may vary in time or speed. DTW operates by storing a reference version of each word in the vocabulary into the database, then compares incoming speech signals with each word and then takes the closest

match. The distortion or edit cost will be calculated between the incoming feature vectors with reference vector for recognition. Let $X(x_1, x_2, \ldots, x_n)$ and $Y(y_1, y_2, \ldots, y_m)$ be the two time series with the length of n and m, respectively, and an $n \times m$ matrix M can be defined to represent the point-to-point correspondence relationship between X and Y, where the element M[i,j] indicates the distance $d(x_i, y_j)$ between x_i and y_j [27] which is the Euclidian distance. In the speech recognition case X can be treated as the reference feature vector and Y be the incoming feature vector for testing. The Dynamic Time Wrapping algorithm compute the similarity or distortion between two time series signal as shown below [28]. Any element M_{ij} in the accumulated matrix indicates the dynamic time warping distance between series $x_{1:i}$ and $y_{1:j}$. Series with high similarity can be effectively identified because the best alignment and matching relationship between two series is defined by the dynamic time distance. The distance threshold was fixed as 200 by trial and error method. The word with minimum cost i.e., will be chosen as recognized word. If the distortion is higher than a threshold system ignores that word.

ALGORITHM - Dynamic Time Wrapping

1. *n = |X|*
2. *m=|Y|*
3. *M[] = X × Y*
4. *M[0,0] =0*
5. *Repeat For i=1 to n with an increment of i= i+1 through Step 7*
6. *M[i,1]=M[i-1,1]+c[i,1]*
7. *END OF FOR*
8. *Repeat For j = 1 to m with an increment of j=j+1through step 10*
9. *M[1,j]=M[1,j-1]+c[1,j]*
10. *END OF FOR*
11. *Repeat For i = 1 to n with an increment of i=i+1 through step15*
12. *Repeat For j = 1 to m with an increment of j=j+1 through step 14*
13. *M[i, j]= c(i, j)+min {M[i− 1, j]; M[i, j − 1]; M[i − 1, j − 1]}*
14. *END OF FOR*
15. *END OF FOR*
16. *return M[]*

V. Speech and Gesture Integrated Environment - Enhanced Isolated Word Recognition System

The conventional speech recognition system is enhanced by incorporating the visual features for improving the accuracy of speech recognition. The input signal from the microphone will be given to the speech recognition module. The relevant feature vectors are extracted from the speech, i.e., MFCC and PSPD features. If user shows gestures (static or dynamic) corresponding to spoken word with hand, that will also be captured by web camera and relevant features will be extracted. Both the speech and gesture recognition systems independently recognize the word corresponding to speech and the gesture respectively. The conclusion of results will be done by the system as discussed below.

Let us assume that user said a word which is a part of the vocabulary and the corresponding gesture is also shown. The feature vector corresponding to the input speech signal is extracted and the distortion vector is calculated. Similarly the distortion vector for gesture also is calculated. The distortion vector returned by both speech and gesture are normalized to a scale of 0 to 1. Let $ds_1, ds_2, ds_3 \ldots ds_n$ be the respective distortion vectors for $W_1, W_2, \ldots W_n$ against the input speech signal. So the list of words along with the distortion corresponding to each word in vocabulary returned by speech recognition module will be $<w_1, ds_1>$, $<w_2, ds_2> \ldots\ldots <w_3, ds_n>$, Similarly the gesture also will return the distortion vector corresponding to the elements in the data base if user issues corresponding gestures. $<W_1, dg_1> <W_2, dg_2> \ldots\ldots\ldots\ldots\ldots\ldots\ldots <W_n, dg_n>$.

The result return by speech recognition engine W_i

$$W_i = \forall i \ \{return(i) \text{ with } \{min< Wi,dsi>\}\} \quad \text{-----1}$$

The result return by gesture recognition engine is W_k

$$W_k = \forall k \ \{return(k) \text{ with } \{min< W_k,dg_i>\}\} \quad \text{-------} \ 2$$

If $i=k$ i.e both the recognizer unit return the same word then, that result will be more authenticated. If $i \neq k$, i.e., two different word indices are returned, then the new word index will be chosen by finding the difference between the speech distortion vector and gesture distortion vector. If the difference is above a predefined threshold value T then select the word index with minimum distortion as follows

$$Wj = \forall i,k \ (Min \ \{<W_i, ds_i> - <W_k, dg_k>\}) \ if \ |ds_i - dg_k| > T \text{---- } 3$$

If the difference between the distortions vectors are less than a predefined threshold system returns both words as a suggested word so that user can pick the desired one. The Flow chart for the speech and gesture integrated environment is shown Fig. 1

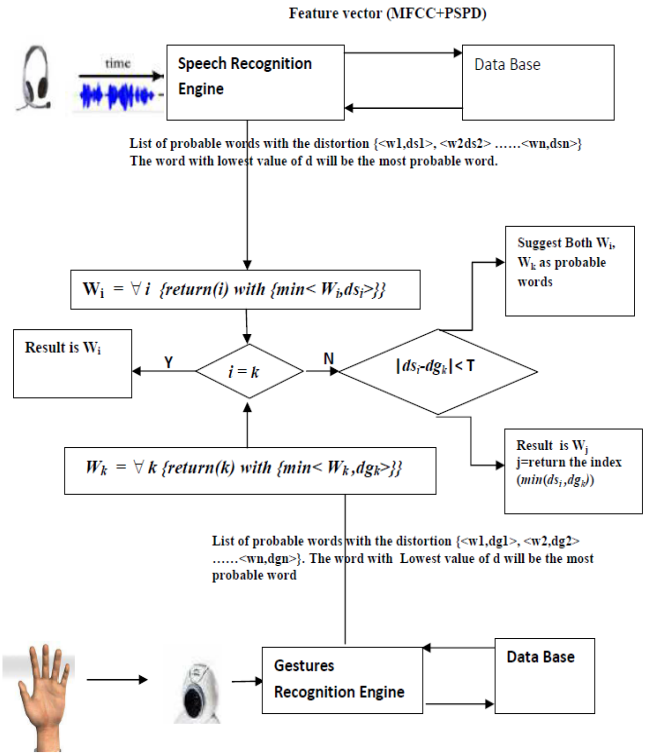

Figure 1. Speech and Gesture Integrated Environment.

VI. SIMULATION EXPERIMENT AND RESULTS

The enhanced isolated word recognition system using visual features, i.e., the feature extracted from static and dynamic gesture is simulated in the study. The forty operational commands, which are used for interacting with computer system and ten digits constitute the database for this experiment. The training data base is created with speech and gesture data acquired from 28 individuals (15 male and 13 female). Thirty speech and gesture samples per word are collected from the individuals, and given to the corresponding system for training. The recognition accuracy of isolated word recognition system without considering the gestural features and with gestural features is tested under different noise conditions. The system is tested with different signal to noise ratios for speech signals by adding Additive White Gaussian Noise(AWGN). Both dynamic and static gestures are used with speech for testing the performance of the system. The dynamic gestures are experimented with special colour (blue) and without special color (skin colour) as discussed in the cited paper [29]. The testing of the system is carried out at different occasions with different signal to noise ratio conditions. Table I shows the data base used for experiment. Table II to VI gives the experimental results under various conditions without gestures and with gestures.

Table I. List of Isolated operational keywords used for experiments.

Operational Commands			
Volume Up	Undo	Open	Start
Volume Down	Zoom-in	Close	Save as
Move down	Zoom-out	Save	Go to
Move Up	Maximize	Sent	Print
Move back	Minimize	Attach	Mail
Move next	Cut	OK	Help
Rotate right	Copy	Find	Properties
Rotate-left	Paste	Restart	Change
Delete	Logoff	Shutdown	Page up
select	Cancel	Redo	Page down

Table II. Overall Recognition accuracy without gestural features under various noise conditions

SNR (Speech Signal +AWGN)	*Clean*	*30dB*	*20dB*	*10dB*	*3dB*	*0dB*
Recognition Accuracy	**91 %**	**80.5%**	**75%**	**62 %**	**41%**	**35%**

Table III. Over all Recognition accuracy with Static Gestures under Different Nose Conditions

SNR (Speech Signal +AWGN)	*Clean*	*30dB*	*20dB*	*10dB*	*3dB*	*0dB*
Recognition Accuracy	**96.7 %**	**92.6%**	**89%**	**88 %**	**87.6%**	**86%**

Table IV. Over all Recognition Accuracy with Dynamic gestures with special colour

SNR (Speech Signal +AWGN)	*Clean*	*30dB*	*20dB*	*10dB*	*3dB*	*0dB*
Recognition Accuracy	**95.4 %**	**91%**	**87.3%**	**85 %**	**83.4%**	**84%**

Table V. Over all Recognition Accuracy with Dynamic gestures without Special colors.

SNR (Speech Signal +AWGN)	*Clean*	*30dB*	*20dB*	*10dB*	*3dB*	*0dB*
Recognition Accuracy	**94 %**	**89.4%**	**84%**	**82.7 %**	**82.1%**	**81.6%**

VII. CONCLUSSIONS

A method for improving the accuracy of automatic speech recognition system by adding non-acoustic parameters are discussed in this paper. The gestures which are co-expressed with speech are considered for improving the accuracy of ASR system in noisy environment. Both dynamic and static gestures are integrated with speech recognition system (isolated word recognition system) and tested in various noise conditions, i.e, signal to noise levels. The addition of visual features provides stable recognition accuracy under different environmental noise conditions for acoustic signals. The approach and method applied for improving the recognition accuracy is to be seen objectively, i.e, no matter the communicated message is recognized by speech recognizer or gesture recognizer but the purpose of communication need to be met. Hence the proposed method provides a natural and effective interaction mechanism for interacting with machines. Further studies can be conducted on the inclusion of non-acoustic features for improving the accuracy of continuous speech recognition system.

VIII. REFERENCES

[1] Wouters, Jan; Vanden Berghe, Jeff "Speech Recognition in Noise for Cochlear Implantees with a Two-Microphone Monaural Adaptive Noise Reduction System"- Ear & Hearing: Journal of American Auditory society. October 2001 - Volume 22 - Issue 5 - pp 420-430. 2001

[2] Sivadasan Kottayi & N.K. Narayanan , A new online secondary path modelling for feedforward ANC systems, International Journal of Electronics Letters (2013): Journal of Electronics Letters, DOI: 10.1080/00207217.2013.858589., 2013.

[3] Nordholm, S. ; Claesson, I. ; Bengtsson, B."Adaptive array noise suppression of hands free speaker input in cars", Vehicular Technology, IEEE Transactions on (Volume:42 , Issue: 4) August 2002

[4] Dong Yu, Li Deng ; Droppo, J. ; Jian Wu ; Gong, Yifan ; Acero, A. "A minimum-mean-square-error noise reduction algorithm on Mel-frequency cepstra for robust speech recognition" Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on DOI: 10.1109/ICASSP.2008.4518541. pp. 4041 - 4044 , 2008

[5] Borgstrom, B.J.; Alwan, Abeer "A Low-Complexity Parabolic Lip Contour Model With Speaker Normalization for High-Level Feature Extraction in Noise-Robust Audiovisual Speech Recognition", IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, Volume: 38, Issue: 6 pp 1273 – 1280, 2008

[6] Hazen, T.J Visual model structures and synchrony constraints for audio-visual speechrecognition IEEE Transactions on Audio, Speech, and Language Processing, Year:, Volume: 14, Issue: 3 pp. 1082 – 1089., 2006.

[7] Maycel Isaac Faraj, Josef Bigun, "Lip Motion Features for Biometric Person Recognition" Book chapter of Medical Information Science Reference, IGI Global, Chapter XVII, pp495-532. , 2009.

[8] Petajan, E. "Automatic lipreading to enhance speech recognition". Global Telecommunications Conference. (pp. 265-272). 1984.

[9] Mase, K., & Pentland, A. "Automatic lip-reading by opticalflow analysis". J. Systems and Computers in Japan, 22(6), 67-76., 1991.

[10] Kittler, J., Li, Y., Matas, J., & Sanchez, M.. Combining evidence in multimodal personal identity recognition systems. Proceedings of the First 48 International Conference on Audio- and Video-Based Biometric Person Authentication, LNCS 1206, pp. 327-334., 1997

[11] Yamamoto, E., Nakamura, S., & Shikano, K.. Lip movement synthesis from speech based on hidden markov models. Journal of Speech Communication, 26(1), 105-115.,1998.

[12] Heracleous. P, Aboutabit. N, Beautemps D." Lip Shape and Hand Position Fusion for Automatic Vowel Recognition in Cued Speech for French", IEEE Signal Processing Letters, MAY 2009,VOL. 16, NO. 5, 2009

[13] Mitra, V; Hosung Nam ; Espy-Wilson, C.Y. ; Saltzman, E. ; Goldstein, L"Gesture-based Dynamic Bayesian Network for noise robust speech recognition", IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). pp.5172-5175, IEEE-DOI 10.1109/ICASSP.2011.5947522, 2011.

[14] Neti, C Potamianos, G. ; Luettin, J. ; Matthews, I. ; Glotin, H. ; Vergyri, D." Large-vocabulary audio-visual speech recognition: a summary of the Johns Hopkins Summer Workshop ", IEEE Fourth Workshop on Multimedia Signal Processing, 2001 , PP 619 – 624, 2001.

[15] Vikramjit Mitra , Hosung Nam, Elliot Saltzman, Louis Goldstein "Recognizing articulatory gestures from speech for robust speech recognition", Journal of . Acoustic. Soc. America . Vol 131 issue 3,pp 2270-2287 March 2012.

[16] Ze Lei, ZhaoHui Gan, Min Jiang, Ke Dong"Artificial Robot Navigation based on Gesture and Speech Recognition", Proc. International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), 2014, pp. 323 - 327, 2014 .

[17] Wu-chun Feng "An integrated multimedia environment for speech recognition using handwriting and written gestures", Proceedings of the 36th Annual Hawaii International Conference on System Sciences, IEEE DOI : 10.1109/HICSS.2003.1174293., , 2003.

[18] Sargin, M.E., Aran, O. ; Karpov, A. ; Ofli, F. ; Yasinnik, Y. ; Wilson, S. ; Erzin, Yemez Y. ; Tekalp, A.M." Combined Gesture-Speech Analysis And Speech Driven Gesture Synthesis", IEEE International Conference on Multimedia and Expo, 2006, pp. 893 - 896, July, 2006

[19] Lei Chen ; Harper, M. ; Quek, F. "Gesture patterns during speech repairs", Proceedings of Fourth IEEE International Conference on Multimodal Interfaces, pp. 155 - 160, DOI 10.1109/ICMI.2002.1166985., 2002.

[20] Lei Yang, Hui Li, Xiaoyu Wu, Dewei Zhao, Jun Zhai.―An algorithm of skin detection based on texture IEEE Image and Signal Processing (CSIP), 2011.

[21] Noor Adnan Ibraheem, RafiqulZaman Khan "Survey on Various Gesture Recognition Technologies and Techniques", International Journal of Computer Applications (0975–8887), Volume 50 – No.7, July 2012, pp. 38–44.

[22] B.J Manikandan, Gowri Shankar, V Anoop, A Datta, V S Chakravarthy: LEKHAK: A System for Online Recognition of Handwritten Tamil Characters. Proceeding of the International Conference on Natural Language Processing (ICON-2002) Vikas Publishing House Pvt. Ltd. pp. 285–291.

[23] P. Prajith, Investigations on applications of Dynamical Instabilities and Deterministic Chaos for Speech signal Processing. PhD Thesis, University of Calicut., 2008.

[24] Thasleema T.M, Computer recognition of V/CV speech units based on linear and non-linear dynamical system models using brain like computing and statistical learning algorithms, PhD thesis, Kannur University, 2012.

[25] Deividas Eringis, Gintautas Tamulevičius, "Improving Speech Recognition Rate through Analysis Parameters" The Scientific Journal of Riga Technical University - Electrical, Control and Communication Engineering Vol 5, pp-61-66., 2014.

[26] Khalid A. Darabkh, Ala F. Khalifeh, Baraa A. Bathech, and Saed W. Sabah "Efficient DTW-Based Speech Recognition System for Isolated Words of Arabic Language" World Academy of Science, Engineering and Technology, Vol:7, 2013.

[27] Pavel Senin "Dynamic Time Warping Algorithm Review", Information and Computer Science Department University of Hawaii at Manoa. http://www2.hawaii.edu/~senin/assets/papers/DTW-review2008draft.pdf, 2008.

[28] B. Plannerer "An Introduction to Speech Recognition ", pp.27-47, March 28, 2005

[29] N.S Sreekanth, N.K Narayanan "Dynamic Gesture Recognition - A Machine Vision based Approach", Proceedings of International Conference on Signals, Networks, Computing and Systems" ICSNCS-2016, Lecture Notes in Electrical Engineering Series , Springer Verlag, Volume 395, 2017 , pp-105-115 held at JNU, New Delhi