Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 76 (2015) 417 – 424

Procedia
Computer Science

2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS 2015)

An Improved Syllabification for a Better Malay Language Text-to-Speech Synthesis (TTS)

Izzad Ramli[a]*, Nursuriati Jamil[a], Noraini Seman[a], Norizah Ardi[b]

[a]Digital Image, Audio and Speech Technology Group, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Malaysia
[b]Academy of Language Studies, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

Abstract

Text-to-speech (TTS) is an important component of robots, humanoids and Internet of Things as a mean of human computer interaction. One of the main components of TTS is text processing that functions as the producer of syllabic speech units to be used in the generation of human-like speech. The naturalness of TTS largely depends on text processing component, particularly word syllabification. Syllabification is a process of segmenting the given text input into sequence syllabic speech units. This paper begins with investigations of previous syllabification technique of Malay language to identify the limitations. An improved syllabification technique is then proposed and compared against the performance of another three known syllabifications. The datasets used comprises 25,000 words collected from Malay language online national newspaper articles and Wikitionary Open Content Dictionary. Word Error Rate (WER) percentage is calculated and our proposed syllabification technique achieved the lowest WER of 2.61% with an accuracy rate of 97.39%.

Peer-review under responsibility of organizing committee of the 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS 2015)

Keywords: Syllabification, Malay language, speech synthesis, TTS

* Izzad Ramli. Tel.: +04-7541296.
*E-mail address:*zadzed89@yahoo.com

Peer-review under responsibility of organizing committee of the 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS 2015)
doi:10.1016/j.procs.2015.12.280

1. Introduction

Speech synthesis is the process of converting written text to spoken audio form, commonly referred to as text-to-speech (TTS). It is widely used as assistive technology for people with wide range of disabilities especially those with speech impairments[1,2]. In recent scenario, TTS is also embedded in service robots and humanoids to interact with intelligent agents[3] and providing short commands[4,5]. In general, TTS consists of two main components, which are text processing and speech generation[1]. The main aim of text processing component is to segment the given text input into sequence of phonemic or syllabic speech units. These speech units are then produced into speech sounds using speech generation component either by synthesis from parameters or by selection of a unit from a large speech corpus[6]. Text processing is the focus of this paper as producing correct sequence of speech units is essential for a natural speech synthesized output[7]. In many languages, speech units may be syllabic or phonemic. However, Malay language is an alphabetic language with salient syllabic structures[8]. Hence, segmentation of text into syllabic speech units is preferred[9, 10]. This process is known as syllabification and this term is used throughout this paper. Inaccurate syllabification remains a challenge in most text processing component of a TTS[11]. Incorrect segmentation of syllable leads to inappropriate pronunciation of speech, thus generating an unnatural synthesized speech. Therefore, this paper reviews previous work of speech synthesis in Malay language to identify the gap in syllabification and proposed an improvement to reduce the word error rate (WER). Section 2 compares related work of syllabification in Malay language. Data sets used in this paper is described in Section 3 followed by explanation of the improved syllable segmentation in Section 4. Evaluation results and discussion are then presented in Section 5 and concluded in Section 6.

2. Related work

Native Malay words are agglutinative alphabetic-syllabic that are based on four distinct syllable structures, i.e. V, VC, CV and CVC[12]. These basic structures can be combined in multiple ways to form a word. For example, a disyllabic word such as '*kita*' (*we*) is represented using CV+CV and a tetrasyllabic word such as '*mesyuarat*' (*meeting*) is composed of CV+CV+V+CVC. Malay words are mostly formed by two or more distinct syllables[13] and a few are mono-syllabic words. Examples of word with various syllabic structures are presented in Table 1.

Table 1. Syllable structures of Malay words.

Type of Syllable	Syllable structure	Example word
Monosyllabic	CVC	gam, cik, sup
	CV	di, ke, si
Disyllabic	CV+CV	beli, cuba, biru
	V+ CV	ini, itu, api
	V+ CVC	amat, ikan, ubat
Trisyllabic	CV+ CV+CV	lelaki, berasa, kerapu
	CV+V+CV	suara, buaya, kuasa
	CV+ CV+V	semua, ketua, deria
Tetrasyllabic	CV+CV+CV+CV	matahari, daripada, serigala
	CV+CVC+CV+VC	perempuan,
	V+ CV+ CV+ CV	apabila, adakala,

Syllabification is a language-dependent process[14] of dividing words into syllables. Each language can have its own set of syllable structure. Therefore, syllabification techniques may vary from one language to another. Previous work on Malay language syllabification shows usage of two techniques: 1) Syllabification based on database[11,15], and 2) Segmentation based on syllable structure[4]. Syllabification based on syllable database requires large storage capacity to store the entire syllable in a language. On the other hand, segmentation based on syllable structure used knowledge rules and phonetics devised by linguists. One of the earliest work of syllabification was done by Zeki & Azizah[15] by selecting the longest characters sequence in a word and comparing the sequence to a syllable database.

The sequence of characters comprised 4 characters beginning from the leftmost character (i.e. first character) of a word. In other words, the selection of characters is in left to right direction. If characters sequence matches with a syllable in the database, the sequence of 4 characters is segmented and considered as a syllable unit. Otherwise, the last character in the sequence is dropped leaving a 3-characters sequence as a syllable unit. The revised syllable unit is again compared against the syllable database for a match. The dropped-and-matched process is continued until a syllable unit match is found in the syllable database. The word is then segmented based on the matched syllable unit. This syllabification technique is simple as only characters are used instead of phonemes. However, incorrect syllabifications are rather rampant. Few incorrect examples are shown in Table 2.

Table 2. Incorrect syllabification using Zeti & Azizah (2001) technique.

Words	Segmented incorrectly	Correct segmentation
[cabaran]	[ca]+[bar]+[an]	[ca]+[ba]+[ran]
[sekarang]	[sek]+[a]+[rang]	[se]+[ka]+[rang]
[perak]	[per]+[ak]	[pe]+[rak]
[mama]	[mam]+[a]	[ma]+[ma]

In 2004, Samsudin et al.[9] adopted a different strategy for Malay words syllabification to develop Malay speech synthesis. In their syllabification technique, each normalized word is converted to consonant (C) and vowel (V) forms and extracted based on 4 syllable structure rules in Malay sound system. The structures are CV, VC, CVC, and V. Syllabification is also done from left to right. Their proposed technique is able to segment words which end with Malay diphthong such as <ai>, <au> and <oi>. Such words are [pant***ai***] and [pul***au***]. Samsudin et al.'s syllabification is also capable of segmenting an unspecified number of English and Arabic loan words which are stored in a database.

Tan & Ranaivo-Malancon[11] introduced another syllabification based on syllable database for an Automatic Speech Recognition (ASR) in 2009. Their method is later adopted by Tiun et al.[10] for syllabification intended for a Malay language speech synthesis. The first steps of syllabification converted grapheme of a word into different sound class such as vowel, diphthong, fricative, affricate, plosive, nasal and glides. Then, the word is segmented by determining the largest possible syllable that can be formed from right to left. A summary of the above mentioned three syllabification techniques are illustrated in Fig. 1.

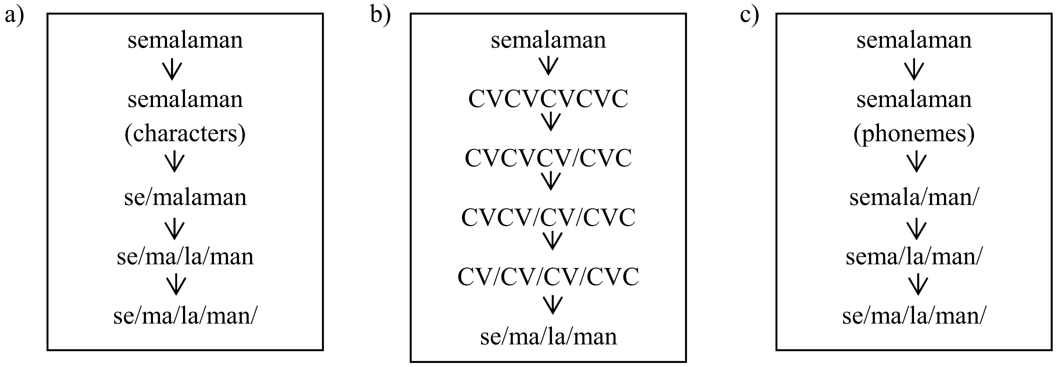

Fig. 1. (a) Zeki & Azizah, 2001[15] (Left to right syllabification); (b) Samsudin et al.[9], 2004 (Left to right syllabification); (c) Tan & Ranaivo-Malancon[11] , 2009 (Right to left syllabification).

3. Data collections

The word datasets used in this paper are randomly selected from online national newspaper articles (i.e. Harian Metro[16] and Berita Harian[17]) and Wiktionary Open Content Dictionary[18]. A total of 25,000 Malay words are collected and their syllable length distributions are graphically represented in Fig. 2 (a). It can be seen from the figure that the highest percentage of syllable length of Malay words is trisyllabic words. These words are further manually segmented to syllable units using Malay syllable structure and stored in a syllable database for syllabification usage. Only 1,444 units of syllables are acquired from these 25,000 words. In syllable database, the distribution of characters in a syllable is show in in Fig. 2 (b). As can be seen, the three characters per syllable scored the highest percentage of 68%. Furthermore, it is also discovered that none of the collected words consists of more than 5 characters per syllable. Due to this fact, we chose a sequence of 5 phonemes sequence in our proposed syllabification technique.

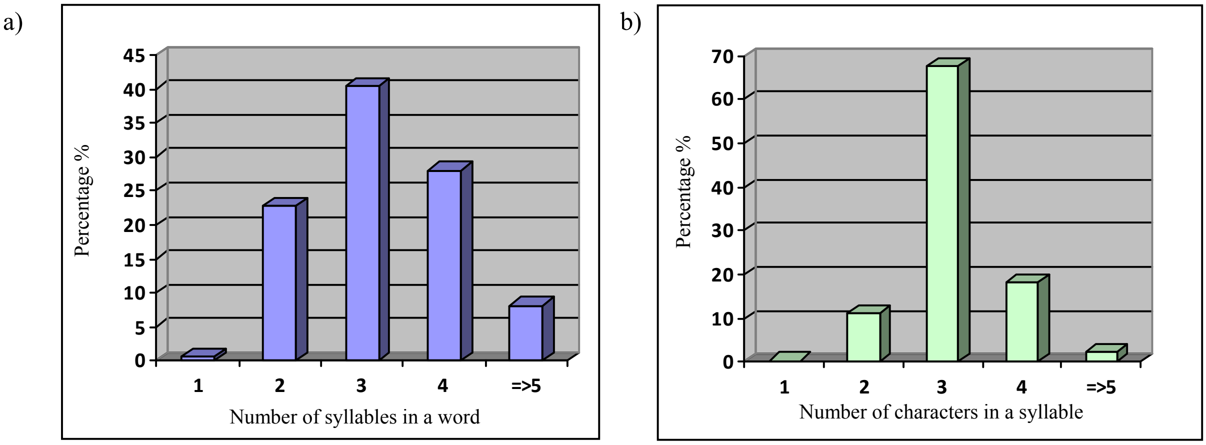

Fig. 2. (a) Tabulated number of number of syllables in a word; (b) Tabulated number of characters in a syllable.

4. Methodology

In this research, syllabification based on syllable database similar to Tan & Ranaivo-Malancon[11] is adopted and improved by adding a specific rule for *MustFront* syllables (refer Table 4). Fig. 3 demonstrates the syllabification technique proposed in this paper.

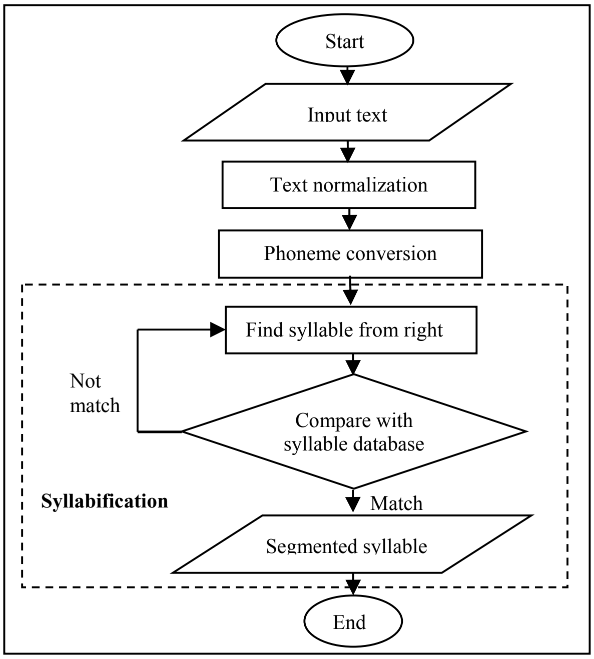

Fig. 3. Text processing flowchart.

4.1. Text normalization

In this paper, we only focus on Malay words comprising alphabetic characters. Therefore, all numbers are converted into alphabetic words in text normalization stage. As an example, number ‘2’ is converted into the word ‘two’. However, all punctuations and abbreviations are ignored and not included in the data set.

4.2. Phoneme conversion

In this stage, all characters in the normalized text are converted to phoneme based on Malay language phonetics. Malay language phonetics consists of 34 phonemes, of which 24 phonemes are of Malay language origin and 10 phonemes are borrowed from English and Arabic language[19]. The 24 original Malay phonemes are further categorized into 6 vowel phonemes and 18 consonant phonemes. The vowel phonemes are /i,e,ə,a,u,o/ and the consonant phonemes are /p, b, t, d, k, g, c, j, m, n, ɲ, ɳ ,s, h, r, l, w, y/. As for the 10 borrowed consonants, 8 consonants are Arabic loan consonants and 2 are English loan consonants. The Arabic loan consonants are (/ z, Ɵ, ž, Ɣ, ð, x, Ç, ʃ/), while the English loan consonants are (/v, f/). At the end of phoneme conversion stage, all characters in the normalized text are mapped to their corresponding Malay language phonemes using Table 3.

Table 3. Characters to phoneme conversion mapping of Malay language.

Character	Phoneme	Character	Phoneme	Character	Phoneme
a	/a/	p	/p/	l	/l/
e	/ə/	r	/r/	m	/m/
i	/i/	s	/s/	n	/n/
u	/u/	t	/t/	q	/k/
o	/o/	w	/w/	v	/v/
b	/b/	y	/y/	kh	/x/
c	/c/	x	/s/	sy	/š/
d	/d/	f	/f/	gh	/ɣ/
g	/g/	z	/ð/	th	/ɵ/
h	/h/	z	/ž/	‘	/Ç/
j	/j/	ny	/ ɲ/		
k	/k/	ng	/ɲ/		

4.3.

4.3. Syllabification

Syllabification of a word starts with selecting 5 phoneme sequences beginning from the rightmost phoneme of the word. The selected phoneme sequence is then compared to the syllable database for a match. If a syllable in the database is matched, the phoneme sequence is identified as a syllable. Otherwise, the rightmost phoneme is dropped from the phoneme sequence leaving a 4-phoneme sequence, and a match is then searched again from the syllable database. The dropped-and-matched process is continued until a syllable unit match is found in the syllable database. The process continues in right to left direction for all phonemes in the word and the word are then segmented based on the matched syllable unit. This syllabification process is adapted from Tan & Ranaivo-Malancon[11]. However, we discovered an anomaly for 8 syllables that are located within a word or at the end of a word. In this paper, these syllables are referred to as *MustFront* syllables (see Table 4) signifying that these designated syllables must be positioned at the beginning of a word. For example, the word /kemaskan/ is wrongly segmented as /ke/ma/skan/ instead of /ke/mas/kan/ because the 4-phoneme sequence /skan/ is identified as syllable in the syllable database. An example of the word that used /skan/ as a syllable is the word /skandal/, and the syllable is located at the beginning of the word. In the case of the word /kemaskan/, the syllable /skan/ is located at the end of the word, hence the 4-phoneme sequence should not be identified as a syllable. In our proposed syllabification technique, we add another rule as shown in Fig. 4.

Table 4. Incorrect syllabifications due to *MustFront* syllables.

MustFront Syllables	False segmentation	Correct segmentation
skan	ke/ma/***skan***/	ke/mas/kan/
kri	ma/**kri**/fat/	mak/ri/fat/
ste	mi/**ste/**ri/	mis/te/ri/
sta	ku/**sta**/	kus/ta/
stik	mi/**stik**/	mis/tik/
klu	ma/**klu**/mat/	mak/lu/mat/
pli	a/**pli**/ka/si/	ap/li/ka/si/
spek	a/**spek**/	as/pek/

```
If a syllable is found in the syllable database
  Check whether syllable == MustFront syllable
  If syllable == MustFront syllables {
    Check its position in the word.
      If position == beginOfWord
           syllable = true; stop
      else {
           newSyllable = truncate leftmost phoneme in syllable
           find a match in syllable database }
    else
    syllable = true; stop
```

Fig. 4. A new rule for *MustFront* syllables.

With the addition of the new rule, if the *MustFront* syllables are located at the end or within a word, the leftmost phoneme is dropped from the phoneme sequence and a search for syllable match from the syllable database is reiterated. As an example, in the case of word /kemaskan/, the leftmost phoneme /s/ is truncated from the syllable /skan/ creating a new syllable /kan/. A match search is reiterated from the syllable database for the syllable /kan/.

5. Result and discussion

In this paper, four syllabifications experiments are conducted using the 25,000 words dataset described in Section 3. The four syllabication techniques are: 1) Zeki & Azizah, 2001[15]; 2) Samsudin et al., 2004[9]; 3) Tan & Ranaivo-Malancon[11], 2009 and 4) Proposed syllabification. The syllabification techniques are evaluated using Word Error Rate (WER) as in Eq. 1 and syllabification results are tabulated in Table 5.

$$WER = \frac{Total\ incorrect\ segmented\ words}{Total\ number\ of\ words} \qquad (1)$$

Table 5. Syllabifications results.

Syllabification technique	Word Error Rate (WER)
Zeki & Azizah, 2001[15]	77.44%
Samsudin et al.[9] , 2004	36.52%
Tan & Ranaivo-Malancon[11], 2009	3.75 %
Proposed syllabification	2.61%

Table 5 shows that the highest WER is produced by Zeki & Azizah[15] and the lowest WER is achieved by our proposed syllabification. Based on the detailed result analysis, Zeki & Azizah[15]'s syllabification structure is unable to cater for complex words such as tetrasyllabic words which comprises 34% of our datasets. On the other hand, Samsudin et al.[9]'s syllabification managed a reduced WER of 36.52% due to the introduction of rules for Malay diphtongs and loan words. A much better result is produced by Tan & Ranaivo-Malancon[11]'s work which achieved only 3.75%. However, with the addition of new rule for *MustFront* syllables in our proposed syllabification technique, WER is further reduced from 3.75% to 2.61%. After further analysis of our proposed technique, the small WER percentage is due to the limitation of syllables in the syllable database. This fact will be one of our concerns for future research.

6. Conclusion

This paper presents the syllabification of Malay word for the purpose of text to speech synthesis. We discovered eight syllables, which we named *MustFront* syllables, that are causing syllabification errors. With the addition of the new rule, our proposed technique managed a WER reduction of 1.10%. However, further improvements are still necessary as several words are still incorrectly segmented. Our next direction is to increase the number of syllables in the syllable database and looking into possibility of integrating root dictionary into the syllabification technique.

Acknowledgements

The research is funded by Research Entity Initiative Fund, Universiti Teknologi MARA, Shah Alam, Malaysia, no: 600-RMI/DANA 5/3/REI (3/2014)

References

1. Tallal P, Miller SL, Bedi G, et al. Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science (New York, NY)*. 1996;271(5245).p.81-84.
2. Terband H, Maassen B, Guenther FH, Brumberg J. Auditory–Motor Interactions in Pediatric Motor Speech Disorders: Neurocomputational Modeling of Disordered Development. *Journal of Communication Disorders*; 2014;(47):17-33.
3. Hönemann A, Wagner P. Adaptive Speech Synthesis in a Cognitive Robotic Service Apartment : An Overview and First Steps Towards Voice Selection. In: *Elektronische Sprachsignalverarbeitung*; 2015.
4. Sugiura K, Shiga Y, Kawai H, Misu T, Hori C. Non-Monologue HMM-Based Speech Synthesis for Service Robots: A Cloud Robotics Approach. *Robotics and Automation (ICRA), IEEE International Conference IEEE*; 2014.p.2237-2242.
5. Ferro F, Marchionni L. REEM: A Humanoid Service Robot. In: *In ROBOT2013: First Iberian Robotics Conference*. International Publishing; 2014.p.521-525.
6. Govind D, Prasanna, Mahadeva. Expressive speech synthesis: a review. *International Journal of Speech Technology*; 2012;16(2):237-260.

7. Raj AA, Sarkar T, Pammi SC. Text Processing for Text-to-Speech Systems in Indian Languages. In: *SSW*; 2007.p.188-193.
8. Lee LW, Low HM, Mohamed AR. A Comparative Analysis of Word Structures in Malay and English Children ' s Stories; 2013;21(1).p.67-84.
9. Samsudin N, Tiun S, Kong TE. A Simple Malay Speech Synthesizer Using Syllable Concatenation Approach. In: *MMU International Symposium on Information and Communications Technologies 2004 (M2USIC 2004)*; 2004.p.4-7.
10. Tiun S, Abdullah R, Tang E. Subword Unit Concatenation for Malay Speech Synthesis; 2011;8(5).p.68-74.
11. Tan T, Ranaivo-Malançon B. Malay grapheme to phoneme tool for automatic speech recognition. *Third International Workshop on Malay*; 2009.p.1-6.
12. Hamdan H. Dasar pendeskripsian sistem fonologi bahasa melayu [Basic description of Malay phonology]. In: *Bunga rampai fonologi bahasa melayu*. Petaling Jaya: Fajar Bakti Sdn. Bhd; 1988.
13. Karim NS, Onn FM, Musa HH, Mahmood, Hamid A. *Tatabahasa Dewan*. Kuala Lumpur: Dewan Bahasa dan Pustaka; 2004.
14. El-Imam Y a. Phonetization of Arabic: Rules and algorithms. *Computer Speech and Language*; 2004;18(4).p.339-373.
15. Zeki AM., Azizah N. A Speech Synthesizer for Malay Language. In: *National Conference on Research and Development in Computer Science (REDECS)*; 2001.p.349-354.
16. Harian Metro. Available at: http://www.hmetro.com.my/.
17. Berita Harian. Available at: http://www.bharian.com.my.
18. Wikitionary Open Content Dictionary. Available at: http://en.wiktionary.org/wiki/Category:Malay_nouns.
19. Zahid I, Omar MS. *Fonetik dan Fonologi*. 2nd ed. PTS akademia; 2012.