Malay Language Speech Recogniser with Hybrid Hidden Markov Model and Artificial Neural Network (HMM/ANN)

H. F. Ong and A. M. Ahmad

***Abstract*—There are many artificial intelligence approaches used in the development of Automatic Speech Recognition (ASR), and hybrid approach is one of them. The common hybrid method in speech recognition is the combination of Hidden Markov Model (HMM) and Artificial Neural Network (ANN). The hybrid HMM/ANN is able to combine the strength of HMM in sequential modeling structure and ANN in pattern classification. Thus, this paper proposed a speaker independent and continuous Malay language speech recogniser by using the hybrid HMM/ANN method. In addition to that, this paper presents a study on Standard Malay's phonetic and phonology to help in the recognition of Malay words. The CSLU toolkit is utilized for building the recogniser, and the experimental results showed that the proposed HMM/ANN model outperformed the conventional HMM model. The performances of the recognisers are measured in term of word accuracy and sentence accuracy.**

***Index Terms*—Artificial Neural Network, Continuous Speech, Hidden Markov Model, Hybrid HMM/ANN, Malay Language, Speaker Independent, Speech Recognition.**

I. Introduction

Automatic speech recognition (ASR) technology allows a computer to identify the words spoken by a person through a microphone or other voice input device. It has long been viewed as a promising alternative for human-computer interaction (HCI) over the traditional keyboard and mouse [1]. In general, the working modes of a speech recognition system can be either speaker-dependent or speaker independent and either isolated-word or continuous speech. Over the years, most of the ASR systems are developed for speaker-independent and continuous-speech recognition to support multi-users that are allowed to speak in more natural way [2-5]. This mean user is no longer needed to train an ASR system to recognise their voice or to speak with pauses between words. In order to achieve that, the limitations such as vocabulary size, noise, and speaker characteristics should be removed.

In the early of 1970's, the Hidden Markov Model (HMM) was first implemented to the speech recognition field by Baker for the Dragon system [6]. Since then, the HMMs have become the dominant technology in ASR. The main advantages of HMM-based systems are the statistical representations of the acoustic speech signal and the stochastic processes that capable of modeling sequential data. However, standard HMMs have some drawbacks in building a large vocabulary speaker independent continuous ASR system. It has poor discrimination power due to unsupervised learning [7] where the model parameters are estimated by maximum likelihood (ML) estimation. Moreover, HMMs require distributional assumptions and typically make first order Markov model assumption for phone or sub-phone states while ignoring the correlation between acoustic vectors. In the 90's, a new type of speech recognition systems have been developed using connectionist methods. According to [8], a connectionist system is which information is represented and processed in terms of the input pattern and strength of connections between units that do some simple processing on their input. The Artificial Neural Networks (ANN) models have been used for connectionist speech recognition but with limited success. This is because, although ANN has a good discriminative power and flexible, it is not tailored for sequential data such as speech [8]. Thus, hybrid HMM/ANN system is proposed to augment ASR performance, where it makes use of the discriminative power of ANN and relies on the temporal aspects of HMM. In a hybrid HMM/ANN architecture, ANN methods can be used for estimating posterior probabilities and training the network, while HMM methods can be used for decoding and language modeling. The Center of Spoken Language Understanding (CSLU) toolkit is one of the toolkits that support the development of HMM/ANN based speech recognition systems.

Apart from the methods applied in ASR systems, phonetic and phonology study also plays an important role for accurate recognition. By determining appropriate phoneme for specific language with correct pronunciation of words will produce a better recogniser. Currently, most of the studies in the literature are based on English speech recognition, and in comparison Malay speech recognition are still limited especially those using hybrid HMM/ANN for speaker independent and continuous speech recognition system. Therefore, this paper aims to apply hybrid HMM/ANN approach for developing a speaker independent continuous speech recogniser with a medium sized vocabulary. Moreover, further study on the appropriate phonemes for Malay language is carried out to augment the speech recognition performance. The experimental results indicate that the word and sentence accuracy for hybrid HMM/ANN

Manuscript received May 30, 2011

H. F. Ong is with the Faculty of Management and Information Technology, UCSI University, 56000 Cheras, Kuala Lumpur, Malaysia (e-mail: hueyfang@ucsi.edu.my).

A. M. Ahmad is with the Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia (e-mail: manan@fsksm.utm.my).

model outperformed the HMM model.

The remainder of this paper is organized as follows. In Section II, we present the Standard Malay phonetics and rules. Section III, we introduce the hybrid HMM/ANN method, while Section IV we discuss on the proposed speaker independent and continuous speech recogniser. The experimental settings and results are then described in Section V. Finally, we conclude our works in Section VI.

II. Standard Malay Phonetic

Over 300 million people in Malaysia, Indonesia, Brunei, Singapore, and Southern Thailand speak Standard Malay (SM) as their first or second language. SM is a member of Western Group of the Austronesian family, and it is the native language for the Deutero-Malays, or normally known as the modern Malays [9].

SM has writing and sound systems different from those of English. The sound system of SM is composed of primary consonant, secondary consonant, vowels and diphthongs. There are 19 primary consonant that can be classified as stops (voiced and voiceless), fricatives (voiced and voiceless), affricates, nasals, lateral, and alveolar still. Besides that, there are 8 secondary consonants, 6 vowels and 3 diphthongs in SM [9, 10]. Phonemes are the basic sounds of a language and any words can be described as a sequence of phonemes (pronunciation). There are about 40 phonemes in English and 36 phonemes in SM. Table I shows the list of phonemes in SM sound system.

TABLE I: Phonemes in Standard Malay Sound System

Phoneme in IPA Symbol	Symbol used in annotating speech waveform	Example of Words
Vowels		
/a/	a	*a*bu, bol*a*, s*a*ya
/e/	e	*e*mak, *e*mpat
/i/	i	*i*ni, *i*snin
/o/	o	b*o*la, *o*tak
/u/	u	*u*lang, b*u*mi, ib*u*
/∂/	e’	*e*nak, m*e*rah
Diphthongs		
/ai/	ai	*ai*r
/au/	au	*au*rat
/oi/	oi	b*oi*kot
Primary Consonants		
/p/	p	*p*ada, *p*ukul
/b/	b	*b*ola, *b*aca
/t/	t	*t*idur, *t*idak
/d/	d	*d*ia, *d*an, *d*i
/k/	k	*k*ami, *k*hamis
/g/	g	pergi, beg
/?/	?	-
/m/	m	*m*akan, *m*alam
/n/	n	*n*asi, *n*aik
/ŋ/	ng	kuni*ng*, te*ng*ah
/ɳ/	n~	*ny*ata
/ts/	ts	*c*awan
/dz/	dz	*c*uti, *j*awab
/s/	s	*s*aya, ma*s*a
/h/	h	*h*ari
/r/	r	*r*inggit, *r*atus
/l/	l	*l*apan, *l*ama
/w/	w	ka*w*an, *w*arna
/j/	j	sa*y*a
Secondary Consonants		
/f/	f	*f*akir
/v/	v	*v*itamin, no*v*el
/δ/	r=	sibe*r*
/ð/	D	*d*aif
/θ/	T	hadi*s*
/z/	z	*z*aman, *z*oo
/ʃ/	S	*sy*arat
/x/	x	*x*enon

In SM, there are 30 rules covered the grapheme-to-phoneme conversion component and 46 rules covered phoneme-to-phonetic component. Listed below are some of the rules applied in this paper:

a) Omission of grapheme r when it occurs in the final position of a word.
- If final r is preceded by a, then omit r and remain a. Example, the word pasar to /pasa/.
- If final r is preceded by i, then omit r and i is substituted by ∂. Example, the word tabir to /tab∂/.
- If final r is preceded by u, then omit r and u is substituted by o. Example, the word tabur to /tabo/.
- When another consonant follows by e, r is deleted. Example kertas to /ketas/.

b) The insertion of glottal stop ?.
- When a word begin with a vowel. Example ambil to /?ambil/.
- Transformation of final k to a glottal stop. Example budak to /buda?/.
- Replacement of prevocalic h to glottal stop. Example hutan to /?utan/.
- To separate to vowels that not a dipthong. Example soal to /so?al/

c) Pronunciation of character e to vowel e or schwa ∂.
- Example nonhomograph word selamat to /s∂ lamat/.
- Example word rela to /r∂ la/.

d) Replacement of grapheme a at the end of a word by ∂.
- Example rasa to /ras∂ /.

e) Diphthong generation rules.
- The grapheme sequence ai equal to /aj/. Example haiwan to /hajwan/.
- The grapheme sequence au to equal /aw/. Example kalau to /kalaw/.
- The grapheme sequence oi to equal /oj/. Example sepoi to /s∂ poj/.

f) Insertion of glide j or w in between sequences.
- The grapheme sequence ia to /ija/.
- The grapheme sequence iu to /iju/.
- The grapheme sequence io to /ijo/.
- The grapheme sequence i∂ to /ij∂ /.
- The grapheme sequence ua to /uwa/.
- The grapheme sequence ui to /uwi /.
- The grapheme sequence u∂ to /uw∂ /.

g) Consonant deletion rules.
- The grapheme sequence ch replaced by phoneme sequences /ts/.
- The grapheme sequence sy replaced by phoneme sequences / ʃ /.

- The grapheme sequence ny replaced by phoneme sequences /ɲ/.
- The grapheme sequence ng replaced by phoneme sequences /ŋ/.

h) Vowel reduction u to o

- Whenever the vowel u is between two consonants and appears as the final syllable. Example peluk to /pelok/.

III. Hybrid HMM/ANN System

Hybrid HMM/ANN is a potential connectionist method for developing ASR system. This method combined the respective properties of ANN and HMM with the aims to solve the limitations in ASR, such as in building continuous and speaker independent recogniser. In this paper, the HMM/ANN method used Markov process to temporally model speech signals, and ANN is used to estimate posterior probabilities for each HMM states and to train the system. Fig. 1 illustrates the typical structures of a hybrid HMM/ANN system. Both methods operate separately on their own layer and play a role in the whole recognition process.

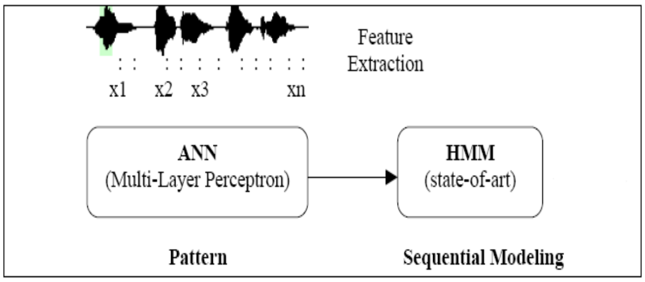

Fig. 1. A typical hybrid HMM/ANN speech recognition system.

The main advantage of HMM is rich of mathematical structure, thus it is able to characterize speech signal in a mathematically tractable way. HMMs are a class of stochastic processes that capable of modeling time-series data and typically defined as a stochastic finite-state-automaton. “Hidden” in HMM is to mean the underlying stochastic process, where the sequence of states is visited and is not directly observable. Observable stochastic process is the output acoustic feature vector at each state. The disadvantages of HMM is poor discrimination power and distributional assumptions, which the first Markov model assumption for phone or sub-phones states while ignoring the correlation between acoustic vectors [15].

On the other hand, ANN typically consist of a number of interconnected processing units, called neurons, which is capable of taking in numbers of input and producing an output. The real power of ANN method is when they combined neurons into multilayer structures called neural networks. Each neuron has input/output (IO) characteristics and implements a local computation or a function. The output of a unit is determined by the I/O characteristics, its interconnection to other units and external inputs. Each connection in the network has a specific weight, and the learning process involves adjustment of weights. Training a network means adapting its connections so that the network exhibits the desired computational behavior for all input pattern. An example of training algorithm used in ANN is the forward-backward algorithm. Moreover, ANN can be trained to produce posterior probability of HMM states when given the acoustic data. Each ANN output will be associated with a specific HMM state. In sum, the advantages of HMM/ANN system are stated as follows [8]:

- Provide a natural structure for discriminative training.
- Provide accurate recognition as no strong assumptions on the statistical distribution of acoustic space.
- Provide ability to model acoustic correlation using contextual information or input.
- Parsimonious use of parameters.
- Better robustness for insufficient training data.
- Efficient in CPU and memory run-time requirements.

IV. The Proposed Speaker Independent and Continuous Malay Language Speech Recogniser

This section discuss on how a speaker independent and continuous speech recogniser is developed using the hybrid HMM/ANN method. The development of the recogniser consists of the training phase and the recognition phase. In the training phase, the recogniser learns the reference patterns representing different speech data. The speech data contains words to be recognised and used to train a neural network. Each reference (training data) is learned from spoken examples stored in form of template or models. While in the recognition phase, unknown input speech is identified by considering the set of possible references. Fig. 2 shows the phases involved in the proposed hybrid HMM/ANN speech recogniser.

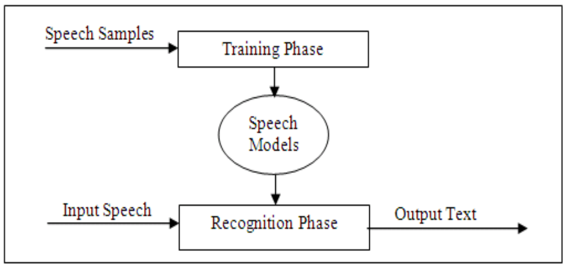

Fig. 2. The phases involved in a hybrid HMM/ANN speech recogniser.

i. *Training Phase*

The training of a recogniser is conducted in advance. Samples of speech data similar to those wanted to be recognize are collected and used to train the network. Once trained, the network will retain the “knowledge” of these samples. The training of the neural network can be done with forward-backward algorithm on a fully connected 3-layer feed-forward network. The processes for training the proposed speech recogniser is illustrated in Fig. 3.

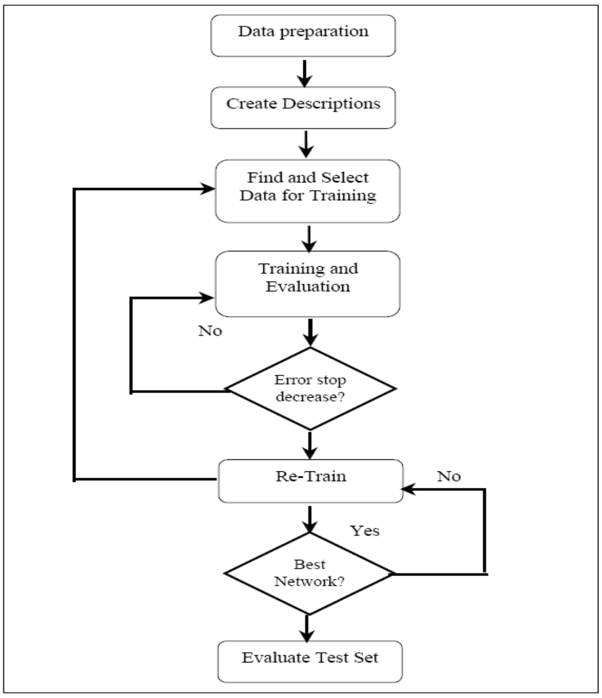

Fig. 3. The processes for training the proposed speech recogniser.

ii. *Recognition Phase*

Fig. 4 shows the recognition framework for the proposed speech recogniser. The differences between this framework and standard HMM method is posterior probabilities are estimated using a neural network instead of a mixture of Gaussians. Using a neural network has better advantage as it does not require assumptions on the distribution or independence of the input data. Moreover, it can easily perform discriminative training and is much faster compare to standard HMM method. The details of the components involved are discussed in the following section.

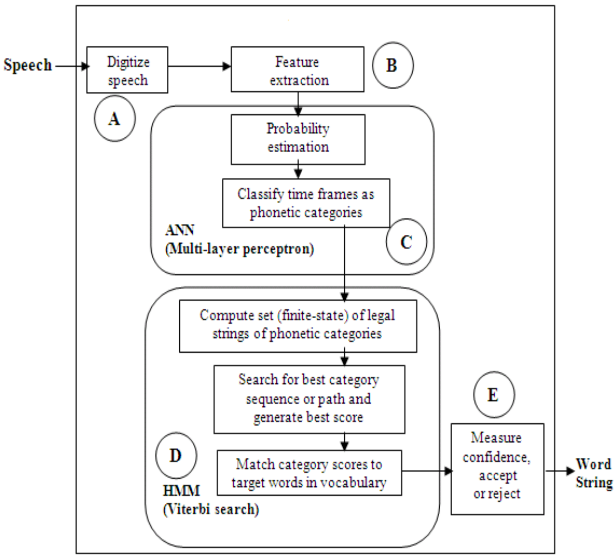

Fig. 4. The recognition framework for the proposed speech recogniser.

A. *Digitize the speech*

Signal is captured by a microphone (or other transducer) and converted into an electrical signal, where the amplitude of the signal corresponds to the magnitude of the original pressure variation. Then the signal is sampled at specific frequency, usually 8000 Hz, so that only a finite number of amplitudes are recorded, stored, or transmitted, for a given period of time. After that, the signal is quantized into one of a discrete number, so that only a finite number of bits are required to represent each sample. This is called Analog-to-Digital (A-to-D) conversion [14].

B. *Feature extraction*

Modify original signal, such that when the result is played, it is still readily recognized by human listeners as the original. There are several ways to modify the signal such as DC-offset removal, frequency components, Perceptual linear predictive (PLP), and energy normalization.

C. *Computing phonemes probabilities using ANN*

A feed-forward neural networks is use to compute phoneme probabilities. The output values of certain units are set externally. These are the inputs to the network representing the frame of speech to be classified. These values are propagated across connections to other units. Each destination unit sums its network input and computes an output value in the range 0 to 1 from this sum. If there are additional layers of units, the process is repeated. Eventually, the propagated activation reaches the final layer of the network, which has no further connections. These units are the output of the net and, ideally, their values represent the probability that the input is from phonemes as in Section II. The major difference between this approach compared to the standard HMM system is the posteriori probabilities are estimated using a neural network instead of a mixture of Gaussians. Using a neural network for estimation is better as it does not require assumptions on independence of the input data and can be easily perform discriminative training [14].

D. *Search algorithm using HMM*

Once the phoneme probabilities are computed, the system use an algorithm called the Viterbi search to find the highest score path. The phoneme probabilities for each successive frame are arranged in a matrix. Then the system finds the paths through the matrixes that give the highest score and match the score with provided target words. The formal steps in the Viterbi algorithm approach for finding the best state sequence are as follows [11]:

Step 1: Set the initial state values

$$\delta_1(i) = \pi_i b_i(O_1) \qquad 1 \le i \le N;$$

$$\psi_1(i) = 0$$

Step 2: Recursion to maximize state sequence

for $2 \le t \le T;\ 1 \le j \le N$

$$\delta_t(j) = \max[\delta_{t-1}(i)\, a_{ij}]\, b_j(O_t) \qquad 1 \le i \le N;$$

$$\psi_t(j) = \operatorname{argmax}[\delta_{t-1}(i)\, a_{ij}] \qquad 1 \le i \le N;$$

Step 3: Termination if all state have been traversed

$$P^* = \max[\delta_T(i)] \qquad 1 \le i \le N;$$

$$i^*_T = \arg\max[\delta_T(i)] \qquad 1 \le i \le N;$$

Step 4: Backtracking to determine optimal state sequence

for t = T-1, T-2,…,1

$$i^*_t = \psi_{t-1}(i^*_{t+1})$$

where N represents the number of states and T represents the length of the observation sequence. The symbol a_{ij} is the state transition probability from state i to state j, while b_i is the observation probability distribution in state i or the probability of an input vector representing each specific phone. On the other hand, π_i is the initial state distribution or the probability of each phone being the first phone of a word.

The algorithm finds the optimal state sequence for the observation sequence. The most likely state is determined at each instance in time (t). The probability is computed based on the probability of being in the previous state (ψ(t-1)(i)), the transition probability from the previous state to the current state and the observation probability of the symbol in the current state [11].

E. Confidence and rejection

Word scores that reflect relative merit are computed. The word score is compared with a threshold. If the score is better than the threshold, then accept the word. If worse, then reject the word.

V. Experiments

Several experiments were conducted to prove that hybrid HMM/ANN is a better method compared to traditional HMM method for building a speaker independent and continuous Malay language speech recogniser. The criteria to evaluate the performance of the speech recogniser are based on the word accuracy and the sentence accuracy. Moreover, the time needed to train the recogniser is taken into consideration.

A. Data Preparation

In this study, the speech data was collected for training set and test set. For the training set, 250 speech data was collected from 5 speakers where their speeches were recorded at 8000 samples/sec sampling rate in waveform files. The selected speakers have a good command in Malay language and are from different genders and races. A total of 250 Malay language sentences with 74 different words were read by the speakers. On the other hand, 30 speech data was randomly taken out from the training set to be tested on the recogniser.

For pre-processing, the speech data needs to be hand labelled. The waveform files are processed by using the Speech View program found in the CSLU Toolkit [12]. The process of annotating the speech waveform can be done at different transcription levels. For instance, the phone-level transcription labels the pronunciation for the uttered words by referring to the phonemes for Standard Malay language (see Section II). Fig. 5 illustrates an example of hand labelled speech data by using the Speech View.

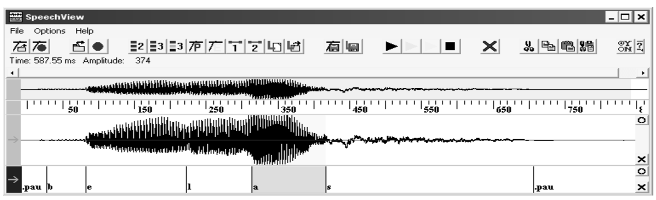

Fig. 5. An example of a recorded speech data ("belas") in waveform.

B. Experiment Setups

The modes for the proposed speech recogniser are speech to text, speaker independent, continuous speech, and medium vocabulary size with simple Malays sentences. This study utilized the CSLU toolkit for waveform analysis and for implementing the hybrid HMM/ANN approach in training and recognition phases with Tcl/Tk scripting language. The CSLU toolkit was designed for research and educational purposes in the area of speech and HCI [12]. It has many modules and tools interacting with each other in development environments called the CSLU-HMM [13] and CLUsh [14]. The system is built on Windows platform with processor 1.0 GHz, or above and memory 512MB, or higher.

To evaluate the performance of the proposed HMM/ANN speech recogniser, it is compared with a conventional Semi-Continuous HMM (SCHMM) recogniser [15]. For fair comparison, the characteristics for both systems are set to be almost the same as shown in Table II. The main differences between the recognisers are the approach, toolkit, and pronunciation used. Moreover, the sample of speakers used in this study is different from the previous work.

TABLE II: Comparison of the Characteristics of hte Conventional HMM and the Proposed Hybrid HMM/ANN Speech Recogniser

Characteristic	Semi-Continuous HMM Speech Recogniser	Proposed Hybrid HMM/ANN Speech Recogniser
Recognize language	Malay	Malay
Recognition mode	Speaker independent Continuous speech	Speaker independent, Continuous speech
Toolkit	Sphinx	CSLU Toolkit
Pronunciation	CMU phone	Standard Malay phone
Total words trained	74	74
Total sentences trained	250	250
Training data (corpuses/sentences)	250	250
Test data (corpuses/sentences)	30 (from training speakers)	30 (from training speakers)
Number of speakers	5	5

C. Experiment Results and Discussions

This study compares the performance of the proposed speech recogniser with conventional recogniser by using the recognition accuracy. Table III shows the results obtained for both systems. The word accuracy is the number of words that were recognized correctly divided by the number of words that were spoken. On the other hand, the sentence accuracy is the number of recognized sentences with no word errors divided by the number of sentences that were spoken.

TABLE III: Comparison of the Results for Different Systems

Speech Recogniser	Word Accuracy (%)		Sentence Accuracy (%)		Training Time (minutes)
	250 Training Data	*30 Testing Data*	*250 Training Data*	*30 Testing Data*	
SCHMM	99.66	92.77	74	74	39
Hybrid HMM/ANN	100	93.85	100	83.33	35

Based on the results in Table III, the speech recogniser using hybrid HMM/ANN approach outperformed the recogniser using SCHMM approach. The word and sentence accuracy on testing data for the hybrid model are 93.85% and 83.33% respectively, which are 1.08% and 9.33% higher compared to the SCHMM model. In addition to that, the training of the hybrid model is 4 minutes faster than the SCHMM model. These show that hybrid HMM/ANN approach is suitable for developing a speaker independent and continuous Malay speech recogniser and is better than conventional HMM approaches. It is also believed that by following the Standard Malay language phone set and rules, the recogniser performance is enhanced compared to the Carnegie Mellon University (CMU) phone set, which is more suitable for English language recognition.

One of the problems encountered on the proposed speech recognition system is decoding words with similar pronunciation such as “masak” with “masa”, “satu” with “sabtu”, “khamis” with “kami”, and “rabu” with “abu”. To solve this problem more training data are needed or larger corpus size. Apart from that, the recogniser performance is dependent on the speakers’ style of speech, where better recognition results are produced if pronunciation of words is clear and accurate.

The developed recogniser can be applied in various applications. For instance, users are allowed to input text to an editor through their speech. Besides that, they can issue speech commands to control the editor, which make their works easier compared to typing in a conventional way. Fig. 6 illustrates an example of prototype created by using the proposed hybrid HMM/ANN speech recogniser.

Fig. 6. An example of speech-to-text editor prototype using the proposed hybrid HMM/ANN speech recogniser.

VI. Conclusions

In conclusion, this paper has proposed the use of hybrid HMM/ANN method for developing a speaker independent and continuous Malay language speech recognition. The recogniser is implemented using the CSLU toolkit with speech data collected from multiple speakers. Standard Malay phonemes and rules are followed to increase the recogniser performance for Malay language. The experimental results showed that the proposed hybrid recogniser outperformed the conventional recogniser in term of word accuracy and sentence accuracy. For future works, the training data size can be increase to produce a quality recogniser. A study to examine the appropriate parameter settings for training model and the best decoding algorithm is suggested. Besides that, further study can be done on the usage of suitable phonemes in Malay words pronunciation for developing effective ASR systems.

References

[1] S. Srinwasan and E. Brown, "Is speech recognition becoming mainstream?," Computer, vol. 35, pp. 38-41, 2002.

[2] C. Burileanu, A. Buzo, C. S. Petre, D. Ghelmez-Hanes, and H. Cucu, "Romanian Spoken Language Resources and Annotation for Speaker Independent Spontaneous Speech Recognition," in Digital Telecommunications (ICDT), 2010 Fifth International Conference on, 2010, pp. 7-10.

[3] M. A. M. Abushariah, R. N. Ainon, R. Zainuddin, M. Elshafei, and O. O. Khalifa, "Natural speaker-independent Arabic speech recognition system based on Hidden Markov Models using Sphinx tools," in Computer and Communication Engineering (ICCCE), 2010 International Conference on, 2010, pp. 1-6.

[4] J. L. Gauvain, L. F. Lamel, G. Adda, and M. Adda-Decker, "Speaker-independent continuous speech dictation," Speech Communication, vol. 15, pp. 21-37, 1994.

[5] K.-F. Lee, "On large-vocabulary speaker-independent continuous speech recognition," Speech Communication, vol. 7, pp. 375-379, 1988.

[6] J. Baker, "The DRAGON system--An overview," Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 23, pp. 24-29, 1975.

[7] G. Tesauro, D. Touretzky, and T. Leen, Advances in neural information processing systems 7: MIT Press, 1995.

[8] H. Bourlard and N. Morgan, Connectionist speech recognition: a hybrid approach: Kluwer Academic Publishers, 1994.

[9] Y. A. El-Imam and Z. M. Don, "Rules and Algorithms for Phonetic Transcription of Standard Malay," IEICE - Trans. Inf. Syst., vol. E88-D, pp. 2354-2372, 2005.

[10] I. Dahaman, Pedoman ejaan rumi bahasa Melayu: Dewan Bahasa dan Pustaka, 1996.

[11] C. E. Steven, "Context Sensitive Optical Character Recognition using Neural Networks and Hidden Markov Models", The Faculty of the Computer Science Department, Master of Science in Computer Science: Rochester Institute of Technology, 1992.

[12] S. Sutton and R. Cole, "The CSLU toolkit: rapid prototyping of spoken language systems," in Proceedings of the 10th annual ACM symposium on User interface software and technology Banff, Alberta, Canada: ACM, 1997.

[13] J. Schalkwyk, P. Hosom, E. Kaiser, and K. Shobaki, "CSLU-HMM: The CSLU Hidden Markov Modeling Environment," Center for Spoken Language Understanding (CSLU), Oregon Graduate Institute of Science & Technology. 2000.

[14] J. Schalkwyk, D. Colton, and M. Fanty, "The CSLUsh Toolkit for Automatic Speech Recognition," Center for Spoken Language Understanding (CSLU), Oregon Graduate Institute of Science & Technology. 1998.

[15] Y. P. Wong, "Applying Semi-Continuous Hidden Markov Model for Malay Speech Recognition", Faculty of Computer Sciences and Information Systems, Bachelor in Computer Science: Universiti Teknologi Malaysia, 2005.

H. F. Ong, photograph and biography not available at the time of publishing.

A. M. Ahmad, photograph and biography not available at the time of publishing.