UNIVERSITI TEKNIKAL MALAYSIA MELAKA

Faculty of Information and Communication Technology

THE INTEGRATION OF SPEECH RECOGNITION TECHNIQUE IN ROLE-PLAYING COMPUTER GAME FOR DOWN SYNDROME CHILDREN MANDARIN LEARNING

Lim Chin Chuan

Master of Computer Science (Multimedia Computing)

2017

THE INTEGRATION OF SPEECH RECOGNITION TECHNIQUE IN ROLE-PLAYING COMPUTER GAME FOR DOWN SYNDROME CHILDREN MANDARIN LEARNING

LIM CHIN CHUAN

A thesis submitted
in fulfillment of the requirements for the degree of Master by Taught Courses
(Multimedia Computing)

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2017

DECLARATION

I declare that this thesis entitled "The Integration of Speech Recognition Technique in Role-Playing Computer Game for Down Syndrome Children Mandarin Learning" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature   :   …………………………………..

Name   :   …………………………………..

Date   :   …………………………..………

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in term of quality for the award of Master by Taught Courses (Multimedia Computing).

Signature   :   ……………………………..

Supervisor Name   :   Dr. Ahmad Naim Che Pee   .

Date   :   ………………...…………

DEDICATION

This thesis is dedicated to Universiti Teknikal Malaysia Melaka, which is a fabulous University in Malaysia. I have really learnt a lot from the gracious academic staffs of this University, I have also received helps and guidance from the Faculty of Information and Communication Technology and Centre for Postgraduate Studies staffs who are extremely helpful and zealous in their career.

ABSTRACT

Down syndrome individuals are known to have difficulties in speech, both pronunciation of words and making sentences. Such problem is caused by the innate physical impairment of their mouth muscles, as well as the cognitive and intelligence impairment. The myths about bilingualism in down syndrome children causing adverse effects on their language development have been debunked, showing that learning more than one language is possible for them without negative consequences. Many applications and courseware have been designed for down syndrome individuals to enhance their language learning experience. Such applications and courseware usually targeting words recognition and reading therapy, but the syntax and expressive language training are rarely focused. In this research, a 2-Dimensional role-playing game is integrated with state-of-the-art speech recognition technology to provide assistive training on Mandarin syntax and short sentences uttering for down syndrome children. The cognitive capabilities of down syndrome children are considered, and the multimedia application design standards for special education are followed. The research outcome is a game designed according to the Waterfall system development life cycle, while the educational contents are planned with hybrid ASSURE and Gerlach-Ely Instructional design model. The resultant game prototype is tested by five typically developing children age between five years old to six years old, and five adults. The results show that majority of the participant shown positive emotion playing the game, which implies that using RPG for speech practices is feasible. Since only the mildly impaired down syndrome children are able to read and speak, this research is intended to cater for the children under this category. Similar emotion trends are expected from the down syndrome children, which means that they would be motivated to speak with better grammar and remember to use the grammar correctly in their daily lives, as the RPG provides visual context for the sentences they speak.

i

ABSTRAK

Adalah diketahui bahawa individu-individu yang dilahirkan dengan sindrom down meghadapi masalah pertuturan, dari segi sebutan perkataan serta pembuatan ayat. Masalah tersebut adalah diakibatkan daripada kesan kecacatan fisikal otot-otot mulut, kecacatan kognitif serta kecacatan kecerdasan. Mitos-mitos berkaitan dengan kesan-kesan buruk yang disebabkan oleh bilingualisme di antara budak-budak sindrom down telah dibukti palsu, dan ditunjukkan bahawa mempelajari lebih dari satu Bahasa adalah boleh dilakukan tanpa sebarang kesan negatif. Banyak aplikasi dan perisian kursus telah dihasilkan untuk membantu individu-individu sindrom down meningkatkan pengalaman pembelajaran Bahasa mereka. Aplikasi-aplikasi serta perisian kursus yang terhasil itu biasanya menyasarkan pengecaman perkataan dan rawatan pembacaan, malah sintaksis dan latihan bahasa ekspresif oleh individu-individu sindrom down jarang diberi tumpuan. Penyelidikan ini menghasilkan suatu RPG 2-dimensi yang menyepaduan teknologi pengecaman pertuturan untuk memberi latihan bantuan bagi melatih sintaksis Bahasa Mandarin and membaca ayat-ayat pendek kepada kanak-kanak yang menghadapi sindrom down. Keupayaan kognitif kanak-kanak sindrom down telah diambil kira, piawaian reka bentuk aplikasi multimedia juga diikuti. Hasil penyelidikan ini adalah satu permainan digital yang dihasilkan mengikuti kitaran hayat penghasilan system "Waterfall" dan model "ASSURE" serta model "Gerlach-Ely" untuk mereka bentuk pengajaran. Penyelidikan ini bermula dengan wawancara awal dan kajian sastera, dan meneruskan dengan kaedah sains reka bentuk. Prototaip yang dihasilkan itu telah diuji oleh lima orang kanak-kanak yang berkembang secara biasa yang berumur lima hingga enam tahun, serta lima orang dewasa. Keputusan yang didapati daripada ujian yang dijalankan itu merujukkan bahawa kebanyakkan penguji-penguji menunjukkan emosi yang positif terhadap permainan tersebut. Oleh itu, ia membayangkan bahawa RPG boleh digunakan dalam latihan pertuturan. Oleh sebab hanya kanak-kanak yang mempunyai sindrom down yang ringan sahaja boleh membaca dan bercakap, penyelidikan ini adalah bertujuan untuk memenuhi keperluan kanak-kanak di bawah kategori ini. Kanak-kanak yang mempunyai sindrom down dijangka menunjukkan trend emosi yang serupa dengan kanak-kanak yang berkembang secara biasa. Maksudnya, kanak-kanak ini akan bermotivasi untuk bercakap dengan tatabahasa yang lebih tepat dan boleh mengingati tatabahasa tersebut dalam kehidupan harian mereka, kerana RPG membekalkan konteks visual bagi ayat-ayat yang mereka baca dengan suara.

ACKNOWLEDGEMENTS

I would love to express my deepest gratitude to my beloved lecturers, Dr. Ahmad Naim Che Pee for his patience, motivation, and immense knowledge in providing guidance which helped me in all the time writing of this proposal. I sincerely thank Dr. Ahmad Naim for his insightful comments and encouragement.

I like to express my regards to the interviewees and participants who have been very cooperative and enthusiastic during the preliminary interviews and prototype testing, that have really been a great help for the completion of this research.

I must also give my very profound gratitude to my fellow classmates for providing me with continuous encouragement and sharing their experience with altruism through the period of my study.

Finally, my uttermost love goes to my family, for supporting me in every possible aspect throughout my life.

The research would have not been completed without the art-works contributors, who have either provided the art-work to the public for free, or provided ideas for the graphics design. The full list of the contributors can be found in Appendix A.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D   -   2-Dimensional

3D   -   3-Dimensional

ANN   -   Artificial neural networks

ASR   -   Automatic speech recognition system

CER   -   Character error rate

DBN   -   Deep belief networks

HMM   -   Hidden Markov Model

ID   -   Instructional Design

KSPK   -   National preschool curriculum standards

LVCSR   -   Large-vocabulary continuous speech recognition

MER   -   Mixed error rate

NN   -   Neural network

NNLM   -   Neural network language model

RPG   -   Role-playing game

SDLC   -   System development life cycle

UI   -   User interface

WER   -   Word error rate

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.0 Introduction

Language learning can be quite challenging for children with down syndrome (Buckley, 1993; Laws and Hall, 2014; Næss et al., 2015). Buckley (1993) points out in her report that language development among these children is usually delayed. Their vocabulary acquisition rate falls between the range of average to slow, and they hardly use good grammar. However, their understanding to the conversational skills seems to be quite well (Peskett and Wootton, 1985; Coggins et al., 1993). Due to the fact that language development among children with down syndrome is not as well as typically developing children, many parents and even professionals would not expect these children to develop a bilingual ability (Buckley, 2002). Therefore, many children develop better language skill in only one language that they speak and practice in rehabilitation centre. However, many research concludes that bilingualism among children with down syndrome is possible and no adverse effects on their language development have been proven, in fact, they perform just as well as their monolingual down syndrome counterparts (Kay-Raining Bird et al., 2005; Cleave et al., 2014; Burgoyne et al., 2016). Which indicates that children born in a multilinguistic country, such as Malaysia, may practice speaking in their own mother tongue with their family members. It is important to note that the mean length of utterance was shorter for down syndrome children, and qualitative analyses revealed expressive language difficulties among them (Feltmate and Kay-raining Bird, 2008). Therefore, if a language

practising system is to be developed, it should aim to encourage the children to speak more often with better grammar.

For special education, integrating multimedia elements in the instructional design (ID) can significantly enhance the learning experience for the children (Khan and Bayoumi, 2015). Multimedia integration in instructional design is therefore one of the most commonly seen approach in courseware developed for this purpose (Husni, 2013). Without reinventing the wheel, multimedia presentation method should consistently be used in courseware or other systems that provide speech training so that the children are not bored and discouraged with the practice. Ng et al. (2014) reviewed three multimedia courseware designed for down syndrome children targeting different areas of learning, two out of three courseware reviewed are specifically design for reading purpose, while another one focuses on numerical skills among Down syndrome children. Nevertheless, the two courseware that focus on reading have only covered three out of four essential skills for language learning, which including reading, writing (typing), and hearing. Still, speaking being one of the fundamental skills in language learning often ignored by researchers and hardly seen in literature. Research can integrate speech recognition together with other multimedia elements to provide a better language learning approach.

As the speech recognition technology advanced, the accuracy of Automatic Speech Recognition system now compare favourably with humans though not reaching the human parity (Graves et al., 2013; Xue et al., 2016; Saon et al., 2017). However, speech recognition in gaming industry is still rarely seen in both serious and conventional games due to the limitation of the underlying technology of speech recognition which constrains its usage in digital games, as a delay in speech recognition occurs to transcribe utterance into text, making it unsuitable to be used for fast-paced games. In contrast to that, a courseware or learning system catered for down syndrome children must be in a slower manner so that

2

the core of learning, which is the knowledge or skill, can be conveyed thoroughly, Therefore, with the current accuracy achievement of speech recognition technology, assumption for the feasibility of using such technology in language learning system can be made. The study by Buckley (1993) suggests that the down syndrome children are more difficult to use proper grammar than the ordinary children, which indicates that the application to be developed must guide the children to speak short sentences, not only individual words.

1.1 Background of the problem

Down syndrome is the most frequent chromosome abnormality in new-borns and the most common cause of mental retardation (Blanco et al., 1998). Children with down syndrome have cognitive disabilities and language delays resulting from trisomy of the 21st chromosome (Cleave et al., 2014). Cleave et al. (2014) suggest in their paper that other research have refuted the position of some interventionists who claimed that language input should be restricted to one language for children with intellectual challenges, in fact, evidence shows that variety of cognitive and linguistic advantages can be resultant from bilingualism in typically developing children. Especially in multilinguistic countries such as Malaysia, learning more languages can mean better communication with family members and others, which can in turn lower their frustration for unable to be understood and increase their self-confidence. Næss et al. (2015) in their paper, suggest that early language intervention should be given high priority as study shows that chronological age significantly correlated with language skills. Therefore, this study will focus on simple Mandarin at levels commensurate with the children's chronological age, as Mandarin being spoken by a significant number of population in Malaysia, but rarely any rehabilitation centre is teaching Mandarin to the children.

Learning begins with student engagement, which in turn leads to knowledge and

understanding (Alonso et al., 2005). This statement suggests that without student engagement in study, learning does not take place. People have been working on this problem for an unknown period of century (Reiser, 2001). "A system of procedures for developing education and training programmes in a consistent and reliable fashion", called instructional design as defined by (Gustafson and Branch, 2002), was produced along with the efforts. Student engagement has been defined as "participation in educationally effective practices, both inside and outside the classroom, which leads to a range of measurable outcomes" (Kuh et al., 2007), and as "the extent to which students are engaging in activities that higher education research has shown to be linked with high-quality learning outcomes" (Krause and Coates, 2008). The Glossary of Education Reform (2016) on the other hand, gives a clearer meaning of student engagement, by referring it to the degree of attention, curiosity, interest, optimism, and passion that students show when they are learning or being taught, which extends to the level of motivation they should learn and progress in their education (Abbott, 2016). The article also suggests that the view on "student engagement" may be defined or interpreted differently from place to place by educators. Multimedia elements which including graphics, videos, animation, sound effects, and text are being integrated in ID to hook students' interest in learning a subject, resulting in widely adopted serious games.

Speech recognition technology has been a topic of great interest to a broad general population since it became popularized in several blockbuster movies of the 1960's and 1970's (Juang and Rabiner, 2005). As stated by Juang and Rabiner, the first system of voice recognition based on isolated digits' representation was built in the 1950s. In the 1970s, IBM took the initiative to create a "voice-activated typewriter" that can transcribe the spoken language into apparent words, which leads to the advanced in language models such as n-gram and Hidden Markov Models (HMMs) that have the core algorithms applied in the statistical modelling (Fung and Schultz, 2008), which are the oldest language models but

4

remain popular. Today speech technologies are commercially available for an interesting range of tasks. These models were trained with thousands of hours of quickly-transcribed audio data and hundreds of millions of words of monolingual text data or millions of words of bilingual text corpora for robust and reliable parameter estimation, which eventually enable machines to respond correctly and reliably to human voices, and provide useful and valuable services.

Speech interaction is seen in digital games and can be tracked back to year 1960s, parallel to the development of voice recognition technology, parallel to the development of voice recognition technology (Allison et al., 2016). In Allison's paper, digital games with voice interaction enabled is said to become ubiquitous in the 1980s. Today, the games featuring speech recognition in language learning are seen in the pronunciation application such as SpeechAce. A recent study developing the "HATLE" courseware integrates speech recognition in pronunciation learning for children with down syndrome shows the feasibility of speech recognition for speech therapy and education (Felix et al., 2016). While true immersion of a player within a game can be achieve when the game behaves closely to reality (Reale et al., 2013) and context is important to the children (Mohid and Mat Zin, 2010; Aziz et al., 2011), this research will focus on the speech recognition in RPG for special education to enhance the students' engagement.

Other than enhancing students' engagement, RPGs also suitable to be used for language learning in special education because RPGs do not have winners or losers in the traditional sense of the terms (Phillips, 1994), which is important to build confidence of the down syndrome children to keep on trying and learning (Yussof et al., 2010). Phillips also points out that RPGs can adopt almost any settings be it detective, fantasy, western, etc., which is useful for the development of language learning application based in different context. Many other reasons why RPGs should be considered as a tool for language learning

5

have been addressed by (Cornillie et al., 2012). However, there are some concerns with the current speech recognition technology in RPGs, most often described as unnatural or uncomfortable (Allison et al., 2016). Through the games available in the market with voice control capability, it is demonstrated that the concerns are the relatively low command-to-action transcription rate, and the response upon valid voice inputs is relatively slow to achieve satisfactory for games of certain genres. Since the RPG theme used in this research is to bring real life context into language learning for children with down syndrome, the game will be fabricated with slower-paced mechanism that is suitable for learning, therefore the impact of the concerns may not apply to this research. However, proper design of the game system also required so that it does not result in frustration and diminishing the students' engagement.

1.2 Problem Statement

Many game-based applications have been developed to help children with down syndrome to read and recognise words, such as MEL-SinD developed by Yussof et al. (2010), AR BACA SindD developed by Ramli and Zaman (2011) and MyCard developed by Lau (2015). However, the applications existed rarely scratch the grammar quotient of the down syndrome children. Down syndrome children have communication problem, not only that the language development among them is delayed, but they can hardly speak with good grammar (Galeote et al., 2013; Burgoyne et al., 2016; Witecy and Penke, 2017). Today, the grammar illiteracy problem among the down syndrome children still prevails, because the practice by down syndrome children to speaking sort sentences is always neglected by educators (preliminary interview, 2017). Other than the shortage of speech therapist in down syndrome centres (Lau, 2015), the stagnated speech recognition advancement in digital games (Allison et al., 2016) is also hindering the down syndrome children from learning

6

through playing. The lags in speech-to-command transcription (Navarro-newball et al., 2014) and highly-diverge speaking style (Buckley, 1993;   Zhang and Ng, 2013) among the down syndrome children, such as making serious grammar mistakes, increase the difficulty of speech recognition. Dehghan et al. (2014) says that "visual memory" has more influence on grammar quotient of the down syndrome children, unfortunately, little attention has been given in literature to integrate speech recognition in RPGs that brings visual context to language learning for down syndrome children. Nonetheless, the down syndrome children must have the ability to comprehend written text and the game developers must understand what presentation manners and learning materials are suitable for these children.

1.3 Research Questions

The research questions are listed as follow:

- How to improve expressive language of down syndrome children?

- How to improve grammar quotient of down syndrome children?

- Is current speech recognition technology suitable for gaming environment?

- How to encourage the use of educational game for down syndrome children?

1.4 Research Objectives

The main purpose of this research is to identify whether speech recognition in RPG is suitable for motivating down syndrome children to speak more often with better grammar by encouraging them to speak in short sentences. This research strives to employ the standards for developing multimedia courseware for special education into RPG to determine whether that helps them to be motivated and to feel comfortable playing the game. Finally, this research is to evaluate the player's feedbacks to determine the room for improvement pertaining to using game for speech practises.

7

1.5 Definition of Terms

The term *cognitive psychology* in this proposal is used to mean "the branch of psychology studying the mental processes involved in perception, learning, memory, and reasoning". The term *courseware* in this proposal is used to mean "educational software designed especially for use with classroom computers". The term *instructional design* in this proposal is used to mean "a field of study that marries education, psychology and communications to create effective teaching plans for groups of students". The term *multimedia* in this proposal is used to mean "using a combination of moving and still pictures, sound, music, and words, especially in computers or entertainment". The term *professionals* in this proposal is used to mean "the educators and researchers working in the down syndrome fields". The term *student engagement* in this proposal is used to mean "the degree of attention, curiosity, interest, optimism, and passion that students show when they are learning or being taught". The term *system* in this proposal is used to mean "information system developed to serve specific purpose". The term *verbal memory* in this proposal is used to mean "the ability to recall something penned or spoken which was already learned". The term *visual memory* in this proposal is used to mean "the ability to visually recall images that have formerly been observed". The term *working memory* in this proposal is used to mean "the part of short-term memory which is concerned with immediate conscious perceptual and linguistic processing".

1.6 Significance of the study

This section will provide brief description on the various significances of the study given the three categories, education, entertainment and social.

Education: The main purpose of this research is to enhance the language learning of down syndrome children. This study complements the state-of-the-art speech recognition

© Universiti Teknikal Malaysia Melaka

technology in RPG as a serious game to improving the students' engagement in learning Mandarin. The participants' emotion is observed and the results are shown in Chapter 4.

Entertainment: Although the purpose of this research inclines more toward language learning for down syndrome children, the technique to integrate natural language recognition in RPG can be extended to normal games to achieve a more natural way of gaming experience, as described in Chapter 1 of the thesis.

Social: The research focuses on RPG with slow mechanism that is suitable for down syndrome children to use, which in turn beneficial for the children with motor difficulties to play with. It has been suggested that playing games has positive mental and health effects to those with physical disabilities, which may help the children to build their social confident and prepare them to their later education.

1.7 Organisation of Thesis

Chapter 1:

This chapter introduces the research subject giving the background for language learning among down syndrome children, instructional design approaches for special education, and automatic speech recognition (ASR) in RPGs for language learning. The objective of this research is also stated in this chapter.

Chapter 2:

This chapter reviews the previous research that have been conducted on different aspects of work, and properly analysed to be aligned with this study. The speech recognition technology is first reviewed, follows by cognitive psychology of down syndrome individuals, instructional design models, system development life cycle and multimedia interface design.

This chapter also presents the findings or contribution by different researchers and deficiencies therein.

Chapter 3:

This chapter presents a brief description of the research methodology used in this research, such as research design, research approach, data sources, data collection techniques, and analytical techniques. The research methodology presented in this chapter is rooted in the previous studies and their achievements. The overall fabrication of the research methodology follows the flow of the literature reviewed.

Chapter 4:

This chapter presents the results from preliminary interview for learner analysis, the children's emotion towards the game for evaluation, and the feedback from adults to determine room for project improvement.

Chapter 5:

This chapter evaluate and summarise the results in Chapter 4, with explanations attempted for the results' pattern.

Chapter 6:

This chapter points out the limitations of this research and the research gaps it did not cover. The recommendations for future research are provided together with the contribution of this study. Finally, a general conclusion summarises the entire research.