

## Hybrid methods of Brandt's generalised likelihood ratio and short-term energy for Malay word speech segmentation

Noraini Seman, Ahmad Firdaus Norazam

Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA (UiTM), Malaysia

---

### Article Info

#### Article history:

Received Jan 21, 2019

Revised Apr 20, 2019

Accepted May 5, 2019

---

#### Keywords:

Brandt's GLR  
Malay pantun  
Segmentation  
Spectral  
Speech recognition

---

### ABSTRACT

Speech segmentation is an important part for speech recognition, synthesizing and coding. Statistical based approach detects segmentation points via computing spectral distortion of the signal without prior knowledge of the acoustic information proved to be able to give good match, less omission but lot of insertion. In this study the segmentation is done both manually and automatically using Malay words in traditional Malay poetry. This study proposed a hybrid method of Brandt's generalized likelihood ratio (GLR) and short-term energy algorithm. The Brandt's algorithm tries to estimate the abrupt change in energy to determine the segmentation points. A total of five Pantun are used in read mode and spoken by one male student in a noise free room. Experiments are conducted to see the the accuracy, insertion, and omission of the segmentation points. Experimental results show on average 80% accuracy with 0.2 second time tolerance for automatic segmentation with the algorithm having no knowledge of the acoustic characteristics.

*Copyright © 2019 Institute of Advanced Engineering and Science.  
All rights reserved.*

---

### Corresponding Author:

Noraini Seman,  
Department of Computer Science,  
Faculty of Computer & Mathematical Sciences,  
UiTM Shah Alam, Selangor, Malaysia.  
Email: aini@tmsk.uitm.edu.my

---

## 1. INTRODUCTION

Humans mainly communicate with each other through words formed from languages. Information exchanges are easily achieved between two individuals that speaks the same language. Computers and digital systems also works as such. A computer can only understand instructions in one's and zero's and thus it is the compilers job to translate high level programming language into these instructions. Automatic Speech Recognition Systems (ASR) aim to "translate" vocal human words in natural language into information usable by computers and digital systems [1]. Speech signal varies greatly based on the context [2]. Even when the same speaker says the same word repeatedly will result in variation, even little, in the speech signal produced. Human communication are complimented with body language and simpler versions of language that better suit two way dialogues [1]. Among these, unclear word boundaries, noise signals, regional and geographical dialects, and speaker variability makes building an accurate ASR system harder.

ASR pre-processing stage will greatly determine the outcome of the later stages. Framing, noise removal, and segmentation are common processes that are done during pre-processing [2]. The focus of this paper is on continuous audio segmentation. Segmentation algorithm can be categorized as follow [3]: First is Metric-based segmentation where audio streams are segmented at the maxima of the distances between neighbouring windows placed at fixed sampling intervals. Second, Decoder-guided segmentation where audio streams are decoded followed by segmentation at silent points generated by the decoder. Third, Model-based segmentation such as the use of Gaussian mixture models. Segments are assumed at locations where there is a change in acoustic class. The incoming stream can be classified by using maximum likelihood selection.

Speech segmentation can be defined as the process of finding the limits (with specific characteristic) in natural spoken language between words, syllables or phonemes. [4, 5]. The main objective of Speech segmentation is to serve other speech analysis problems such as speech synthesis, data training for speech recognizers, or to fabricate and label prosodic databases. Therefore, it can be viewed as a vital sub-issue for various fields in speech analysis and research. [6, 7]. The traditional approach handling this issue is by manual segmentation of speech, which is generally performed by specialized phoneticians. However, this method is based on listening and visual judgment on required boundaries which makes it inconsistent and time consuming. [8, 9]. Another method which is considered very convenient is an automatic segmentation. The speech can be automatically segmented into sub word units which are defined acoustically. [10] In Automatic Speech Recognition ASR systems, segmentation can be performed: (i) At the system training stage, when segmentation is applied to the training set recordings. (ii) At the recognition stage [5].

## 2. SEGMENTATION TECHNIQUES

Several well-established segmentation techniques have been proposed by previous researchers, such as in [10], that audio segmentation is performed using segment features. The proposed technique uses a log-linear segment model to determine the segmentation of the input audio stream [11]. First, the audio data is processed with a speaker independent acoustic model [12]. The decoding process will hypothesis the locations of sentence start and end. The resulting segments are also clustered and used in Constrained maximum likelihood linear regression (CMLLR) feature transformations and maximum likelihood linear regression (MLLR) mean transformations. The experimental results in [10] shows that the framework is applicable for various segments, boundary features, and for different change point detection methods.

The Hidden Markov Model (HMM) is one of the highly-used segmentation techniques. A refined HMM algorithm was tested for segmenting a Chinese corpus [11]. The method is carried out in 3 steps:

1. Obtain initial segmentation marks using HMM with forced alignment.
2. Create a super vector for each boundary of this database by placing acoustic vector near the boundary. The pseudo-triphone formed from the boundary are classified using a classification and regression tree (CART) where the pseudo-triphone are clustered into smaller number of classes. Then each leaf node on the CART is used to train a Gaussian Mixture Model (GMM).
3. For each labelled sentence, attempt to refine the boundary of each segment. Using the HMM boundary obtained above, compute the likelihood of this frame contains the actual boundary. The optimal boundary is assumed to be the frame that has maximum likelihood of the GMM model associated with the CART leaf node for the pseudo-triphone.

Experimental results in [5] shows that the refined HMM is more accurate than the standard HMM segmentation.

The Brandt's generalized likelihood ratio (GLR) method aims to detect discontinuities in homogenous segment of the speech signal models using statistics to detect sequentially abrupt changes in the parameter of the model [11-14]. The signal  $Y_n$ , is described using an autoregressive model  $M$ , such that

$$M(A, \sigma)f(x) = \begin{cases} Y_n = \sum_{i=1}^p a_i Y_{n-1} + e_n \\ \text{var}(e_n) = \sigma^2 \end{cases} \quad (1)$$

where  $e_n$  is a zero-mean noise with variance  $\sigma^2$ .

Assume the audio signal is windowed as in Figure 1.

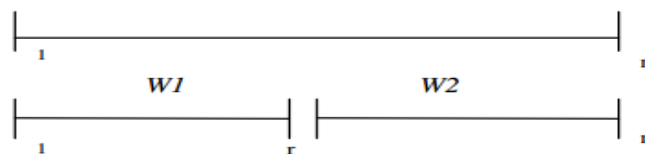


Figure 1. Location of three windows in Brandt's GLR

where  $W1$  is described the signal  $(Y_1, \dots, Y_r)$  and  $W2$  describes the signal  $(Y_{r+1}, \dots, Y_n)$ . A jump is detected at  $r$  if  $D_N(r) \geq D_0$  if

$$D_n(r) = -(n - r) \text{Log} \sigma_2^2 - r \text{Log} \sigma_1^2 + N \text{Log} \sigma_0^2 \tag{2}$$

and  $D_0$  is the predefined threshold.

This research will focus on the use of Brandt’s GLR algorithm on segmenting Malay words. The Malay language is one of many Austronesian languages which also includes languages such as Pilipino and Tagalog [15]. It is officially used in Malaysia and Singapore. Some aspects of the language are influenced by English and some words are directly borrowed from English itself. ASR system for Malay is challenging due to various regional dialects used and the occasional use of English words in sentences. Structures for the Malay syllables exists in the form of V, CV, CVC, CCV, CCVC, and CCCVC [16]. C’s are consonant and V’s are vowels. Syllable borders have two significantly different energy clusters which are visually noticeable however when pronounced continuously and closely together, false abrupt changes might occur depending on the speaker’s utterance style [17]. For read mode, the patterns are quite obvious and easily detected as only minimal amount of noise is present but in spontaneous mode, the presence of background noise, talking pace, and interference from other speaker may cause difficulty during segmentation [18].

### 3. METHODOLOGY

The framework of this research is shown in Figure 2. Speech signal data in the form of Malay Poems or known as *Pantun* is used to test the Brandt’s GLR. The *Pantun* represents read mode audio data where audio recording is done in controlled environments and the speaker controls his/her manner of speech to ensure clear pronunciation and fixed reading pace [19, 20]. The Poems (Pantuns) are framed per sentence. No windowing method is used due to the Brandt’s GLR only working on detecting energy difference in time domain. Further processes are explained in detail in the following sections.

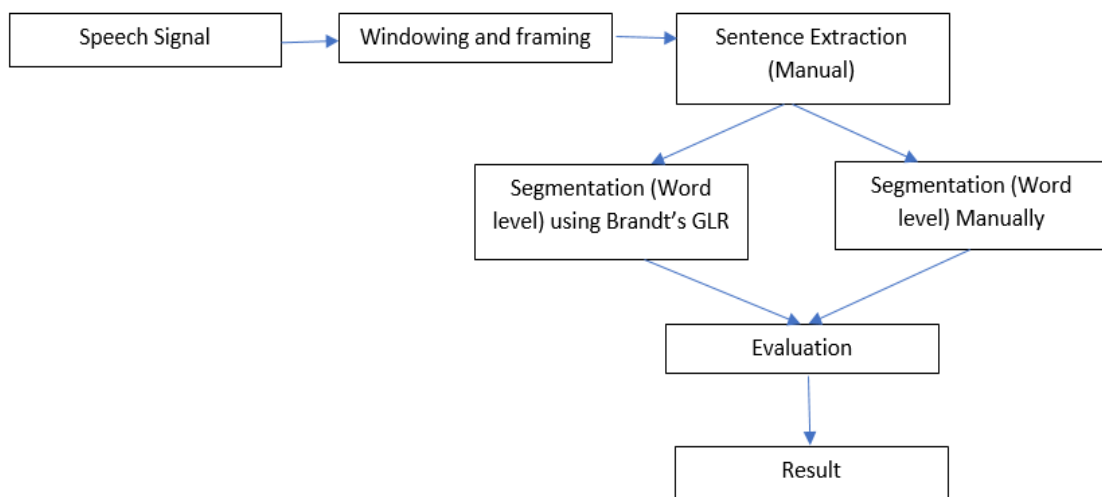


Figure 2. Research framework

#### 3.1. Data collection

As mentioned earlier, the data used during testing are in the form of traditional Malay Poems (Pantun) read by a male speaker in a noise free room. Instances of modern poems, as well as other types of poems, can be found in all sorts of printed and electronic documents including books, newspapers, magazines, and websites [21]. An instance of modern Malay poems can represent a complete poem or a poem portion. Five (5) poems (pantuns) of 10 sentences long each are first manually segmented. Segmentation both manually and automatically is done on frames consisting of one sentence long cut from the Pantuns. Manual segmentation is performed using the wavesurfer program as shown in Figure 3.

Segmentations are done on word by word by word basis visually by observing the waveform while listening to the audio as well. Referring to Figure 3, the energy difference between words are visually obvious since the data is in read mode and there are clear silences in between words.

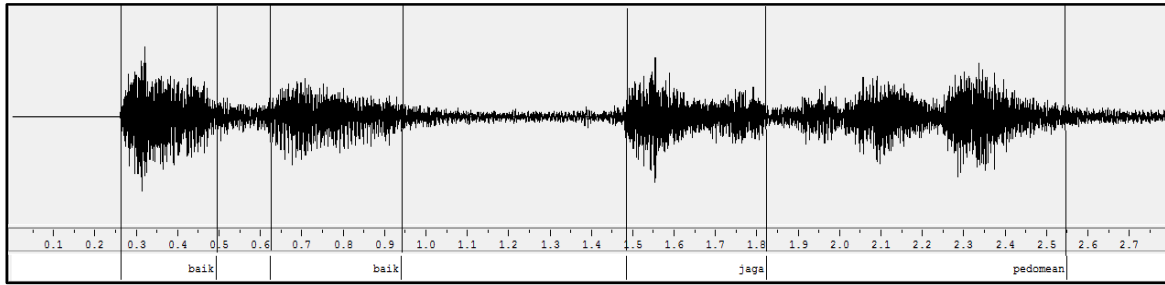


Figure 3. Manual segmentation using on the sentence “baik-baik jaga pedoman” using the wavesurfer program. Words segmented are “baik”, “baik”, “jaga”, “pedoman”

### 3.2. Brandt’s GLR algorithm

Noise filtering is not required since little to no noise is present in the data. The data is first framed into individual frames of one (1) sentence each [22-24]. Then the Brandt’s GLR ratio is calculated for each sample as follow, using the windows in Figure 4.

1. The covariance of  $W_1$  and  $W_2$  are calculated using brute force calculations.  $W_1$  will start at  $r = 1$  and will grow until  $n$  and the variance values of values for both  $W_1$  and  $W_2$  are calculated.
2. The variation of  $D_n$  is then calculated using equation (1) and graph  $D_n$  is plot.

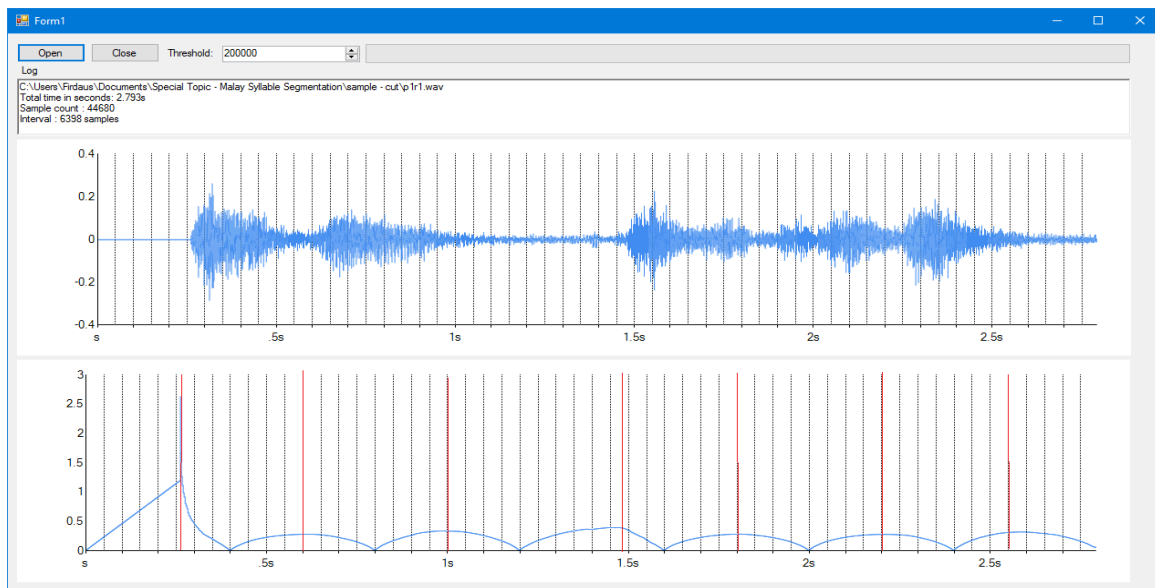


Figure 4. Top shows the signal waveform for sentence “baik-baik jaga pedoman” Bottom shows variation of  $D_N$ . The red lines show the points of segmentation

The segmentation points are then acquired by observing the highest ratio calculated by the Brandt’s GLR. Resulting segmentation points from the Brandt’s GLR method are compared to the reference segmented points from the manual segmentation that was done using the wavesurfer program by us. The measurement criterion is adapted from [25]. Let  $K = \{K_1, K_2, \dots, A_n\}$  and  $R = \{R_1, R_2, \dots, R_n\}$  be the segmented points in seconds obtained from the Brandt’s GLR and manual segmentation respectively. For each  $K_j$ , the corresponding point  $R_{kj}$  is determined by the time instance closes to that of  $K_j$ . Thus a sequence  $R_k = \{R_{k1}, R_{k2}, \dots, R_{kn}\}$  is build to compare both segmentations.

Omission can be detected as when point in  $R_k$  is not in  $K_j$  and insertion when points in  $K_j$  is not in  $R_k$ . Number of similar points in both  $R_k$  and  $K_j$  are calculated as Match,  $m = (m/p * 100)$  where  $p$  is the number of points in  $R$  [26, 27]. Accuracy is calculated using,  $accuracy = ((m/p+n) * 100)$  which is affected by the number of insertions. The Brandt’s GLR method will be evaluated in terms of the number of omission and insertions, matches, and accuracy [28].

### 3.2 Short-term energy algorithm

The energy parameter has been used in speech segmentation since the 1970's [29]. This algorithm was adopted and modified to better locate the beginning and ending of speech points for the isolated spoken Malay utterances and will be discussed in detail. This is a two-step search algorithm where the absolute energy (AE) for a coarse search is first used [19]. The speech signal was first divided into 50% overlapping frames of 10ms and then passed through a rectangular window [30, 31]. The AE was computed by summing the absolute magnitudes of speech samples in each frame as shown in (3).

$$E_s = \sum_{n=m-N+1}^m |s(n)w(m-n)| \quad (3)$$

where,  $w(m)$  rectangular window,  $N$  length frame duration ending at  $n=m$  and  $m$  speech samples overlapping at 10 ms.

The mean and standard deviation of the AE measure is first computed during the first 50ms of the speech, assuming there is only background noise in that interval [22]. This information was further used to compute the peak energy (IMX) for the entire interval in each speech sample and the silence energy (IMN) [23, 24]. Subsequently the IMX and IMN were used to set two energy thresholds: upper threshold ( $T_u$ ) and lower threshold ( $T_l$ ) according to (4).

$$T_l = IMN \left( 1 + 2_{\log_{10}} \left( \frac{IMX}{IMN} \right) \right) \quad (4)$$

The upper threshold ( $T_u$ ) will be computed as in (5) and (6).

$$W_L = \frac{\sum_i E_n(i)}{\sum_i 1} \quad (5)$$

$$T_u = T_l + 0.25(W_L - T_l) \quad (6)$$

where,  $W_L$  word length,  $I$  is index of all frames, having  $E(i) > T_l$

Therefore, upper level for average energy is set to 0.25 based on experimental findings in case of high noise [25].

## 4. RESULTS AND ANALYSIS

In the experiments, the Brandt's GLR is applied on periodic frames of 0.4 seconds. The algorithm was tested on frames of 0.2 seconds but was found to create high amount of insertion thus lowering the accuracy of the segmentation. At 0.2 seconds, Brandt's GLR produces twice the amount of segmentation points compared to the reference segmentation points as shown in Figure 5.

Each *pantun* is read in a controlled rhythm where each of the words in each sentence is read approximately 0.5 seconds apart from each other. Therefore 0.4 second framing is relatively effective for this type of segmentation. Pantun five (5) shows the worst accuracy as a lot of the words are made of prefixes such as *dihati* and *membuang*. Insertion occurs in between the prefix and the word hence lowering the accuracy of it. All of the other *pantun*s manage to be segmented with 80% accuracy with a 0.2 second tolerance.

Nine out of 20 of the data managed to be 100% segmented and overall result is presented in Table 1. And the 5<sup>th</sup> sentence from pantun two (2) achieved 100% segmentation within 0.1 second time tolerance. In that sentence, "*suka hati kumbang yang terbang*", none of the words contains prefixes and suffixes, and contains no more than two syllable per words. Suffixes and prefixes can sometimes be captured as new words. For example, in the 3<sup>rd</sup> sentence of pantun five (5), the prefix "*membu*" in "*membuang*" was captured as a separate word. 0.4 second frames are chosen as it manages to segment words that are two syllables long without over-segmenting. This however will cause over-segmentation in words that are three syllables or more which is commonly due to the presence of prefixes or suffixes.

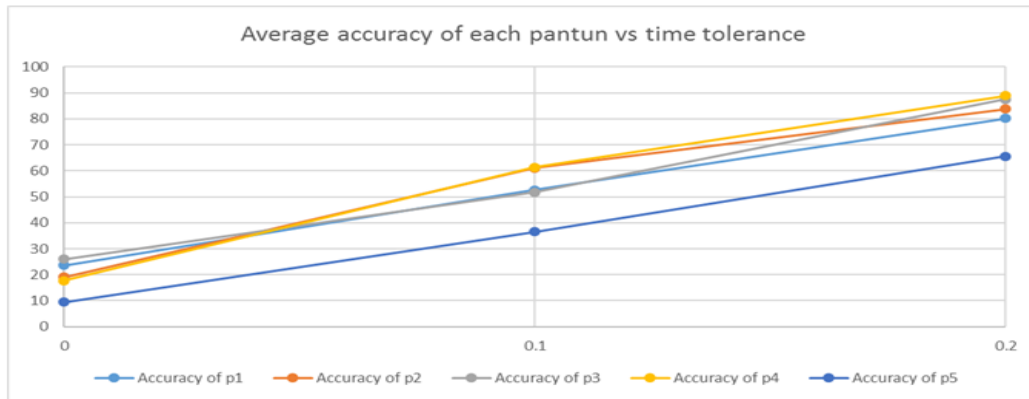


Figure 5. Average accuracy of each *pantun* vs time tolerance

Table 1. Overall segmentation results for five (5) *pantun*

Time tolerance (seconds)	0 second				0.1 second				0.2 second			
	p (auto)	m (match)	n (miss)	accuracy (%)	p (auto)	m (match)	n (miss)	accuracy (%)	p (auto)	m (match)	n (miss)	accuracy (%)
Poem 1 Sentence 1	7	5	2	55.56	7	6	1	75.00	7	7	0	100.00
Poem 1 Sentence 2	8	2	6	14.29	8	4	4	33.33	8	5	3	45.45
Poem 1 Sentence 3	7	1	6	7.69	7	3	4	27.27	7	7	0	100.00
Poem 1 Sentence 4	7	2	5	16.67	7	6	1	75.00	7	6	1	75.00
Poem 2 Sentence 1	6	2	4	20.00	6	5	1	71.43	6	6	0	100.00
Poem 2 Sentence 2	8	3	5	23.08	8	5	3	45.45	8	6	2	60.00
Poem 2 Sentence 3	7	2	5	16.67	7	3	4	27.27	7	6	1	75.00
Poem 2 Sentence 4	7	2	5	16.67	7	7	0	100.00	7	7	0	100.00
Poem 3 Sentence 1	7	2	5	16.67	7	4	3	40.00	7	6	1	75.00
Poem 3 Sentence 2	7	4	3	40.00	7	5	2	55.56	7	7	0	100.00
Poem 3 Sentence 3	6	2	4	20.00	6	5	1	71.43	6	6	0	100.00
Poem 3 Sentence 4	7	3	4	27.27	7	4	3	40.00	7	6	1	75.00
Poem 4 Sentence 1	7	3	4	27.27	7	6	1	75.00	7	7	0	100.00
Poem 4 Sentence 2	7	2	5	16.67	7	5	2	55.56	7	7	0	100.00
Poem 4 Sentence 3	7	0	7	0.00	7	4	3	40.00	7	5	2	55.56
Poem 4 Sentence 4	7	3	4	27.27	7	6	1	75.00	7	7	0	100.00
Poem 5 Sentence 1	7	1	6	7.69	7	5	2	55.56	7	5	2	55.56
Poem 5 Sentence 2	9	2	7	12.50	9	3	6	20.00	9	7	2	63.64
Poem 5 Sentence 3	8	1	7	6.67	8	4	4	33.33	8	6	2	60.00
Poem 5 Sentence 4	5	1	4	11.11	5	3	3	37.50	5	5	1	83.33

Figure 6 shows the reference segmentation of the word “*membujang*” and Figure 7 shows how the algorithm did it. The word “*membujang*” was read 0.6 seconds long, which was captured by two separate GLR frames thus causing over segmentation.

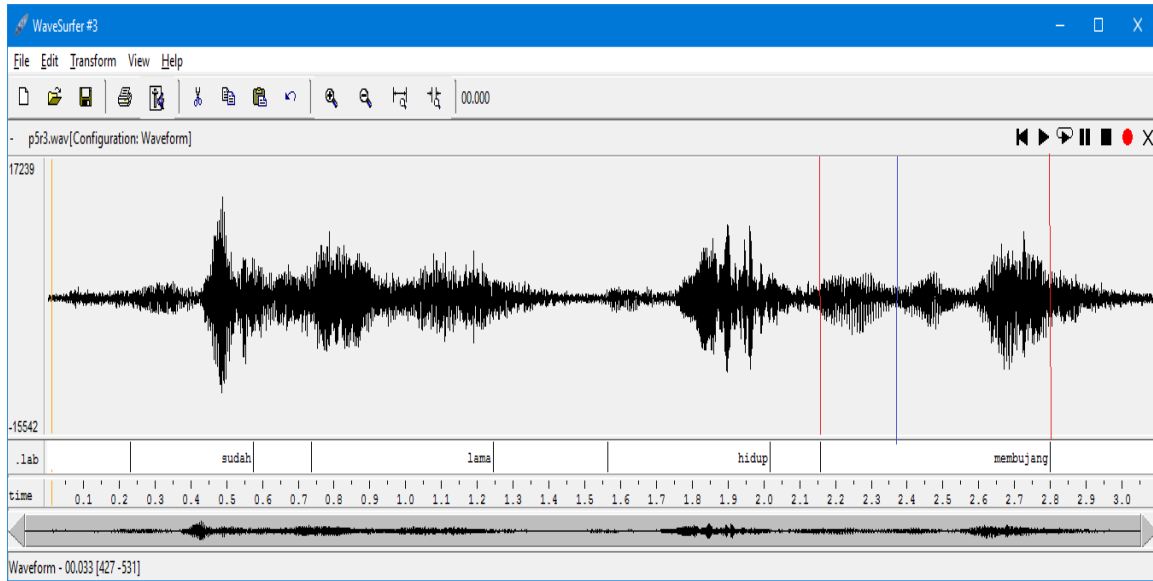


Figure 6. Waveform for the sentence “*sudah lama hidup membujang*”, the red line shows the start of the word “*membujang*” and the blue line marks the end of the prefix “*mem*”

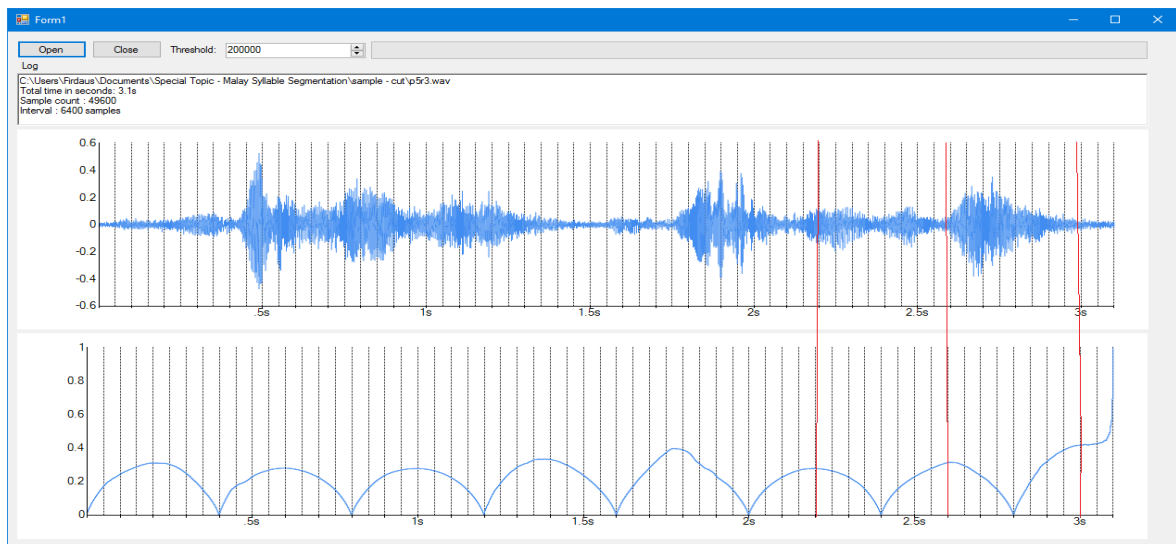


Figure 7. Automatic segmentation of the sentence “*sudah lama hidup membujang*”. Red line shows segmentation points

### 5. CONCLUSION

Four out of five of the pantun managed to be segmented with 80% accuracy. However, it is to be noted that all the data is in read mode and was recited in a controlled rhyme thus making the segmentation process a lot simpler than if to be done on spontaneous speech where there will be multiple speakers which all speak at different pace. Salam recommended in [7] to use higher order of autoregressive model to purposely cause over segmentation and to remove the insertions using Neural Network. This might also help with segmentation of spontaneous data. To test this would be our future goal.

### ACKNOWLEDGEMENTS

The authors would like to thank Universiti Teknologi MARA (UiTM) for the research funding and support via grant number 600-IRMI/PERDANA 5/3 BESTARI (107/2018).

## REFERENCES

- [1] Forsberg, M., "Why is speech recognition difficult?," *Chalmers University of Technology*, 2003.
- [2] M. Cutajar, J. Micallef, O. Casha, I. Grech, and E. Gatt, "Comparative study of automatic speech recognition techniques," *IET Signal Processing*, vol. 7, no. 1, pp. 25-46, 2013.
- [3] Chen, Scott and Ponani Gopalakrishnan, "Speaker, environment and channel change detection and clustering via the bayesian information criterion," *Proc. DARPA Broadcast News Transcription and Understanding Workshop*, vol. 8, 1998.
- [4] M. A. Al-Manie, M. I. Alkanhal, M. M. Al-Ghamdi, N. Mas-torakis, A. Croitoru, V. Balas, E. Son, and V. Mladenov, "Auto- matic speech segmentation using the arabic phonetic database," in *WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering*, no. 10, World Scientific and Engineering Academy and Society, 2009.
- [5] R. Makowski and R. Hossa, "Automatic speech signal segmentation based on the innovation adaptive filter," *International Journal of Applied Mathematics and Computer Science*, vol. 24, no. 2, pp. 259-270, 2014.
- [6] A. Cherif, L. Bouafif, and T. Dabbabi, "Pitch detection and for- mant analysis of Arabic speech processing," *Applied Acoustics*, vol. 62, no. 10, pp. 1129-1140, 2001.
- [7] M. Sharma and R. Mammone, "Subword-based text-dependent speaker verification system with user-selectable passwords," in *Acoustics, Speech, and Signal Processing, ICASSP-96. Conference Proceedings, 1996 IEEE International Conference on*, vol. 1. IEEE, 1996, pp. 93-96.
- [8] J. P. van Hemert, "Automatic segmentation of speech," *IEEE Transactions on Signal Processing*, vol. 39, no. 4, pp. 1008-1012, 1991.
- [9] Y.-S. Lee, K. Papineni, S. Roukos, O. Emam, and H. Hassan, "Language model based arabic word segmentation," in *Proceed- ings of the 41st Annual Meeting on Association for Computa-tional Linguistics*. Association for Computational Linguistics, vol. 1, pp. 399-406, 2003.
- [10] Rybach, David, et al., "Audio segmentation for speech recognition using segment features," *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009.
- [11] Jarifi, Safaa, Dominique Pastor, and Olivier Rosec. "Brandt's GLR method & refined HMM segmentation for TTS synthesis application," *Signal Processing Conference, 2005 13th European*. IEEE, 2005.
- [12] R. Andre-Obrecht, "Automatic segmentation of continuous speech signals," *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86*, pp. 2275-2278, 1986.
- [13] Noor Aina Zaidan and Md Sah Hj Salam, "Emotional speech feature selection using end-part segmented energy feature", *Indonesian Journal of Electricakl Engineering and Computer Science (IAES)*, 2015.
- [14] "Segmentation of Malay Syllables in connected digit speech using statistical approach," *International Journal of Computer Science and Security*, vol. 2, no. 1, p. 23, 2008.
- [15] Tan, Tien-Ping, et al., "MASS: A Malay language LVCSR corpus resource," *Speech Database and Assessments, 2009 Oriental COCODA International Conference*. IEEE, 2009.
- [16] Rasaan, O., "Segmentation and clustering methods for a new speech recognition architecture," *Master Thesis, Helsinki University of Technology*, 2007. [Online]. Available: [http://www.acoustics.hut.fi/publications/files/theses/rasaanen\\_mst.pdf](http://www.acoustics.hut.fi/publications/files/theses/rasaanen_mst.pdf).
- [17] Mermelstein, P., "Automatic segmentation of speech into syllabic units," *J. Acoust. Soc. Am.*, vol. 58, pp. 880-883., 1975. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/1194547>
- [18] Zhang, T. and C. Kuo, "Hierarchical classification of audio data for archiving and retrieving," *Proceeding of Acoustics, Speech and Signal Processing on 1999 IEEE International Conference, Mar. 15-19, IEEE Xplore Press, USA*, pp. 3001-3004, 1999. [Online]. Available: DOI: 10.1109/ICASSP.1999.757472.
- [19] Abdulrahman Almuhareb, Waleed A. Almutairi, Haya Altuwaijri, Abdulelah Almuarak and Marwa Khan, "Recognition of Modern Arabic Poems", *Indonesian Journal of Electricakl Engineering and Computer Science (IJECS)*, pp. 454-464, 2015.
- [20] Young, S., "A review of large vocabulary continuous speech," *IEEE Signal Process. Mag.*, vol. 13, pp. 45-57, 1996. [Online]. Available: DOI: 10.1109/79.536824.
- [21] Jarifi, S., D. Pastor and O. Rosec, "Brandt's GLR method and refined HMM segmentation for TTS synthesis application," *Proceedings of the European Signal Processing Conference, Sept. 4-8, Antalya, Turkey*, pp. 1-4, 2005. [Online]. Available: <http://www.ee.bilkent.edu.tr/~signal/defevent/pape rs/cr1491.pdf>.
- [22] Michele, B. and V. N. Igor. *Detection of Abrupt Changes: Theory and Application*. 1st Edn. Prentice Hall, Englewood Cliffs, New Jersey, ISBN: 10: 0131267809, pp: 528, 1993.
- [23] Rabiner, L. R. and M. R. Sambur, "Some preliminary experiments in the recognition of connected digits," *IEEE Trans. Acoust., Speech Signal Process*, vol. 24, pp. 170-182, 1976. [Online]. Available: [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=1162794](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1162794)
- [24] Obrecht, R. A., "Automatic segmentation of continuous speech signal," *IEEE Trans. Acoust., Speech Signal Process*, vol. 36, pp. 29-40, 1988. [Online]. Available: DOI: 10.1109/29.1486.
- [25] Prasad, V., Kamakshi and H. A. Murthy, "An automatic algorithm for segmenting and labeling a connected digit sequence," *Proceeding of the International Conference on Spoken Language Processing*, pp. 87-90, 2000. [Online]. Available: [http://www.iscaspeech.org/archive/icslp\\_2000/i00\\_3087.html](http://www.iscaspeech.org/archive/icslp_2000/i00_3087.html).
- [26] Runqiang, Y., Z. Yiqing and Z. Yisheng, "Automatic speech segmentation combining HMMbased approach and recurrence trend analysis," *Proceeding of the International Conference on Acoustic, Speech and Signal Processing, May 1419, IEEE Xplore Press, USA.*, pp: 797-800, 2006. [Online]. Available: DOI: 10.1109/ICASSP.2006.1660141.



- [27] David, L., "Learning spectral-temporal dependencies using connectionist networks," *Proceeding of the International Conference on Acoustic, Speech and Signal Processing, Apr. 1114, IEEE Xplore Press, USA*, pp. 418-421, 1998. [Online]. Available: DOI: 10.1109/ICASSP.1988.196607.
- [28] Abdulla, W. H., V. Kecmanand and N. Kasabov, Speech-background classification by using SVM technique. 2003. [Online]. Available: [http://www.aut.ac.nz/resources/research/research\\_institutes/kedri/downloads/pdf/abdkeckasiconip2003.pdf](http://www.aut.ac.nz/resources/research/research_institutes/kedri/downloads/pdf/abdkeckasiconip2003.pdf)
- [29] Salam, M. S., D. Mohamaed and S. H. Salleh, "Segmentation of Malay syllables in connected digit speech using statistical approach," *Int. J. Comput. Sci. Sec.*, vol. 2, pp. 23-33, 2008. [Online]. Available: <http://www.cscjournals.org/Journals/IJCSS/Volume2/Issue1/IJCSS-26.pdf>
- [30] Salam, M. S., D. Mohamad and S. Hussain, "Neural network speaker dependent isolated Malay speech recognition system: Handcrafted Vs genetic algorithm," *Proceeding of the 6th International Symposium on Signal Processing and its Applications, Aug. 13-16, IEEE Xplore Press, USA*, pp: 731-734, 2001. [Online]. Available: DOI: 10.1109/ISSPA.2001.950252
- [31] Master, T., *Practical Neural Network Recipes in C++*. 1st Edn., Academic Press, New York, ISBN: 10: 0124790402, p. 493, 1993.

## BIOGRAPHIES OF AUTHORS



**Noraini Seman** received the Bachelor degree in Computer Science from Universiti Putra Malaysia (UPM), Malaysia in 1999; the MSc degree from Queensland University of Technology (QUT), Australia (2002), and the Ph.D. degree from the Universiti Teknologi MARA (UiTM), Malaysia (2012). She is presently Head of Academic Program Accreditation under Curriculum Affairs Unit and senior lecturer at Department of Computer Science, Faculty of Computer and Mathematical Sciences. Her research interests include AI application to digital signal processing problems, speech summarization, machine translation and machine learning techniques in speech recognition technology.



**Ahmad Firdaus Norazam** received the M.E degree in Electrical Engineering from Queens University, Kingston, ON, Canada, in 2016. Currently, he is a Computer Engineer and previously, he worked in various industrial positions in research areas such as LMDS wireless communications and image processing with a focus on data fusion signal applications. Some of his research interests include adaptive signal estimation and nonlinear signal processing.