Journal of Network and Innovative Computing
ISSN 2160-2174 Volume 2 (2014) pp. 366-372

Assessing Usability and Fun in *MyCard:* Malay Automatic Speech Recognition for Articulation Training

Lau Kum Hoe [1], Syariffanor Hisham [2] and Noor Azilah Draman @ Muda [3]

[1] Faculty of Information and Communication Technology,
UniversitiTeknikal Malaysia Melaka
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
petersky86@hotmail.com

[2] Faculty of Information and Communication Technology,
UniversitiTeknikal Malaysia Melaka
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
*syariffanor@utem.edu.m*y

[3] Faculty of Information and Communication Technology,
UniversitiTeknikal Malaysia Melaka
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
azilah@utem.edu.my

***Abstract*: The main objective of this study is to investigate the feasibility of Malay automatic speech recognition (ASR) engine as a tool in speech rehabilitation among children with Down syndrome (DS) in Malaysia. Three children with DS were asked to use an application called *MyCard* during the evaluation session. Integrating with Malay ASR engine, *MyCard* was designed to enable the children with DS to practice their Malay articulation skills. Identifying which parts of *MyCard* causing the usability and fun problems on the children and therefore needed to improve would eventually increase the feasibility of the application in the speech rehabilitation for children with Down syndrome. Therefore, an adaptive coding scheme was used to assess the feasibility of *MyCard* based on usability and fun aspects. Based on the findings, the result suggests that the recording scene was appeared being most trouble for the children in this study.**

***Keywords*: *DEVAN, Down syndrome, Evaluation, Usability, Fun*.**

I. Introduction

Speech and language learning of children with Down Syndrome(DS) occurs at a slower rate compared to other children of the same age. Therefore, children with DS often take longer time in learning how to talk; in some cases, they may not be able to talk in intelligible manner throughout their lifetime [1]. Learning how to convey verbally and properly is crucial as it helps an individual relating to one another in society. Regardless of abilities, every child needs to learn to be independent by socializing with their peers. Therefore, improving the speaking and language skills of DS is essential in promoting an independent living for these children.

One way to improve speech and language development of DS is to provide these children with stimulus that can stimulate their learning desire. Information and communication technology (ICT) is able to fulfill this objective by presenting learning contents (in the form of multimedia application) via interactive plays and reliable instant feedback. This can offer greater opportunity to bypass the learning difficulties among children with DS. Creative activities help to improve the confidence and motivation for children with DS [2]. Likewise, these interactive applications could also help to improve the speech and language difficulties among children with DS.

In the study presented in this paper, the researchers aimed to investigate the feasibility of Malay automatic speech recognition (ASR) application as a tool for speech rehabilitation among children with DS in Malaysia. A prototype Malay ASR known as *MyCard* was developed in this study and three children with DS took part in the evaluation of the prototype. The feasibility of *MyCard* was sought based on the difficulties that the children encountered and their emotions were observed while interacting with the prototype Malay ASR.

Usability study can help researchers and developers in identifying problems that is useful for future development of an application. As a result, this will directly help in improving user satisfaction and feasibility of *MyCard* as a tool for speech rehabilitation among children with DS. This paper aims to

discuss the usability and fun problems encountered by the children with DS when interacting with *MyCard*. The discussion starts with some background and literature review of related fields, description of *MyCard*, experiment procedure, analysis and discussion of the results and finally conclusion.

II. Literature Review

Children with DS differ from normal children in various aspects such as cognitive, sensory perception and processing, language acquisition as well as gross and fine motor skill development. Furthermore, due to the stereotype of DS, they often require longer time to complete a task compare to normal children. Thus, children with DS are more likely to have more interaction problems while interacting with technologies including assistive tools. There are a few aspects that must be taken into account in proposing interactive system for children with DS.

Study shows that recruiting participants with special needs can be challenging due to the fact that children with disabilities often associated with nature and attitudes that are hard to expect, and impulsive decisions and actions [7]. Nielsen [8] suggests that three to five participants is acceptable in a usability study among people with special needs. The increasing number of sample will not only increase the costs but also prolong the evaluation session, which can be troublesome to participants with special needs. Therefore, some technical and ethnical measures were considered in conducting evaluation of *MyCard* with three children with DS.

In this study, the usability and fun problems of *MyCard* were assessed based on the structured coding scheme [4]. The coding scheme was originally called DEVAN method, a tool for detailed video analysis of user test data [5]. The method makes use of a table format to represent interactions at multiple levels of abstract to assess the usability problems in products targeted for adults. Barendregt and Bekker [6] later adapted the DEVAN method to assess the usability and fun problems in games among children. Macedo and Trevisan [4] later adapted this method to observe children with DS interaction with games particularly in eliciting usability and fun problems in games.

III. Description of Tools

A. Coding Scheme

MyCard was developed to provide articulation training for children with DS which differs from the game being assessed in earlier work by Macedo and Trevisan [4]. Therefore, in conducting usability and fun assessment of *MyCard*, some of the breakdown indications from the original list were removed or refined. The final version of the coding scheme used in this study is shown in Table I, and the researchers in this study used it as a guideline to observe the behaviors and actions of the children with DS when they were interacting with *MyCard*.

Table 1. Coding Scheme (Adapted from [4])

Description	Definition
Breakdown Indication Types Based on Observed Actions with MyCard	
Wrong Action	An action does not belong in the correct sequence of actions. An action is omitted from the sequence. An action within a sequence is replaced by another action. User performs a wrong action unintentionally.
Intentional Wrong Action	The user knows that the action is wrong, but still performs this action only to have fun.
Help	The user could not proceed without help from Teacher. The Teacher intervenes to prevent serious problems.
Dislike	The user indicates dislike something. The user does not cooperate.
Puzzled	The user indicates not knowing how to proceed.
Impatience	The user shows impatience by moving his/her body while waiting the application to response.
Passive	The user stops interacting and does not perform the expected action
Execution Problem	The user indicates not knowing how to pronounce the word. The user makes inconsistent speech errors. The user still makes speech errors after receiving assistances from the Teacher. The user stops in the middle suddenly when he or she is pronouncing
Perception Problem	The user indicates not able to hear or see something clearly, not understanding how to proceed.
Random Actions	The user performs random actions.
Bored	The user indicates being boring by sighing, yawning or acting sleepy

B. MyCard

MyCard is a card-based game learning application developed to provide articulation training for children with DS [3]. It utilized the speech automatic speech recognition (ASR) to access the speech pronunciation skills of DS. Figure 3. shows the description of each scene from the application.

Usually, determining whether certain aspects of *MyCard* cause usability or fun problem is always unclear. For example, it is hard to determine whether problem occurs when the child is concentrate watching the animation and at the same time showing behavior that could signal usability problem, like staring on the screen or passive. Therefore, some actions are anticipated by the researchers and these actions serve as insight for detecting the usability and fun problems during the assessment of *MyCard.* Table 2 lists the actions expected that the children would perform when he or she is interacting with *MyCard*

Table 2. Expected Actions Towards MyCard

Scene	Expected Actions
Menu	The child will click or select spontaneously word he or she likes to practice. (As the teacher(s) would select the word for the participants to participate, the screen was omitted during the analysis.)
Animation	The child will spontaneously imitate the word when he or she watch the animation.
Recording	The child will spontaneously repeat the word when microphone icon is click.
Feedback	The child will react accordingly to the score being displayed, i.e. showing disappoint or satisfy when low or high score was displayed repectively.

C. Word List

In the meantime, *MyCard* can only recognize seven isolated Malay common words. These isolated words are seemed ideal to practice for children with DS because these words are often appeared in the daily activities for these children. All the words are structured in two syllables format, with different combination of consonants and vowels. Table 3 shows these isolated Malay words, with the level of difficulties increases from top to the bottom, recognized by the application.

Table 2. Isolated Word List

Word	Syllable Structure
Adik (brother)	V-CVC
Emak (mother)	V-CVC
Suka (like)	CV-CV
Makan (eat)	CV-CVC
Sakit (pain)	CV-CVC
Nuah (fruit)	CV-VC
Askar (army)	V-CVC

IV. Procedure

Three children with DS, aged between 6 and 12 years old (mean = 10 years old), were recruited for the usability study of *MyCard*. None of the children had any previous experience with using *MyCard*. The limited number of children recruited in this study is due to fact that one of the features in *MyCard* was to provide speech assessment for user, and the children with DS must have some basic sound imitation skills in order for them to interact with *MyCard*. This had imposed a technical problem when recruiting the children with DS for the usability study as many of these children did not meet this eligibility due to severe speech difficulties.

The usability study was conducted as individual session, and the children were accompanied by their teachers. The teachers were allowed to provide assistance but they were advised to keep their interaction with their child as minimal as possible. Before the study took place, the researchers demonstrated how to use *MyCard* and later each child was given the opportunity to explore the application with assistance. Each session was videotaped and the researchers obtained three videos of approximated 60 to 90 minutes for the evaluation of *MyCard*.

V. Analysis of the Test Session

The analysis started with transcribing the video and logging the behaviors as well as actions observed during the evaluation session. The behaviors and actions were transcribed into a multicolumn table; one for each participant. It is noted that some similar usability and fun problems were reoccurred when the children were interacting with *MyCard* throughout the video. Therefore, instead of transcribing the video in full length, 20 minutes of these videos were transcribed.

Interval for video transcribing must be determined first before video transcribing begins. If the interval was too short, some actions or behaviors of the children would not be observed. However, if the interval was too long, some actions or behaviors of the participants would be overlooked. In this study, the interval used for video transcribing is between two to four seconds. The reasons behind this are: 1) if transcribing the video based on 1 second the video segment only contains single action of user which is “meaningless” because human-computer interaction is a bidirectional process; and 2) children with DS often have different interaction problems which may delay their responses when they are interacting with the interactive system. Based on the pilot study, it seemed four seconds is ideal for the delay among children in this study.

Once the videos were transcribed inside a multi-column table, locating the breakdown indications will follow next. The procedure is simply go through the multi-column table and ‘one-by-one’ locating the behaviors based on the breakdown indications from Table 1. the actions anticipated from Table 2. This leads to the creation of interaction table as shown Figure 1. The left-most column represents the time stamp when an action (second column) performed by the child when he or she was interacting with the application. The third column is the

contextual information and *MyCard* status when the action was performed by the child. The right-most column is the breakdown indication coded for the actions or behaviors observed on column two or three. For instance, passive was coded in the right column when the participant was observed staring on the screen without performing any action (00:02:30).

Figure 3 shows the overall total occurrence of each indication found in the animation, recording and feedback scenes of *MyCard*. The result shows that the recording scene being most difficult to the children as the number of assistance provided by the teachers to the children was high. This happened mainly due to the fact the children could not spontaneously repeat the word.

Time Stamp	Log	Context	Breakdown Indications
00:02:30	Looking on the screen	**[Trial 1 – *Adik* Animation starts]**	Passive
00:02:55	Turn to facilitator and smile	Animation stops	
00:02:58	Looking on the screen ("Tak dengar")		Perception Problem
00:03:00	Looking on the screen	Recording starts	Passive
00:03:04	Assistance - repeating		Help
00:03:07	Smile ("Take dengar")		Puzzled
00:03:10	Assistance – repeating	Recording starts	Help
00:03:12	Repeat the word		
00:03:15	Assistance – instructing		Help
00:03:18	("Huh")		Puzzled
00:03:24	("Tak dengar")		Puzzled
00:03:30	Looking on the screen	Recording starts	Puzzled

Figure 1 Interaction Table

Once coding was completed, the occurrence of each indication observed from each participant was keyed in separately in Microsoft Excel document. The goal is to determine which aspects (animation, recording, and feedback) are more likely to cause usability and fun problems to the children. Therefore, each indication was organized according to Animation, Recording, and Feedback aspect of *MyCard*.

VI. Result

The overall score of usability and fun problems was done by simply counting all the breakdown indications observed from each participant as shown in Figure 2. The result shows that Participant 1 (P1) and Participant 2 (P2) have indicated more usability and fun problem compared to Participant 3 (P3). Most of the usability and fun problems among P1 and P2 were observed during the recording scene. This indicates that the recording scene appeared being most trouble for P1 and P2.

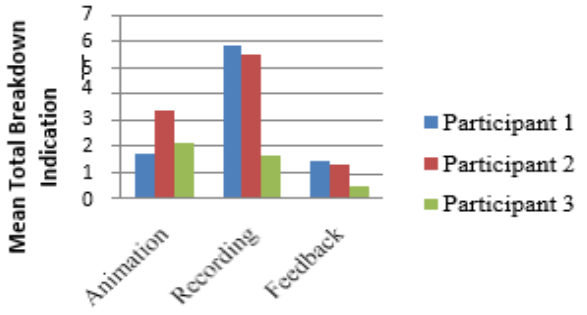

Figure 2 Total Mean Usability and Fun Problem

The breakdown indications like *Puzzled* and *Passive* were also found frequently occurred during recording scene.

In video analysis, it was observed that the participants could not repeat the word when the application did not provide an audio display during recording of the scene. Consequently, the children remained passive by staring or pointing to the scene during the recording scene.

Analyzing the breakdown indications in the animation scene was most confusing. For example, coding the *passive* indication was confusing when the children were watching the animation. In the video analysis, the children did not necessary imitate the animation displayed – except P1 who always imitate the animation. Therefore, the researcher found it was hard to determine whether the children had usability and fun problems during the animation scene.

In some occasions, the children imitated the word while animation was playing but they made various and inconsistent pronunciation errors. As a result, the teachers offered assistances to the children during the animation scene. This is problem associated to the children's articulation skill which is not relevant to the usability and fun problems of the *MyCard*. The children showed less usability and fun problems during the feedback scene. In the video analysis, the children often could not pronounce the word precisely during the recording scene which resulting in low score. Meanwhile, the children responses indicated *dislike* or *passive* behaviors.

Screen	Description
	Menu scene – User selects the word he or she like to practice in this screen. Audio – No audio displays in this scene.
	Animation scene – The screen displays the animation and pronunciation of the word selected. The color of text will change as the animation is playing to highlight the pronunciation of the syllable. Audio – The pronunciation of the word, e.g. "A-dik".
	Recording scene – The scene contains three icons programmed to provide different funtion. This screen allows the user to record his or her voice by clicking the microphone icon. The mouth icon will playback the animation and pronunciation of the word selected. The tick/cross icon will analyze the pronunciation of the user after the user had made the pronunciation during the recording. Audio – No audio displays in this scene.
	Recording scene – After the microphone was clicked, the listen sign will start to play by putting her hand behind the ear. The ear of the agent will also start to animate for five seconds. Audio - No audio display in this scene.
	Feedback Scene – This scene displays the score of the user. One star indicates low score; whereas, five stars indicate high score. Auditory feedback like 'uh-oh' and applauding are used to reflect low and high score respectively. Audio – Applauding for high score (4 to 5 stars), and warning sound ('uh-oh') for low score (1 star).

Figure 3 *MyCard*

VII. Discussion

Based on the adapted coding scheme, we found that the recording scene is the most difficult task in *MyCard* for the children with DS. This happened mainly due to the fact that the children were confusing with the task required during the recording scene. In developing an application for children with special needs, the interface should be easy enough in order to encourage the children in a simple and efficient way. If this did not happen, the children can get confuse which hindered their interaction with the application.

The main drawback of the recording scene is that the interface does not provide audio display. In the video analysis, it was observed that the children could not associate the displayed 'listen sign' on the screen. It also confusing because there is no sound was played with animation. As a result, the children stopped interacting with *MyCard* during the recording scene.

The combination of visual and audio display seems having an impact on the way the children interact with *MyCard*. For example, in the animation and the feedback scene, it was observed that the children were more likely to get attracted by audio and visual feedback when both feedbacks were displayed together. This indicates that audio and visual display together can stimulate the curiosity and desire of the children to interact with the application.

Grounded on these, the recording scene of *MyCard* needs to be improved by adding a sound. For instance, repeating the word's pronunciation when the 'listen sign' is playing tends to stimulate the children to imitate the word. In addition, to this, another way to improve the usability and fun problems of the recording scene is by adding a real-time audio or visual feedback when the children pronouncing the word [9]. For instance, when the children start pronouncing a word, a visual effect is display on real-time from the scene. As the feedback is displayed on real time, it tends to capture the interest of the children when he or she is interacting with the application. Therefore, audio and real-time feedback displays can help to improve the usability and fun problems particularly for children with DS in this study. Figure 4 illustrates the summary of the findings from the activities comparisons.

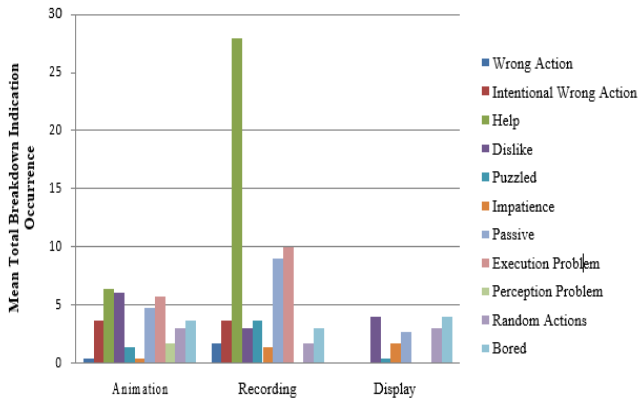

Figure 4. Usability and Fun Problems of *MyCard*

Acknowledgment

This research was funded under FRGS (Project No: FRGS/2010/FTMK/TK-02/4-F0090) and this paper was funded under FRGS (Project No: FRGS/2/2013/ICT02/FTMK/02/5/F00188) by Ministry of High Education.

References

[1] M. Bray, "Speech production in people with Down syndrome," *Down Syndrome Research and Practice*, 2008. [Online]. Available: http://www.down-syndrome.org/reviews/2075/. [Accessed: 12-Jul- 2011].

[2] J. Feng, J. Lazar, L. Kumin, and A. Ozok, "Computer Usage by Children with Down Syndrome," *ACM Trans. Access. Comput.*, vol. 2, no. 3, pp. 1–44, Mar. 2010.

[3] L. K. Hoe, S. Hisham, C. Technology, U. Teknikal, and H. T. Jaya, "Self-paced Speech Rehabilitation for Down Syndrome Using Malay Automatic Speech Recognition (ASR) Center of Advanced Computing Technologies (C-ACT)," in *Proceedings of Malaysian Technical Universities Conference on Engineering & Technology (MUCET)*, 2013, no. December, pp. 3–4

[4] I. Macedo and D. G. Trevisan, "A Method to Evaluate Disabled User Interaction: A Case Study with Down Syndrome Children," pp. 50– 58, 2013.

[5] A. P. O. S. Vermeeren, K. den Bouwmeester, J. Aasman, and H. de Ridder, "DEVAN: A tool for detailed video analysis of user test data," *Behav. Inf. Technol.*, vol. 21, no. 6, pp. 403–423, Jan. 2002.

[6] W. Barendregt and M. M. Bekker, "Developing a coding scheme for detecting usability and fun problems in computer games for young children.," *Behav. Res. Methods*, vol. 38, no. 3, pp. 382–9, Aug. 2006.

[7] Y. Ma, J. Feng, L. Kumin, and J. Lazar, "Investigating User Behavior for Authentication Methods," *ACM Trans. Access. Comput.*, vol. 4, no. 4, pp. 1–27, Jul. 2013.

[8] "Heuristic Evaluation: How-To: Article by Jakob Nielsen." [Online]. Available: http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/. [Accessed: 04-Aug-2014].

[9] J. Hailpern, "Encouraging speech and vocalization in children with autistic spectrum disorder," *ACM SIGACCESS Access. Comput.*, no. 89, pp. 47–52, Sep. 2007

Author Biographies

Lau Kum Hoe is deep learning enthusiast who loves to build personalized apps including mobile and web. His latest mobile app allows the music lovers to download and play their favorite songs provided by the music provider. In spare time, I'm working on some R&D works related to Deep Learning. Skills: Dart, Flutter, Python, PyTorch, JavaScript, PHP, SQL.

Syariffanor Hisham currently works at the Centre of Advanced Computing Technology, Universiti Teknikal Malaysia Melaka (UTeM). Syariffanor does research in the area of Human-computer Interaction focusing on people with special needs and Computing in Social science, Arts and Humanities. Her current works mostly in the area of brain computing and the uses of ICT for children/adult with specific learning difficulties such as dyslexia and ADHD.

Noor Azilah Muda received her B.Sc. and M.Sc. degree in 1998 and 2001, respectively. She is currently pursuing her PhD in pattern recognition area focusing on music analysis. She is now serving the Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia as a Senior Lecturer in Software Engineering Department. Her current research interests include deep learning, artificial immune system