Recognition Test On Highly Newly Robust Malay Corpus Based On Statistical Analysis For Malay Articulation Disorder

Mohd Nizam Mazenan, Tan Tian Swee, Sarah Samson Soh

IJN-UTM Cardiovascular Engineering Centre, Medical Implant Technology Group (MediTEG), Material Manufacturing Research Alliance (MMRA), Department of Biotechnology & Medical Engineering, Faculty of Biosciences and Medical Engineering (FBME), Universiti Teknologi Malaysia (UTM), 81310 Skudai Johor, Malaysia
Email: tantswee@biomedical.utm.my

Abstract— **In designing the Malay language database for articulation disorder, the priority is more on Malay alveolar target words where the important set of words had been used for therapy training exercise especially for the patient at Sekolah Kebangsaan Pendidikan Khas (SKPK), Johor Bahru [9]. The use of manual or traditional technique by speech-language pathologist (SLP) at SKPK is not efficient anymore because it can lead to time consuming and require a lot of involvement of SLP for each therapy session for the ratio of 2:1000 of SLP to patient. Therefore this paper describe the computerized technique that been use in speech recognition where few experiment had been conducted in the process of building the Computer-based Malay Language Articulation Diagnostic System that can be use specifically for speech articulation disorder. The technique use for statistical and processing the word behind this system is Hidden Markov Model (HMM). From the total 108 target words that been collected, few words been selected to run the experiment by using voice sample of real patient. The experiment results shows the accuracy of the recognition rate has achieved about 97% from the overall sample and few words can be claimed as "major spoken" mistake that always happen in speech articulation disorder case. The experiment regarding to voice sample evaluation had also been done to find the total accuracy rate for Malay alveolar consonants.**

Keywords; HMM; Articulation Disorder; SLP; Malay Language Vocabulary.

I. INTRODUCTION

World Health Organization (WHO) statistic shows that, speech and language disorder affects at least 3.5% of human population communication skills [6]. Patient with speech disorder are currently diagnosed manually by the SLP or speech therapist (ST) to confirm their speech problem in most hospital and clinic at Malaysia [13]. This manual process is often referring as a traditional method which may be considered by many SLPs to embrace 'traditional' approaches [14]. This method sometimes lack of accuracy, time consuming and require a lot of involvement of SLP for each session where the ratio is about 1:800 upon SLP and patient in SKPK alone. The common practice in speech therapy is simply having clients repeat words over and over again as a way to correct their spoken phonemes. All the clinician would do was to ask the client to repeat words after her. That would go on for an hour and sometimes it would lead to error elimination during the process [20].

A computer-based speech therapy early diagnosis system is still new in Malaysia and most of the computerized systems available now are in foreign language such as English [19]. Therefore, the proposed of this research is to come up with computerized technique that use speech recognition application and contains specific design of Malay language corpus that can served Malay speech articulation disorder. The development of this system require a correct set of Malay vocabulary database to cover the articulation case in Malaysia and the right technique in speech recognition for higher accuracy in early screening and diagnosis process. The use of HMM in recognition process is the main key for this system development.

II. METHODOLOGY

A. Malay Vocabulary Design

Based on previous observation at SKPK, Malay word for therapy are about 108 words which cover the Malay alveolar and Malay plosives target word. For the target word in the first phase of system development, the target words been use is focusing more on Malay alveolar vocabulary database. Based on the traditional speech articulation therapy in Malaysia hospital, a few target words have been selected. The selected target word is used to diagnose and also to train the patient with speech alveolar disorder. Alveolar consonants are articulated with the tongue against or close to the superior alveolar ridge [15], which is so called because it contains the alveoli (the sockets) of the superior teeth. There are 7 consonants involved in speech articulation in Malay alveolar (d, t, n, z, s, r, l). Some of Chinese speaking people have difficulties pronouncing these 7 consonants [12]. Most assessment instruments focus on the step where the target word will be arranged at the suitable positions such as start (initial), middle (medial) and end (end) based on the need of the patient [3]. Table I show the list of the target words that been use for the experiment towards the system corpus development. For the early experiment, the selected word been chosen is about 50 target words which cover the set of Malay alveolar word.

978-1-4799-6801-5/14/$31.00 ©2014 IEEE

TABLE I. MALAY ALVEOLAR TARGET WORDS

Alveolar	Start	Middle	End
R	rambut	zirafah	seluar
	rumah	kertas	motor
	rusa	jari	leher
D	duit	sudu	jasad
	dagu	baldi	abjad
	dahi	hidung	kad
T	tangan	katil	rambut
	tisu	bantal	biskut
	televisyen	motor	monyet
S	sudu	tisu	kertas
	seluar	wisel	malas
	syampu	biskut	nenas
L	leher	telinga	katil
	lampu	seluar	bantal
	lembu	malas	wisel
Z	zip	beza	ustaz
	zirafah	lazat	lafaz
	zoo	liza	mumtaz
N	nenas	telinga	tangan
	nyamuk	bantal	lapan
	nganga	hidung	cawan

B. *Speech Therapy Process*

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 [18]. As we can see, most of the words in a therapy session are using object thing rather than abstract. This is because, object imagery are easier to visualize where abstract thing is hard to visualize especially for children [4, 8]. Most of the SLP are using visual image followed by words that represent that object. Then the patient will speak what he or she had seen. The flow of this diagnostic can be maintain whether we are using manual or computerized technique. This example flow has been show in Fig. 1 below that has been taken from therapy center at Malaysia.

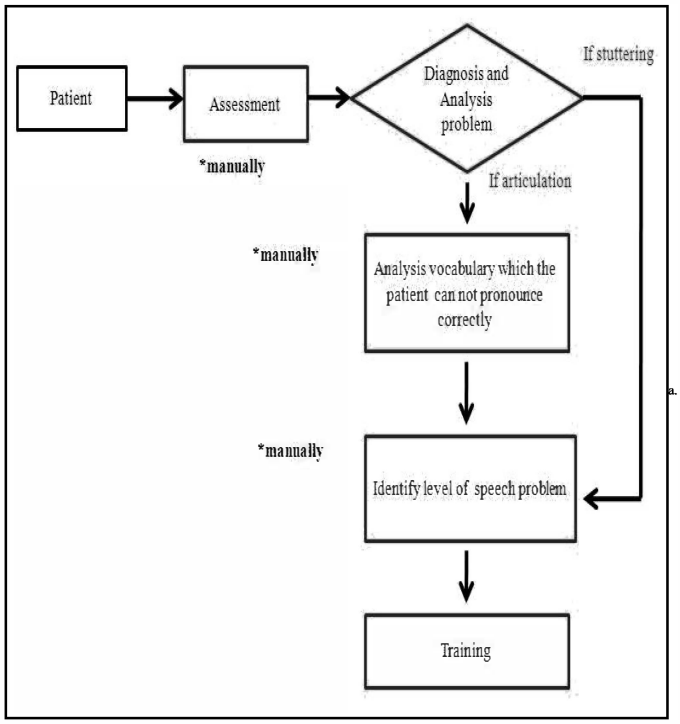


Fig. 1. Conventional Speech disorder diagnosis flow.

Most of the SLP are using visual or image to conduct a testing for each patient. Based on observation at SKPK, SLP are using card that contains a picture with the word of the object. The patient will be asked to say the object loud and SLP will record the score based on their experience and hearing. Then they will make a table of score which there are some calculation where the final result will be conclude the level and type of speech disorder for that particular patient. This step is sometime been call the articulation speech assessment. This assessment is to help the speech therapist to understand the person's articulatory disability before the speech therapist design an effective therapy strategies that consist of exercise or training for that patient in order to cure their specific speech problem. According to James and Michael [10], the assessment involves also the investigation of anatomical and physiological of the organs of the speech production. Some of the patient will need small surgery to cure their problem before they can begin the therapy session [2]. The therapy session may take a month or in a worst case scenario it may took years of repeated procedures which depends on the type and severity of the speech and language problem of the patient [5]. For this research, the use of computerized technique such as speech recognition may help the therapy process more effective and achieve higher accuracy in analysis and diagnosis the speech signal produce by the patient.

C. *Malay Corpus Architecture*

In the designing of Malay corpus for Malay articulation early diagnostic system, there are five step used in this research. First step design the target word. The target word is being selected and designed according to the needs of the speech therapy system. Second step is voice recording where the sample voice is being recorded by using GoldWave software in WAV format. After that, third and fourth step is to transform WAV to SND and transformation again from SND to MFCC model. The last step, is building the HMM Model. The architecture of this database design been shown in Fig. 2. The technique use for statistical and processing the word behind the Computer-based Malay Language Articulation Diagnostic System is actually Hidden Markov Model (HMM). 12 mel-frequency cepstral coefficients (MFCCs) been used as a recognition feature. It is been consider that, MFCC has characteristics of the human auditory system and commonly used in the automatic speech recognition systems [1].

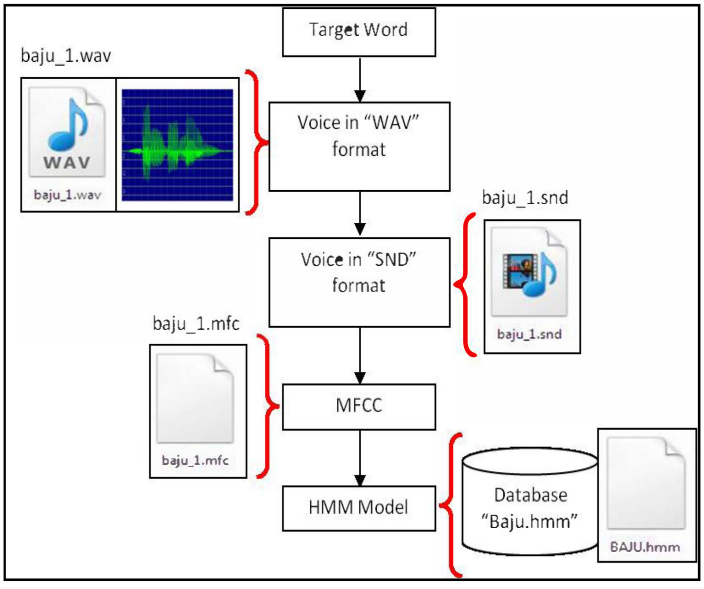


Fig. 2. Malay database design architecture.

For the speech unit database that emphasize on phonemes, there are three important parts in designing the database that is the speech wave, speech unit transcription file and speech unit's speech feature, MFCC [17]. The development of this Computer-based Malay Language Articulation Diagnostic System is towards the direction of segmentation unit in the phoneme or phonology level that can achieve higher accuracy in term of recognition of speech compare to the word based or syllable based level. An experiment for this level needs to be done to prove the concept. Based on previous experiment by Siivola et al., 2003, the overall result for word error rate (WER) in three trigram language model: the baseline word model, syllable model, and morphemes model shows that morphemes model obtained lower WER percentage compare to others model. This experiment is based on Finnish language where it contains high number of distinct word forms. Table II shows this result.

TABLE II. SPEECH RECOGNITION RESULTS USING TRIGRAM MODELS

Lexical Unit	WER	ToER	LER
Word	56.4%	-	13.8%
Syllable	43.9%	18.1%	10.9%
Morphemes	31.7%	20.3%	7.3%

D. Pattern Recognition Based on HMM

The process of speech recognition is starting from the speech signal in an acoustic waveform been converted into digital signal. After that the signal had been model by equation as in (1).

$$P(\vec{O}_t \mid (W_{t-1}, W_t, W_{t+1})) \tag{1}$$

Where the P is the probability, O as the sequence of individual observations and W is the sequence of words. The signal modeling is to highlight important and represent enough speech signal information. Next step is pattern matching that denoted in (2).

$$[W_t^i, P(\vec{O}_t, \vec{O}_{t-1}, \ldots \mid W_t^i)] \tag{2}$$

Pattern matching is been done in a sequence of individual observation for the important feature in the signal data. The next step is to apply search algorithm where the process is to uncover the word sequence $W = W_1 W_2, \ldots, W_m$ that has the maximum posterior probability $P(W \mid O)$ [7, 11]. Equation (3) as shown.

$$P(W_t^i \mid O_t) = \frac{P(O_t \mid W_t^i) P(W_t^i)}{P(O_t)} \tag{3}$$

Where $P(W_t^i)$ is the language model and $P(O_t \mid W_t^i)$ is the acoustic model probability. The last step is where the recognize symbols took place denoted in (4)

$$P(S \mid O) = \arg\max \mid_T \prod_i P(W_t^i \mid (\vec{O}_t, \vec{O}_{t-1}, \ldots)) \tag{4}$$

III. EXPERIMENT AND RESULT

A. Malay Vocabulary Database Analysis

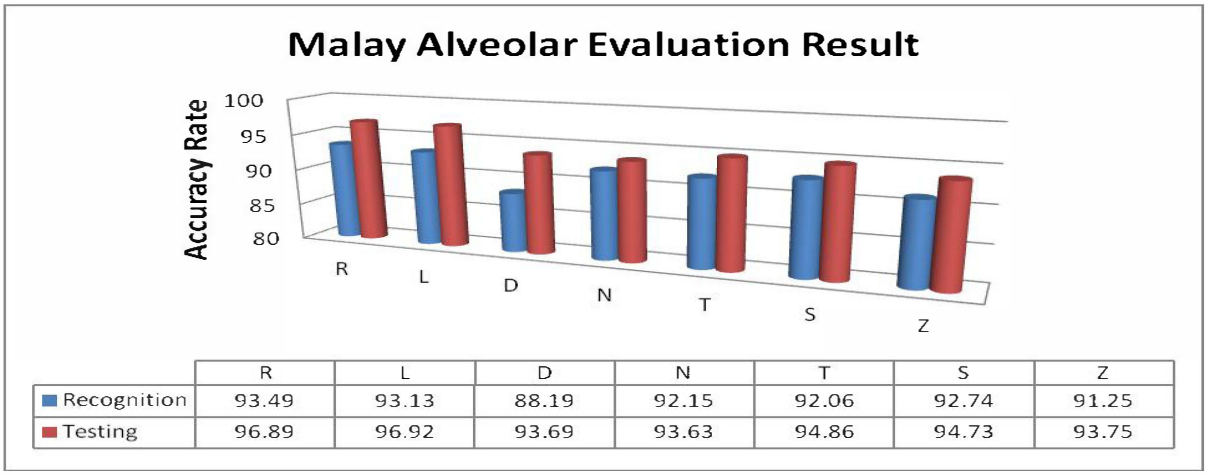
Few experiments have been conducted by using HTK (Hidden Markov Model Toolkit). The first step begins with the total experiment data of Malay vocabulary database where the data of Malay word been collected from SKPK. Thus, this chapter will discuss the Malay vowel evaluation, Malay consonant test, Malay alveolar performance, Malay plosives accuracy rate and the overall accuracy rate for the system. For the first experiment, the speech sample data consists of 100 words taken from 20 patients. The speech sample is in a form of .wav file. Each patient data contains five Malay language words that were *Abjad*, *Baju*, *Dadu*, *Empat* and *Gajah*. The main objective of this experiment is to test the Malay database where the word that been use must be suitable for the case of articulation problem. For this experiment, the total of word is only 5, but for the large scale experiment onwards, this data will be added more for better result. For the overall data in this study, there are about 108 words that cover all the Malay words that enough for speech disorders therapy. The result of the analysis from the sample data shows that the accuracy has achieved about 97% from the overall sample. Table 3 shows the analysis result.

TABLE III. ANALYSIS RESULT OF WORD ACCURACY

WORDS	SCORE / 20	%
ABJAD	19	95
BAJU	18	90
DADU	20	100
EMPAT	20	100
GAJAH	20	100

B. Voice Sample Evaluation

The experiment of evaluating the voice sample of Malay alveolar recognition and testing shows that, the Malay alveolar recognition shows the highest rate in accuracy rate for consonant "R", that is about 93.49%. For the lowest accuracy rate occur in consonant "D" that is about 88.9%. The accuracy rate for consonant "R" is nearly as high as consonant "L". The explanation behind this is because, both of this consonant do replace on another during speaking process. In this experiment, the sample is taken from patient that is been suspected of having articulation disorder where the target word "*Roti*" is pronounced as "*Loti*". It shows that consonant "R" is being replaced with consonant "L". Based on observation, normally consonant "R" will be replaced by consonant "L" and very seldom consonant "L" replacing the consonant "R" during speaking occurs.



	R	L	D	N	T	S	Z
Recognition	93.49	93.13	88.19	92.15	92.06	92.74	91.25
Testing	96.89	96.92	93.69	93.63	94.86	94.73	93.75

Fig. 3. Malay alveolar consonants accuracy rate

For the Malay alveolar testing experiment, the result show consonant "L" has the highest accuracy rate where it reached about 96.92% in the evaluation. A consonant "N" has the lowest accuracy rate for testing that is about 93.6%. And again, Consonant "R" gets 96.89%, nearly as high as the consonant "L" where the same explanation is because there is a conflict between these two consonant. Malaysian Chinese speaking people have trouble pronouncing these two consonants where "R" is pronoun as "L". This is because, they had the difficulty to differentiate and to perform the Malay plosives pronunciation due to the used of Chinese dialect such as Mandarin daily, that did not much using phoneme "L" and "D". Based on the result, 65% of Chinese speaking people have the tendency to pronoun the Malay plosives word wrongly [12]. Fig. 3 shows the evaluation result of accuracy rate for Malay alveolar consonants in a graph illustration.

IV. Conclusion and Future Work

Currently, the used of manual technique in speech therapy is not relevant anymore as the process of this "traditional" technique may lead to time consuming, lack of accuracy in analysis and requires a total involvement of SLP for each therapy session. Therefore, the use of computerized technique by applying speech recognition in the Computer-based Malay Language Articulation Diagnostic System can gain higher accuracy in term of voice signal analysis and diagnostic. Based on early experiment, word can be used as a parameter to test the patient level of accuracy in their pronunciation before the analysis can be made. This statement is based on experiment and a survey made by collaboration with SLP at SKPK. For the future works, the experiment needs to be conduct to find the best way to get the actual result to determine the patient are having articulation disorder whereby the segmentation technique in the phonology level is been use instead of word based or syllable based level. The different of each level can be seen as a solving method for different type of speech disorder problem between word, syllables & phoneme based.

Each of this different type may produce different result in term of emphasizing the parameter targeting. The right combination of this set of type, level and language model in speech recognition process will help the speech recognition system to produce better result with higher accuracy in analysis and the classification of the disorder type for each patient may be accurate. This can help SLP to reduce their diagnosing time for each patient and they can focus more on the rehabilitation process of the patient.

Acknowledgment

The authors gratefully acknowledge the Ainuddin Wahid scholarship provided by Universiti Teknologi Malaysia and the research grant provided by Research Management Centre (RMC), sponsored by Ministry of Higher Education (MOHE), Malaysia. Vot: Q.J130000.2545.04H41, and GUP Universiti Teknologi Malaysia, Johor Bahru, Malaysia.

References

[1] Axelsson, E. Bjo¨rhall.: Real Time Speech Driven Face Animation. In Master Thesis: The Image Coding Group, Linko¨ping University, Linko¨ping, (2003)

[2] B. Otto.: Language Development in Early Childhood. In Pearson Allyn Bacon Prentice Hall edition, p. 334-336 (2006)

[3] Bauman-Waengler J.: Articulatory and Phonological Impairments: A Clinical Focus. Pearson Higher Ed, (2012)

[4] Bell, N.: Visualizing and Verbalizing: For language comprehension and thinking. In Paso Robles, CA: Academy of Reading Publications, (1991)

[5] Canvin R.: Speech and Language Therapy, Bupa Health Information Team. http://www.bupa.co.uk. August (2012), Accessed on 22 January 2013

[6] Bowen C.: Children's Speech Sound Disorders. In New York: John Wiley & Sons, 2009, (2009)

[7] Chen B.: Search Algorithms for Speech Recognition. Retrieved from http://berlin.csie.ntnu.edu.tw/Courses/SpeechRecognition/Lectures2011/SP2011F_Lecture10_Search Algorithms.pdf, (1998), Accessed on 5 February 2013

[8] Gambrell, L.B., Jawitz, P.: Mental imagery, text illustrations, and children's story comprehension and recall. In Reading Research Quarterly, p. 28 (1993)

[9] Sekolah Kebangsaan Pendidikan Khas (SKPK), Johor Bahru, Speech therapist.

[10] James, W. G., Michael, S. H.: Introduction to Human Communication. In Dubuque, IA: Wm C. Brown Publishers, (1992)

[11] Juang, Biing-Hwang.: Hidden Markov Models. In Encyclopedia of Telecommunications, (1985)

[12] Liboh H.: Computer-Based Malay Speech Analysis System for Speech Therapy System. In Master Degree Thesis: Universiti Teknologi Malaysia, Skudai, (2011)

[13] Ooi C. A., Yunus J.: Computer-based System to Assess Efficacy of Stuttering Therapy Techniques. In 3rd Kuala Lumpur Int. Conference on Biomedical Engineering 2006 p. 374-377, (2007)

[14] Powers, M.H.: Clinical Educational Procedures In Functional Disorders Of Articulation. In L.E. Travis, (Ed.). Handbook of speech pathology and audiology, Englewood Cliffs, N.J.: Prentice-Hall (2007)

[15] R. Wyatt, D., Sell, J., Russell S, A., Harding, K., Harland and E. Alberfl.: Cleft Palate Speech Dissected: A Review Of Current Knowledge And Analysis. In British Journal of Plastic Surgery, No. 49(3), pp. 143-149. (1996)

[16] Siivola, Vesa, Teemu Hirsimäki, Mathias Creutz and Mikko Kurimo.: Unlimited Vocabulary Speech Recognition Based On Morphs Discovered In An Unsupervised Manner. In Proc. Eurospeech, vol. 3, pp. 2293-2296. (2003)

[17] Swee, Tan Tian and Sheikh Hussain Shaikh Salleh.: Corpus-based Malay Text-to-Speech Synthesis System, Communications. In 14th Asia-Pacific Conference on. IEEE, (2008)

[18] Tan, T.S., Liboh, H., Ariff, A.K., Ting, C.M., Salleh, S.H.: Application of Malay Speech Technology in Malay Speech Therapy Assistance Tools. In Intelligent and Advanced Systems ICIAS 2007. International Conference on p. 330-334, IEEE. (2007)

[19] Ting H. N., Yunus,J., Vandort, S., Wong L. C.: Computer-based Malay Articulation Training for Malay Plosives at Isolated, Syllable and Word Level. In Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on, vol. 3, p. 1423-1426. IEEE, (2003)

[20] Van Riper, C.: Eliciting Sounds: Techniques and Strategies for Clinicians. Personal correspondence to Wayne Secord. In: Secord et al (2007) Second Edition. Clifton Park, NY: Thomas Delmar Learning. (P. viii), (1993)