Application of Malay Speech Technology in Malay Speech Therapy Assistance Tools

Tian-Swee Tan, Helbin-Liboh, A. K. Ariff, Chee-Ming Ting and Sh-Hussain Salleh
Center for Biomedical Engineering
Faculty of Biomedical Engineering and Health Science
Universiti Teknologi Malaysia
81310 Skudai
Johor, Malaysia
E-mail: tantianswee@hotmail.com, niquelia84@yahoo.com, amarism@yahoo.com, cmting1818@yahoo.com, hussain@fke.utm.my

***Abstract* — Malay Speech Therapy Assistance Tools (MSTAT) is a system which assists the therapist to diagnose children for language disorder and to train children with stuttering problem. The main engine behind it is the speech technologies; consist of speech recognition system, Malay text-to-speech system and Malay Talking Head [1-4]. In this project, speech recognition system utilizes the Hidden Markov Model (HMM) technique for evaluating speech problem for children such as stuttering. The voice pattern of the normal and speech disorder children are used to train the HMM model for classifying the problem of speech disorder. Thus, the system is localized with local dialect voice database which focus on local Malay dialect that is currently not available in market. Besides that, the system also utilizes Malay text-to-speech system and Talking Head to guide and lead the children or parents to follow the diagnostic process. This feature makes the system more interactive and not only single direction of communication.**

***Keywords*— Speech therapy, speech recognition, HMM**

I. INTRODUCTION

Speech and language therapy is focused on spoken and written human communication and is concerned with prevention, diagnosis and treatment of deviations of the normal communicational behavior children and adults [5]. Speech therapy is a clinical field concerned with disorders of human communication. Speech and language disorders refer to problems in communication and related areas such as oral motor function. These delays and disorders range from simple sound substitutions to inability to understand or use language or use oral motor mechanism for functional speech and feeding. Speech disorders refer to difficulties producing speech sounds or problems with voice quality.

Stuttering is one of the serious problems focused on in speech pathology. It occurs in about 1% of population and has found to affect four times as many males as females [6]. Stuttering has been considered a heritable disorder since the 1930s [7]. Stuttering is the condition in which the flow of speech is broken by abnormal stoppages, repetitions, or prolongations of sound and syllable. Individuals who stutter can learn to control their speech fluency by shaping the tempo, loudness, effort, or duration of their utterances. The most commonly utilized techniques of facilitating fluency are called fluency shaping and stuttering modification [6]. Thus, this project is to design a training software for assisting speech therapist in diagnosing the Malaysian children in speech fluency. Figure 1 shows the diagnosis method of MSTAT. It utilizes the intelligent dialog system simulating therapist questioning and answering to the subject and lead the children step by step using talking agent in diagnosing their speaking problem. The client will speak the sample words, for example "Sembilan" (which is nine in Malay Language). Then, the system will evaluate the speech input to find out either it stutters or not.

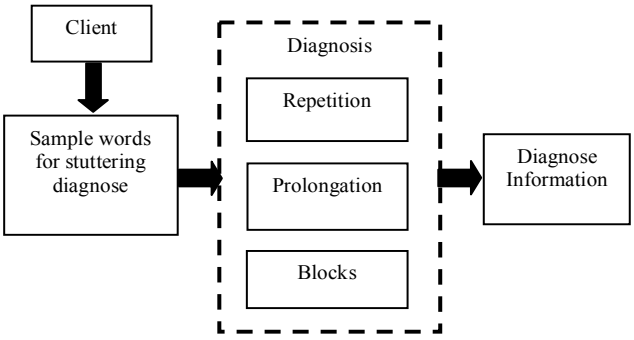

Figure 1: Diagnosis Methods

II. SPEECH STUTTERING

Stuttering is a speech disorder characterized by certain types of speech disfluencies, such as sound repetitions, which are frequent enough to be disruptive [8]. According to Wingate (1964), part of the reason for so many definitions of stuttering is that its cause has yet to be identified. There have been many proposals as to the cause, including genetics, physical impairments, and psychosomatics. With no known cause, definitions of stuttering are based on the symptoms exhibited by stutterers.

The most common symptom is the occurrence of speech disfluencies, which are widely used to define and identify stuttering. Most methods for stuttering assessments require a count of the disfluencies in a patient's speech. Some features of stuttering that are more subjective in nature include overall tenseness, hyperarticulation, and inappropriate word stress, which all contribute to a vague sense of unnaturalness of the stuttered speech.

1-4244-1355-9/07/$25.00 @2007 IEEE

Authorized licensed use limited to: UNIVERSITY TEKNOLOGI MALAYSIA. Downloaded on January 12, 2009 at 03:58 from IEEE Xplore. Restrictions apply.

Each child is totally different from one another. Therefore they will have a different combination of stuttering behaviours [9-13]. Table 1 shows the stuttering behaviours occur in children who are stutter.

Table 1: Stuttering Behaviours (Preschool Stuttering)

Stuttering Behaviours	Details
Repetition	Can occur at the start, middle, or end of a word or sentence. Commonly occur at the beginning of the sentence.
Blocks	Occurs at the start of the word, occurs when there is a stoppage of airflow. Therefore, no sound comes out.
Prolongation	Occur on sounds in any position in the word or sentence and can vary in length.
Interjections	Known as fillers, frequent insertion of words like 'um' and 'ah'.

III. MALAY SPEECH THERAPY ASSISTANCE TOOLS (MSTAT)

There are a number of diagnosis methods for producing a stuttering severity rating for a patient. All methods require samples of the patient's speech, which might be gathered in real time through direct interaction with the subject. Most methods for stuttering assessments require a count of the disfluencies in a patient's speech [8]. To diagnose stutter, the client are required to read and speak the selected words.

Figure 2 shows the architecture of MSTAT and talking head. The stuttering training tools utilizes a PC equipped with microphone and loudspeakers. The system provides the ability to record, match voice, detect stutter, and also allow client to playback their speech utterances.

This software uses speech recognition to automate the disfluency count measure. The output of these tools are the total number of disfluencies, the total number from each disfluency class, the number of words or syllables spoken, the calculated frequency of disfluencies and the calculated frequency of each disfluency class. Obviously, to provide a count of total disfluencies and a count of disfluencies from each disfluency class, the system has to both detect and classify the disfluencies in the speech sample.

Speech recognition systems have much to offer for identifying disfluencies. First, stuttering events are word-based, so with the correct sequence of words identified, it is a simple matter of scanning through the sequence to find stuttering events like word repetitions and phrase repetitions. Furthermore, with the correct sequence of phonemes identified, and knowledge of how phonemes group into words, it would be a simple matter to find sound repetitions [8].

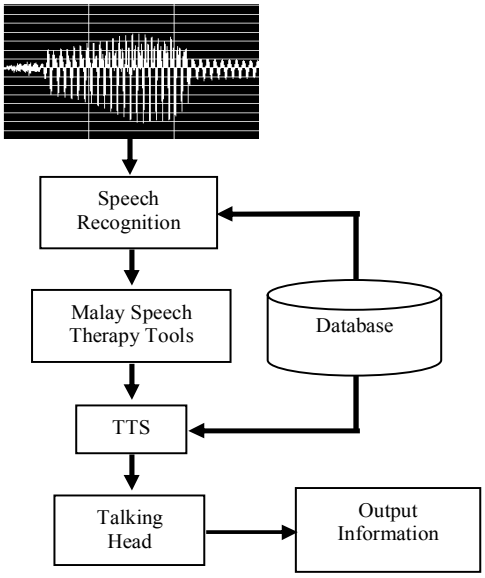

Figure 2: The architecture of Malay Speech Therapy System.

Malay Speech Therapy Assistance Tools and talking head

Figure 3 shows the relationship between MSTAT and Talking Head. MSTAT is integrated with talking head in stuttering diagnostic.

Figure 3: MSTAT and Talking Head

Authorized licensed use limited to: UNIVERSITY TEKNOLOGI MALAYSIA. Downloaded on January 12, 2009 at 03:58 from IEEE Xplore. Restrictions apply.

IV. HMM MODELS

HMMs use a Markov process to model the changing statistical characteristics that are only probabilistically manifested through actual observation. The observed speech signal is assumed to be a stochastic function of the state sequence of the Markov chain. The state sequences itself is hidden. We consider two types of HMM which differentiated by its observation probability functions so called discrete HMM (DHMM) [15] and continuous density HMM (CDHMM) [16,17].

Figure 4 shows the type of HMM we consider in this paper. It is a 5 state left-to-right model.

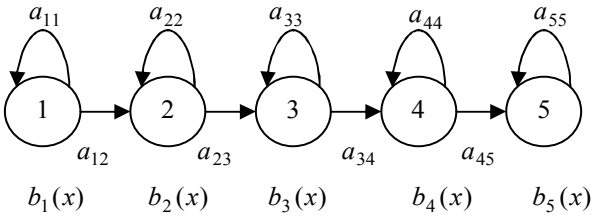

Figure 4: Representation of left-to-right HMM.

The parameters which characterize the HMM of figure 4 are the following: (1) *N,* number of states in the model; (2) π_i ,initial state probability vector; (3) $A = \lfloor a_{ij} \rfloor, 1 \le i, j \le N$, the state transition matrix where a_{ij} is the probability of making a transition from state *i* to state *j*; and (4) B, the observation probability function associated with each state j. For DHMM, $B = \{b_j(O_k)\}$, discrete observation probability distribution.

For CDHMM, $B = \{b_j(x)\}$, continuous observation probability density function where $1 \le j \le N$ .Here O_k represents discrete observation symbols and *x* represents continuous observations (usually speech frame vectors) of K-dimensional random vectors. The overall HMM model, λ is represented by (π, A, B) .

The most general representation of the pdf of CDHMM, for which a re-estimation procedure has been formulated, is a finite mixture density of the form

$$b_j(x) = \sum_{m=1}^{M} c_{jm} \mathrm{N}(x, \mu_{jm}, \Sigma_{jm}), \quad 1 \le j \le N$$

,where *x* is vector being modeled, c_{jm} is the mixture coefficient for the *m*th mixture component in state *j* and N is any log-concave or elliptically symmetric density (eg. Gaussian density), with mean vector $\mu_{jm} = [\mu_{jmd}]$ and covariance matrix $\Sigma_{jm} = [\Sigma_{jmde}]$ for the *m*th mixture component in state *j*, for $1 \le d, e < D$, where *D* is the number of dimensions in feature vectors. There are two form of Σ_{jm} , namely diagonal matrices (with assumed zero correlation between different element of observation vector) and full covariance matrices. Use of diagonal matrices consume less training data and time while providing better performance than full covariance matrices.

The CDHMM considered in this paper use multivariate Gaussian mixture densities with diagonal matrices and 4 mixture components, M=4. The multivariate Gaussian density is given by,

$$N(x, \mu_{jm}, \Sigma_{jm}) = \frac{1}{\sqrt{(2\pi)^n \left|\Sigma_{jm}\right|}} e^{-\frac{1}{2}(x_t - \mu_{jm})^T \Sigma^{-1} (x_t - \mu_{jm})}$$

Feature extraction

The feature extraction used in the system front end to extract important features from speech signals to reduce the data size before used for training and recognition. Our front end uses Mel-cepstral frequency coefficient extraction (MFCC) [18]. The front end parameterized an input speech signal into a sequence of output vectors of MFCC features. A block diagram of the front end is shown in Figure 5.

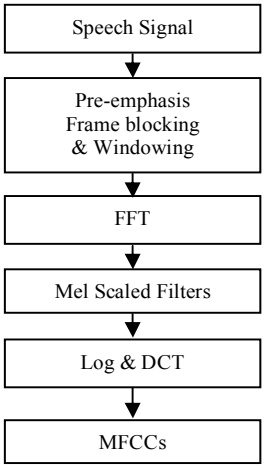

Figure 5: Block diagram of MFCC front end.

In this paper, the sampled speech signals are pre-emphasized with filter. Then, the waveform is blocked into frames. Each frame is multiplied by a Hamming window.

In the process of MFCC extraction, the DFT spectrums are filtered by triangular windows, which are arranged in mel-scale (designed to approximate the amplitude frequency response of human ear). Next, log compression is put on the output of each filter. Finally, Discrete Cosine Transform (DCT) with order 12 is applied to decorrelate feature vector of 12 MFCCs.

V. SCORE EVALUATION

Speech signal will be evaluated using HMM and the scores will be taken for analysis. For this project, 20 samples of normal speech data and 15 samples of artificial stutter speech data are taken. From these samples of data, 10 samples of each normal and artificial stutter speech are used to generate a speech model for each normal speech data and artificial stutter speech data. After the threshold has been set, the other 5 samples of data for each normal and artificial stutter speech are used to test the models. If the score is greater than the threshold, then the result for diagnose is normal. If the score is less than the threshold, the result for

Authorized licensed use limited to: UNIVERSITY TEKNOLOGI MALAYSIA. Downloaded on January 12, 2009 at 03:58 from IEEE Xplore. Restrictions apply.

diagnose is a stutter. Figure 6 shows the speech signal score evaluation.

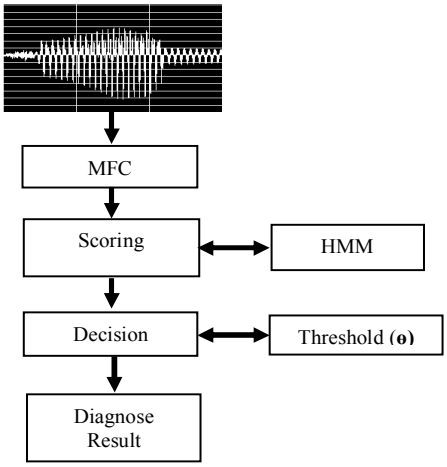

Figure 6: Speech signal score evaluation

Figure 7 shows the normal speech signal pattern while figure 8 shows the artificial stutter signal pattern for speech signal.

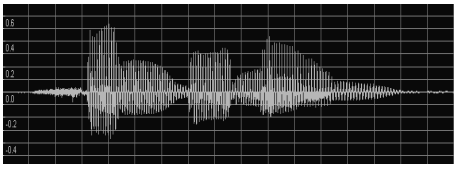

Figure 7: Normal Speech Signal Pattern

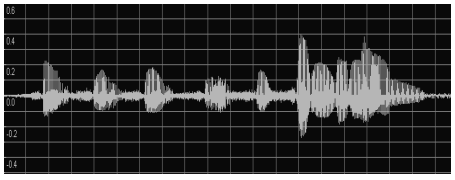

Figure 8: Artificial Stutter Signal Pattern

Figure 9 shows the diagnosis scoring graph of the system. Threshold has been set to get the score result for the correct percentage of recognition rate in normal speech and artificial stutter speech.

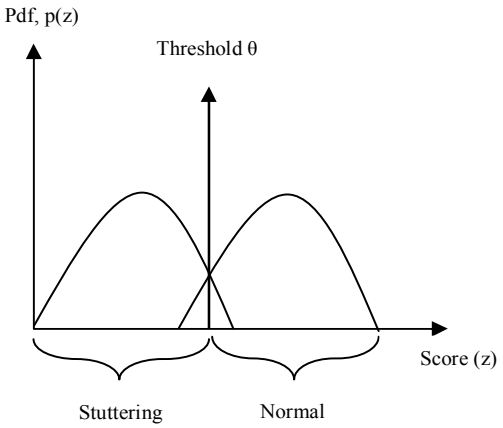

Figure 9: Diagnosis Scoring Graph

VI. RESULTS AND DISCUSSIONS

Test subjects were selected from final year students of Faculty of Electrical Engineering, UTM Skudai. A total of 10 subjects participated, 7 males and 3 females. The participants are required to record word "Sembilan" which mean nine in Malay Language for 20 samples of data for the normal speech and 15 samples of data for the artificial stutter speech. Table 2 shows the results of scoring for normal signal while table 3 shows the score result for artificial stutter signal.

Table 2: Score Result for Normal Speech Signal

Subject	Normal 1	Normal 2	Normal 3	Normal 4	Normal 5
A	-24.5291	-24.309	-24.4748	-23.2062	-24.0997
B	-22.7922	-23.5707	-23.2987	-23.2902	-23.5428
C	-23.5188	-23.5448	-23.8240	-23.8468	-23.9396
D	-24.2263	-23.8672	-23.6559	-24.1884	-23.7044
E	-24.7221	-24.1693	-24.5919	-24.4264	-24.4909
F	-23.3529	-23.1219	-23.1897	-22.8659	-23.2586
G	-23.7277	-23.7113	-24.7342	-23.9714	-23.6783
H	-23.3952	-23.4722	-23.6283	-23.2523	-23.3971
I	-23.2282	-23.7414	-23.4254	-24.3017	-24.9617
J	-23.283	-24.6071	-25.0663	-24.1904	-24.1225

Table 3: Score Result for Artificial Stutter Speech Signal

Subject	Stutter 1	Stutter 2	Stutter 3	Stutter 4	Stutter 5
A	-24.1695	-26.27	-27.5381	-24.8392	-24.5303
B	-26.0134	-26.9843	-26.8738	-24.9187	-26.1228
C	-27.0025	-27.6848	-27.5281	-27.5301	-27.1578
D	-27.0083	-28.4988	-28.0277	-28.0697	-27.8136
E	-25.7746	-25.5725	-25.367	-26.9716	-26.8124
F	-26.6427	-26.9438	-24.8921	-26.2839	-27.5183
G	-26.7827	-23.8984	-24.2606	-26.0668	-25.5724
H	-28.0607	-27.3932	-27.7846	-27.6121	-26.9984
I	-26.1426	-25.8951	-25.7458	-25.5457	-24.4738
J	-24.9843	-25.2556	-27.3584	-26.3006	-27.567

The score results from table 2 and table 3 are used to get the correct percentage of recognition rate for each subject. Each subject will get the correct percentage of recognition rate on the normal and artificial stutter speech.

From the result of each subject, the average percentage of correct recognition rate will be taken. Table 4 below shows the average percentage of correct recognition rate for normal speech is 96%, while for the artificial stutter speech is 90%.

Table 4: Correct Percentage of Recognition Rate

		Percentage of Recognition Rate	
Subject	Threshold θ	Normal	Stuttering
A	-24.7439	100%	60%
B	-24.7439	100%	100%
C	-24.7439	100%	100%
D	-24.7439	100%	100%
E	-24.7439	100%	100%
F	-24.7439	100%	100%
G	-24.7439	100%	60%
H	-24.7439	100%	100%
I	-24.7439	80%	80%
J	-24.7439	80%	100%
Average Percentage		**96%**	**90%**

Authorized licensed use limited to: UNIVERSITY TEKNOLOGI MALAYSIA. Downloaded on January 12, 2009 at 03:58 from IEEE Xplore. Restrictions apply.

VII. CONCLUSIONS

Malay Speech Therapy Assistance Tools (MSTAT) is a system which assists the therapist diagnostic not just for the stutter but also can help the speech therapist to evaluate or keep track with their clients. Our speech recognition system utilizes the Hidden Markov Model (HMM) technique for evaluating speech problem stuttering children. The voice pattern of the normal and stutter children are used to train the HMM model for classifying the problem of speech stuttering. Meanwhile, the system also utilizes Malay text-to-speech system and Talking Head to guide the stutter children during diagnostic process. The average percentage of correct recognition rate for normal speech is 96%, while for the artificial stutter speech is 90%. We hope and believe that our software tool will improve the effectiveness in stuttering diagnostic.

ACKNOWLEDGMENT

This research project is supported by CBE (Center of Biomedical Engineering) at Universiti Teknologi Malaysia and funded by Minister of Higher Education (MOHE), Malaysia under grant "Support Vector Machines (SVM) As a Large Margin Classifier For Machine Learning Algorithm" Vot 78029.

REFERENCES

[1] Tan, T. S., Sheikh, H. and Aini, H. Building Malay Diphone Database for Malay Text to Speech synthesis System Using Festival Speech Synthesis System. *Proc of The International Conference on Robotics, Vision, information and Signal Processing 2003*. January 22-24. Penang, Malaysia: ROVISP03, 634-648.

[2] Tan, T. S. and Sheikh H. Building Malay TTS Using Festival Speech Synthesis System. *Conference of The Malaysia Science and Technology*, September 2-3. Johor Bahru, Malaysia: MSTC 2002, 120-128.

[3] Tan Tian Swee (2003). *The Design and Verification of Malay Text To Speech Synthesis System*. Master Thesis. Univeristy Technology Malaysia, 2003.

[4] Tan Tian Swee, Sheikh Hussain Shaikh Salleh, Aini Hussain(2002). "Building Malay Audio Visual Engine in Developing Audio Visual Teaching Module Speech Recognition Toolkit (ATMSR)", 22-25 July 2002 2nd *World Engineering Congress Sarawak*, Malaysia.

[5] Leen Cleuren (2003). "Speech Technology in Speech Therapy?", State of the Art and Onset to the development of a Therapeutic Tool to Treat Reading Difficulties in the First Grade of Elementary School. SLT Internship at ESAT-PSI Speech Group: 2-3.

[6] Selim S.Awad (1997). The application of digital speech processing to stuttering therapy. *Instrumentation and Measurement Technology Conference, 1997*. May 19-21. Ottawa, Canada: IMTC.1997, Page(s) 1361 - 1367 vol.2.

[7] Anu Subramanian, Ehud Yairi (2006). Identification of traits associated with stuttering. Journal of Communication Disorders 39 (2006) :200–216.

[8] Kristy Hollingshead, Peter A. Heeman (2004). Using a Uniform-Weight Grammar to Model Disfluencies in Stuttered Read Speech: A Pilot Study. Technical Report. Center for Spoken Language Understanding OGI School of Science & Engineering Oregon Health & Science University.

[9] Peter Reitzes (2006). *Five Fun Activities to Practice Pausing with Children Who Stutter*. Journal of Stuttering, Advocacy and Research. Vol 1 Page(s): 102 – 110.

[10] Gerrie Nachman (2006). *Learning and Discovering: A Parent's Journey with Stuttering*. Journal of Stuttering, Advocacy aqnd Research. Vol 1 Page(s):111 – 113.

[11] Eric Jackson (2006). A Stutterer's Perspective: A Stutterer's Challenge. Journal of Stuttering, Advocacy and Research, 1 (2006). Vol 1 Page(s): 114 – 118.

[12] John Coakley (2006). My Journey with Stuttering. Journal of Stuttering, Advocacy and Research, 1 (2006). Vol 1 Page(s): 90 – 93.

[13] Phil Reed, Peter C. Howell, Steve Davis, & Lisa A. Osborne (2007). *Development of an operant treatment for content word dysfluencies in persistent stuttering children: Initial experimental data.* Journal of Stuttering, Advocacy & Research, 2 (2007). Vol 2 Page(s): 1 – 13.

[14] Rabiner. L. R. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, In Proc of the IEEE, 77, 257-286.

[15] 15S.E Levinson, L. R. Rabiner, M. M. Sondhi. 1983. Speaker Independent Isolated Digit Recognition Using Hidden Markov Model, In Proceedings of ICASSP, Boston: IEEE, pp.1049-1052.

[16] Bahl, L. R., Brown, P. F., de Souza, P. V., and Mercer, R. L. 1983. Speech recognition With Continuous Parameter Hidden Markov Models, In Proceedings of ICASSP. New York: IEEE, pp. 40-43.

[17] Rabiner, L. R., Levinson, S. E. and Sondhi, M. M. and Juang, B-H. 1985. Recent developments in the application of hidden Markov models to speaker-independent isolated word recognition, In Proceedings of the ICASSP, Tampa, FL, pp. 9-12.

[18] Tuzun, O. B., Demirekler, M. and Nakiboglu, K. B. Comparison of Parametric & Non-Parametric Representations of Speech for Recognition. 1994. In Proc of 7th Mediterranean Electrotechnical Conf. Turkey: IEEE, 65-68.

Authorized licensed use limited to: UNIVERSITY TEKNOLOGI MALAYSIA. Downloaded on January 12, 2009 at 03:58 from IEEE Xplore. Restrictions apply.