Research Journal of Recent Sciences E-ISSN 2277-2502
Vol. **5(5),** 47-52, May **(2016)** *Res. J. Recent Sci.*

Emotion Detection based on the Hidden Markov model Chain Speech Recognition

Safi Seyyed Mohammad[1*] and Aynehband Meghdad[2]
[1]Department of Computer Engineering, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran
[2]Department of Computer Engineering, Hendijan Branch, Islamic Azad University, Hendijan, Iran
m_saifi85@yahoo.com

Available online at: www.isca.in, www.isca.me
Received 3rd August 2015, revised 7th December 2015, accepted 7th January 2016

Abstract

Detection of user mode is one of the main arguments used in all systems such as expert systems, as a significant Parameter. Hidden Markov model is one of the most important models in speech recognition, that the several strong researches confirmed this method. This research attempts to recognize person's voice based on the mathematical model of the emotional detection by recognizing its voice.

Keywords: Mode Detection, (HMM) Hidden Markov Model, Emotion Detection, Speech Recognition Systems.

Introduction

Expert system needs speech recognition tools for interactive behavior to the user as input method; this Research Article is an attempting on human's behaviors recognition. There are many studies have been conducted and several systems have been proposed based on this plan. Van den Broek E. and Westerin recognized the person's emotion by using a system based on the Galvanic Skin Response(GSR) and signals electromyography (EMG) during the 24-hour that was taken at every two minutes and used them as context of pervasive systems. Signals was measured during playback of different films to implement this system[1] (Figure-1).

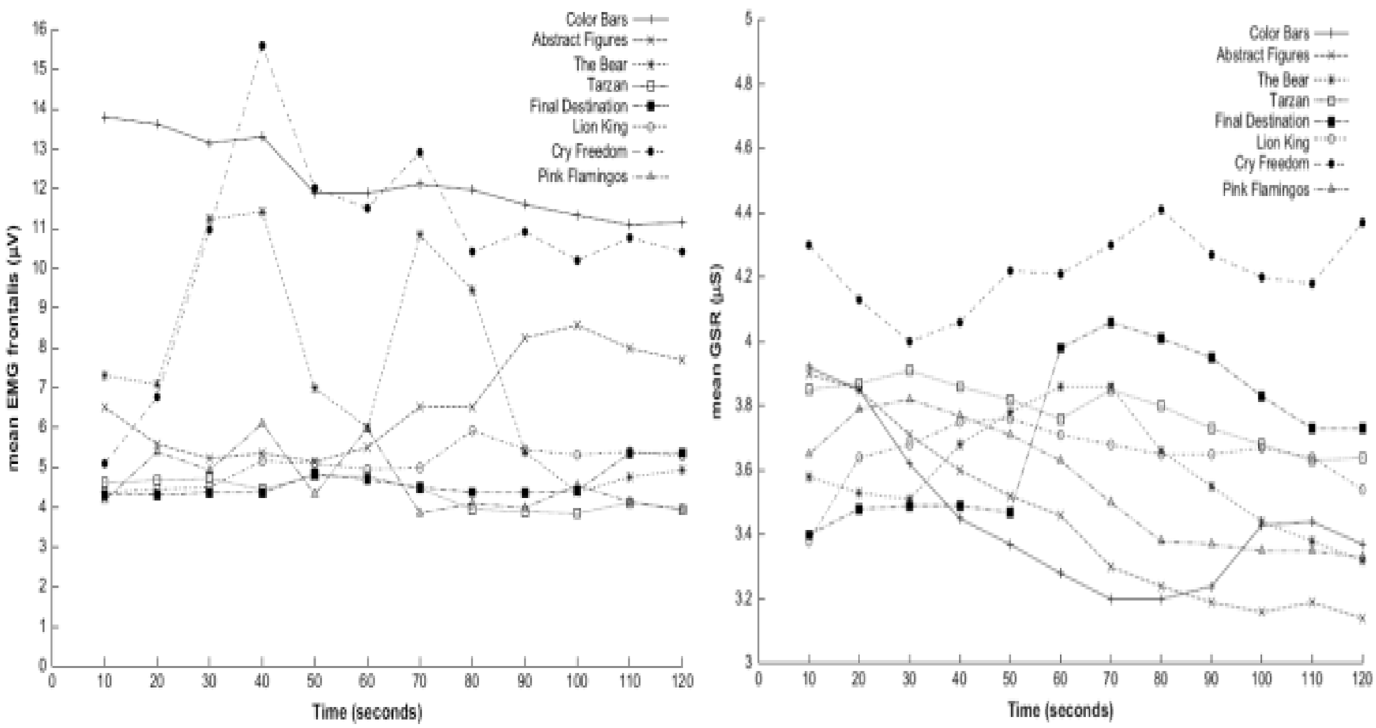


Figure-1
GSR, EMS Changes and distribution during playing movies for the emotion detection[1]. Hong J*et al*., invented messaging system that by wearable sensors could detect the activity and emotion of ones and electronically published them[2]

Hong J. *et al*. invented messaging system that by wearable sensors could detect the activity and emotion of ones and electronically published them[2].

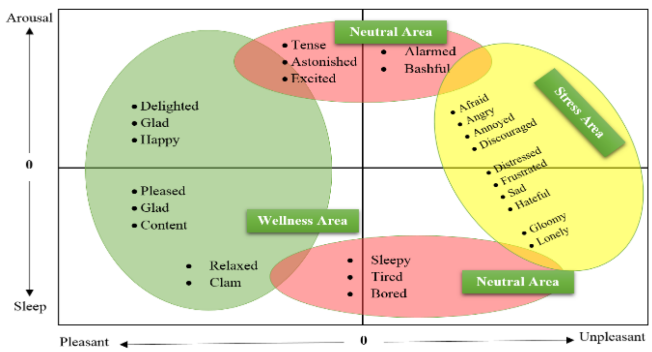


Figure-2
Classification of individual states[3]

You et al., presented condition detecting system that requires speech recognition system with over 95% efficiency. In this method, at first, person voice convert to text and text had been invested according to that terms in mental state and had scaled in a two-dimensional environment and then person emotion was recognized (Figure-2). Then recognized case will use pervasive computing input value[3].

Hidden markov model: When we can't see states in markov chains, we have to use a specific model of markov that named HMM (hidden markov model). Dynamic network based on the Bayesian can be a most simple example of this model. LE. Baum with his team defined a prefect mathematical pattern for this model. Also Ruslan L. and co-workers' attempt has a very similar method on optimizing a filtering in the nonlinear format that focused on process modeled by stochastic can be named as the debut explanation functions based on the forward-backward template.

Basic markov chains have some states can be observable for visitors and consequently probability of move from state are the essential parameters for solving problems. In the Hidden Markov model there aren't any visible state that we can find states. We have to used generated value of model to specific value in the favorite time's model state.

A Markov chain (Discrete-Time Markov Chain or DTMC) named after Andrey Markov is a mathematical system that undergoes transitions from one state to another on a state space. It is a random process usually characterized as a memoryless: the next state depends only on the current state and not on the sequence of events that preceded it. This specific kind of "memorylessens" is called the Markov property. Markov chains have many applications as the same as statistical models of real-world processes.

A Markov chain is a sequence of random variables like X_1, X_2, X_3, ... with the Markov property, namely that, given the present state, the future and past states are independent. Formally,

$$\Pr(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \Pr(X_{n+1} = x \mid X_n = x_n) \quad (1)$$

If both sides of the equation are well defined. The possible values of X_i form a countable set S called the **state space** of the chain. Markov chains are often described by a directed graph, where the edges are labeled by the probabilities of moving from one state to the other states.

Variations: Continuous-time Markov processes have a continuous index. Time-homogeneous Markov chains (or stationary Markov chains) are processes that whole range of n parameter. There aren't any dependency between n and transition's probability.

$$\Pr(X_{n+1} = x \mid X_n = y) = \Pr(X_n = x \mid X_{n-1} = y) \quad (2)$$

A Markov chain of order m (or a Markov chain with memory *m*), where *m* is finite, is a process satisfying

$$\Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_1 = x_1) = \Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_{n-m} = x_{n-m})\ for\ n > m \quad (3)$$

In other words, the future state depends on the past *m* states. It is possible to construct a chain (Y_n) from (X_n) which has the 'classical' Markov property by taking as state space the ordered *m*-tuples of *X* values, i.e. $Y_n = (X_n, X_{n-1}, \ldots, X_{n-m+1})$.

Implementation: In the sampling phase, 30 persons say 10 words in the common, sleepy and angry emotion and save them into the WAV files, and then the samples are named according to Table-1.

Table-1
Guidance of the sampled file name

Symbol	**Description**
pXXX	Person number
cXXX	Sampled emotion number
sXXX	Sampled term number

Preprocess blocks: Preprocessing block in both state and speech recognition system and also in training and testing phases have the same behavior, therefore, in this section have been investigated separately.

Sound sensor (microphone) inputs after crossing blocks of feature extraction and vector quantization, which are converted into a numerical vector, can be used into the hidden Markov model statistical system.

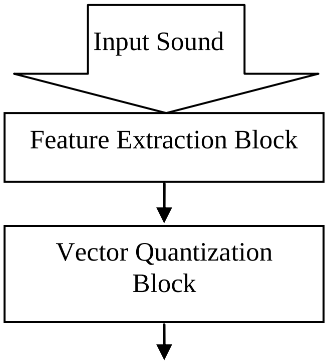


Figure-3
Diagram of sound preprocessing

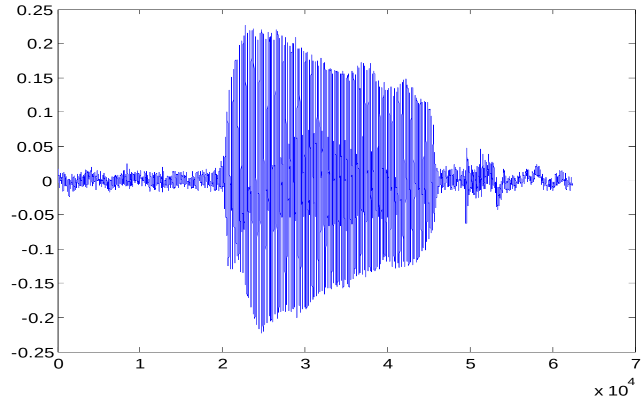


Figure-5
Diagram raw signal sampled

Feature Extraction Block: The block goal is to convert a raw sound signal (Figure-5) to a range of different features that are smaller than the raw signal, but the samples resolution can be acceptable.

There are various methods can be implemented in this phase, but a statistical method that based on frequency domain named Cepstral analysis method, was employed. After separation of the extracted data into multiple frames Pearson correlation coefficients make Cepstral Coefficients and the coefficients as the features of the system are considered, and then sent to the vector quantization unit[4]. The feature extractions block structure can be observed in Figure4. Audio signal inputs into block and Cepstral coefficients are abroad.

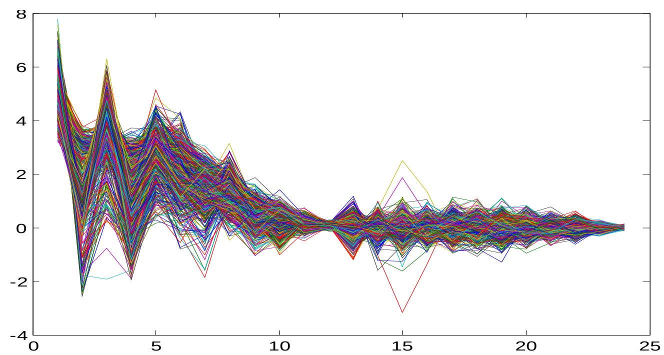


Figure-6
Diagramof sampled signal after feature extraction phase

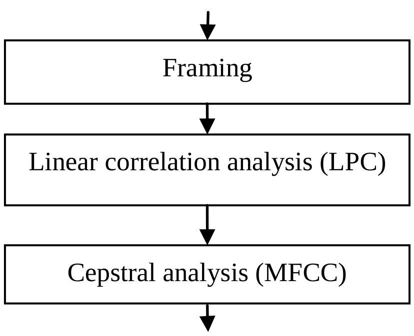


Figure-4
Workflow of feature extraction Block

Table-2
Parameters used in the implementation of the preprocessing

Parameter	Value
Frame Number	320
Frame spacing	80 ms
Iteration number ink-means	500
Cepstral Coefficient number	12
Cluster count created by k-means	10
LPC	12

Vector quantization block: To reduce the load on CPU and reduces the storage space required, we are reducing the number of input vector by this block In this block, the feature vector inputs and then apply the k-means clustering algorithm based on the factor K (number of clusters) to increase the system processing duration. The K value is higher, greater accuracy and speed decreases and vice versa. Sampled audio format without compressing files (WAV) and the bit rate is 705 kbps sampled. The population aged 50 years and has collected samples of both types of sex.

For each sample, the input signal after passes the feature extraction block makes a sequence of observations. This trail is a large range of system variables required and training time greatly increases. So a large number of data in the training and testing phase needs the heavy processing ,parameters vary with time, so the extracted feature vectors, are not the same size for each sample. Posed problems should be reduced to a fixed number of vectors to the ability to find practical use. We are in need of a new vector; K parameter is the length of the vector.

Table of parameters used in the processing power of the system and the expected time in the training phase of the system parameters are subject to change. These parameters can be reduced by reducing the load of processing system.

Sample diagrams of p001_c001_s001.wav. Raw signal in Figure-5, and the feature extraction result in Figure-6, and vector quantization process result can be found in Figure-7.

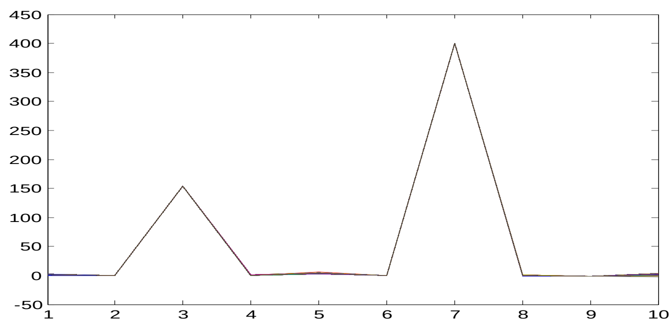


Figure-7
Chart sampled signal after feature extraction phase and vector quantization

Training System Phase: In training phase, input sounds according to the emotions index, are divided into three groups. Each group participates in ahidden Markov model training.

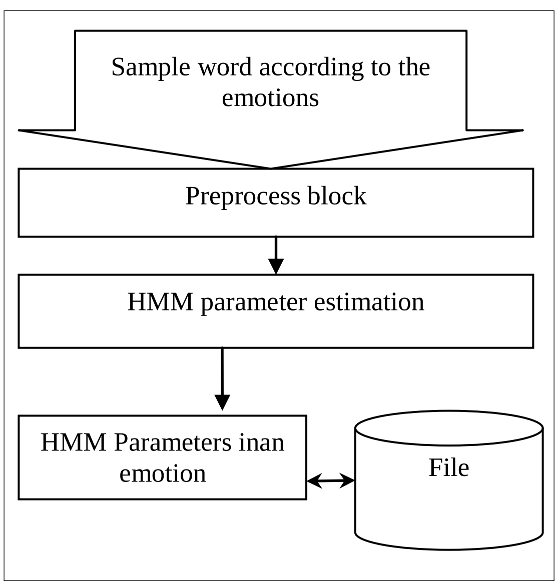


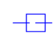
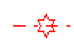
Figure-8
Block diagram of the training (emotion detection)

A hidden Markov model randomly generated, and every voice input already has passed thepre-processing block, by the forward-backward algorithm, hidden Markov model amounts to a better place. All samples of this practice will be train and eventually will be stored. This is done for the next emotion. This block diagram can be seen in Figure-8.

Testing system phase: In this phase, the hidden Markov model is trained for each emotion. Voice input after passing through the pre-processing block, a single hidden Markov model, is analogous to a forward algorithm and the maximum value is chosen as recognized emotion.

Implementation results: In the test phase, for each of the twelve samples tested, six were in the first stage, the data are of very low quality and there was a noise, and then the next six are data that a good quality and low noise made have. Table-2 shows the symbols used in the sampling emotion, is ideal that stars symbols be in a line in front of angry mode (the thirdaxisy), triangle symbols be in front ofa line of sleepy emotion (second centric y) and square symbols be in front of a line of normal emotion (about the first axis, y), and the displacement of each iconic symbol other than the location indicator is a wrong emotion recognizing by this system.

Table-3
Guide to the symbols used in the result diagrams

Symbol	**Emotion**
	Normal emotion sample
	Sleepy emotion sample
	Angry emotion sample

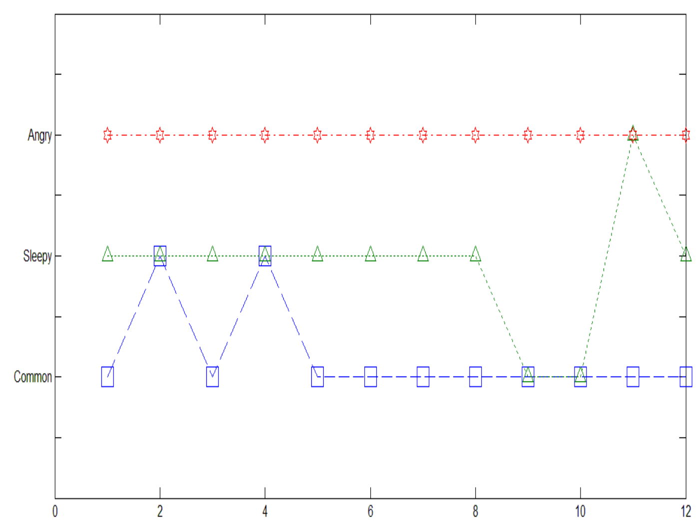


Figure-9
Results of per 36 samples tested implementation of the first phase of "two" word

In Figures-9 and 10, the horizontal axis represents the number of experimental data in three cases, and the vertical axis represents the detected emotion.

Table-4
Results of emotion detection

Learning type		Testing type		Emotion detection result	
Data	**Noise**	**Data**	**Noise**	**True %**	**False %**
Single word	Without noise	Single word	Without noise	96	4
Single word	Without noise	Single word	Noisy and Without noise	67	33
Single word	Noisy and Without noise	Single word	Without noise	67	33
Single word	Noisy and Without noise	Single word	Noisy and Without noise	75	15
Multi words	Without noise	Single word	Without noise	96	4
Multi words	Without noise	Single word	Noisy and Without noise	66	34
Multi words	Noisy and Without noise	Single word	Without noise	75	15
Multi words	Noisy and Without noise	Single word	Noisy and Without noise	74	16
Average				**77**	**23**

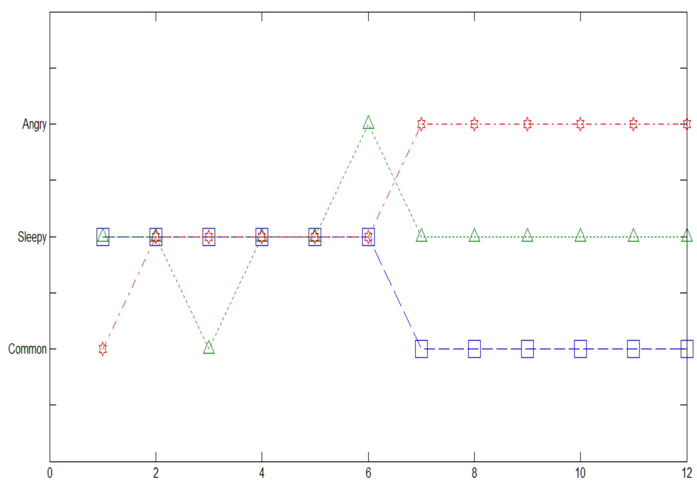


Figure-10
Results of implementation of the second phase of the experiment, for 36 instances of the word "four"

Conclusion

One of the most important issues in the implementation of this system is the right way of sampling. In this implementation of some words, a breakdown of normal and sleepy emotion was very difficult and did not make a significant difference between some of the samples; because practical sampling was very difficult when the person is sleepy, In instances where there is no much difference between the samples, no appreciable difference in performance with other models, But the words being pronounced manner than they are different conditions, The different samples in different states, the greater the difference between this system and other systems is very sensible.

Acknowledgement

This paper has been derived from Seyyed Mohammad Safi research plan in Islamic Azad university Ahvaz branch.

References

1. Van Den Broek E. and Westerin J. (2009). Considerations for emotion-aware consumer products. *Elsevier Applied Ergonomics*, 40(6).
2. Hong J., Yang S. and Cho S. Cona (2010). MSN: A context-aware messenger using dynamic Bayesian networks with wearable sensors. *Elsevier Expert Systems with Applications*, 37(6).
3. Yoo H., Kim M., Kwon O., 2011, Emotional index measurement method for context-aware service, *Elsevier Expert Systems with Applications*, 38(1) (2011)
4. Rabiner L. and Biing Hwang J. (2001). Fundamentals of Speech Recognition. ISBN-13: 978-0130151575, 69-139.
5. Alpaydin E. (2010). Introduction to Machine Learning. The MIT Press, ISBN-13: 978-0-262-01211-9.
6. Huang X., Acero A. and Hon H.W. (2001). Spoken Language Processing: A Guide to Theory, Algorithm and System Development. ISBN-13: 978-0130226167.
7. Loke S. (2006). Context-Aware Pervasive Systems: Architectures for a New Breed of Applications. ISBN-13:

978-0849372551.

8. Martin T.B., Nelson A.L. and Zadell H.J. (1964). Speech recognition by feature abstraction techniques. *Tech. Report AL-TDR-64-176, Air Force Avionics Lab.*

9. Mc Quaid H., Goel A. and McManus M. (2003). Designing for a pervasive information environment: The importance of information architecture. Conference HCI, Designing for Society Bath, UK, Proceedings Volume 2.

10. Myers C.S. and Rabiner L.R. (1981). A level building dynamic time wraping algorithm for connected word recognition. IEEE Trans. Acoustics, Speech, Signal Proc., ASSP-29: 284-297.

11. Nagata K., Kato Y. and Chiba S. (1963). Spoken digit recognizer for Japanese language. NEC Res. Develpo., No. 6.

12. Olson H.F. and Belar H. (1975). Phonetic Typewriter. *J. Acoust. Soc. Am.,* 28(6).

13. Paul D.B. (1989). The Lincoln robust continuous speech recognizer. ICASSP 89, Glasgow, Scotland, 449-452.

14. Górriz J.M., Ramírez J., Lang E.W., Puntonet C.G. and Turias I. (2010). Improved likelihood ratio test based voice activity detector applied to speech recognition. *Elsevier Speech Communication,* 52(7).

15. Rabiner L.R. (1989). A tutorial on Hidden Markov Models and selected applications in speech recognition. *IEEE,* 77(2)

16. Rabiner L.R., Levinson S.E., Rosenberg A.E. and Wilpon J.G. (1979). Speaker independent recognition of isolated words using clustering techniques. *IEEE Trans. Acoustics, Speech, Signal Proc.,* ASSP-27, 336-349.

17. Rastegari E., Rahmani A.M. and Setayeshi S. (2008). Pervasive Computing In Healthcare Systems. *International Journal of Biometrics and Bioinformatics (IJBB),* 3(4).

18. Reddy D.R. (1966). An approach to computer speech recognition by direct analysis of the speech wave. *Tech. Report No. C549, Computer Science Dept., Stanford Univ.*

19. Sakai, T., and Doshita, S., The phonetic typewriter, information processing 1962, *IFIP Congress, Munich*. (1962)

20. Sakoe H. (1979). Two level DP matching—A dynamic programming based pattern matching algorithm for connected word recognition. *IEEE Trans. Acoustics, Speech, Signal Proc.,* ASSP-27: 588-595.

21. Sakoe H. and Chiba S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoustics, Speech, Signal Proc. ASSP*26 (1).

22. Suzuki J. and Nakata K. (1961). Recognition of Japanese vowels: Preliminary to the recognition of speech. *J. Radio Res. Lab,* 37(8).

23. Tappert C.C., Dixon N.R., Rabinowitz A.S. and Chapman W.D. (1971). Automatic recognition of continuous speech utilizing dynamic segmentation, dual classification, sequential decoding and error recovery. *Rome Air Dev. Cen, Rome, NY, Tech Report TR-71-146.*

24. Amano A., Aritsuka T., Hataoka N. and Ichikawa A. (1989). On the use of neural networks and fuzzy logic in speech recognition. *Int. Joint Conf. Neural Networks,* (301).

25. Van Kleek M.K. (2003). Info: An Architecture for Smart Billboards for Informal Public Spaces. *UBICOMP, Seattle, WA.*

26. Weintraub M. et al. (1989). Linguistic constraints in Hidden Markov Model based speech recognition, ICASSP 89, Glasgow, Scotland, 699-702.

27. Bridle J.S. and Brown M.D. (1979). Connected word recognition using whole word templates. *Inst. Acoust. Autumn Conf,* (25).

28. Zimmermann H.J. (1996). Fuzzy set theory and its applications. *Kluwer Academic Publishers. Boston/ Dordrecht/London, Third edition.*

29. Zue V., Glass J., Phillips M. and Seneff S. (1989). The MIT summit speech recognition system: A progress report. DARPA Speech and Natural Language workshop, 179-189.

30. Matiko J.W., Beeby S.P. and Tudor J. (2014). Real time emotion detection within a wireless sensor network and its impact on power consumption. *IEEE Wireless Sensor Systems,* IET, 4(4).

31. Munezero M., Montero C.S., Sutinen E. and Pajunen J. (2014). Are They Different? Affect, Feeling, Emotion, Sentiment, and Opinion Detection in Text. 5(2).

32. Yeh Huann Goh, Raveendran P. and Jamuar S.S. (2014). Robust speech recognition using harmonic features. *IEEE Signal Processing, IET*8(2).

33. Wand M., Janke M. and Schultz T. (2014). Tackling Speaking Mode Varieties in EMG-Based Speech Recognition. *IEEE Biomedical Engineering, IEEE Transactions on,* 61(10).