Pronunciation Variations and Context-dependent Model to Improve ASR Performance for Dyslexic Children's Read Speech

Husniza Husni, Zulikha Jamaludin

Graduate Department of Information Technology
College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok, Kedah
{husniza;zulie}@uum.edu.my

ABSTRACT

Focusing on the key element for an ASR-based application for dyslexic children reading isolated words in Bahasa Melayu, this paper can be an evidence of the need to have a carefully designed acoustic model for a satisfying recognition accuracy of 79.17% on test dataset. Pronunciation variations and context-dependent model are two main components of such acoustic model. This model adopts the most frequent errors in reading selected vocabulary, which are obtained from primary data collection and analysis. The analysis gives the most frequent spelling and reading errors as vowel substitution with over 20% of total errors made.

Keywords
Dyslexic children, automated speech recognition (ASR), reading, speech error pattern, Bahasa Melayu vocabulary.

1.0 INTRODUCTION

The objectives of this paper are 1) to recognize the patterns of spelling and reading errors in *Bahasa Melayu* vocabulary; and 2) to use the pattern in modeling a context-dependent pronunciation model for ASR. Currently, the demand for ASR technology to help children read has increased significantly (Steidl, Stemmer, Hacker, Noth, & Nieman, 2003). Such technology has been seen as an alternative way of teaching reading to children especially those who suffer from a neurological and developmental condition called dyslexia.

Dyslexia is a condition that impedes phonological awareness, which is strongly related to reading ability especially in the letter- sound correspondence area. Despite reading, dyslexia also causes problems in other skills such as writing, spelling, and motor skills as well as memory and cognition.

The following section shall be attributed to a brief introduction to dyslexia and phonological deficits theory. Next, this paper outlines methods to select and collect suitable vocabulary and illustrates the most frequent errors emerged from the analysis performed on the gathered data. Later, an ASR engine is trained and tested on datasets of dyslexic children's read speech of isolated words and how context-dependent and pronunciation variations could increase ASR performance significantly. The final section concludes the paper.

2.0 RELATED STUDIES ON DYSLEXIA, READING, AND ASR

Dyslexic children suffer from dyslexia, a condition that affects the ability to progressively learn to read, spell, and write due to deficits in phonological origin. A solid body of research has concluded that the phonological-based deficit is the major contributor towards this condition (Frost, 2001; Lundberg, 1995; Shaywitz, 1996; Snowling, 2000; Wolf, 1999; & Ziegler, 2006). The International Dyslexia Association (IDA) defines it as a neurological learning disability that affects the ability to accurately or fluently recognize words and have poor spelling and decoding abilities, normally causes problems in reading comprehension as well as reduced reading experience that holds back vocabulary and background knowledge expansion (International Dyslexia Association, 2006)

Projects such as the Colorado Literacy Tutor, CoLiT (http://www.colit.org/) with its component, the CSLR Reading Tutor Project are aiming at providing computer-aided reading instruction for children to enhance reading with collaborations with public schools (http://cslr.colorado.edu/beginweb/reading/ reading.html). Another example of such project to improve reading amongst children is LISTEN's Reading Tutor (Banerjee, Beck, & Mostow, 2003).

These major projects use ASR as the key technology. ASR is used to track reading while the children are reading aloud and allow for interaction between the user and the application via speech (e.g. asking questions). Pronunciation accuracy is also provided for feedback. ASR technology has the potential to enhance reading ability for normal children and it is also a good tool for helping those with dyslexia in reading as reported by previous studies (Hagen, Pellom, Vuuren, & Cole, 2004; Nix, Fairweather, & Adams, 1998; Raskind, & Higgins, 1999; Williams, Nix, & Fairweather, 2000).

ASR is found to offer such effect to dyslexic children as it can remediate the problems that concerns with phonological awareness through multi-sensory experience (Raskind, &

Higgins, 1999; Williams, Nix, & Fairweather, 2000; Higgins, & Raskind, 2000). The multi-sensory experience is created as the child read aloud a word and that particular word be displayed on the computer screen. This involves senses at least in terms of articulation and speech production, hearing, and visual.

3.0 DATA COLLECTION AND ANALYSIS METHODS

The intention of this study is using ASR for training/teaching dyslexic children to read in *Bahasa Melayu* due to the importance of this language in the Malaysian education scenario. Therefore, the language corpus has to be introduced and incorporated into ASR. Currently, there is ASR-based research in *Bahasa Melayu* but none were designed for training and teaching dyslexic children to read and instead focusing more on digit recognition, such as evidenced in Md Sah, Dzulkifli, and Sheikh Hussain (2001). The vocabulary needs to be chosen carefully to serve as stimuli for data collection.

The vocabulary chosen for ASR to train with is based on Malaysian public school syllabus, focusing on level one (standard one, two, and three) common words. The vocabulary consists of 114 words which have been carefully selected and used as stimuli. The words contain all syllable patterns (consonant-vocal pair) that make up valid words in *Bahasa Melayu*. Random cluster sampling technique is used for word selection where each syllable pattern is regarded as a cluster. Common words that appear in level one text book and *Buku Panduan Pelaksanaan Program Pemulihan Khas (Masalah Penguasaan 3M)* are therefore listed in the clusters accordingly. The words in the list are then randomly selected to represent their corresponding clusters and thus serve as stimuli. A total of ten dyslexic children, as young as 7 years old to 14 years old whose reading level are similar, are participated in the study. The participants are required to read aloud into a head-mounted microphone each of the 114 words prompted separately. While the participants are reading aloud the word, recording is done simultaneously to obtain a speech file (.wav).

Once all ten participants completed their reading and recording sessions, the data collected are tabled which include all reading mistakes made during data collection. The errors are then grouped into suitable categories. Phonological-based spelling error categories of Sawyer, Wade, and Kim (1999) are used to guide the groupings of the errors made.

The analysis performed on the data found that the most frequent spelling and reading error pattern made is *vowel substitution* with 20% of occurrences of all errors. This finding supports the study done by Sawyer, Wade, and Kim (1999) on phonological-based error patterns in English, which gave vowel substitution as the most frequent error made. Table 1 illustrates the findings.

Table 1: *Error patterns by category and their frequency of occurrences in dyslexic children's reading and spelling.*

Category of Errors	*n*	%
Substitutes vowel	1286	20.34
Omitted consonants *	786	12.43
Nasals (*m, n*)	770	12.17
Substitutes consonants *	577	9.13
Omits vowel	511	8.03
Substitutes word	384	6.07
Adds consonants	363	5.74
Substitutes with non-words	272	4.3
Reversals	268	4.24
Incorrect sequence	224	3.54
Omits syllable	167	2.64
Liquids (*l, r*)	156	2.47
Substitutes vowel with consonant / consonant with vowel **	143	2.26
Substitutes nasals for liquid	124	1.96
Adds vowel	124	1.96
Syllable Division Confusion	94	1.49
Adds syllable	74	1.17

* excludes *m, n, l ,r*
** *if: substitution of a vowel with a consonant (**excluding** m, n, l, r) or substitution of a consonant (**including** m, n, l, r) with a vowel*

4.0 ACOUSTIC MODELING

Only the words with the highest percentile of error categories are considered to be modeled and further trained. The words considered are those that fall under the 'substitute vowel', 'omit consonant', 'nasals', and 'substitute consonant' categories. The categories are considered based on their percentile as shown in Table 1. The categories are considered not only because of their high contribution to reading errors but also because they represent general categories for which every dyslexic child could attempt to. This allows an ASR engine be tuned accordingly to support more dyslexic children.

The context-dependent pronunciation modeling is done manually. The pronunciation model is thus constructed using manual, hand-coded transcription of the selected words citations into their correspondence Worldbet phones. Worldbet is the ASCII phonetic symbols that include phonetic alphabet of the world's languages in a systematic way (Hieronymus, 1993). For example, the transcriptions in Table 2 are for the word *abang* (older brother), *ibu* (mother), and *bapa* (father) respectively in Worldbet.

Table 2: *Examples for the transcriptions of four words namely abang, ibu, bapa, and nyata.*

Word	Worldbet
abang	`A bc b A N`
ibu	`i: bc b U` or `I bc b U`
bapa	`bc b A pc ph A`
nyata	`n~ A tc th A`

Each of the read words in the selected category together with the actual words (the stimuli) are transcribed according to the words' correct pronunciations (i.e. how they sounds phonetically) and represent them in Worldbet. The errors are also included in the lexical model. This conforms to suggestions in Nix, Fairweather, and Adams (1998) and Williams, Nix, and Fairweather (2000) that the errors produced are also regarded as and included in the active lexicon to increase recognition accuracy.

4.1 Context-dependent Lexical Model

Context-dependent model is specified for modeling the lexicon as it can significantly improve the ASR performance. The model is based on BM's phonetics and phonology system (Indirawati, & Mardian, 2006). Figure 1 depicts the vowel sounds in BM, adapted from (Indirawati, & Mardian, 2006).

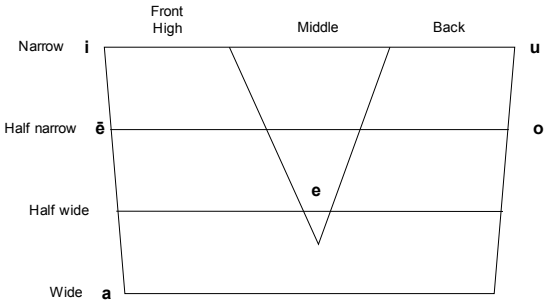

Figure 1: Vowel sound classification in BM.

For the purpose of modeling the lexicon aiming to achieve high accuracy, the vowels are modeled as having three sub-phonetic parts. This means, for example, the letter 'a' which produces the sound `A` (Worldbet) is depending upon its left context, its middle context, and its right context. See Figure 2. For semi-vowels 'w' and 'y' and vibrate letter 'r', they depend upon their left and right context. Finally, all the other consonants are defined as having only one part or context-independent since all consonant in BM are always pronounced in the same way.

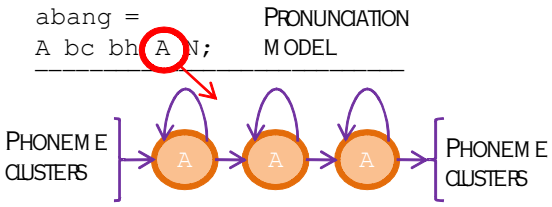

Figure 2: Context-dependent model for vowel 'a'.

The vowels are modeled as dependent upon its three parts because unlike consonants, vowel speech signals are often slightly different even for the same phoneme. The difference, although very little, does make a significant impact towards recognition. Figure 3 illustrates vowel 'a' from *bawang*.

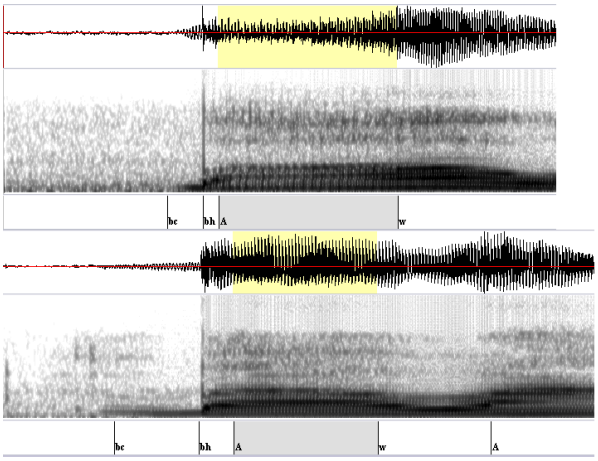

Figure 3: 2-D spectrogram of one single phoneme `A`, which differs even from the same word *bawang*.

4.2 Pronunciation Variation Adaptation

Pronunciation variation is also considered in the lexical model to include the variations produced by the children while reading aloud the selected vocabulary. The variations here include the reading errors. Instead of treating them as a separate lexicon, they can be modeled as pronunciation variations of their respective target words. For example, the errors produced when reading the word 'ayat' includes its correct form, 'ayah' (consonant subsitution) and 'aya' (consonant omission). Therefore, the pronunciation model for this word is given by

```
ayat = (A j A tc t|h)|(A j A);
```

where the pronunciation variation is allowed by having the OR operator (|).

The pronunciation variations also follow the simple rules adapted from Noraini, and Kamaruzaman (2008). This rule here (Table 3), also considers the deletion of phonemes in every word model. The rules are adapted with the results obtained from an analysis performed of recognition results.

Table 3: *The pronunciation variations.*

Character	pronunciation variations
b	p OR d OR m OR omitted
r	t
a	u
e	I u
j	c
k	g OR omitted
g	omitted

5.0 RESULTS AND ANALYSIS

Given this active lexicon, an ASR-based engine is trained on the selected speech samples. HMM/ANN is the chosen method for their performance. For that, CSLU Toolkit is used. A feed-forward, 3 layer network is used consisting of 130 input units and 200 hidden units for a standard feature of the toolkit, and 77 output units based on the vector file created.

The speech files and transcription files are used by dividing them into 3 datasets – training set, development set, and testing set. All files are exclusively for one dataset only. A total of 188 speech files are used for training, 53 files for development, and 48 for testing. Training dataset is for use in training the network and weight adjustment purposes. The goal is to learn about the general properties of the training data as much as possible. The development set is a dataset used to evaluate the network ability to recognize phonetic categories while the testing dataset is used to evaluate the network's performance.

The same data and acoustic model is used in the first training and results (after force-alignment is performed). The resulting percentile for the development set is 52.54%, whereas for testing set is 70.91%. However, after the lexicon refinement considering the variability as mentioned in the previous section, the recognition accuracy percentage is increased slightly more than 9%. It gives the result of 77.36% for the development set and 79.17 for the testing set.

6.0 CONCLUSION

This study concludes three answers. Firstly, most frequent errors illustrate and support that phonological deficit is the major factor for reading disabilities in dyslexics. The errors, when analyzed, named *vowel substitution* as the most frequent for both spelling and reading error patterns.

Secondly, careful and suitable lexical model (based on data collection and analysis results) can yield better recognition. This study showed that the use of context dependent lexical model in conjunction with pronunciation variation adaptation give lower word error rate (100% - accuracy% = WER), which means better recognition accuracy.

Finally, it is also obvious that simple phonetic refinement by adapting pronunciation variations has great impact towards increasing the recognition accuracy. This is true especially when dealing with phonetically similar words.

7.0 ACKNOWLEDGEMENTS

Thank you to special education teachers (*Bahasa Melayu*) of SK Taman Tun Dr. Ismail (2), Kuala Lumpur and SK Jalan Datuk Kumbar, Alor Star for the positive cooperation received throughout the entire data collection period. A special thank you also goes to Assistant Professor John-Paul Hosom of CSLU, OGI, USA for his consistent help.

REFERENCES

Banerjee, S., Beck, J., & Mostow, J. (2003). Evaluating the Effect of Predicting Oral Reading Miscues. *Proceedings of the EUROSPEECH 03*, Geneva, Switzerland.

Frost, J. (2001). Phonemic Awareness, Spontaneous Writing, and Reading and Spelling Development from a Preventive Perspective. *Reading and Writing: An Interdisciplinary Journal, 14*, 487 - 513.

Hagen, A., Pellom, B., Vuuren, S. V., & Cole, R. (2004). Advances in Children's Speech Recognition within an Interactive Literacy Tutor. *Proceedings of HLT-NAACL*, Boston Massachusetts, USA, 2004.

Hieronymus, J. L. (1993). *ASCII Phonetic Symbols for World's Languages: Worldbet*. Bell Labs Technical Memorandum.

Higgins, E. L., & Raskind, M. H. (2000). Speaking to Read: The Effects of Continuous vs. Discrete Speech Recognition Systems on the Reading and Spelling of Children with Learning Disabilities. *Journal of Special Education Technology, 15*, 19 - 30.

Indirawati, Z. & Mardian, S. O. (2006). *Fonetik dan Fonologi: Siri Pengajaran dan Pembelajaran Bahasa Melayu*, Kuala Lumpur: PTS Professional Sdn. Bhd.

International Dyslexia Association. (2006). *What is Dyslexia?* Retrieved Mar 30, 2007, from http://www.interdys.org/servlet/compose?section_id=5&page_id=95.

Lundberg, I. (1995). The Computer as a Tool of Remediation in the Education of Students with Reading Disabilities: A Theory-Based Approach. *Learning Disability Quarterly, 18*(2), 88-99.

Md Sah, S., Dzulkifli, M., & Sheikh Hussain, S. S. (2001). Neural Network Speaker Dependent Isolated Malay Speech Recognition System: Handcrafted vs. Genetic Algorithm. *Proceedings of the International Symposium on Signal Processing and its Applications (ISSPA)*, Kuala Lumpur.

Nix, D., Fairweather, P., & Adams, B. (1998). Speech Recognition, Children, and Reading. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, Los Angeles, U.S.

Noraini, S. & Kamaruzaman, J. (2008). Acoustic Pronunciation Variations Modeling for Standard Malay Speech Recognition, *Journal of Computer and Information Science, 1*, 4, 112-120.

Raskind, M. H., & Higgins, E. L. (1999). Speaking to Read: The Effects of Speech Recogition Technology on the Reading and Spelling Performance of Children with Learning Disabilities. *Annals of Dyslexia, 49*, 251 – 281.

Sawyer, D. J., Wade, S., & Kim, J. K. (1999). Spelling Errors as a Window on Variations in Phonological Deficits among Students with Dyslexia. *Annals of Dyslexia, 49*, 137 – 159.

Shaywitz, S. E. (1996). Dyslexia. *Scientific American,* 98 - 104.

Snowling, M. J. (2000). *Dyslexia* (2nd ed.). UK: Blackwell Publishers.

Steidl, S., Stemmer, G., Hacker, C., Noth, E., & Nieman, H. (2003). Improving Children's Speech Recognition by HMM Interpolation with an Adults' Speech Recognizer. *Lecture Notes in Computer Science*. Springer Berlin/Hiedelberg.

Williams, S. M., Nix, D., & Fairweather, P. (2000). Using Speech Recognition Technology to Enhance Literacy Instruction for Emerging Readers. *Proceedings of the 4th International Conference of the Learning Sciences*, Mahwah, NJ.

Wolf, M. (1999). What Time May Tell: Towards a New Conceptualization of Developmental Dyslexia. *Annals of Dyslexia, 49*, 3-28.

Ziegler, J. (2006). Do Differences in Brain Activation Challenge the Universal Theories of Dyslexia? *Brain and Language, 98*, 341-343.