Proceedings of the International Conference on Man-Machine Systems (ICoMMS)
11 – 13 October 2009, Batu Ferringhi, Penang, MALAYSIA

Overview of Automatic Stuttering Recognition System

Lim Sin Chee, Ooi Chia Ai, Sazali Yaacob
Universiti Malaysia Perlis
Jalan Kangar-Arau
02600 Jejawi Perlis

Abstract- **Stuttering is a speech disorder. The flow of speech is disrupted by involuntary repetitions and prolongation of sounds, syllables, words or phrases, and involuntary silent pauses or blocks in communication. Stuttering is an interest subject of researchers from many various domains such as speech physiology & pathology, psychology, acoustic and signal analysis. Thus there are many researchers have been done previously. This paper presents an overview of previous works on automatic stuttering recognition system. Normally, classification of speech disorder is difficult and complicated. However some classification techniques associated with stuttering are commonly recognized. This paper review on classification techniques are utilized in automatic stuttering recognition for evaluating speech problem for stutterers. Some previous works discussed the different steps involved in recognizing stuttered speech from speech samples. This paper compares different classification techniques proposed by previous researchers. Classification techniques used in previous works are Artificial Neural Networks (ANNs), Hidden Markov Model (HMM) and Support Vector Machine (SVM).**

I. INTRODUCTION

Speech will not always without disruptions, which lead to dysfluency while communicate. Dysfluency in a speech can be normal or pathological. Stuttering is one of the serious problems found in speech pathology. It occurs in about 1 % of population and has found to affect four times as many males as females [1-4]. The types of dysfluencies that [5-7]employed are: interjections (extraneous sounds and words such as "uh" and "well"), revisions (the change in content or grammatical structure of a phrase or pronunciations of a word as in "there was a young dog, no, a young rat named Arthur"), incomplete phrases (the content not completed), repetition, prolonged sounds (sounds judged to be unduly prolonged), broken words (words not completely pronounced)[5, 6, 8, 9].

Initially, stuttering is a neurologic trait that may involve specific abnormalities of speech motor control in the brain [1]. The muscles of the stutterer's larynx that control the opening and closing of the vocal cords is malfunctioning, making normal speech impossible. Stutterer has less blood flow; there is either increased or decreased electrical activity in regions of the brain involved in speech production. Thus, stuttering cannot be permanently cured; it may go into remission for a time, or stutterer can learn to shape their speech into fluent speech with the appropriate speech pathology treatment.

Stuttering assessment is essential to assess the improvement of the stutterers during speech therapy. Stuttering assessment is used to detect types of dysfluencies, especially repetitions. This is because stuttering is commonly associated with repetitions [5, 6]. Phrase-repetitions, word-repetitions and syllabic-repetitions are one of the elements of stuttering. Word-repetition is a prevalent feature of early stuttering.

Conventionally, stuttering assessment is done by counting number of dysfluent words as a proportion of total words in a passage. Stuttering assessment is also done by measuring the time of dysfluencies and comparing with the duration of the entire passage. However, it is time-consuming and it results in poor agreement with different judges on the same material.

Therefore, automatic stuttering recognition system is used to automate the dysfluency count and type of dysfluencies classification, thus providing an objective and consistent measurement. Such approach can support Speech Language Pathology (SLP) by doing tedious routine works and allowing more time for therapeutic session between SLP and stutterers. This paper is organized as follows: Section II is an overview of previous work on automatic stuttering recognition system using different feature extraction algorithms and different classifies. Section III is a review and comment about advantage and disadvantages of previous works. Opinion of the author of this paper will be included as well. Finally, section IV presents discussion and conclusion on this paper. Automatic stuttering recognition system using different classifiers are being compared in following sections.

II. AUTOMATIC STUTTERING RECOGNITION SYSTEM

Stuttering recognition is an ideal example of multidisciplinary research. In the present day, there are a large number of research works focused on the automatic stuttering recognition and classification, by methods of acoustic analysis, feature extraction, neural network and statistical method. This section presents an overview of previous works found in the literature which concentrates on how the automatic stuttering recognition is being performed, how they design the experiments and evaluate their results. Table I lists summary of several related research works chronologically, and gives a rough idea of the variety approaches that used in existing literature.

A. ANNs

The ANNs are mathematical model or computational models that try to develop intelligent systems¸ some inspired by biological neural network. Researchers from many scientific disciplines are designing ANNs to solve problems in pattern recognition, prediction, optimization, associative memory and control[10]. Neural networks play an important role both in the speech[11] and the speaker recognition[12] and are an irreplaceable tool in distinguishing between similar signals[13-15]. In recent years, ANNs are widely used in many ways in stuttering recognition, such as recognition of prolongation and repetition in stuttered speech, classification of fluent and dysfluent in stuttered speech.

Peter et al. have several papers using ANNs as classification technique. The ANNs model is used to detect the stuttered events. The particular stuttered events to be located are repetitions and prolongations. This is because the repetitions and prolongations are ubiquitous in the speech of stutterers [16] and as a result of this, it is widely used by SLP. In this work[17], they proposed stuttering recognition using ANNs were achieved by marking which words are repeated, prolonged and others types of dysfluencies. The networks are automatically adjusted to bring the mapping between input and output into correspondence [18]. Each word is judged as fluent, repetitions, prolongations or other dysfluency categories. The training is achieved by linking the acoustic input representations either onto repetitions or prolongations output through a layer of hidden units [19]. The basic architecture of multilayer feed forward neural network is shown in fig. 1.

In 1995, they [17] used autocorrelation function and envelope parameters as the input vector of ANNs, yielding a best accuracy of around 80%.

In 1997, they [9, 20] done their research in depth by employed 12 children who were stutter. The recordings were made as the child read the “Arthur the rat” passage. Speech was recorded on DAT tape and transferred digitally to computer and was down-sampled to 20Hz for further processing. They used the fragmentation measure, spectral measures and part work duration and energy for differentiating between fluent and dysfluent words. The ANNs correctly identified 78.01% of the dysfluent (combination of prolongations and repetitions) words.

Geetha et al.[13]presented a research on classification of childhood dysfluencies using ANNs. They used 10 variables to make the distinction between normal non-fluency and stuttering. The variables such as age, sex, type of dysfluency, frequency of dysfluency, duration, physical concomitant, rate of speech, historical, attitudinal, and behavioral scores, family history. 25 children were used to train the ANNs, and 26 children were used for predicting the ANNs for the same sets of variables. They presented ANNs could predict the classifications of normal non-fluency and stuttering with 92% accuracy.

Czyzewski et al. in 2003 [14] approached the recognition task based on detection of stop-gaps, discerning vowel prolongations, detection of syllable repetitions. They employed 6 fluent speech samples and 6 with stop-gaps speech samples. They used ANNs and rough set to detect stuttering events. They presented that better scores were obtained using rough set-based system yielded an accuracy of more than 90% than the ANNs with average accuracy equal to 73.25%.

Szczurowska et al in 2006 [21] described the neural networks tests on ability of recognition and categorizing the non-fluent and fluent speech samples. The research materials were recordings that taken from 8 stuttering speakers. They analyzed the recordings by using FFT512 with the use of 21 digital 1/3-octave filters of centre frequencies between 100 Hz and 10 kHz. They applied Kohonen and Multilayer Perceptron Networks to recognize and classify fluent and dysfluent. The authors achieved a best result of 76.67% with the best network, built of 171 input neurons, 53 neurons in hidden layer and 1 output neuron.

In 2008, K.M Ravikumar et al [6] proposed an automatic detection method for syllable repetition in reading speech for objective assessment for stuttered dysfluencies which has four stages comprising of segmentation, feature extraction, score matching and decision logic. A 150 words Standard English passage was selected for preparing the database. Speeches collected in databases are segmented manually. 12MFCC was being employed as feature extraction algorithm and the recognition system was based on neural networks (Perceptron). Perceptron is used to make decision whether a syllable is repeated or not. Perceptron was the first iterative algorithm for learning linear classification. The basic idea behind Perceptron is shown in fig. 2. If distinct parameters are separated, do not move. If not, move it to the left. If the pattern is correctly classified, do nothing. 83 % of accuracy was achieved by them with 10 speech samples. Out of 8 samples were used for Perceptron classifier training while the remaining 2 samples are for testing.

In 2009, Świetlicka et al[22] presented a research concerning on automatic detection of dysfluency in stuttered speech. They employed 8 stuttering people for recordings. 59 fluent speech samples and 59 non-fluent speech samples were obtained from the recordings. 21 digital 1/3 octave filters of centre frequencies between 100 Hz and 10 kHz were used to analyzed the speech samples. These parameters of the speech samples were used as an input for the Networks. They applied Multilayer Perceptron (MLP) and Radial Basis Function (RBF) networks to recognize and classify fluent and non-fluent in speech samples. They yield classification correctness for all networks ranged between 88.1% and 94.9%. Undoubtedly, ANNs can be used as a tool in speech analysis both of the fluent and non-fluent speaker.

B. HMMs

HMM is a stochastic model that captures the statistical properties of observed real world data[23]. In speech recognition, speech signal could be viewed as a piecewise stationary signal or a short-time stationary signal[24]. Thus, HMMs are widely used in speech recognition, especially in stuttering recognition to recognize speech dysfluency such as prolongation and repetition [25, 26].

TABLE I. SUMMARY OF SEVERAL RESEARCH WORKS ON AUTOMATIC STUTTERING RECOGNITION SYSTEM, DETALING IN NUMBER OF PATIENTS IN THE DATABASE, THE ACOUSTIC FEATURES EMPLOYED AND THE CLASSIFICATION METHOD

First Author	Year	Database	Features	Classifier	Best Results (%)
Howell [17]	1995	-	Autocorrelation function and envelope parameters.	ANNs	≈ 80%
Howell[9, 20]	1997	12 Speakers	Duration, energy peaks, spectral of word based and part word based.	ANNs	78.01%
Geetha [13]	2000	51 Speakers	Age, sex, type of dysfluency, frequency of dysfluency, duration, physical concomitant, Rate of speech, historical, attitudinal, and behavioral scores, family history.	ANNs	92%
Nöth[7]	2000	37 speakers	Duration and frequency of dysfluent portions, speaking rate	HMMs	-
Czyzewski [14]	2003	6 normal speech samples + 6 stop-gaps speech samples	Frequency, 1st to 3rd formant's frequencies and its amplitude	ANNs & rough set	73.25% & ≥ 90.0%
Szczurowska [21]	2006	8 speakers	Spectral measure (FFT 512)	Multilayer Perceptron (MLP), Kohonen	76.67%
Wiśniewski [25]	2007	38 samples for prolongation of fricatives + 30 samples for stop blockade + 30 free of silence samples	Mel Frequency Cepstral Coefficients (MFCC)	HMMs	70%
Wiśniewski [26]	2007	-	Mel Frequency Cepstral Coefficients (MFCC)	HMMs	Approximately 80%
Tian-Swee [2]	2007	15 normal speakers + 10 artificial stuttered speech	Mel Frequency Cepstral Coefficients (MFCC)	HMMs	96%
Ravikumar[6]	2008	10 speakers	Mel Frequency Cepstral Coefficients (MFCC)	Perceptron	83%
Ravikumar[5]	2009	15 speakers	Mel Frequency Cepstral Coefficients (MFCC)	SVM	94.35%
Świetlicka [22]	2009	8 stuttering speakers + 4 normal speakers (yields 59 fluent speech samples +59 non-fluent speech samples)	Spectral measure (FFT 512)	Kohonen , Multilayer Perceptron(MLP), Radial Basis Function(RBF)	88.1% - 94.9%

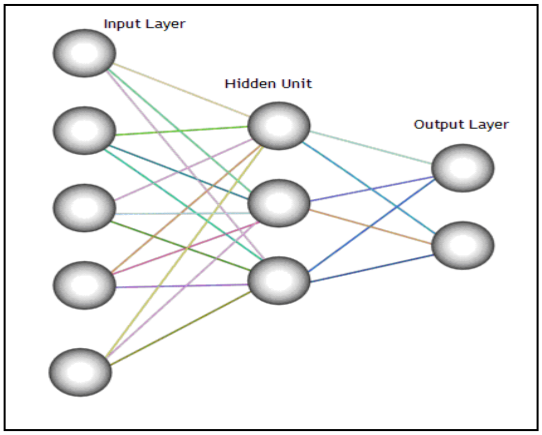

Fig.1, The basic architecture of multilayer feed forward neural network.

In 2000, Nöth et al. [7] combined the work of SLP and speech recognition system to evaluate the degree of stuttering during therapy session. This system can perform statistical analysis such as counting and classification of typical repetitions, pauses and phoneme duration. The measurable factors of stuttering used to classify the degree of stuttering are frequency of dysfluent portions in the speech, duration of the dysfluencies and speaking rate.

The database consists of 37 patients with stuttering symptoms in that either reading all or the beginning of a passage. Results of word and phoneme accuracies of the stuttered text in relation to the number of detected dysfluencies showed a correlation coefficient of up to 0.99. However, systems seem to have over-estimated the reading errors. Nöth et al. [7]perceived the results still crude and needs more experiments in the future, especially with stutterers belonging clearly either the repetition or blocking type.

Wiśniewski et al [25, 26] in 2007 have presented 2 papers about an automatic detection system using HMM as the classification technique. In [23] they employed 38 samples for prolongation of fricatives recognition model, 30 samples for stops blockade of recognition model and 30 samples for summary model. All the samples were down-sampled to 22050 Hz. The samples were parameterized using MFCCs. According to their result, the best recognition accuracy was achieved for summary models with free silence equal to 70%. In their next paper, they [26] proposed an automatic detection system fully concentrate on recognition of prolonged fricative phonemes with HMM as classification method. Before analysis, the acoustic signal requires to be parameterized. They used MFCC as the set of parameters. The sampling frequency of the sound samples were 22050 Hz and all the records were normalized to the same dynamic range – 50dB. They showed a best result of approximately 80%.

Tian-Swee et al [2] presented automatic stuttering recognition system utilizes HMM technique to evaluate speech problem for children such as stuttering. Speech Therapy Assistance Tools (MSTAT) is a system assists SLP to diagnose children for language disorder and to train children with stuttering. The voice pattern of the normal and speech disorder children are used to train the HMM model. HMM use Markov process to model the changing statistical characteristics that are only probabilistically manifested through actual observation [2]. Speech signal is assumed as a stochastic function of the state sequence of the Markov Chain. The state sequences itself is hidden. There are 2 types of HMM differentiated by its observation probability functions called discrete HMM (DHMM) and continuous density HMM (CDHMM). Fig. 3 shows the type of HMM is used by them. It is a 5 state left-to- right model.

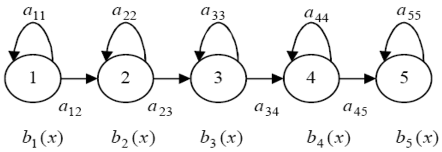

Fig. 3, Representation of left-to-right HMM [2].

The database consists of 20 samples of normal speech data and 15 samples of artificial stutter speech data. 10 samples of each normal and artificial stutter speech were used to generate a speech model respectively. Remaining 5 samples of normal speech data and artificial stuttered speech data were used to test on HMM models. If the score is greater than threshold's value, it is diagnosed as normal and vice-versa. An average percentage of correct recognition rates were 96% achieved by normal speaker and 90% for the artificial stutter speech.

There are 3 related works in 2007 used same feature extraction method and classification techniques, but the way they implement their work is different such as sampling frequency, number of states of HMMs, codebook size(training data) and testing sets. All the criteria will result a different accuracy on stuttering recognition.

C. SVM

SVM is a powerful machine learning tool is widely used in the field of pattern recognition. SVM optimization problem attempts to obtain a good separating hyper-plane between two classes in the higher dimensional space [5, 27, 28]. Fig 4 shows decision algorithm in SVM.

To date, SVM was used as classification tool in stuttering recognition. Ravikumar et al [5] proposed automatic detection method for syllable repetition in reading speech for objective assessment for stuttered dysfluencies which has four stages comprising of segmentation, feature extraction, score matching and decision logic same as their previous work. As an improvement to their previous work, the decision logic was implemented using SVM to classify between fluent and dysfluent speech. The Fifteen speech samples were collected from 15 adults who stutter, 12 samples were used for training. The remaining three samples were used for testing. The system yielded 94.35% accuracy which is higher than their previous work.

I. REVIEW

There are some comments on previous works regarding the reliability of the results due to small amount of training data. The number of training data should be increased to improve the accuracy. Example, there are only 8 speech sample employed by Ravikumar [6] to train the system and it was tested with only 2 speech samples although they achieved 83% of accuracy but the reliability of the result is questionable. The number of training data should be increased to improve the accuracy. In addition, database in findings [5, 6, 14]were created without taking into consideration the distribution of features such as gender, age or origin.

Moreover, some of the systems were tested with artificial stuttered speech example, Tian-Swee at el [2] employed 10 artificial stuttered speeches samples to train the system, and used only 5 artificial stuttered speech data to test the system. Despite yielding a high accuracy as 90%, the accuracy was being questioned due to the reason that system was not tested on speech samples taken from actual stuttering clients. Besides that, the paper did not include details of how artificial stuttered speech being generated. Finding [2]did not classify stutterers into different types of stuttering, although it was mentioned as one of the objectives in the research.

In addition, the reliability of the results can be achieved through cross-validation and confidence intervals. Cross-validation is a technique for estimating the accuracy of a classifier[29]. Thus, by applying it, the performance of the recognition system can be measured.

III. CONCLUSION

Stuttering is a disorder of speech communication. In last 2 decades, there are a lot of researches being done on stuttering recognition. There are 3 major classifiers used to classify types of dysfluencies and also between stutterers and non stutterers. Three classifiers are ANNs, HMMs and SVM. Each classifier provides different accuracies where HMM is proven to give highest accuracy 96%, then SVM yielded 94.35%, while ANNs achieved an accuracy of 94.9% [2, 5, 22].

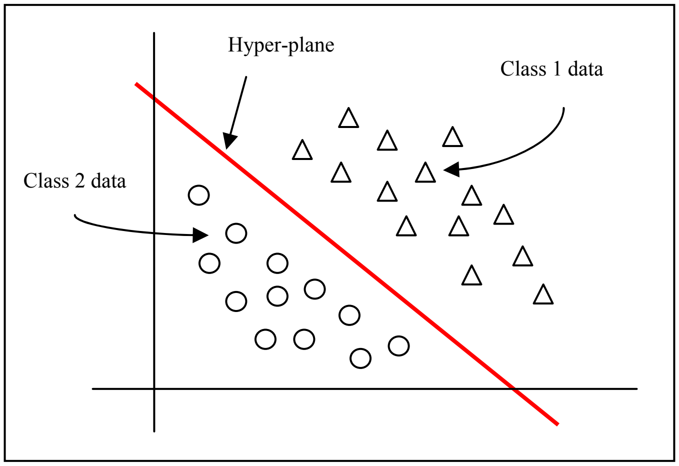

Fig. 4, Decision algorithm in SVM.

In this paper, a review was presented and highlights the strengths and weaknesses of earlier works. Accuracy of stuttering recognition may be improved by using other feature extraction algorithm like Linear Predictive Cepstral Coefficient (LPCC) or (Perceptual Linear Prediction) PLP or other classifier like *k*-nearest neighbor (*k*-NN) or determinant analysis which may improve the accuracy and overall performance of stuttering recognition system.

ACKNOWLEDGEMENT

This research project is supported by E-Science Fund from Ministry of Science, Technology and Innovation (MOSTI).

REFERENCES

[1] S. S. Awad, "The application of digital speech processing to stuttering therapy," in *Instrumentation and Measurement Technology Conference, 1997. IMTC/97. Proceedings. 'Sensing, Processing, Networking'., IEEE*, 1997, pp. 1361-1367 vol.2.

[2] T. Tian-Swee, L. Helbin, A. K. Ariff, T. Chee-Ming, and S. H. Salleh, "Application of Malay speech technology in Malay Speech Therapy Assistance Tools," in *Intelligent and Advanced Systems, 2007. ICIAS 2007. International Conference on*, 25-28 Nov 2007, pp. 330-334.

[3] Ooi Chia Ai and J. Yunus, "Computer-based System to Assess Efficacy of Stuttering Therapy Techniques," in *Proceeding of 3rd Kuala Lumpur International Conference on Biomedical Engineering*, Kuala Lumpur, 2006, pp. 395-398.

[4] Ooi Chia Ai and J. Yunus, "Overview of a Computer-based Stuttering Therapy," in *Regional Postgraduate Conference on Engineering an Science (RPCES 2006)*, Johore, 2006, pp. 207-211.

[5] K. M. Ravikumar, R.Rajagopal, and H.C.Nagaraj, "An Approach for Objective Assessment of Stuttered Speech Using MFCC Features," *ICGST International Journal on Digital Signal Processing, DSP,* vol. 9, pp. 19-24, 2009.

[6] K. Ravikumar, B. Reddy, R. Rajagopal, and H. Nagaraj, "Automatic Detection of Syllable Repetition in Read Speech for Objective Assessment of Stuttered Disfluencies," in *Proceedings of World Academy Science, Engineering and Technology*, 2008, pp. 270-273.

[7] E. Nöth, H. Niemann, T. Haderlein, M. Decher, U. Eysholdt, F. Rosanowski, and T. Wittenberg, "Automatic stuttering recognition using hidden Markov models," 2000.

[8] P. Howell, S. Sackin, and J. Au-Yeung, "Assessment Procedures for Locating Stuttered Events," in *Proceedings of the Second World Congress on Fluency Disorders*, 1998.

[9] P. Howell, S. Sackin, and K. Glenn, "Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: I. Psychometric procedures appropriate for selection of training material for lexical dysfluency classifiers," *Journal of Speech, Language, and Hearing Research,* vol. 40, p. 1073, 1997.

[10] A. K. Jain, M. Jianchang, and K. M. Mohiuddin, "Artificial neural networks: a tutorial," *Computer,* vol. 29, pp. 31-44, 1996.

[11] W. Chen, S. Chen, and C. Lin, "A speech recognition method based on the sequential multi-layer perceptrons," *Neural Networks,* vol. 9, pp. 655-669, 1996.

[12] K. Farrell, R. Mammone, and K. Assaleh, "Speaker recognition using neural networks and conventionalclassifiers," *IEEE Transactions on speech and audio processing,* vol. 2, pp. 194-205, 1994.

[13] Y. V. Geetha, K. Pratibha, R. Ashok, and S. K. Ravindra, "Classification of childhood disfluencies using neural networks," *Journal of fluency disorders,* vol. 25, pp. 99-117, 2000.

[14] A. Czyzewski, A. Kaczmarek, and B. Kostek, "Intelligent processing of stuttered speech," *Journal of Intelligent Information Systems,* vol. 21, pp. 143-171, 2003.

[15] L. Leinonen, J. Kangas, K. Torkkola, and A. Juvas, "Dysphonia detected by pattern recognition of spectral composition," *Journal of speech and Hearing Research,* vol. 35, p. 287, 1992.

[16] M. Wingate, *The structure of stuttering: A psycholinguistic analysis*: Springer Verlag, 1988.

[17] P. Howell and S. Sackin, "Automatic recognition of repetitions and prolongations in stuttered speech," in *Proceedings of the First World Congress on Fluency Disorders*, 1995, pp. 372-374.

[18] P. Howell, J. Au-Yeung, S. Sackin, K. Glenn, and L. Rustin, "Detection of supralexical dysfluencies in a text read by children who stutter," *Journal of fluency disorders,* vol. 22, pp. 299-307, 1997.

[19] J. MacClelland and D. Rumelhart, *Parallel distributed processing: Explorations in the microstructure of cognition*: The MIT Press, 1988.

[20] P. Howell, S. Sackin, and K. Glenn, "Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: II. ANN recognition of repetitions and prolongations with supplied word segment markers," *Journal of Speech, Language, and Hearing Research,* vol. 40, p. 1085, 1997.

[21] I. Szczurowska, W. Kuniszyk-Jozkowiak, and E. Smolka, "The application of Kohonen and Multilayer Perceptron Networks in the speech nonfluency analysis," *Archives of Acoustics,* vol. 31, p. 205, 2006.

[22] I. Świetlicka, W. Kuniszyk-Jóźkowiak, and E. Smołka, "Artificial Neural Networks in the Disabled Speech Analysis," in *Computer Recognition System 3*. vol. 57/2009: Springer Berlin / Heidelberg, May 12, 2009, pp. 347-354.

[23] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE,* vol. 77, pp. 257-286, 1989.

[24] H. Hammady, S. Abdou, M. Shahin, and M. Rashwan, "An HMM System for Recognizing Articulation Features for Arabic Phones," 2008, pp. 125-130.

[25] M. Wiśniewski, W. Kuniszyk-Jóźkowiak, E. Smołka, and W. Suszyński, "Automatic Detection of Disorders in a Continuous Speech with the Hidden Markov Models Approach," in *Computer Recognition Systems 2*. vol. 45/2008: Springer Berlin / Heidelberg, October 18, 2007, pp. 445-453.

[26] M. Wiśniewski, W. Kuniszyk-Jóźkowiak, E. Smołka, and W. Suszyński, "Automatic Detection of Prolonged Fricative Phonemes with the Hidden Markov Models Approach," *Journal of Medical Informatics & Technologies,* vol. 11/2007, 2007.

[27] R. A. El-Khoribi, "Support Vector Machine Training of HMT Models for Land Cover Image Classification}," *ICGST*

International Journal on Graphics, Vision and Image Processing, GVIP}, vol. 8, pp. 7-11, December 2008.

[28] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data mining and knowledge discovery,* vol. 2, pp. 121-167, 1998.

[29] M. Mullin and R. Sukthankar, "Complete cross-validation for nearest neighbor classifiers," in *Proceedings of the Seventeenth International Conference on Machine Learning*, 2000, pp. 639 - 646