ASR technology for children with dyslexia: Enabling immediate intervention to support reading in Bahasa Melayu

Husniza Husni, Zulikha Jamaludin

(College of Arts and Sciences, University Utara Malaysia, Kedah 06010, Malaysia)

Abstract: Reading is an essential skill towards literacy development, and should be provided so that children can master the skill at their early ages. For dyslexic children, mastering the skill is a challenge. It has been widely agreed that the theory behind such difficulties in reading for dyslexic lies in the phonological-core deficits. Support has been given in many ways to dyslexic children to teach them reading through using various multi-sensory methods and using computer-based applications that include animated characters and text-to-speech (TTS) technology. In such applications, although stimulating, it requires the children to call for help by clicking on the custom-made buttons on the computer screen, often, such an application requires the dyslexic children to be aware of their mistakes and be able to judge when help is needed. They are just reluctant to ask the computer for help. Hence, such technology does not provide immediate intervention to correct any reading failure. It is therefore worth to look at the promising automatic speech recognition (ASR) technology to provide such intervention. Hence, this paper gives an overview of the use of ASR to facilitate immediate reading intervention which is the key element of remediating reading among dyslexic children. For such intervention to work, data on reading mistakes and patterns are observed and collected in audio format. The data serve as training and testing samples for an ASR to train on. An observation was carried out in two public schools participated in the study to record dyslexic children's reading in Bahasa Melayu (BM) and observe error patterns and their behaviors toward reading. A total of 10 dyslexic children are involved and a total of 6384 utterances from a set of selected words have been gathered and analyzed. Data are grouped into error type categories and the analysis performed gives "vowel substitution" as the most frequent error made (20%). The significant findings can be of interest of special education teachers or parents to devise and use suitable approach to correct reading mistakes often made by dyslexic children. The findings also contribute to the development of a suitable and well-tuned ASR model focusing on dyslexic children reading aloud in BM.

Key words: automatic speech recognition; dyslexia; reading; Bahasa Melayu

1. Introduction

In Malaysia, dyslexia, a condition that impedes the ability to read, spell and write, has gained serious attention from the government through the Ministry of Education, private institutions, and researchers as well as parents. It is estimated that up to 10% of children are dyslexics and very few of them receive proper, suitable teaching method for remediation. Despite having private organizations such as Malaysia Dyslexia Association that

Husniza Husni, M.S., College of Arts and Sciences, University Utara Malaysia; research fields: speech recognition, affective engineering.

Zulikha Jamaludin, College of Arts and Sciences, University Utara Malaysia research fields: multimedia, interaction design, automatic speech recognition (Malay language/Bahasa Melayu).

offers private tuition for dyslexics, the Malaysian government is currently running 30 public schools equipped with special dyslexia classes to provide help for the affected children. Help is essentially needed for these children to facilitate their learning experience, especially when they have difficulties in basic learning skills—reading, spelling, and writing.

Reading is a skill that contributes to knowledge building and sharing. For dyslexic children, mastering this skill is a challenge due to their difficulties. Many have agreed that the theory behind such difficulties is the phonological-core deficits theory (Lundberg, 1995; Shaywitz, 1996; Wolf, 1999; Snowling, 2000; Frost, 2001; Ziegler, 2006). Support has been given in many ways to dyslexic children to teach them to read from teaching using multi-sensory methods to using computer-based applications that include animated characters and text-to-speech (TTS) technology. Table 1 presents a list of software for use to aid the learning process of a dyslexic, adopted from Husniza and Zulikha (2008).

Table 1 Software as technological support to facilitate learning for dyslexic children

Software	Tech.	Function
Go phonics	CD-ROM	Software to teach reading to dyslexic children based on Orton-Gillingham method, provides test and assessment
Language tune-up kit	CD-ROM	Software to teach reading to dyslexic children based on Orton-Gillingham method, teaches grammar, punctuations, and the rules
TextHelp	TTS	Word processing support, suggest spelling when typing, read aloud writing for checking
Kurzweil 3000	TTS	Users can scan books/reading material and it will read out to them
Clicker 4	TTS	Word processing support, read aloud once done typing, able to “speak out” a word/letter upon user request
Readplease	TTS	“Read” aloud text from web pages/emails
Helpread	TTS	Read-along software while user are reading
WordQ	TTS	Writing tool (typing), suggest words, provide speech feedback
Dragon naturally speaking	ASR	Dictation software
Via voice	ASR	Dictation software

In such applications, although stimulating, it requires the children to call for help by clicking on the custom-made buttons on the computer screen. Such an application requires the dyslexic children to be aware of their mistakes and be able to judge when help is needed. This is a major disadvantage because they are often unaware of mistakes made when reading. They are just reluctant to ask the computer for help. Hence, such technology does not provide immediate intervention, which is the key in teaching them to read, for correction. It is therefore worth to look at the promising ASR technology to provide such intervention.

2. Immediate intervention via ASR

As mentioned, immediate intervention is the key to teach dyslexic children to read. Immediate intervention means that whenever dyslexics read incorrectly or stumble upon a problem when reading feedback should be given immediately. For example, if a dyslexic read “aku” (I or me in English) as “aki” (error in substituting the vowel “u” with “I”), a reading facilitator or teacher must immediately correct his/her reading. Correction must be made quickly so that the child realizes the mistake that has been made and learn the correct pronunciation of that particular word.

ASR technology is the key technology to provide such support to reading-aid applications. ASR technology simply, is a technology that enables computer or machine to recognize spoken attributes of a human. This can be

achieved by using various recognition methods for which hidden Markov model is the state-of-the-art method. Projects such as the *Colorado Literacy Tutor*, *CoLiT* (http://www.colit.org/) and project *Listen's Reading Tutor* (Mostow, Roth, Hauptmann & Kane, 1994; Mostow & Beck, 2003) are aiming at providing computer-aided reading instructions for children to enhance reading. These major projects use ASR as the key technology. ASR is used to track reading while the children are reading aloud and add for an interactive application using speech, which enable users to ask question to the application. In addition, pronunciation accuracy is also provided for feedback. Figure 1 illustrates the ASR-enabled immediate intervention to help dyslexic children to read.

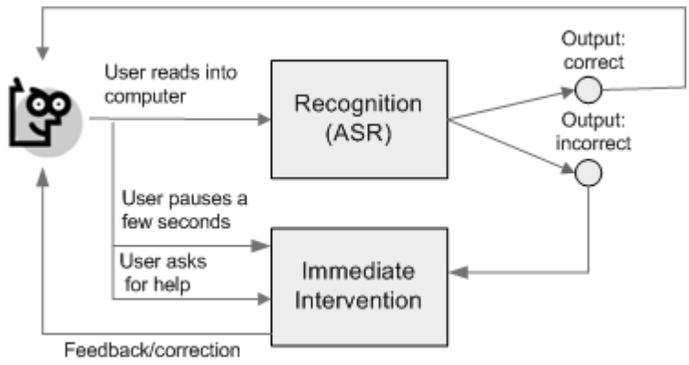

Figure 1 A general scheme of immediate intervention in speech recognition application to facilitate learning to read correctly for dyslexic children

A user reads into the computer (using an attached microphone) and the read speech is input to the recognition process to generate an output. If the output is correct, the user proceeds his/her reading. Once a mistake is made, the immediate intervention module is invoked, allowing for feedback that informs the user of the mistake. In some application, immediate intervention is automatically provided when the application detects a pause (a few seconds interval) that normally occur when the user hesitates to read certain words (Mostow, Roth, Hauptmann & Kane, 1994; Williams, Nix & Fairweather, 2000). This often happens when the user has no knowledge of how to pronounce the word or is unsure of the pronunciation. To get feedback, the user can ask for help by means of clicking on the help button provided within an application.

By providing the immediate intervention as support, ASR technology has the potential to enhance reading ability for normal children and is a good tool for helping dyslexic children to read (Nix, Fairweather & Adams, 1998; Williams, Nix & Fairweather, 2000; Hagen, Pellom, Vuuren & Cole, 2004). Furthermore, ASR is found to offer multi-sensory experience to dyslexic children as means of teaching (Raskind & Higgins, 1999; Higgins & Raskind, 2000). The multi-sensory experience is created as the child read aloud a word and see it being displayed on the computer screen. This involves using senses such as in terms of articulation and speech production, hearing, and visual.

ASR technology is particularly useful in one-to-many teaching. In other words, it is suitable to be used in a big classroom where one language teacher is teaching a number of students. For example, in Malaysia, dyslexic children need to attend and learn in accordance to the primary school syllabus regardless of their condition. During BM classes, they attend a special language class where a teacher is teaching about five or six students at the same time. The situation reduces immediate intervention given to the children as the teacher needs to monitor more than one students at the same time. Although the number is small, these children often need one-to-one

instruction for better benefit. Therefore, it is promising that by using ASR-based application, the immediate intervention can continue facilitating each student.

3. Data collection

To materialize such an intervention for dyslexic children reading in BM, the first step is data collection of the children's reading isolated words. For the purpose of recognition of read speech, samples of speech of children reading aloud were collected.

3.1 The stimuli

A total of 114 isolated, single words in BM are gathered across 23 syllable patterns of this particular language, ranging from simple formation of consonant-vowel (CV) forms to somewhat complex CV forms. These words serve as stimuli in order to record audio form of dyslexic children's reading and later obtained their reading error patterns.

3.2 The subjects

To obtain audio data of dyslexic children's reading, ten dyslexic children are recruited from two public schools that offer special dyslexia classes for these children under the dyslexia program by the Ministry of Education Malaysia. The ten subjects who posses similar reading level (problems at word recognition level) are suggested by their teachers who are acquainted with their reading levels.

3.3 The process

Each subject is required to read aloud into a computer via a standard headphone with a microphone attached to it. The recordings of their reading are performed on separate, individual sessions with one subject in one session. They are required to read all 114 words out loud in a single session for seven sessions all. However, some of them need more than one session to read all the stimuli presented and some only managed to complete 2-3 sessions due to various problems such as time constraint and recording venue. A subject is prompted a word and asked to read the word out loud into the computer. Recording is performed simultaneously to capture the reading into audio format. While recording is performed their reading errors are observed.

4. Data analysis

A total of 6384 utterances from a set of selected words have been gathered. Data analysis is performed by transcribing all utterances into their corresponding transcripts, which include all mistakes made. Errors made are grouped into corresponding categories adapted from Sawyer, Wade and Kim (1999). From there, the most frequent error patterns are obtained, which result in vowel substitution, SV (slightly more than 20%) as the most frequent error patterns made. Error category that ranked second is consonant deletion, OC (12%) followed by errors in nasals, N (replace, delete or add nasals letter, m and n) with also 12% of frequency. The third most frequent error made is of consonant substitution, SC (9%). Figure 2 illustrates the results of other error categories with respect to BM—omit vowel (OV), substitute word (SW), add consonant (AC), substitute non-word (SNW), reversals (Rev.), incorrect sequence (IS), omit syllable (OS), liquid (L), substitute vowel with consonant or vice-versa (SVC), substitute nasal for liquid (SNL), add vowel (AV), syllable division confusion (SDC), and add syllable (AS).

It is worthy of being noted that this finding is somewhat similar to that of Sawyer, et al (1999) where vowel substitution is the most frequent errors made. Unlike Sawyer, et al., who present consonant substitutions and omissions collectively as the second largest category of errors, this study shows that consonant deletion and nasals are the second most frequent errors made when the children read aloud single words in BM. Consonant

substitution, which is believed to have been related to “possible misperception of similar sounds” as claimed by Sawyer, et al., does not involve substitutions of voiceless consonant. Substitutions of voiceless consonant letters appeared to be the biggest mistake which contribute to the number of substitution errors. Unlike English, BM has no voiceless consonant letters. Even the language examined are different, there is certain similarities in terms of the results which exhibit that the most frequent errors made is of vowel substitutions. For both English and BM, vowels are represented by the letters “a”, “e”, “i”, “o” and “u”.

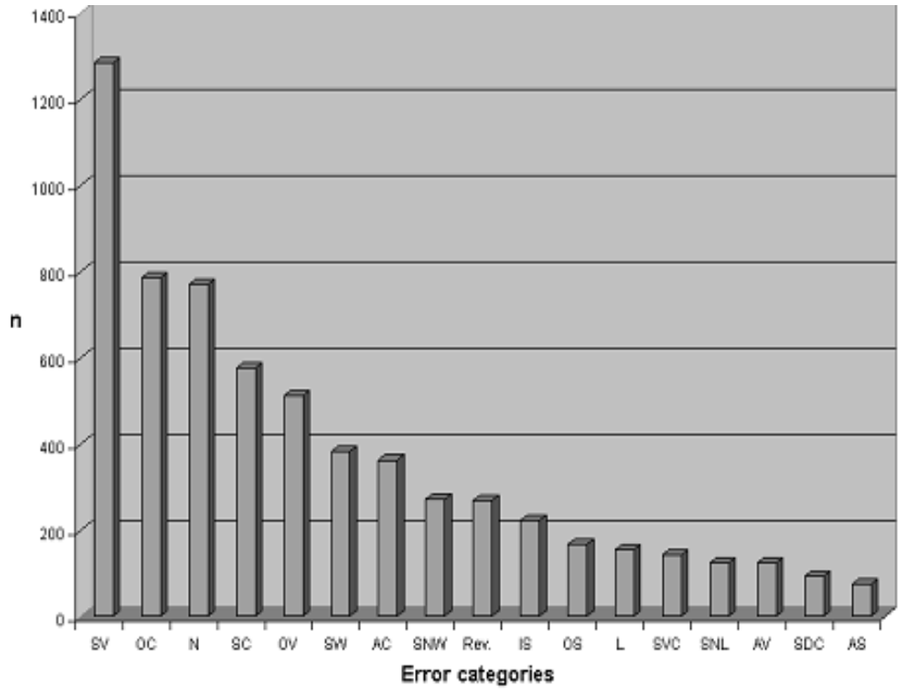

Figure 2 Error categories and number of errors, *n* made by subjects reading selected single, isolated BM words

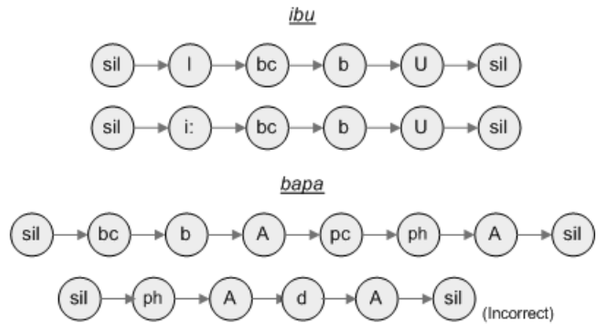

Figure 3 Example of pronunciation models of common words in BM, ibu and bapa

The obvious difference is the way that assignment of errors of suitable categories. In this study, the word is examined letter by letter, which means each word could contain zero, one or more errors. For example, the word “abang” (meaning brother) is read as “adangan”. “Adangan” is a non-word in BM but it is not assigned into “substitute with non-word” category but instead assigned to reversal category of which the letter “b” is reversed to

form the letter “d”. And the letters “a” and “n” at the end of the word, although look like additions of a vowel and a consonant, forms a valid post-fix in BM, which is the syllable “an”. Hence, this word’s errors are assigned to two error categories namely “reversal” and “add syllable”. Although Sawyer, et al., did not explicitly outline the means of classification, each error is classified into one category, for example, “jem” read as “drum” and is assigned into consonant substitution and “ship” read as “sep” for consonant deletion.

Next, the most frequent error categories are selected and transcripts of words that fall under these categories are chosen to be modelled in terms of their pronunciations. These pronunciation models are to be incorporated as vocabulary for the ASR engine. Example pronunciation models are depicted in Figure 3. The pronunciation models model the words “ibu” (mother) and “bapa” (father) respectively. The pronunciations are transcribed using Worldbet phonetic transcription scheme (Hieronymus, 1993).

In Figure 3, “sil” represents silence before and after the word being spoken. In the word “ibu”, two models are considered—one without the prolonged sound of “I” and the other for a prolonged pronunciation of the same sound. As for “bapa”, two models are also taken into consideration. The first model models the correct pronunciation and the second one models the incorrectly read word as “pada” (meaning “to” in English). The pronunciation models help determine the recognition process accuracy. Noteworthy, the pronunciation models of words that belong to the error categories also include all miss-pronunciations. It is recommended by Nix, Fairweather and Adams (1998), and Williams, Nix and Fairweather (2000) to incorporate the reading mistakes for the ASR engine training to obtain a more accurate ASR training results.

5. Discussion and future work

Through the incorporation of the pronunciation models, which include also the incorrect pronunciations, immediate intervention shall be realized. This way, whenever a reading error is made when a dyslexic child is reading aloud words from the vocabulary, the ASR application shall be able to provide a more accurate suggestion of correct word for useful feedback. Noteworthy, such intervention could not have been materialized without employing ASR as the key technology.

One particular challenge in gathering the incorrect readings for modelling the pronunciations lies in data collection that involves dyslexic children reading aloud. Since reading is a hassle, they are often reluctant to participate or they simply provide an uncommitted reading (not paying attention to instructions, playing, etc.), which yield errors that are not suppose to surface during their actual reading, i.e. they perform worse than their actual ability. One way to overcome this challenge is to use computer-based recording method as opposed to performing manual recordings. Using computers are just attractive enough to capture the children’s attention (Olson & Wise, 1992; Russell, et al., 1996; Lerner, 1997). Therefore, using a simple computer-based application that is able to prompt the words and record the readings of a subject can help motivate them to read by the installment of excitement and fun (using animated characters, colorful presentations, etc.). In addition, this recording method could improve the time taken to record each subject’s readings, which can be rather slow.

The future work concerns with the ASR prototype development that uses the vocabulary (pronunciation models of words). The prototype is of a recognition engine for training the vocabulary using suitable ASR method. The most popular method is Hidden Markov Model (HMM) as the dominant technique for speech recognition. However, this prototype shall be developed using the hybrid of HMM and Artificial Neural Network (ANN). The hybrid method incorporates the advantages of ANN, i.e. the excellent classification ability with HMM for faster

recognition and better recognition accuracy.

6. Conclusion

Immediate intervention is the key in teaching dyslexics to read. It allows them to learn from the process of making mistake when reading and immediate correction given as feedback. By incorporating immediate intervention in computer-based applications to teach the children to read, they can benefit more from the applications as they can receive corrective feedback immediately after incorrect reading. This provides a way to enable interactive application that is promising to give motivation in reading activities. To enable such immediate corrective feedback in computer-based applications, ASR technology is the key by means of incorporating vocabulary that consists of not only the correct, target words but also the incorrectly read words. Inclusion of the incorrect words read by the children into the vocabulary can improve recognition accuracy as well as the accuracy of suggestions of correct word for feedback. It is therefore important to be able to suggest correct word for any incorrectly read word to dyslexic for an enhanced learning experience for the children.

References:

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., Van Vuuren, S. & Cole, R. (2004). Advances in children's speech recognition within an interactive literacy tutor. *Proceedings of HLT-NAACL*, Boston Massachusetts, USA.

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.

Husniza, H. & Zulikha, J. (2008). A retrospective and future look at speech recognition applications in assisting children with reading disabilities. *Proceedings of the World Congress on Engineering and Computer Science*, Berkeley, San Francisco, USA.

Lerner, J. (1997). *Learning disabilities: Theories, diagnosis, and teaching strategies (7th ed.)*. Boston: Houghton Mifflin Company.

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.

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

Mostow, J. & Beck, J. (2003). When the rubber meets the road: Lessons from the in-school adventures of an automated reading tutor that listens. Proceedings of *the Conceptualizing Scale-Up: Multidisciplinary Perspectives*, Park Hyatt Hotel, Washington, D.C.

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, USA.

Olson, R. K. & 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, 4*, 107-144.

Raskind, M. H. & 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, 49*, 251-281.

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 (Vol. 1)*, 176-179.

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 November). Dyslexia. *Scientific American*, 98-104.

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

Williams, S. M., Nix, D. & Fairweather, P. (2000). Using speech recognition technology to enhance literacy instruction for emerging readers. In Fishman, B. & O'Connor-Divelbiss (Eds.). *Fourth International Conference of the Learning Sciences*, Mahwah, NJ: Erlbaum, 115-120.

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.

(Edited by Max and Jean)