International Journal of Pure and Applied Mathematics
Volume 118 No. 24 2018
ISSN: 1314-3395 (on-line version)
url: http://www.acadpubl.eu/hub/
Special Issue

http://www.acadpubl.eu/hub/

Automated Speech-Image Coding: Hardware Interface Realization

Dr. Ibrahim Patel
[1]Associate Prof.
Dept of ECE.
B. V. Raju Institute of Technology,
Narsapur, Medak, Telengana, India
ptlibrahim@gmail.com,

May 1, 2018

Abstract

Voice communication is the most effective mode of communication used by common people.The limitation of inability to speak isolates the deaf and dumb people from the community. In case a normal individual is unaware of sign language then the interaction becomes more difficult.The basic problem in developing a system which bridges the gap between able and the disable is that the sign language is constraint to vocally disabled community. To overcome these difficulties an electronic interface system is developed which provides interaction by means of coding and display interface. The developed hardware goes through an algorithmic approach to improve estimation accuracy for speech interface; a hybrid model which interfaces speech code with image mapping technique and a visual interface logic to generate equivalent sign sequence using knowledge database system. The work was developed by using Raspberry Processor with Python tool.

***Key Words*:**Python tool, Raspberry pi, Finger Spelling Recognition System,SNR.

1 WORK ETHICS

Technology is one possible solution to remove various hindrances and benefit the disabled in all types of communication modes. Visual communication consists of three major components:

Finger Spelling Recognition System such as BoltayHaath It has a limitation of One Way of Communication.

Word Level Sign Language such as Instrumented Gloves by Australians - developed systems shown limitation towards Intelligibility which were lost due to quantization distortions and spatial resolution reduction and also it demands larger bandwidth.Vocabulary limitations, accurate rate of cue symbol generation, limitations towards handling of dynamic speech patterns are some of the limitations seen by the researchers.

Non-Manual Features Methods such as Lip Reading, Action-to-Speech(A2S) etc. Degrades speech perception due to reduced frequency resolution and fails to discriminate between different speech units at low SNR and put architectural limitations.The limitations with respect to synchronization between speech, cue symbol and display were largely seen in todays developed modules. It can be concluded that there is an absence of natural and efficient hardware and it is a fact that there is no effective hardware interactive machine mechanism is available for transforming information from a common man to a disable and vice-versa.This work has been developed by taking a note of all the previous drawbacks and is developed to provide interaction from normal individual to vocally disabled individual without the knowledge of sign language.

2 SYSTEM FLOW CHART

The system flow chart figure 1 developed for the said work depicts the entire process of processing of speech signal to hardware interfacing. Part A depicts the speech processing up to mapping and part B depicts its interfacing with hardware.

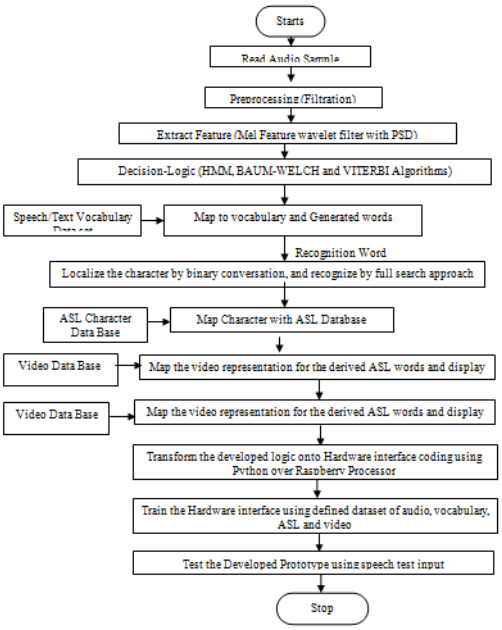

Fig.1: Operational Flow Chart

3 SYSTEM ARCHITECTURE

The proposed system is developed in four basic units:User interface unit, auditory interface unit, Processing Unit, Display unit. The system architecture of the proposed hardware system is outlined below figure. 2:

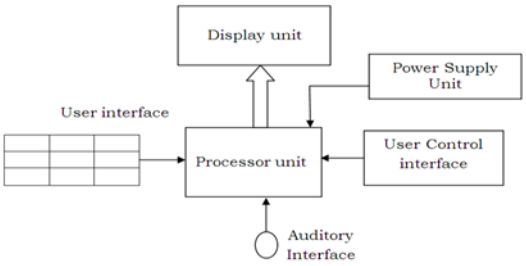

Fig. 2: System architecture for Developed hardware unit.

4 OPERATIONAL ALGORITHM

The proposed steps of the algorithm are

- Step 1: initialize the power supply.
- Step 2: reset the setup.
- Step 3: press train button from control interface to switch the unit to training operation.
- Step 4: pass the training audio samples from MIC interface.
- Step 5: buffer the cue symbols to processor memory.
- Step 6: train the vocabulary characters to the system.
- Step 7: reset for testing mode.
- Step 8: press enable button to initialize.
- Step 9: take i/p of speech test from MIC interface.
- Step 10: pass to the processing unit for detection.
- Step 11: the features from the passed speech signal is extracted and mapped.
- Step 12: detected speech is displayed as text.
- Step 13: cue mapping is performed.
- Step 14: corresponding cue symbols in characters are displayed.
- Step 15: words are formed and buffered video display is shown.

5 EVALUATION

- Speech Signal Interfacing

  Test speech sample interfaced for the modeled unit

- Recognition Process

  The recognized text information for a given test speech signal DO YOU HAVE WATER, is displayed on the interfaced Display unit. The illustration of this process is presented in figure 3.

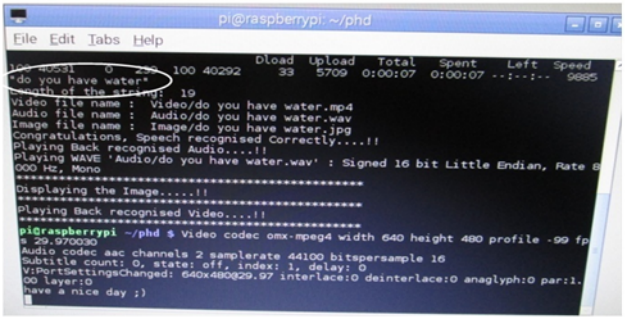

Fig. 3: Display showing recognized speech signal

- Video Illustration

Fig. 4: Obtained video Sequence for the word DOYOU HAVEWATER

6 IMPLEMENTATION FLOWCHART

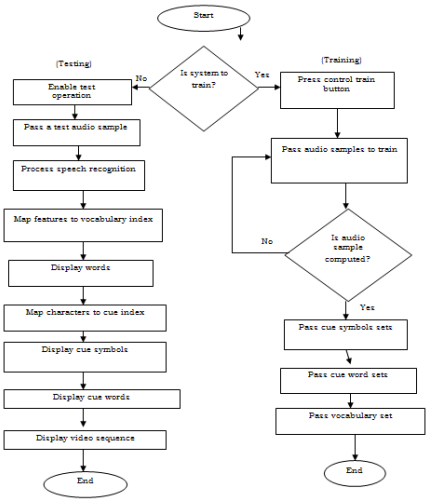

Fig. 5: Operational flow chart for the developed system

7 IMPLEMENTATED MODULE AND FINDINGS

To model the proposed system architecture, the hardware interface unit developed is shown in figure 6. A user interface unit is developed as illustrated in figure 6 for the testing of the developed speech recognition and transformation unit for communication interface. In the developed prototype the hardware units are developed as an integrated model of sub processing units. The system is developed with the following sub section units.

Fig. 6: Developed Hardware prototype unit

Analog Acoustic Signal is applied to the input of the sound card of the system which digitizes the speakers speech waveform at a sampling rate of 8 KHz with sampling period of 10 sec. This digitized speech is stored in memory.Analog video image signal representing speakers face is captured by a frame grabber card where it is digitized at a rate of 30 video images per second. This digitized image is stored in memory. Bandwidth requirement is met by using a 32 bit wide peripheral connect interface (PCI) which offers data rate of 120 MB per second by using DMA.Synchronization between the displayed cues and the speakers speech and/or facial expression allows deaf people to understand a speaker more easily because the cue is displayed substantially at the same time as when the cued syllable is pronounced. Digital speech samples are continuously processed by main processor to recognize the phones associated with successive segments of the speech of the speaker. The resulting stream of recognized phones is mapped to a sequence of cue labels, each specifying a synthetic cue shape and position. After mapping, the cue sequence is used to generate composite video images which are played back.Speech elements initially delivered by a speaker are recognized. A sequence of video images are displayed showing the speaker delivering the speech elements. The displayed sequence of video images is delayed relative to the initial delivery of the speech elements by the speaker. In conjunction with displaying the sequence of video images, an image of one of the cues corresponding to the recognized speech elements is displayed with a timing that is synchronized to a visible facial action.

8 CONCLUSION AND FUTURE SCOPE

For the realization of suggested automated speech recognition and cue symbol generation a progressive coding scheme based on integrated model of Markov speech coding and image processing is suggested. The implementation of the suggested work is evaluated over various speech samples with the approach of Markov modeling. A spectral feature for the speech signal is taken from a real time recording environment and its equivalent coding approach is developed. The speech recognition system uses a modified spectral feature based on MFCC codes with sub band based coding approach for feature training and its recognition. a hardware realization over raspberry processor with visual interface is developed.A audio interface for processing to interface with user interface and controlling is developed. The functional validation to such coding is made under different test inputs and the validation is found accurate in recognition and displaying.

The developed system is outlined for the usage of automated cue symbols generation using automated speech-image coding system. The process of image mapping for remote application is proposed in this work.A extension for long range presentation in wired or wireless mode can be developed with medium effects. Additionally the work can also be extended for hardware designing of the developed approach for real time applications.

References

[1] Aleem Khalid, Ali M, M. Usman, S. Mumtaz, Yousuf Bolthay-Haath Paskistan sign Language Recgnition CSIDC 2005.

[2] Kadous, Waleed GRASP: Recognition of Australian sign language using Instrumented gloves, Australia, October 1995, pp. 1-2, 4-8.

[3] Zhang Jie; Huang Zhitong; Wang Xiaolan; Selection and analysis of HMM's state- number in speech recognition 12-16 Oct. 1998 Page(s):641 - 645 vol.

[4] D. E. Pearson and J. P. Sumner, An experimental visual telephone system for the deaf, J. Roy. Television Societyvol.16, no. 2. pp. 6-10, 1976.

[5] Guitarte Perez, J.F.; Frangi, A.F.; Lleida Solano, E.; Lukas, K. Lip Reading for Robust Speech Recognition on Embedded Devices Volume 1, March 18-23, 2005 Page(s): 473 476

[6] G. S. Sperling, Bandwidth requirements for video transmission of American sign language and finger spelling, Scienc vol. 210, pp. 797-799, 1980.

[7] A. N. Netravali and J. 0. Limb, Picture coding: A review, Proc. IEEE. vol. 68, pp. 366-406, Mar. 1980.

[8] Yunxin Zhao, Maximum likelihood joint estimation of channel and noise for robust speech recognition Acoustics, Speech, and Signal Processing, 2000.ICASSP '00.Proceedings. 2000 IEEE International Conference Volume 2, 5-9 June

[9] K. Woo, T. Yang, K. Park, and C. Lee, Robust voice activity detection algorithm for estimating noise spectrum, Electronics Letters, vol.36, no. 2, pp. 180181, 2000.

[10] T. Masuko, K. Tokuda, T. Kobayashi, and S. Imai, Speech synthesis using HMMs with dynamic features, in Proc. ZCASSP-96, May 1996, pp. 389-392.

[11] Suebvisai, S.; Charoenpornsawat, P.; Black, A.W.; Woszczyna, M.; Schultz, T. Thai Automatic Speech Recognition Acoustics, Speech, and Signal Processing, 2005.Proceedings. (ICASSP '05). IEEE International Conference Volume 1.

[12] SantoshKumar, S.A.; Ramasubramanian, V. Automatic Language Identification Using Ergodic HMM Acoustics, Speech, and Signal Processing, 2005.Proceedings. (ICASSP'05).IEEE International Conference Vol1, March18-23, 2005 Page(s): 609-612

[13] V. Ramasubramanian, A. K. V. SaiJayram, and T. V. Sreenivas Language identification using parallel sub-word recognition an ergodic HMM equivalence., In Proc. Eurospeech, pages 13571360, Geneva, Switzerland, Sep 2003.

[14] M. A. Zissman Automatic language identification using Gaussian mixture and hidden Markov models. In Proc. ICASSP, pages 399402, Apr 1993.

[15] Nakamura, S.; Markov, K.; Nakaiwa, H.; Kikui, G.; Kawai, H.; Jitsuhiro, T.; Zhang, J.-S.; Yamamoto, H.; Sumita, E.; Yamamoto, S. The ATR Multilingual Speech-to Speech Translation System Page(s): 365- 376

[16] Yamamoto, E.; Nakamura, S.; Shikano, K. Lip movement synthesis from speech based on hidden Markov models Automatic Face and Gesture Recognition, Third IEEE International Conference on 14-16 April 1998 Page(s):154 159

[17] Devarajan, M.; FanshengMeng; Hix, P.; Zahorian, S. A. HMM-neural network monophone models for computer-based articulation training for the hearing impaired Acoustics, Speech, and Signal Processing, 2003. (ICASSP '03). 2003 IEEE International Conference on Volume 2, 6-10 April 2003 Page(s): II - 369-72

[18] Zahorian S., Zimmer M., MengF(2002) Vowel Classification for computer-based visual feedback for speech training for the hearing impaired, International Conference on Spoken Language Processing

[19] C. J. Leggetter and P. C. Woodland. Speech Adaptation of HMMs Using Linear Regression. Technical Report TR. 182, Cambridge University, 1994.