The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1328-7265.htm

Improving ASR performance using context-dependent phoneme models

Husniza Husni and Zulikha Jamaludin
UUM College of Arts and Sciences, Universiti Utara Malaysia, Sintok, Malaysia

Abstract

Purpose – The purpose of this paper is to present evidence of the need to have a carefully designed lexical model for speech recognition for dyslexic children reading in Bahasa Melayu (BM).
Design/methodology/approach – Data collection is performed to obtain the most frequent reading error patterns and the reading recordings. Design and development of the lexical model considers the errors for better recognition accuracy.
Findings – It is found that the recognition accuracy is increased to 75 percent when using context-dependent (CD) phoneme model and phoneme refinement rule. Comparison between context-independent phoneme models and CD phoneme model is also presented.
Research limitations/implications – The most frequent errors recognized and obtained from data collection and analysis illustrate and support that phonological deficit is the major factor for reading disabilities in dyslexics.
Practical implications – This paper provides the first step towards materializing an automated speech recognition (ASR)-based application to support reading for BM, which is the first language in Malaysia.
Originality/value – The paper contributes to the knowledge of the most frequent error patterns for dyslexic children's reading in BM and to the knowledge that a CD phoneme model together with the phoneme refinement rule can built up a more fine-tuned lexical model for an ASR specifically for dyslexic children's reading isolated words in BM.

Keywords Speech recognition equipment, Reading, Dyslexia, Children (age groups), Malaysia
Paper type Research paper

1. Introduction

The demand for automated speech recognition (ASR) technology to help children to read has increased significantly due to the potential that ASR has (Steidl *et al.*, 2003; Raskind and Higgins, 1999; Higgins and Raskind, 2000). Such technology has been seen as an alternative way of teaching reading to children. In fact, ASR is the key towards an automatic reading tutor where it is used to "listen" to the readings, track the reading, and detect miscues (Mostow *et al.*, 1994; Russell *et al.*, 1996; Hagen *et al.*, 2004; Nix *et al.*, 1998; Williams *et al.*, 2000; Duchateau *et al.*, 2006; Li *et al.*, 2007, 2008; Liu *et al.*, 2008).

With the advancement of the ASR technology in teaching and training children to read, its potential could be manipulated to provide help for children especially those with 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. In favor of the fact that reading is the key towards knowledge acquisition, help should be provided to these children from

Thank you to special education teachers (*Bahasa Melayu*) of SK Taman Tun Dr Ismail (2), Kuala Lumpur, and SK Jalan Datuk Kumbar, Alor Setar 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.

Journal of Systems and Information Technology
Vol. 12 No. 1, 2010
pp. 56-69

1328-7265
DOI 10.1108/13287261011032652

using conventional teaching methods for dyslexics to using the ASR-based application as teaching aid and support tools.

This paper thus provides the first step towards materializing an ASR-based application to support reading for Bahasa Melayu (BM), which is the first language in Malaysia. The objectives of this paper are:

- to recognize the patterns of spelling and reading errors in BM vocabulary;
- to use the pattern in modeling a context-dependent (CD) pronunciation model for ASR; and
- to investigate the effect(s) of CD modeling as opposed to context-independent (CI) modeling on recognition accuracy.

With that, this paper is organized as follows: Section 2 is attributed to a brief introduction to dyslexia and its relation to reading in BM and ASR. The next section, section 3 briefly introduces BM and its system. Section 4 outlines methods to select and collect suitable vocabulary that illustrates the most frequent errors emerged from the analysis performed on the gathered data. Section 5 describes the lexical modeling, focusing on modeling the lexicon with respect to dyslexic children's pronunciation model. Later in section 6, an ASR engine is trained and tested on datasets of dyslexic children's read speech of isolated words. Section 7 discusses how CD and pronunciation variations increase ASR performance significantly. The final section, section 8 concludes the paper by presenting its limitation and future directions.

2. 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 impediment (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). Due to their difficulties, they produce a relatively high phonetically reading and spelling errors when single word is of concern. Thus, reading is always a major hurdle for dyslexics to be able to learn at schools. Although there are special crafted methods for teaching dyslexics to read, which often resort to using multi-sensory experience, the children are often self-withdrawn from the learning process. To motivate them, some element of fun and excitement need to be instilled so that the learning process continues. And, using computers is just interesting enough and thus exciting for them to be engaged in (Russell *et al.*, 1996; Lerner, 1997; Olson and Wise, 1992).

With the advancement in educational technology, research has progressed towards using ASR to help children to read. Projects such as the Colorado Literacy Tutor, CoLiT (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 *et al.*, 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 potential tool for helping those with dyslexia in reading as reported by previous studies (Hagen *et al.*, 2004; Nix *et al.*, 1998; Raskind and Higgins, 1999; Williams *et al.*, 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 and Higgins, 1999; Williams *et al.*, 2000; Higgins and 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 and not to mention the arousing dimension a computer has onto the children.

The aforementioned ASR is mostly for English readers. Only a few works involved languages other than English such as Dutch and Mandarin (Duchateau *et al.*, 2006; Liu *et al.*, 2008). Currently, there are ASR-based research in BM but none were designed for training and teaching dyslexic children to read and instead focusing more on digit recognition involving adults speech such as evidenced in Md Sah *et al.* (2001) and Sheikh Hussain *et al.* (2000).

3. The BM language

BM is the official language in Malaysia and serves as the medium in national schools. Although using the same 26 Latin alphabet as in English to construct words the phonological system differs from that of English (Indirawati and Mardian, 2006). For example, the letter "*c*" is pronounced as /tS/ and never /*k*/. BM has a more transparent grapheme-phoneme correspondences and hence more systematic when compared to English (Liow and Lee, 2004). Thus, BM shares similar granularity-transparency dimension as German or Italian (Lee, 2008).

Lee (2008) asserted that the most common method of teaching word reading in BM is via spelling the segmented syllables by sounding out the letter names of a particular syllable in a word. The syllables are constructed of consonant (C) and vowel (V) where the base structures involve V, VC, CV, and CVC to generate more complex structures such as CVC + CV + CVC (e.g. *maklumat* meaning information). BM words also involves prefixes and suffixes such as "me-", "pe-", "-an", and "-kan" constructing words such as *penjelasan* (explanation) from its root word *jelas* (explain).

As mentioned, the transparent feature of BM phoneme-grapheme correspondences requires its readers to rely heavily on phonological awareness for reading and spelling (Liow and Lee, 2004). Consequently, reading in BM too is a challenge for dyslexic children whose phonological system is impeded. Because teaching BM involves the knowledge of the phoneme-grapheme correspondences in order to spell out the syllables correctly and read the word, dyslexic children could found it such a difficult task even for simple, familiar words such as *apa* (what) and *baca* (read). For a dyslexic child to be able to read a single word they too need to be able to spell the word before they can pronounce it. Thus, help should be provided in whatever means necessary so that they are not left too far behind in education.

4. Data collection and analysis methods

The intention of this study is to use ASR-based application to support dyslexic children to read in BM due to the importance of this language in the Malaysian education system. Therefore, the first step before it could be materialized is by gathering the language corpus to be introduced and incorporated into ASR. Thus the vocabulary needs to be chosen carefully to ensure that they meet the level of the target dyslexic children's who read at word recognition level.

The vocabulary chosen for the ASR to train on is based on the Malaysian primary school syllabus, focusing on level one (standard one, two, and three) of common words. To compose the vocabulary literature review and discussions with special education teachers were conducted. The discussions are needed in order to obtain the suitable words with respect to dyslexic children and BM context.

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 BM. 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* (a guidebook for special development program for pupils in primary schools published by the Malaysian Ministry of Education that targets on the recuperation of reading, writing, and doing math) are therefore listed in the clusters accordingly. The clustered words in the list are then randomly selected and thus serve as stimuli.

A total of ten dyslexic children, as young as seven years old to 14 years old whose reading levels are similar, participated in the study. The participants are required to read aloud into a head-mounted microphone each of the 114 words prompted randomly. While the participants are reading aloud the word, recording is performed simultaneously to obtain the speech file (.wav). Figure 1 summarizes the data gathering process and its deliverables. The data gathering process is performed in seven sessions held in different days for each participant.

Once all ten participants completed their reading and recording sessions, the data collected were tabled which include all reading mistakes produced during data collection. The errors were then grouped into predefined categories. Phonological-based spelling error categories of Sawyer *et al.* (1999) are used to guide the groupings of the errors made. The categories are "substitute vowel", "substitute consonant", "omit vowel", "omit consonant", "nasals", "liquids", "incorrect sequence", "reversals of letters", and "substitute word". BM-surfaced error categories are also introduced to cope with errors that do not fall into either of these categories as presented in Table I (marked

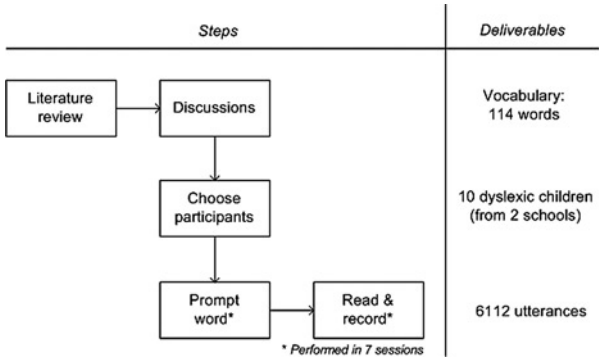


Figure 1.
The steps in data gathering process and the delivered outputs

Error types	*n*	%
Substitutes vowel	1,286	21.25
Omits consonants[a]	786	12.99
Nasals (*m*, *n*)	770	12.73
Substitutes consonants[a]	577	9.54
Omits vowel	511	8.44
Substitutes word	384	6.35
Adds consonants	363	6.00
Reversals	268	4.43
Incorrect sequence	224	3.70
Omits syllable	167	2.76
Liquids (*l*, *r*)	156	2.58
Substitutes vowel with consonant/consonant with vowel[b,n]	143	2.36
Substitutes nasals for liquid[n]	124	2.05
Adds vowel[n]	124	2.05
Syllable division confusion[n]	94	1.55
Adds syllable[n]	74	1.22

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

Notes: [a]excludes *m, n, l, r*; [b]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; [n]new BM-based categories

with [n]). For example, the *bunga* (flower) is read is incorrectly read as "bun-ga", which suggests that an error occurred when the syllables in the word is incorrectly divided. The correct division should be "bu-nga" and therefore, this type of error is assigned to a BM-based category called "syllable division confusion".

The analysis performed on the data found that the most frequent spelling and reading error pattern made is *vowel substitution* with 20 percent of occurrences of all errors. This finding supports the study of Sawyer *et al.* (1999) on phonological-based error patterns in English, which gave vowel substitution as the most frequent error made followed by consonant substitutions and omissions collectively. Table I illustrates the findings.

5. The lexical 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 I. 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 most probably would attempt to do.

The CD pronunciation modeling is performed 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 II are for the word *abang* (older brother), *ibu* (mother), *bapa* (father), *nyata* (real), and *suka* (like) respectively in Worldbet.

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 the suggestions by Nix *et al.* (1998) and Williams *et al.* (2000) that the errors produced are also regarded as and included in the active lexicon to increase recognition accuracy.

5.1 CD phonetic models

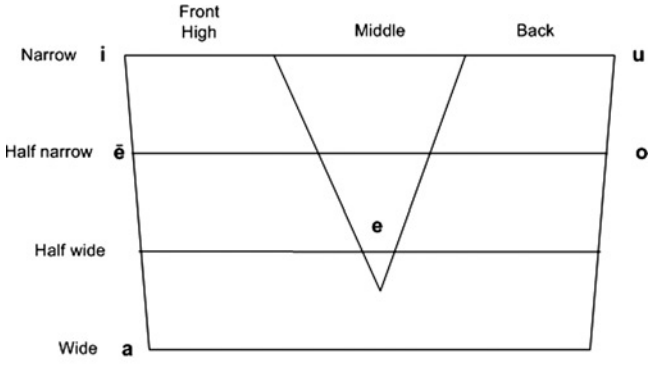
One way to increase recognition accuracy is by using CD modeling for modeling the pronunciation models of the lexicon. CD phonetic models are used when modeling the acoustic model as it can significantly improve ASR performance. CD modeling models the pronunciation of a phone with respect to its surrounding context that take into consideration the influence that a preceding phone and a following phone has upon the current phone to be pronounced. The CD model is constructed based on BM's phonetics and phonology system (Indirawati and Mardian, 2006). Figure 2 depicts the vowel sounds in BM, adapted from Indirawati and Mardian (2006).

For the purpose of modeling the acoustic model 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 3). 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 CI since all consonant in BM are always pronounced in the same way.

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 4 illustrates vowel "*a*" from *bawang*.

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*
suka	*s U* kc kh *A*

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



Source: Indirawati & Mardian (2006)

Figure 2.
Vowel sound classification in BM

5.2 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 substitution) and "aya" (consonant omission). Therefore, the pronunciation model for this word is given by:

$$\text{ayat} = (\text{A j A tc t|h}) \,|\, (\text{A j A})$$

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

The pronunciation variations adapt pronunciation variations from Noraini and Kamaruzaman (2008), which is observed from recognition results. The rule as presented in Table III also considers the deletion of phonemes in every word model. The table shows the letters and their corresponding pronunciation variants. For illustration,

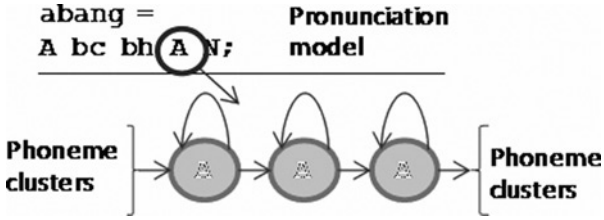


Figure 3.
CD model for vowel "*a*"

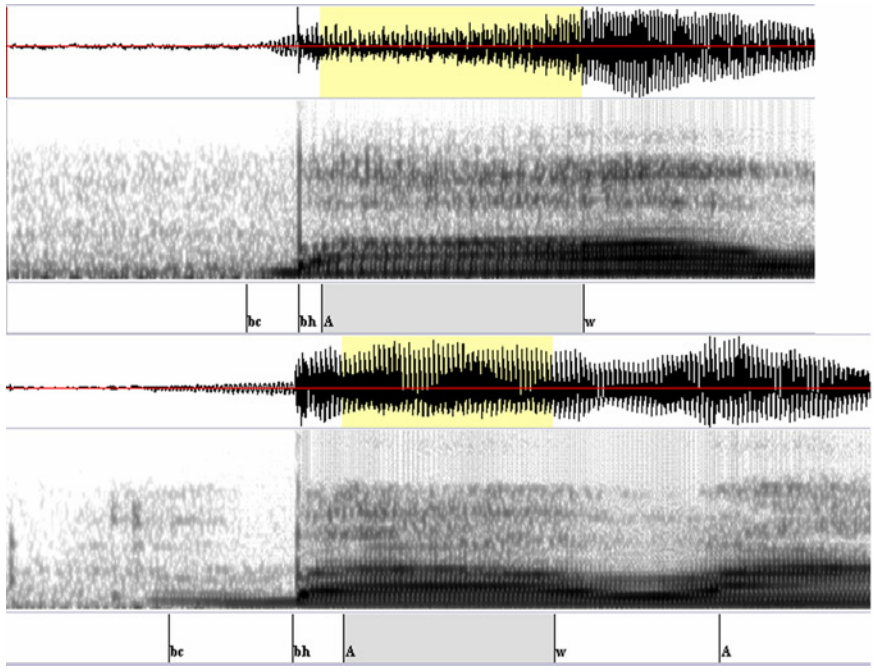


Figure 4.
2-D spectrograms of one single phoneme A, which differs even from the same word *bawang*

consider the letter "*b*". Whenever the letter "*b*" occurs in a word the pronunciation variants are modeled into the pronunciation of "*b*" by allowing its pronunciation to be that of "*p*" or "*d*" or "*m*" or omission of the letter completely from the word.

The pronunciation variations presented in Table III are used as a *phonetic refinement* rule where for example, the letter b is phonetically represented as having the phoneme p OR d OR m OR it is simply omitted from the word. To illustrate the rule, consider the following example of phonetic refinement of the letter "p" presented in Worldbet for the word *abang*. The refinement rule applied gives:

$$\text{abang} = \text{A} \quad (\text{bc} \quad \text{bh}) \ |\text{m}| \ (\text{dc} \quad \text{dh}) \ | \ (\text{pc} \quad \text{ph}) \quad \text{A} \quad \text{N}$$

where the letter "b" is refined to allow for "variations" of "p", "d", and "m". The refinement rule is applied for each word in the lexicon that contain the letters as listed in Table III and the lexical model is then used for training and testing the recognizer built.

It is important to note that the focus of this recognition is to recognize either correct or incorrect reading. Thus, it is not trying to determine whether or not the recognized utterance is a valid word. According to the pronunciation model of the word *abang* in referenced to the previous example, *adang* (Worldbet: *A* dc dh *A N*) is considered as a pronunciation variation. Even though *adang* is a valid word in BM, it is still considered as incorrect reading because the target word is *abang*.

6. Training and testing

Given the lexical model, an ASR-based engine is trained on the selected speech samples. The hybrid HMM/ANN is the chosen training method for their performance (Renals *et al.*, 1994; Franco *et al.*, 1994; Yan *et al.*, 1997; Trentin and Gori, 2003; Trentin and Gori, 2001; Cosi, 2000; Rigoll and Willett, 1998). For training using the hybrid method, Centre of Spoken Language and Understanding (CSLU) has developed a toolkit called CSLU Toolkit, which is available for free for research purposes. For this study, the CSLU Toolkit is used. A feed-forward, three 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 network is as illustrated in Figure 5. The number of input nodes and hidden nodes follows the standard feature of the toolkit whereas the number of output nodes is given by the maximum number of categories to be trained, which is automatically generated by the toolkit prior to training.

The speech files and transcription files are used by automatically dividing them into three datasets – training set, development set, and testing set. The 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. All files are

Character	Pronunciation variations
b	*p* OR *d* OR *m* OR omitted
a	*U*
e	*I* OR *u*
j	*C*
k	*g* OR omitted
g	Omitted

Table III.
The pronunciation variations

exclusively for one dataset only. A total of 188 speech files are used for training, 53 files for development, and 48 for testing.

The training process requires the use of the lexical model and the language model, together with the description files needed to train a network using CSLU Toolkit. The files are an info file for training dataset, a corpora file, and a parts file (defining the CI and/or CD of phonemes). These files are used with the speech files and their corresponding transcriptions to train the network. Figure 6 illustrates the training process until satisfying recognition rate is achieved.

The training is performed on the training dataset with specification provided by its info file, using all the files as shown in Figure 6 for 30 iterations or cycles. Once completed, the accuracy of the trained network is measured in order to determine whether or not the performance is of satisfaction. To obtain the measure, the development dataset is used, specified in an info file, which describes the development dataset. Since the study involves phonetically similar vocabulary (due to the children's difficulties), the recognizer needs high accuracy at the phoneme level. Current state-of-the-art phoneme recognition gives around 70 to 75 percent (J.-P. Hosom, 2009, pers. comm., 27 March) and thus it is defined to be of satisfaction. If the results fail to accomplish the satisfaction rate, force-alignment is performed and training is conducted once more. Force-alignment is a process where the phoneme files are generated automatically and the original, hand crafted phoneme files are not used during training. The process iterates until the performance of satisfaction is achieved. When it is, the test dataset is used to validate the network's final performance and the results are presented.

7. Results and analysis

The testing is performed using the test dataset to measure the recognition accuracy and for comparing the CI and CD models. For that purpose, the same lexicon is also modeled using CI modeling where every phoneme, vowels and consonants, is independent of its contexts. Noteworthy, the test dataset used to perform the evaluation on both CI and CD models is the same. After training with CI model, the recognition accuracy on the development and test dataset are considerably poor and far from satisfaction. With less than 50 percent of recognition rate, it implies that the possible judgment of deciding whether a word can be successfully recognized is not reliable. Consequently, the accuracy for recognizing phonetically similar vocabulary is very much jeopardized and thus confirms to the expectation that CI models perform worse when compared to CD models.

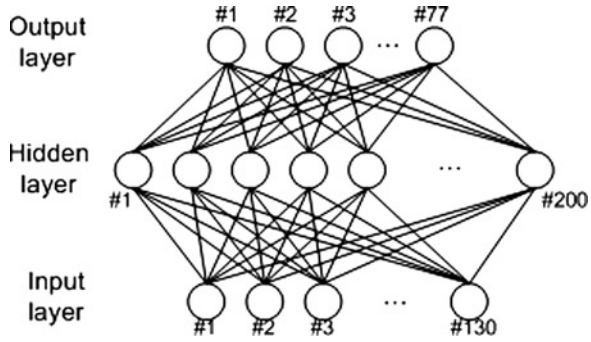


Figure 5.
The feed-forward neural network architecture used

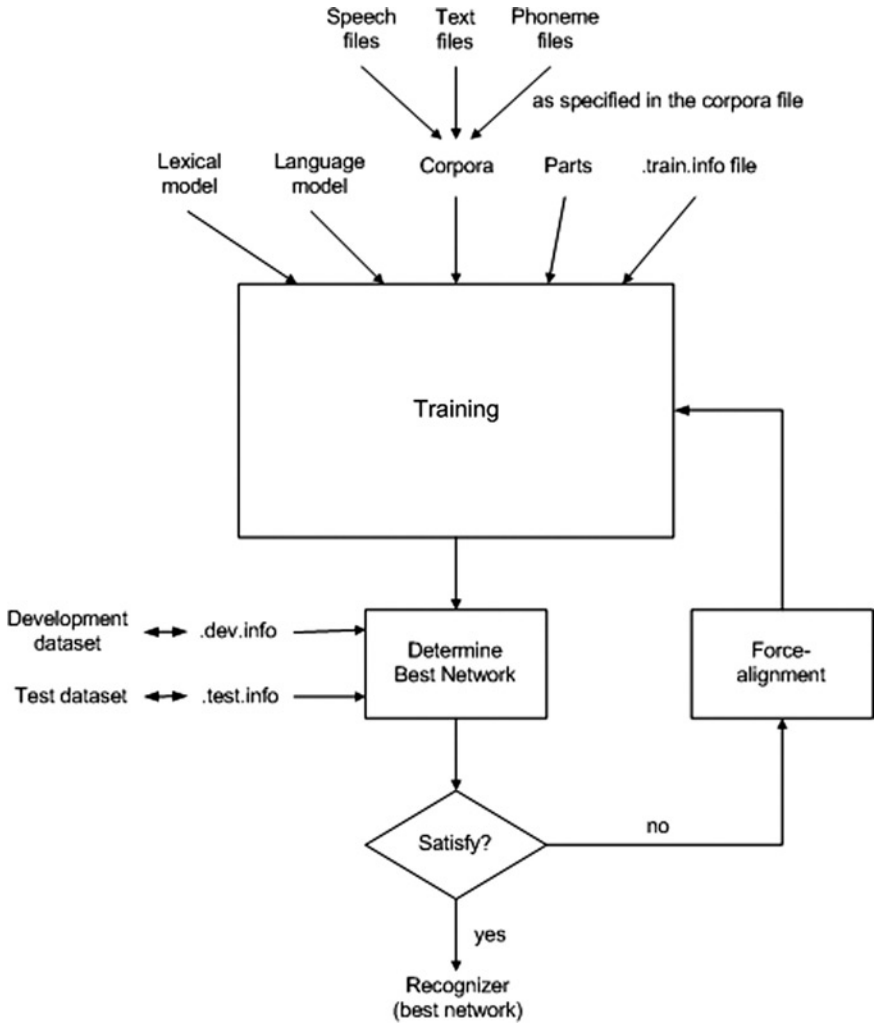


Figure 6.
The overview of the training process using CSLU Toolkit

As expected, the CD model manages to increase the performance to a significant rate. The resulting percentile for the development set is 52.54 percent, whereas for testing set is 70.91 percent that is more than 20 percent increment on the test set relative to that of CI model. However, after the *phonetic refinement* considering the variability as mentioned in and listed in Table III, the recognition accuracy rate is increased significantly. It gives the result of 77.36 percent for the development set and 75 percent for the testing set. Table IV depicts the results for 30 iterations when tested on the development set. The best network that gives the highest recognition rate is selected for evaluation on the test dataset. In this case, the network on iteration 28 (see Table IV) provides the highest percentile of 77.36 percent and thus it is chosen for testing on the test dataset.

Note that the development dataset is used to evaluate the network's ability to recognize phonetic categories mainly for cross-validation whereas the test dataset is used to evaluate the network's performance. Thus, the network on iteration 28 is

Iteration	#Words	SubError (%)	Word accuracy (%)
30	53	26.42	73.58
29	53	24.53	75.47
28	53	22.64	77.36
27	53	26.42	73.58
26	53	24.53	75.47
25	53	24.53	75.47
24	53	26.42	73.58
23	53	24.53	75.47
22	53	28.30	71.70
21	53	28.30	71.70
20	53	24.53	75.47
...	...	...	...
5	53	33.96	66.04
4	53	32.08	67.92
3	53	41.51	58.49
2	53	49.06	50.94
1	53	52.83	47.17

Table IV.
The recognition result on development dataset

selected for its best performance on the development set and the test dataset is fed into this network for further evaluation on the overall accuracy. Table V illustrates the final result and Figure 7 summarizes the findings.

Referring to Figure 7, clearly CD modeling helps increase the recognition accuracy to a significant figure of 70.91 percent and thus perform better than CI modeling. The positive effect of taking into considerations the contexts that influenced the articulation of a single phoneme suggests to the manipulation of such information to be included in modeling the language and lexical models of an ASR. After all, the ultimate goal of an ASR is to strive for better recognition accuracy, especially when phonetically similar vocabulary is a real challenge for ASR. Hence, the refinement rule introduced into the lexicon helps to boost up the performance to 77.36 percent on the development set and

Iteration	#Words	SubError (%)	Accuracy (%)
28	48	25.00	75.00

Table V.
The final output evaluated on test dataset

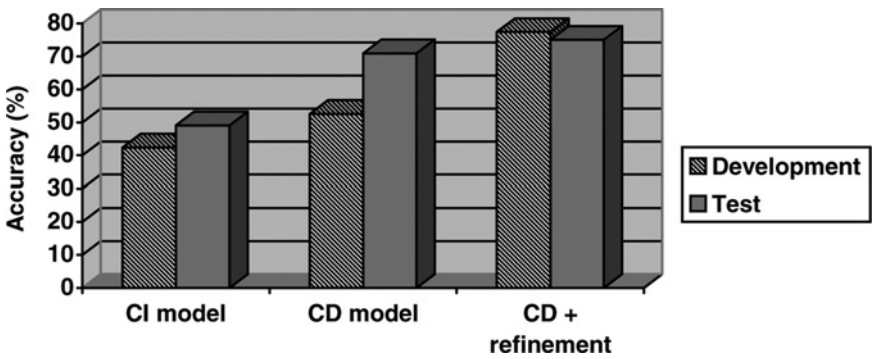


Figure 7.
Results comparison between CI model, CD model, and CD and refinement model

75 percent on the test set resulting in a satisfying figure for speech recognition that handles phonetically similar vocabulary of dyslexic children's reading isolated words in BM. A 5 percent increase in the recognition accuracy for test dataset is a positive reflect on the outcome of CD modeling and phonetic refinement and suggest to a potential way of modeling the language and lexical model for phonetically similar vocabulary that includes mispronunciations for reading-oriented ASR.

8. Conclusion and future directions

This study concludes three answers. First, the most frequent errors recognized and obtained from data collection and analysis illustrate and support that phonological deficit is the major factor for reading disabilities in dyslexics. The errors namely "vowel substitution", "consonant omission", "nasals", "consonant substitution", similarly replicate that of English as mentioned. Second, a careful and suitable lexical modeling is modeled based on the most frequent errors obtained by adapting the mispronunciations into the lexical model, which models the every phoneme to its corresponding context clusters. Third, as expected, CD modeling does improve the recognition accuracy significantly as opposed to the CI modeling by more than 20 percent. The CD pronunciation modeling applied with phoneme refinement strategy manages to demonstrate that better recognition accuracy could be achieved satisfyingly. Obviously, the phoneme refinement by adapting pronunciation variations of place and manner of articulation of a phoneme has a positive impact towards increasing the recognition accuracy. This is true especially when dealing with phonetically similar words where the final rate obtained is 75 percent on test dataset.

A more robust modeling of the ASR lexicon is needed to include the variations of acceptable pronunciations of a word under the influence of the reader's dialect and accent. Even though reading requires only standard BM to be pronounced, dealing with children is rather challenging. Despite the difficulties that children' speech entails to ASR, which researchers in the field have unanimously agree, dyslexic children basically read either in a hasty or playful manner that often resulted in reading a word in their dialect context rather than in standard BM. Low quality of speech is often produced when these children are asked to commit in the data collection process to obtain their read speech. Since reading is difficult to the children, reading thus becomes the least favorite activity and so they can easily be de-motivated and worn out by the process. So, it is seen important that in future the data collection is performed in an automatic way by using a computer as it can help increase their attention and bring in some fun factor to the activity.

Future works include corpus collection and continuous speech recognition for sentence reading. The corpus collection is for gathering more BM reading corpus and their mispronunciations from the children by using automated tool as mentioned so that the data collection process is less time consuming and more attractive to the participants. The language model, which currently supports only discrete recognition, could be modified and enhanced to cater for continuous recognition for the purpose of recognition of dyslexic children reading a phrase or a sentence in BM.

References

Banerjee, S., Beck, J. and Mostow, J. (2003), "Evaluating the effect of predicting oral reading miscues", *Proceedings of the EUROSPEECH 03, Geneva*.

Cosi, P. (2000), "Hybrid HMM-NN architectures for connected digit recognition", *IEEE Transactions on ASSP*, Vol. 5, pp. 85-90.

Duchateau, J., Wigham, M., Demunyck, K. and Van hamme, H. (2006), "A flexible recognizer architecture in a reading tutor for children", *Proceedings of ITRW on Speech Recognition and Intrinsic Variation*, *Toulouse*, pp. 59-64.

Franco, H., Cohen, M., Moran, N., Rumelheart, D. and Abrash, V. (1994), "Context-dependent connectionist probability estimation in a hybrid hidden markov model-neural net speech recognition", *Computer Speech and Language*, Vol. 8, pp. 211-22.

Frost, J. (2001), "Phonemic awareness, spontaneous writing, and reading and spelling development from a preventive perspective", *Reading and Writing: An Interdisciplinary Journal*, Vol. 14, pp. 487-513.

Hagen, A., Pellom, B., Vuuren, S.V. and Cole, R. (2004), "Advances in children's speech recognition within an interactive literacy tutor", *Proceedings of HLT-NAACL*, *Boston, MA*.

Hieronymus, J.L. (1993), *ASCII Phonetic Symbols for World's Languages: Worldbet*. Bell Labs Technical Memorandum, AT&T Bell Laboratories, Murray Hill, NJ, available at: www.ling.ohio-state.edu/~edwards/WorldBet/worldbet.pdf (accessed 30 May 2008).

Higgins, E.L. and 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*, Vol. 15, pp. 19-30.

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

International Dyslexia Association (2006), "What is dyslexia?", available at: www.interdys.org/servlet/compose?section_id=5&page_id=95 (accessed 30 March 2007).

Lee, L.W. (2008), "Development and validation of a reading-related assessment battery in malay for the purpose of dyslexia assessment", *Annals of Dyslexia*, Vol. 58, pp. 37-57.

Lerner, J. (1997), *Learning Disabilities: Theories, Diagnosis, and Teaching Strategies*, 7th ed., Houghton Mifflin Company, Boston, MA.

Li, X., Deng, L., Ju, Y. and Acero, A. (2008), "Automatic reading tutor on hand held devices", *Proceedings of the Interspeech 2008*, *Brisbane*, pp. 1733-6.

Li, X., Ju, Y., Deng, L. and Acero, A. (2007), "Efficient and robust language modeling in an automatic children's reading tutor systems", *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'07)*, *Honolulu, HI*, pp. IV-193-IV-196.

Liow, S.J.R. and Lee, L.C. (2004), "Metalinguistic awareness and semi-syllabic scripts: children's spelling errors in malay", *Reading and Writing: An Interdisciplinary Journal*, Vol. 17, pp. 7-26.

Liu, C., Pan, F., Ge, F., Dong, B., Zhao, Q. and Yan, Y. (2008), "Application of LVCSR to the detection of chinese mandarin reading miscues", *Proceedings of International Conference on Natural Computation*, *Vol. 5*, pp. 447-51.

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*, Vol. 18 No. 2, pp. 88-99.

Md Sah, S., Dzulkifli, M. and 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*, pp. 731-4.

Mostow, J., Roth, S., Hauptmann, A.G. and Kane, M. (1994), "A prototype reading coach that listens", *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94), Seattle, WA*, pp. 785-92.

Nix, D., Fairweather, P. and Adams, B. (1998), "Speech recognition, children, and reading", *Proceedings of the ACM Conference on Human Factors in Computing Systems*, *Los Angeles, CA*, pp. 245-6.

Noraini, S. and Kamaruzaman, J. (2008), "Acoustic pronunciation variations modeling for standard Malay speech recognition", *Journal of Computer and Information Science*, Vol. 1 No. 4, pp. 112-20.

Olson, R.K. and Wise, B.W. (1992), "Reading on the computer with orthographic and speech feedback: an overview of the colorado remediation project", *Reading and Writing: An Interdisciplinary Journal*, Vol. 4, pp. 107-44.

Raskind, M.H. and Higgins, E.L. (1999), "Speaking to read: the effects of speech recognition technology on the reading and spelling performance of children with learning disabilities", *Annals of Dyslexia*, Vol. 49, pp. 251-81.

Renals, S., Morgan, N., Bourlard, H., Cohen, M. and Franco, H. (1994), "Connectionist probability estimators in HMM speech recognition", *IEEE Transaction on Speech Audio Processing*, Vol. 2 No. 1, pp. 161-74.

Rigoll, G. and Willett, D. (1998), "A NN/HMM hybrid for continuous speech recognition with a discriminant nonlinear feature extraction", *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Seattle, CA*, pp. 9-12.

Russell, M., Brown, C., Skilling, A., Series, R., Wallace, J., Bonham, B., *et al.* (1996), "Applications of automatic speech recognition to speech and language development in young children", *Proceedings of the 4th International Conference on Spoken Language ICSLP'96, Philadelphia, PA*, Vol. 1, pp. 176-9.

Sawyer, D.J., Wade, S. and Kim, J.K. (1999), "Spelling errors as a window on variations in phonological deficits among students with dyslexia", *Annals of Dyslexia*, Vol. 49, pp. 137-59.

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

Sheikh Hussain, S.S., Ahmad, Z., Zulkarnain, Y., Rahman, S. and Lim, S.C. (2000), "Implementation of speaker identification systems by means of personel computer", *Proceedings of TENCON 2000, Kuala Lumpur, Vol. 1*, pp. 43-8.

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

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

Trentin, E. and Gori, M. (2001), "A survey of hybrid ANN/HMM models for automatic speech recognition", *Neurocomputing*, Vol. 37, pp. 91-126.

Trentin, E. and Gori, M. (2003), "Robust combination of neural network and hidden markov models for speech recognition", *IEEE Transactions on Neural Network*, Vol. 14 No. 6, pp. 1519-31.

Williams, S.M., Nix, D. and 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*, pp. 115-20.

Wolf, M. (1999), "What time may tell: towards a new conceptualization of developmental dyslexia", *Annals of Dyslexia*, Vol. 49, pp. 3-28.

Yan, Y., Fanty, M. and Cole, R. (1997), "Speech recognition using neural networks with forward-backward probability generated targets", *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-97, Munich*, Vol. 4, pp. 3241-4.

Ziegler, J. (2006), "Do differences in brain activation challenge the universal theories of dyslexia?", *Brain and Language*, Vol. 98, pp. 341-3.

Corresponding author
Husniza Husni can be contacted at: husniza@uum.edu.my