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.

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 menyepadukan 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

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

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

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

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

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.

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

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.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

Different theories have been proposed for the development of special educational application or courseware. This review will focus on five major themes which emerge repeatedly throughout the literatures reviewed, that have significant impacts on the study. These themes are: speech recognition technology, cognitive psychology, instructional design models, system development life cycle (SDLC), and multimedia in special education. Although the literature presents these themes in a variety of contexts, this report will primarily focus on their application on game-based language learning for down syndrome children.

2.1 Literature

2.1.1 Speech Recognition Technology

Speech recognition technology is the fundamental of this research. The review of speech recognition technology in this study addresses the fundamental theories of ASR including the language models and ASR system training, follow by the literature review on google speech recognition system and its recent accuracy achievement on Mandarin recognition since it is available for public use under Webkit website browsers at no extra costs while being one of the most advanced in speech recognition in terms of utterance to text transcription accuracy. The review is to show the viability of integrating current speech

recognition technology in game-based educational application.

When it comes to automatic speech recognition, a few theories were proposed, such as the artificial neural networks (ANN), Hidden Markov Models (HMMs), n-gram method and deep belief networks (DBNs) (Arısoy et al., 2012; Adel et al., 2013; Sarikaya et al., 2014; Wei et al., 2016; Xue et al., 2016). According to Arısoy, recent studies have shown larger interest in neural network language models (NNLM) since they show superiority over the conventional n-gram language models in both perplexity and word error rate. Wei et al. however, suggest that ANN for speech recognition are more reliable than HMMs in terms of classification performance.

ASR system training refers to the training of ANN in machine learning. Training can be divided into supervised and unsupervised training in which both inputa and outputa are provided for supervised training, whereas only inputs are provided for unsupervised training. Deep multi-layer neural networks have many levels of non-linearities allowing them to compactly represent highly non-linear and highly-varying functions (Larochelle et al., 2009). Huang et al., (2013) in their study on multilingual deep neural network uses unsupervised pretraining method as it does not involve the language-specific softmax layer and therefore can be carried out easily without any modification of the existing tool.

Similar to other big names in speech recognition innovation such as Microsoft, Google adopts multiple neural network acoustic models in their speech recognition system (Xiong et al., 2016; Protalinski, 2017). Xiong et al. (2016) claim that with their hybrid model trained on Switchboard corpus, they achieve 5.8 percent word error rate (WER), reaching the human parity. While Protalinski (2017) reported that Google has achieve 4.9% WER in 2017. However, their achievement with Mandarin corpora is unclear. An evaluation performed on the Microsoft Chinese search shows a average character error rate (CER) around 14 percent (Ye et al., 2016). Ye et al. suggest that heavy accent has impact on the

CER, e.g. the Guangdong accent causes the CER to rise above 14.4 percent. It is not known whether the enunciation difficulties among the down syndrome children is equally comparable to the regional accent. For recognition of children's voice, a study compared several ASR engine including Google and Microsoft's (Kennedy et al., 2017), found that Google ASR always out-perform other engines. Therefore, Google ASR is anticipated to work best in the game development of this study.

2.1.2 Cognitive Psychology

Understanding the cognitive psychology of down syndrome children is vital for the development of multimedia learning systems that gear towards these children. The review of cognitive psychology in this study focuses on the working memory of the down syndrome children, which is the one mechanism that contributes to acquisition and processing of language (Gathercole and Baddeley, 2014), so that the method of context presentation that is suitable to be use by these children can be understood, and therefore producing an efficient language learning application, for the case of this research, a speech practising game.

Deficit of verbal working memory among down syndrome children may contribute to the delay in their language development (Chapman and Hesketh, 2001; Dehghan et al., 2014). Working memory involves the temporary storage and manipulation of information for complex cognitive activities (Baddeley, 2003). As mentioned in Baddeley's paper, the working memory can be further divided into phonological loop - verbal memory, visuospatial sketchpad - vision memory, episodic buffer and central executive. For information gaining such as learning a language, phonological loop and visuospatial sketchpad are involved before the information being processed by the episodic buffer and central executive. Baddeley (2013) also suggests that children with good immediate verbal memory are proved to be better at language learning than those with short spans, measured

by both vocabulary and acquisition of syntax or grammar. Unfortunately, it is also suggested by other researchers that the verbal abilities in the down syndrome population are usually lower than visuospatial abilities (Chapman and Hesketh, 2001; Lanfranchi et al., 2009; Dehghan et al., 2014). Therefore, visuospatial sketchpad must be emphasized as an alternative approach for the acquisition of syntax.

Visuospatial sketchpad can be divided into three subsystem – visual, spatial-sequential and spatial-simultaneous (Mammarella et al., 2008; Frenkel and Bourdin, 2009; Lanfranchi et al., 2009). Visual memory is the ability of the brain to remember what is presented to the eyes, spatial-sequential memory is the brain's ability to remember the presented visual details in correct sequence, whereas spatial-simultaneous is the brain's ability to remember two or more visual details presented at the same time. To examine the performance of individuals with down syndrome on specific visuospatial memory components, Lanfranchi et al. (2009) compared 34 typically developing children with 34 down syndrome children on four visuospatial working memory tasks: in two of these are spatial-sequential, and in another two spatial-simultaneous. They concluded that it might be the brain's processing speed that affect the performance of individuals with down syndrome on spatial simultaneous tasks, that is, lower ability to move quickly between different points to include details in their working memory.

Although must be emphasized, it is important to note that visuospatial component of working memory is selectively impaired in individuals with down syndrome (Pulina et al., 2015). Frenkel and Bourdin (2009) show in their study that down syndrome children performed better in spatial-sequential working memory than the typically developing children before the age of 6. The study by Lanfranchi et al. (2009) shows similar results. Mammarella et al. (2008) point out that visuospatial working memory measures reach an asymptote at about 11 years and do not develop much further in adulthoods. Other than

utilising the spatial-sequential working memory, parents can lead their down syndrome children in spatial-simultaneous working memory training, which shows positive effects on memory performance (Pulina et al., 2015).

By understanding that down syndrome children are slower in brain's processing speed, the feasibility of using RPG for speech practice is further justified. This is because with RPG, the down syndrome children can see the sequence of their daily situations and remember to use correct sentences in that particular situation. For example, events in the RPG are always triggered by other events. For a speech practising RPG, is the sentences uttered will trigger an event. With the written guiding sentences, the down syndrome children can remember the sentences and grammar associated with one situation to another. The options enabled to be chosen in RPG allow the parents to lead their children in spatial-simultaneous working memory training. For instance, in the bathroom, the option allowed are to bathe and to brush teeth. These two actions are presented in simultaneous fashion to be chosen from. With more practices, the down syndrome children can become more familiar with choices in their daily lives and they can recall the sentences to say.

2.1.3 Instructional Design Models

As briefly introduced in Chapter 1, ID is important to ensure that the contents of certain learning materials are being delivered to the students. Instructional design models on the other hand, is defined by Silvern (1960 cited in AECT, 1977) who was one of the pioneers in ID models (Gustafson and Branch, 2002), as "graphic analogue representing a real-life situation either as it is or as it should be". The review of instructional design models in this study focuses on the ID models for special education, which is to address the systematic process for contents delivery.

According to Reiser and Dempsey (2012), all ID models practically contains the core

elements of the ADDIE process, wherein ADDIE is an acronym for analyse, design, develop, implement, and evaluate. However, despite being the core of all ID models, it has been criticised for being outdated, ineffective and inefficient in leading to the best instructional solutions (South, 2008). While there are many ID models to choose from, some educators still prefer the use of ADDIE framework in their design for down syndrome education, such as the AR BACA SindD developed by Ramli and Zaman (2011). This section briefly describes three popular ID models apart from ADDIE framework that have either been based or with potential for special education, which are ASSURE model as adopted by Hassan et al. (2017), Keller's ARCS model as adopted by Yussof et al. (2010), and Gerlach-Ely model as adopted by Osman and Lee (2014) for its flexibility to include media (Isman et al., 2005). The models that are adopted in this research is later summarized and further discussed at the end of this section.

The ARCS Model was developed by Keller for diagnosing problems associated with learning motivation based on the correlation between performance and learning motivation (Huang et al., 2004). The four elements of this model are attention, relevance, confidence and satisfaction, which are said to be effective at enhancing motivation of down syndrome children (Yussof et al., 2010). According to Huang et al. (2004), Keller's primary ARCS Model is based on the interaction between instructional materials and learners. Attention refers to the learner's response in perceiving the provided instructional materials. Relevance helps learners associate their prior experiences with the instructional materials being presented. Confidence stresses to provide meaningful experiences during the learning process to build positive expectation in the learners towards the learning task. Satisfaction is achieved when learners are can use their newly acquired knowledge or skills and receive feedback that leads to positive attitudes towards the learning task. Keller's ARCS model of motivation is more of an aim to achieve rather than being a model. The development of the

RPG for down syndrome children should have the attributes of the ARCS model, which can be achieved by proper game planning and design.

The ASSURE Model is said to be the most suitable and most convenient model for integrating the theories of education technology and research with practice, as it focuses on instruction goals achievement by selecting and making the best multimedia tools, as well as by encouraging the learners to interact and participate (Megaw, 2006; Chen & Chung, 2011 cited in Karakis et al., 2016). ASSURE is the acronym for analyse learners, state objectives, select media and materials, utilise media and materials, require learner participation, and evaluate and revise. Karakis et al. show that a computer game activity design using ASSURE and ARCS models has positive effect on learners' attitude. For this research, the main focus is on the usability of the game rather than emotional contact with the learners, there for it might be slightly modified to be used as a hybrid with the Gerlach and Ely Model.

Different from the two models mentioned above that are sequential processes, Gerlach and Ely Model follows a more flexible sequential pattern that incorporates a parallel process in its second phase. The phases of this model comprise of three parts - design, develop and evaluate. The original model was to explain each element of teaching and learning while establishing a relationship to the media of instruction (Grabowski, 2003). As the Gerlach and Ely Model focuses more on the instructional materials and resources without identifying the instructional problems (Lee, 2011), it is rarely seen to be implemented alone, but in a hybrid system. Lee (2011) suggests that the combination of Gerlach and Ely Model with the Kemp model can overcome the weaknesses of both models. In other words, using hybrid ID models is possible and may produce better results at delivery of knowledge or skills. Therefore, this research adopts both ASSURE and Gerlach-Ely Model in hybrid form.

Before modifying the ASSURE model for integration with Gerlach and Ely Model, the elements of this model should be understood so that those element that are suitable to be

used for the development of RPG can be selected and utilised. ASSURE model follows a logical flow of its acronym, starting from "Analysis", which is the study of learners' characteristics, including their ages, learning style, prior knowledge or entry competencies, and learning abilities (Megaw, 2001; Faryadi, 2007). A general analysis of these characteristics can provide some guidance to the designers in selecting instructional methods and media. In this research, for example, acknowledging the cognitive abilities of down syndrome children and their physical capabilities are vital prior to selecting the methods for knowledge conveyance. According to Megaw (2001), it is critical to evaluate these characteristics so that the learners can successfully complete the learning process and gain new knowledge.

The second component of the ASSURE model is to "state the objectives of instruction", which also involves systematic planning and procedure (Callison, 2002). In this component, the ASSURE model focuses on the objectives related to the teaching and learning such as the sequence of learning activities, learners' behaviour and their learning conditions (Megaw, 2001; Faryadi, 2007). However, the objectives of this research are more to identify the usability of RPG in speech practising, especially for down syndrome children, it is important to make the acquisition of knowledge happen "invisibly" in an enjoyable manner, so that these children are not intimidated and discouraged by failing the defined objectives. Therefore, the objective selection of the ASSURE model is non-applicable in this situation and should be altered.

The third component in ASSURE model is to "select media". At this stage, media materials such as graphics, text and font styles, audios and videos must be selected. Three media selection process are stated by Megaw (2001), which are "deciding on the appropriate method for the stated learning tasks", "choosing a media format that is suitable for applying the method and finally selecting", and "modifying or designing specific materials within the

media format selected". With proper selection of media, learners' attention can be captured and sustained, thus effective learning can occur. In this research, the media is tailored to suit the down syndrome children's preferences based on previous studies, as discussed in the later section, such as using outlined-font and interesting graphics.

The fourth component is to "utilise materials", in other words, to decide ways to present the courseware or product to the learners. Megaw (2001) suggests that "5Ps" should be followed at this stage, which are first: preview the materials to ensure that the materials selected are appropriate; second: prepare the materials and determine the sequencing of instruction; third: prepare the environment, for instance, a product that implements audio files must utilise a speaker; fourth: prepare the learner, as to inform them about the expectation of the courseware or product; and finally: provide the learning experience. For this research, Internet connection, computer, microphone, and a suitable website browser that supports the Webkit speech recognition must be utilised after the first two "Ps". On the other hand, the RPG in this research should inform the parents what is the expected outcome so that they can guide their children to complete the game.

The fifth component of the ASSURE model is to "require learner participation", wherein adequate practice opportunity must be provided for the students, as well as to give feedbacks for improvements. For this study, the main concept is to have the learners to speak according to the sentences being displayed. Since parent or adult guidance is expected, feedbacks are expected from their parents or guardians.

The final component of ASSURE model is to "evaluate and revise", which is essential in designing good instruction. This evaluation should cover for the entire instruction in terms of learners' objective achievement, instructional process and media selected (Megaw, 2001). This model is suitable to be used to design instruction for down syndrome children because it was created based on cognitive theories of learning, which is

emphasized in this study.

For Gerlach and Ely model on the other hand, there are ten elements to this model, which are illustrated in Figure 2.1.

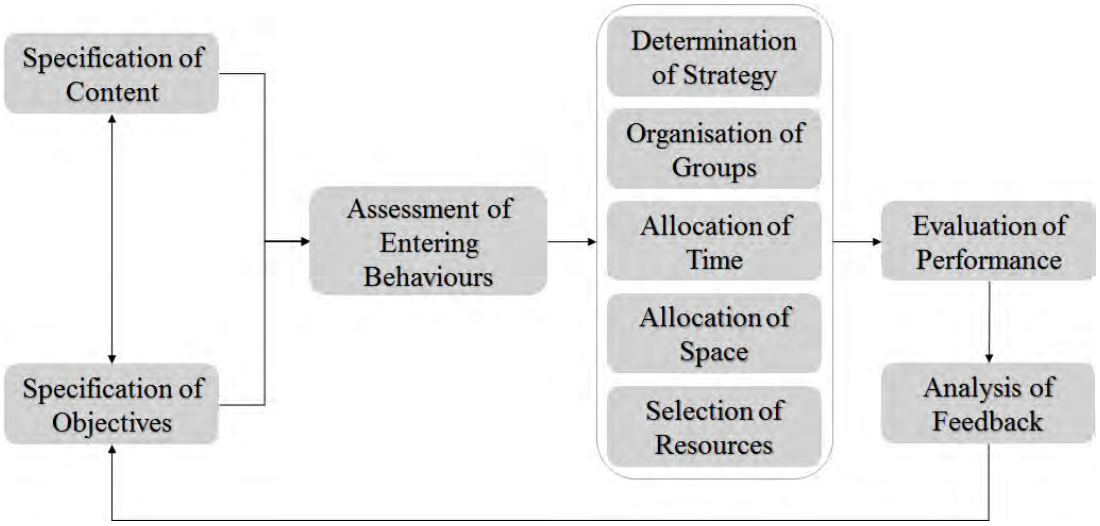

Figure 2.1: Gerlach and Ely Model

The central focus of Gerlach and Ely model is on systematic planning, which emphasises on clearly defining teaching goals and how to reach these goal (Forest, 2016).

The first two elements are interconnected, which are specification of content and specification of objectives. This model acknowledges that under many circumstances, choosing the objectives before defining the content is not critical, which is true for this research as well. In this study, what the down syndrome children can learn with the game is flexible, and the objectives are shaped according to the earlier analysis of learners. At the same time, new content may be specified for the defined objectives.

The third element is to assess the learners' prior knowledge and skill level, which is similar to the first component of ASSURE model. Hence, the ASSURE model can be integrated to this model from here, since ASSURE model provides more systematic way of learner analysis.

The fourth to ninth element are interdependent, which means that "decision concerning

one of these elements will impact on the others, and will often influence the range of viable options available with respect to the others." (Forest, 2016). For instance, a small group size may require small space for learning, and the amount of resources are dependent on the strategies predetermined. The allocation of time may also have potential influence on the allocation of space, etc. At this stage for the research, it appears that some elements are not applicable for the purpose such as organisation of groups, allocation of time and allocation of space since they depend solely on the individual learners in this case. Therefore, the third and fourth components of ASSURE model are to be followed instead.

The tenth element of the Gerlach and Ely model concerns about evaluation of learners' performance. Before this begins, the fifth component of the ASSURE model can be followed, which is, for the case of this research, to allow the learners to interact with the RPG. After the learners' participation, evaluation of learners' performance may start. Finally, the feedbacks from the evaluation be analysed and revise to determine room for improvement to the game.

Gerlach and Ely model is a resource-focused model, which makes it suitable to be used in this research, since the educational content is to be presented using game approach, which is rich in multimedia elements. Using this model will ensure smooth integration of instructional design with the game development as it provides clear stages to designing the instruction. However, Gerlach and Ely model is insufficient in this research because it is not a user centred model, whilst this study is catered for learners with special needs. Therefore, ASSURE model is integrated with Gerlach and Ely model in this case.

2.1.4 System Development Life Cycle

A disparate form of systematic process that involves in the product development is the SDLC. SDLC is a software cycle that deals with various parts and phases of methods

and strategies to develop, design and maintain the software project to ensure that all the goals, objectives, functional and user requirement are met (Ragunath et al., 2010; Arora and Arora, 2016). This research is based on design science methodology, wherein the SDLC plays a crucial role other than instructional design. This section introduces and compare some popular SDLC models, including Waterfall, Spiral, V-Model and Agile models to contrast their usability for this research.

The advantages and disadvantages of these models as discussed in (Ragunath et al., 2010; Balaji, 2012; Mahalakshmi and Sundararajan, 2013; Arora and Arora, 2016). are tabulated in Table 2.1.

Table 2.1: Advantages and disadvantages of Waterfall, Spiral, V and Agile Models

Models	Advantages	Disadvantages
Waterfall	- Simple to implement. - Phases do not overlap, more efficient management. - Works very well for small projects.	- No evaluation on each phase - high risk and uncertainty. - Very rigid project scope. - User approval and project usability can only be determined at the end of cycle.
Spiral	- High amount of risk analysis. - Efficient for large projects. - Software is produced early in the life cycle.	- Does not work well for smaller projects. - Costly due to repeated processes.
V	- Easy to use. - Planning and designing happen before coding.	- Very rigid and low flexibility. - No early prototypes are produced. - Requirements documents and test

	- Requirement changes is possible at any phase.	documentation needs to be updated when changes happen. - Not suitable for short term projects.
Agile	- Incorporate changes easily. - Continuous inputs from clients. - Costs changes can be handled well.	- Not suitable for large projects. - Vague planning date. - Experienced programmers needed to make decision. - Team involvement is highly essential

From the table above, waterfall appears to be the simpler to implement compare to other models. This is because the phases are in sequence rather than overlapping and each step only go through once, as illustrated in Figure 2.2. Since there is only one cycle to be completed and the development is to be stopped immediately, it is said to suit for small project. On the other hand, due to its one iteration and non-overlapping nature where evaluation is only done at the fourth stage, projects developed using this model consist high level of latent risk and uncertainties, such as the compatibility issues. At the same time, this model allows very rigid project scope, that is, no changes shall be made after the design phase begins, feedbacks from the users will not be considered as recommendation for project improvement. Without user involvement in any intermediate stages between analysis and maintenance, user approval and usability of the project can only be identified at the end of system development.

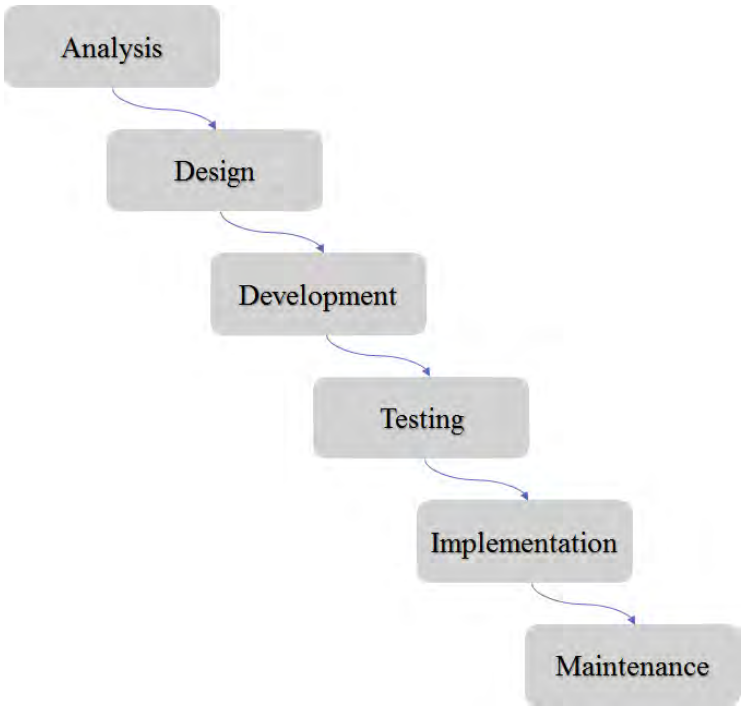

Figure 2.2: Waterfall Model

Despite waterfall model has many disadvantages, it is still adopted in this study. This is because as shown in table 2.1, only waterfall model and agile model are suitable for small and short term project. However, teamwork is an essential element in agile model, in which experts from all fields, such as the programming experts and content experts, must ready to respond to the changing requirements, even in late development stages (Balaji, 2012). This is due to the multiple iterations where phases run in parallel (Arora and Arora, 2016).

For the spiral model, it is not suitable for small and short term projects because in involves the repeating of the sequential iteration in spiral (Ragunath et al., 2010; Arora and Arora, 2016), which imposes no limit on the repetition. As the risk analysis induces large overhead costs on the project development such as time and budget, applying it on short term projects is not feasible. The V-shaped model, however, is not suitable for short term project although it is also a sequential, one cycle model because it emphasises testing on all stages (Ragunath et al., 2010). With more testing steps involve, more time must be allocated. Ergo,

the only suitable model to be applied in this study is the waterfall model, wherein the development process are mostly planned by using Gantt chart (Ragunath et al., 2010; Arora and Arora, 2016), that is, to follow a step-by-step process. The progress flows from top (Analysis) to bottom (Maintenance) like a waterfall, as illustrated in Figure 2.2.

For the waterfall model, different formulations exist to argue which step should come before another. While analysis and design are always the first and second steps of the model, the following steps are varied based on different proposals. The first model was definition was made by Winston W. Royce in 1970 as a five steps model in the sequence: Analysis, Design, Implementation, Verification, Maintenance. The more recent waterfall models was as illustrated in Figure 2.2 (Balaji, 2012; Arora and Arora, 2016). In this research, the latter waterfall model is followed. The first element in the waterfall model is analysis, which is to gather all the user requirements. The second phase is can be broken up into logical design and physical design sub-phases (Hughey, 2009). The information collected in the analysis phase is translated into logical design independent of hardware and software. Once the logical design is complete, it is transformed it into physical design, which is dependent on the hardware and software technologies. On the next phase, the system begins its development in small “units”, and testing is done for each of these units for its functionality, also known as the unit testing. Then, the entire system is tested for faults and failures. Upon no failure detected, the system can be implemented to the end users. Maintenance is provided when issues are found at the users’ environment, such as improper requirements determination or changes in the users' requirements (Hughey, 2009).

2.1.5 Multimedia in Special Education

Multimedia elements plays important role in this research. The review of multimedia for special education in the study focuses on the standards for designing multimedia

applications for down syndrome children. The review is to outline the factors to be considered when developing the application so that it does not cause confusion or distraction to the students.

Multimedia approach is proven to be very effective in courseware design for special education (Khan and Bayoumi, 2015). Nevertheless, a set of standards should be considered when designing application or courseware for students with special needs. Generally, most of the existing design methods focus on visual and auditory approaches, which involve graphics, text and sounds (Ng et al., 2014). Courseware designed for down syndrome learners must take the learners' learning capabilities into account by considering all possible disabilities of the learners such as different degrees of intelligence quotient that may fall within the range of 20-70, hearing loss, otitis media and eye disease that may hinder their learning process (Bull, 2011), wherein the lower IQ is said to be the most prevalent symptom amongst the down syndrome children. Research suggests that courseware designed for down syndrome children must be able to catch and maintain their attention to increase their motivation (Ng et al., 2014; Khan and Bayoumi, 2015). Inclusion of multi-interactions techniques and different types of activities are preferable.

Down syndrome students are usually visually and auditorily impaired, therefore the interface of the courseware, including the font size and images, must be ideal for them to use, and the application must equip with clear voice delivery and the appropriate voice intonation (Shima et al., 2014). Some features of visual and auditory to be considered while developing courseware for visually and auditorily impaired students with down syndrome wherever applicable, are summarised in table 2.2 (Lazar, 2007; Mohid and Mat Zin, 2010; Aziz et al., 2011).

Table 2.2: Standards for developing courseware for visually and auditorily impaired children

Elements	**ID principles**
Audio	Clear and concise to make sure that the hearing-impaired learners have accessed to the full range of speech sounds (Aziz et al., 2011).
	Use sound effects to enhance learners' understanding (Aziz et al., 2011) .
	Avoid using background music as the learners cannot differentiate the background music from desired information (Aziz et al., 2011).
Video	Shows the real situation related to the sound (Lazar, 2007).
	Display combination of expressions and emotions together with the audio (Mohid and Mat Zin, 2010).
Formatting style and text	Use font size of 18-point and above, and font type of serif font (Aziz et al., 2011).
	Combination of font colour and background colour must be highly contrast, such as yellow and dark green, black and white (Aziz et al., 2011).
	Use bright coloured and outlined font (Aziz et al., 2011).
	Avoid animated text and font decoration (Aziz et al., 2011).
Graphic and animation	Minimized the use of graphics in one screen (Aziz et al., 2011).
	Avoid overwhelming animation (Lazar, 2007).
	Emphasized on the size of graphics (Aziz et al., 2011).
	Use real life photographs and whimsical illustration (Lazar, 2007).
	Use relevant and interesting animation for beginner users to create the relationship of sound between the animation and video (Lazar, 2007).

Three application user interfaces (UIs) are reviewed, represented by a Malay reading application for down syndrome children – MEL-SindD (Yussof et al., 2010), a Cantonese application for individuals with intellectual disabilities - 龍情訓練坊 2 (Long Qing Xun Lian Fang 2) (Hong Kong Lutheran Social Service, 2014), and a English conversation therapy application for individuals with communication difficulties - Conversation Therapy (Tactus Therapy Solutions Ltd., 2017), to compare with the multimedia standards above, as well as to relate with the visuospatial working memory.

Figure 2.3: MEL-SindD user interface

The UI of MEL-SindD uses font size of 18-point and above but with sans-serif font type, which is not recommended in the standards. Although the texts are outlined, their colours do not contrast with the background. Texts are not animated but used dark colour. The graphics in one screen is minimized, with large and relevant graphic on the screen. Relevant animation is used such as the rain drops to describe the word “hujan” or “rain”. The graphic used is neither whimsical nor real life.

Figure 2.4: 龍情訓練坊 2 (Long Qing Xun Lian Fang 2) user interface

The UI of 龍情訓練坊 2 (Long Qing Xun Lian Fang 2) uses unknown font size, but clearly below 18-point. Since the standards did not mention the recommended font type for Chinese characters, the standard font was used. Although the header characters are outlined, its colour does not contrast with the background, and very difficult to read. The graphics in one screen is minimized, but the size of graphics is not emphasized. There is also no animation involved in the application, making it looks dull and uninteresting. Real life pictures are used in the application.

Figure 2.5: Conversation Therapy user interface

The UI of Conversation Therapy uses font size of 18-point and above but also with sans-serif font type. The texts are not outlined, but the colour does contrast with the background. There is no animation involved in the interface to capture attention. The graphics in one screen is minimized, with large main graphic on the centre of the screen. Real life pictures are used.

The preference for 2D or 3D games to play, however, varied from person to person. Nevertheless, a study by (Prena, 2014) shows that the games that down syndrome children play are always require minimal cognitive skills, as described by the parents. Apart from better looking, no other advantages have been specified to be more beneficial for the down syndrome children, instead, it creates unnecessary cognitive load on the programme users (Kyritsis et al., 2013; Prena, 2014). From the perspective of spatial-memory impacts, research has found no different in subject's performance on the memory tasks (Cockburn, 2004). As a learning-centred RPG, the main purpose for the game development is to provide a more effective way to convey knowledge, by omitting any known obstacles that may hinder the learning process. An overall comparison of learning gain using 2D and 3D representation shows that 2D model delivered better learning outcomes (Richards and Taylor, 2015), which is a solid reason for this research to incorporate 2D representation in the game design rather than 3D.

2.1.6 Two-Dimensional Role Playing environment

A traditional 2D RPG incorporates a top-down view where the players play the role of the "hero", which is also featured in this study. Normally, computer games involve simulation, rendering, non-playing characters' motion planning, prototyping and proximity queries (Kenneth E. et al., 2001). However, since the game in this research is a fully guided educational RPG, only translation animation is involved, triggered by different options at

players' utterance. Multiple platforms exist to facilitate 2D RPG creation, such as Unity 2D engine and RPG makers. While these engines are powerful in their own ways, many have limited support to speech recognition, which is essential to the game development of this research. Since speech recognition in game is still rarely focused as discussed in Chapter 1, many RPG engines such as the RPG makers do not support this functionality at all. With the more well-known Unity engine, it only supports speech recognition on Windows 10 Operating System. To minimise the computing system requirement, HTML5 with jQuery is based instead, by implementing the Webkit Speech API which allows players to play the game online with a Webkit-core website browser.

2.2 General Area of Research

The general area of this research is to integrate speech recognition in a 2-dimensional role-playing environment to facilitate the down syndrome Mandarin learning for children, so that they can improve their speaking confidence and grammar. Many digital games exist for children to learn as well as many applications designed for special education. However, very rare has hitherto been integrating digital games with special education, even though the idea of "learning while playing" has been popular for a while, the research area of this study also includes the identification of multimedia standards that are not intrusive to the down syndrome children and to utilise their short-term memories by using the sequential natural of RPG, so that they are able to grasp on the language content.

2.3 Underlying Theory

The theories underlay in the literatures include the theory of speech recognition and its recent achievement in terms of accuracy, that highlights its usability for Mandarin recognition in digital games for education purpose. The theories also include the methods to attract attention of down syndrome children, such as using adequate amount of animation and interesting graphics. The theory of instructional design, SDLC and their impacts to special education also seen in the past research. Lastly, the theory of cognitive psychology of down syndrome children is included in the literatures, which shows that the less impaired visuo-sequential memory could be used to facilitate language learning.

2.4 Theoretical Framework

Several language models were mentioned in the literature such as ANNs, HMMs, DBNs and n-gram models to give a solution to recognising natural language, such as speech. It is seen in the literature that Microsoft that employs multiple NN models has achieved the accuracy that is viable to be integrated in digital games due to its high accuracy.

The three subsystems of visuospatial working memory were mentioned in the literature, including visual, spatial-sequential and spatial-simultaneous. While utilising the relatively stronger spatial-sequential working memory in the language teaching role-playing game, spatial-simultaneous working memory can be improved with parents' guidance, which means to convey the knowledge with spatial-sequential approach, while showing the context with spatial-simultaneous approach.

Several ID models and SDLC models were discussed in past studies, such as ADDIE framework, ARCS Model, ASSURE Model, Gerlach-Ely Model, Waterfall Model, Spiral Model, V-Model and Agile Model. These models act as the guideline for this research to produce an effective language learning game for down syndrome children.

The multimedia standards are useful to make sure that the students are comfortable using the application, learning can happen in a non-distracting and non-confusing manner. It also ensures that the visualisation does not incur overwhelming memory burden on the students using the application.

2.5 Hypotheses

Computer role-playing-game with speech recognition ability can be used for Mandarin learning, by guiding the users to speak with better grammar respect to context of the game's background.

2.6 Conclusion

Some of the major contributions of these studies include the use of neural network language models to improve speech recognition accuracy even for heavy accent articulation. Although hybrid neural network models have shown superior performance over conventional neural network models, Microsoft is using hybrid conventional neural network models in the ASR technology. The studies also compare the verbal working memory and visuospatial working memory in down syndrome population with the typically developing population. These studies then address the part of visuospatial working memory that should be emphasized for down syndrome language learning since it has been confirmed that their

visual memory is also impaired, only less severely than their verbal memory. The studies also outlined the general standards for the development of special education applications. The models for instructional design and system development are given in the literature so that the effective applications or games can be developed for teaching and learning.

From the reviewed literature, it is seen that the devoid of knowledge lies between the working memory and multimedia content suitable for down syndrome children. It is not known whether 2D visualisation or 3D visualisation is easier to be processed and digested by the memory impaired individuals. The previous studies also lack focus on bringing context to language learning, which could enhance grammar quotient of down syndrome individuals. In this research, RPG is used to bring context to language learning, especially to encourage the children to speak in sentences, enforcing the use correct grammar to be a habit in their expressive language. 2D visualisation is chosen in this research because it is a better approach to convey knowledge than 3D visualisation. It is also less resource thirsty and thus, can be implemented and accessed easily from lower end computers. The literatures reviewed also lack the involvement of speech recognition in RPG environment.

In conclusion, the literatures reviewed suggest that speech recognition accuracy is now comparable to human parity and therefore could be used in digital games even for language learning. In addition, the literatures reviewed highlight the effects of working memory impairment to language learning and give the direction to focus when teaching language to the down syndrome children. The general standards and the models for instructional design and system development should also be considered when developing multimedia application or games for individuals with down syndrome.

CHAPTER 3

RESEARCH METHODOLOGY

3.0 Introduction

This chapter defines the research methods used to conduct the study. The methods of necessary data and information collection to address the research objectives and questions are presented and analysed. Reasons and justifications for the research design, research approach, data sources, data collection techniques, and analytical techniques used are given.

3.1 Research Design

The research philosophy of this research is interpretivism, wherein different individual may have different level of acceptance to the application to be produced. Interpretivism philosophy is based on observation of the outcomes of the previous studies to weakly predict the possible outcome with certain system design interface. No general rules may be deducted from the research itself that is true for everyone. This philosophy is adopted as the research follows the design science research method.

Since the design science research focuses on building of theory through creating and evaluating a prototype, the research approach to be used in this research is inductive. First and foremost, it is observed that the grammar quotient of down syndrome individuals in weak, and they have expressive language difficulties. However, no study has been found in the literature that strives to improve their weakness by utilising their relative strength of visual memory and encourages them to practice speaking in sentences, which manifests a

gap in the knowledge. A prototype is developed that base on RPG and speech recognition system using HTML5 and jQuery, and the interface that base on the multimedia design standards for down syndrome children.

Strategy of this research is qualitative. The research aims to analysing user experience for using the artefact in language learning. The variables such as the context of languages, multimedia design standards and navigation methods are controlled. The results are evaluated based on the users' acceptance level, which makes this a cross-sectional study, and the Waterflow Model is adopted in the game prototype development.

3.2 Prototype Development

The development of prototype follows the hybrid ASSURE and Gerlach-Ely Models for instructional design, and the Waterfall Model for the game design. The ASSURE and Gerlach-Ely Model are combined as illustrated in Figure 3.1.

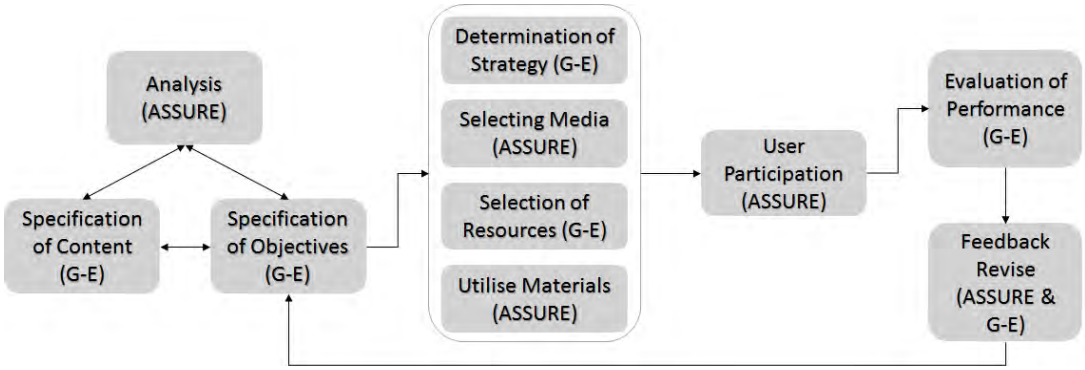

Figure 3.1: Hybrid ASSURE and Gerlach-Ely Model

At the first step, a brief objective of the research is formulated. Next, content was specified based on the analysis from the preliminary interviews. Then, the objectives and suitable contents are specified. The data collection method is discussed in section 3.3. After the objectives are determined, the strategy to achieve them is planned. The strategy used in this research is by developing an RPG for the down syndrome children in Malaysia. After

that, media is selected for the purpose of instruction, where in this case, the media selected are digital images and text illustration, sound, and speech recognition. Material Utilised are computer with display monitor, Webkit-core website browser on Windows Operating System (tested with Maxthon Web Browser and Google Chrome), microphone and loud speaker. Learners participation is in terms of interaction with the game. The learners' performance will be evaluated by the speech recognition system and feedback be given in terms of different colours text-outline, as shown in Figure 3.2 (a) and Figure 3.2 (b).

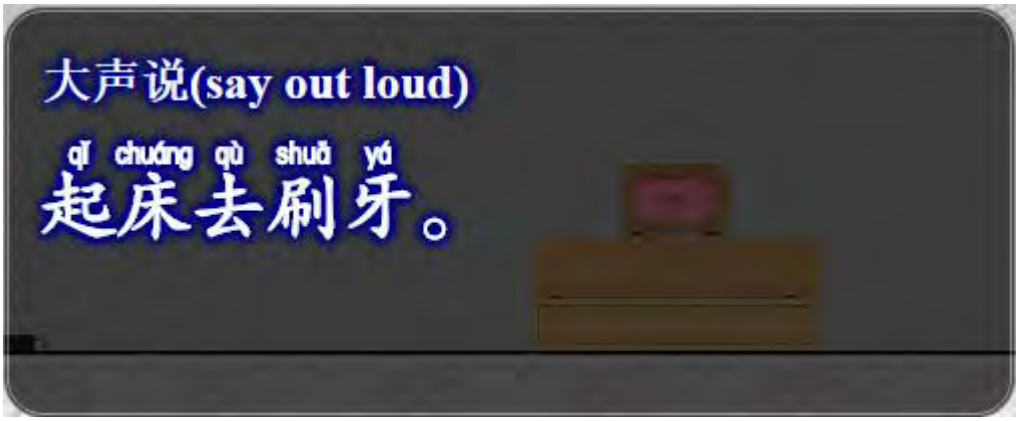

Figure 3.2 (a): Instruction without speech input detected

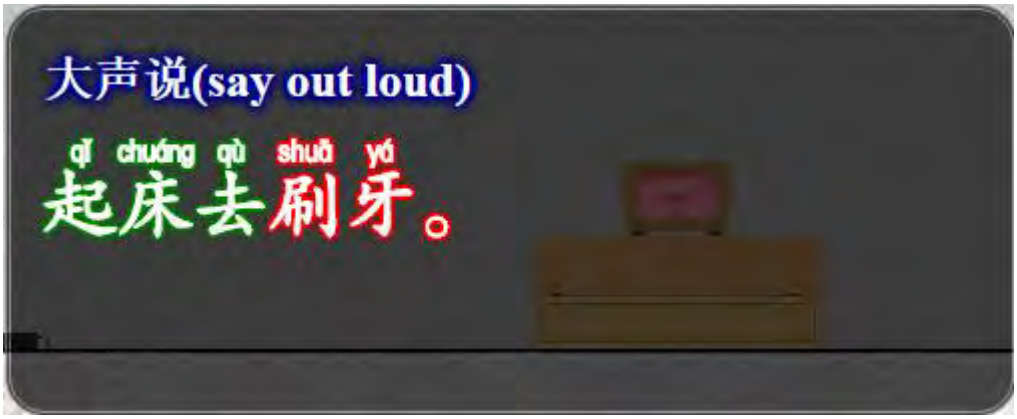

Figure 3.2 (b): Instruction with three correct and two wrong utterance

For the game's logical design, a game details were planned and represented in storyboard based on the analysed inputs from the instructional design process. The completed game will be consisting nine individual scenes, while the prototype contains six individual scenes. The NPC locations and the options allowed for the game were decided at

the planning stage. The storyboard is illustrated in Figure 3.3, representing the storyline of the prototype.

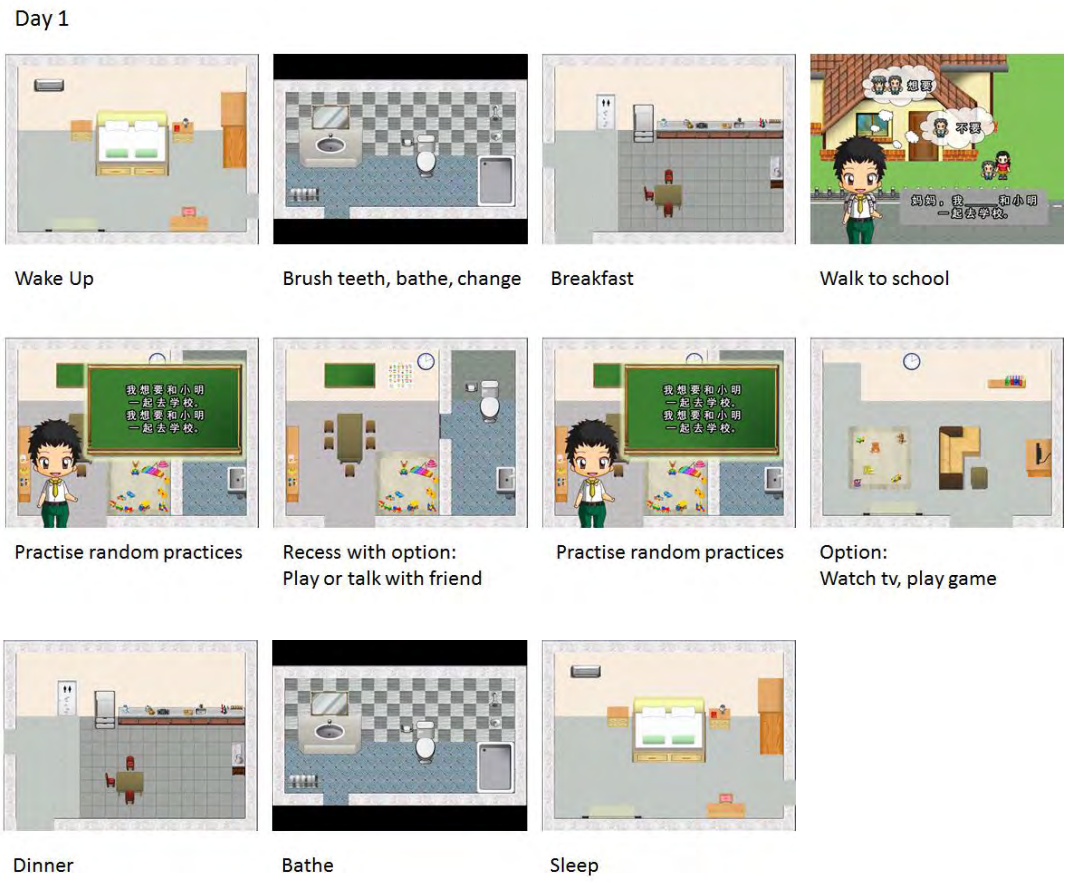

Figure 3.3: Storyline of game prototype on storyboard

The design features the player's gender option where the players can choose which gender they like to play with at the beginning so that it creates a personal connection between the players and the game. At specific scenes, the players can choose the sequence of the tasks to be completed. For instance, the players can choose to brush their teeth or to take a bath first at the bathroom, at the second scene on Figure 3.3. This is to create a sense of autonomy to the players so that they can relate the game better to their daily lives. The game includes a seven-day situation to be completed. The story begins with the main character waking up in the morning, follows by the instruction pop-up. Upon the correct match of players' utterance with the on-screen instruction, the main character is navigated to the bathroom for

next actions. As mentioned earlier, the players can decide whether to take a bath or brush their teeth first by reading out the instruction in short sentence. On the completion of the first activity, players are left with the other unchosen option. When two actions are accomplished, the players must say the instruction to navigate the main character to have his or her breakfast at the kitchen. After having the breakfast and prepared to attending school, the players can choose whether to walk alone with his or her mother, or walk together with friend. At this stage, the choices will affect the endings, either an ending with friendship or become an independent person. Both endings are positive so that the players are not discouraged. At school, longer sentences are to be spoken by the players, which are educational on specific subject commensurate for children of 8-year-old and below. The school is divided into two session with a break in between, where the player can choose to talk with friends or play with friends. The options also affect the endings. When the class ended and the main character returned home, he or she can choose to either bathe, watch television or play with toys in any sequence. Finally, the main character goes back to sleep after having dinner and the first day ended.

For the days to come until the fifth day, the game will follow the flow of day one, but with branches to be triggered by other events, for instance, the friend will be sick on day three, and the player can choose to visit or not to visit. If visit, then the player can choose to go to the friend's house after school for the remaining days to achieve the friendship ending. At day six and day seven, which no school to attend, the player can visit to the park on both days and go to the beach with family on day seven. The idea to have outdoor activities is to motivate and encourage the down syndrome children to do exercise which are beneficial to their motor skills. After the logical design is completed, the physical design is deployed using HTML5, CSS3, javascript and jQuery. The development began with positioning and styling of the scenes, masks, items, and texts. The dimension of the game's window is set to be at

960-pixel width and 720-pixel height. The font decoration is minimized with only text outline to contrast the bright colour font, as recommended by the design standards discussed in chapter 2. The speech recognition and comparison functions is included in a separated Javascript file as follow.

```
function startDictation() {
  if (window.hasOwnProperty('webkitSpeechRecognition')) {
      var recognition = new webkitSpeechRecognition();
      recognition.continuous = true;
      recognition.interimResults = false;
      recognition.lang = "cmn-Hans-CN";
      recognition.start();
      recognition.onresult = function(event) {
          var last = event.results.length - 1;
          var res = event.results[last][0].transcript;
          document.getElementById("demo").innerHTML =  'Result received: ' +
res + '.';
          speechRes = res;
          splitSpeech = speechRes.split("");
          if(day>0){
              compareSpeech();
          }
      };
      recognition.onerror = function(event) {
          document.getElementById("demo").innerHTML = 'Error occurred in
recognition: ' + event.error;
```

```
            recognition.stop();
        };
        recognition.onend = function() {
            recognition.start();
        };
    }
}
function compareSpeech(){
    $("#words").html("");
    for(var i=0;i<splitWord.length;i++){
        if(splitSpeech[i]==splitWord[i]){
            $("#words").append("<span style='text-shadow:0 0 4px green,0 0 4px green,0 0 4px green,0 0 4px green,0 0 4px green,0 0 4px green,0 0 4px green'>"+splitSpeech[i]+"</span>");
        }
        else{
            $("#words").append("<span style='text-shadow:0 0 4px red,0 0 4px red,0 0 4px red,0 0 4px red,0 0 4px red,0 0 4px red,0 0 4px red'>"+splitWord[i]+"</span>");
        }
    }
    var patt1 = /[，,。??]/g;
    for(var i=0;i<theWord.length;i++){
        var clearword = theWord[i].replace(patt1,"");
        if(speechRes==clearword){
            theWord = [];
```

```
                theWord[0] = speechRes;
                speechRes = "";
                $("#words").html(clearword);
                hideRead();
                break;
            }
        }
    }
```

From the function above, the recognition language is defined as Mandarin Chinese before the recognition starts. The function is modified so that it splits the detected results into array of individual words to be compared and return feedbacks to the players. In the compare function, the words stored in the array are compared with the instruction sentence. First, special characters such as the comma and question marks are remove from the sentence in the backend and stored in another array. Next, the instruction sentence is removed from the display. Then, green colour outlined characters that correspond to the correctly uttered words and red colour outlined characters that correspond to the falsely uttered words are appended to the display. Hence, a feedback unit is produced.

The transition animation and the scenes switch of the game are fully defined by the x and y position on the coordinated system. Once the players' voice matches the instruction, the path finding function will be triggered, and once the transition animation stops, another function will be called, or the global variable be updated. The code below shows a part of the path finding function for illustration.

```
if(_sceneStartEnd == "schoolSitOut"){
    //6ms each px
```

```
$("#player_char").animate({"left": "240px"}, 60, "linear", function(){
    $("#player_char").animate({"top": "624px"}, 1134, "linear", function(){
        if(day==1 && actNum==8){
            posNow = "livOut";
            loadScene("living");
            optAllow = 3;
        }
        else{
            posNow = "schoolOut";
        }
    });
});
}
```

The overall game testing is performed after the game is developed. Since the testing focuses on the functionality of the game, it is tested with different website browsers that are compatible with the game – Google Chrome and Maxthon browser. On google chrome, the speech recognition only works with https address prefix. On the other hand, Maxthon browser works well with and without the https address prefix. However, this game prototype is not implemented and distributed to the public.

3.3 Data Collection Method

The dependent variable is student engagement, which can be deduced from the effectiveness of the teaching approaches. An effective approach means that it successfully transfers the knowledge from educators to students, which indicates that the students are engaged in the study. The independent variables are students' ages, teaching approaches, and

content being taught. The data was collected from preliminary interview with a parent with down syndrome child to understand the teaching and learning approaches used in down syndrome rehabilitation centres; and a preschool teacher for language content, as well as the teaching and learning approaches used in typical kindergarten. A set of questions were asked during the interview regarding the content being taught at down syndrome rehabilitation centre and kindergarten, students' ages, teaching approaches and the effectiveness of current approaches. Questions regarding the down syndrome characteristics and the language used for teaching and communication at the rehabilitation centre were also asked. Purposeful-expert sampling method is used.

3.4 Data analysis method

The data collected are excerpted, analysed, classified and tabulated. From these tables, the patterns emerge in the data are found. From these patterns, deviation is identified and explanation to these atypical patterns is attempted.

CHAPTER 4

RESULTS AND FINDINGS

4.0 Introduction

The section presents a complete account of the results obtained in the study, which include the preliminary interview findings and the emotion evaluation of children participated in the game testing.

4.1 Results

The data from the preliminary interview findings are reduced and tabulated in table 4.1 to contrast between kindergarten and rehabilitation centre on their teaching approach and the teaching materials as well as some fundamental differences.

Table 4.1: Preliminary Interview Findings

Elements	Kindergarten	Rehabilitation Centre
Students' ages	- 5-6	- Below 7
Language	- Mandarin (in Chinese class)	- English (No Chinese subject)
Content	- Follow National preschool curriculum standards (KSPK). - To prepare the students for primary education.	- No specific standards, curriculum taught is personalised, based on the degree of students' cognitive impairment and

		their physical capabilities. - To prepare the students for primary education
Teaching approaches	- No text books - Play cards to recognize words. - No speaking practice involve.	- No text books - Play cards, Lego and craft for all learning. - No speaking practice involve.
Effectiveness of current approaches	- Effective	- Depends on the degree of cognitive impairment and capabilities. - Improvement is seen (both physical and speaking abilities)
Down syndrome characteristics	- Not applicable	- Intelligence impairment - Visual difficulties - Speech difficulties - Hearing (Unclear as the testing method is just on their response to sound stimulation)

As seen from Table 4.1, both typical kindergarten and down syndrome rehabilitation centre incorporates no-text-book and learning through playing method to teach children of eight years old and below. The interviewees agreed that this method is very effective to catch and sustain the children's attention and providing them high motivation to learn. For the

down syndrome children, playing games helps them to improve the speaking skills.

Five Chinese-speaking children of age five to six years old, four Chinese-speaking adults and one non-Chinese-speaking adults have participated in the prototype testing. The results gathered are in the form of video clips, pictures, and verbal description. Figure 4.1 (a) shows a six years old boy playing the game, and Figure 4.1 (b) shows a six years old girl playing the game. So far, the game has shown no bias to players' gender.

(a) (b)

Figure 4.1 (a): A six years old boy playing the game

Figure 4.1 (b): A six years old girl playing the game

After the testing, the adult participants were asked about their feeling during the testing, while the children's emotion was observed and described by the adults who were the teachers and the children's guardians. The participants were also asked if they feel the game's operation was easy. Table 4.2 shows their feedbacks on the ease to operate the game, and Table 4.3 shows their emotion towards the game during the test.

Table 4.2: Participants' feedback on ease to operate the game

	Very easy	Easy	Inconvenient
Children	2	3	0
Adults	1	2	2
Total	**3**	**5**	**2**

From this table, it is seen that majority of the participants found the game easy to operate, especially with the children. Two out of five children said that the game was very easy, while the other three children said it was not inconvenient, but sometimes requires them to speak more than once for one sentence to trigger an action. None of the children found the game to be inconvenient to play with. For the adults, one out of five said that the game was very easy, and two adults said the game was easy. However, two adults found the game to be inconvenient for them to use.

Table 4.3: Participants' emotion towards the game during the test

	Interesting	No special feeling	Boring
Children	4	1	0
Adults	2	3	0
Total	**6**	**4**	**0**

This table shows that majority of the participants found it interesting to play with the game. Four out of five children are described to be interested and engaged by the game, but one child was said to be not really interested in the game. On the other hand, two of the adult participants said that they are interested in the game, including the non-Chinese-speaking adult. Three other adults however, said they do not have special feeling towards the game. One of these two adults added, she thinks this game may be interesting to children. Overall, no participants have reported to be bored by the game.

CHAPTER 5

DISCUSSION

5.0 Introduction

The section provides the discussion of results in Chapter 4. By analysing the results and observing the pattern, theoretical and practical implications are revealed.

5.1 Discussion

Although no children felt inconvenient to play the game, not all of them found the game very easy for them to use. Based on the adults' descriptions, the game does respond to the speech, but only when the pronunciation is clear. Consequently, this symptom may bother the children as the game does not respond to their utterance sometimes. However, from the video recording of gameplay by the non-Chinese-speaking adult participant, it is seen that incorrect pronunciations are sometimes recognisable by the speech recognition system. This implies that age is a factor that affects the speech recognition accuracy, which has been a research-focus topic by Google's researchers (Liao et al., 2015). Although the intelligibility of the speech recognition system on children's voices is not as comparable to adult voices, the results have shown that the speech recognition system still works for the more common vocabularies that are often spoken by children.

Although the ASR is said to work better with adults' voices, more adult participants have reported that the game is inconvenient to operate. This is because the game was always prepared for the children to play by the adults, and the initial preparation could appear to be

troublesome under certain conditions. For instance, one of the participants encounter a problem as displayed in Figure 5.1 while testing the game on a desktop. Although the problem was resolved, it stirred impatient in the children who were waiting for the game and lead to frustration of the adult.

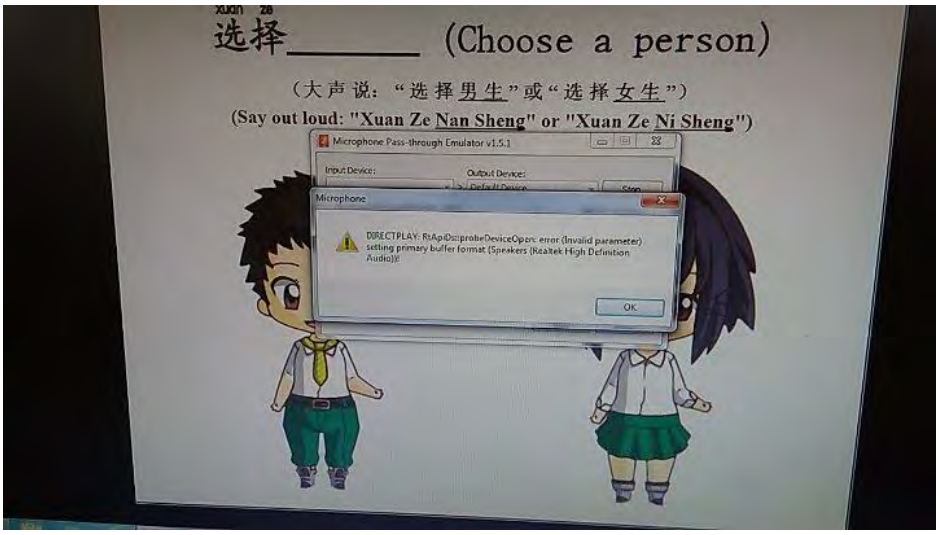

Figure 5.1: Microphone input problem encountered during prototype testing

Since the Google ASR system must work under https connection with Google Chrome website browser, the game must be accessed with this header. However, since the certificate use for this game is not recognised by Google Chrome, the connection warning in is likely to be reported. While for most cases the page can still be accessed if the users want to, there are times when no proceed option is available to the users, such as the case illustrated in Figure 5.2.

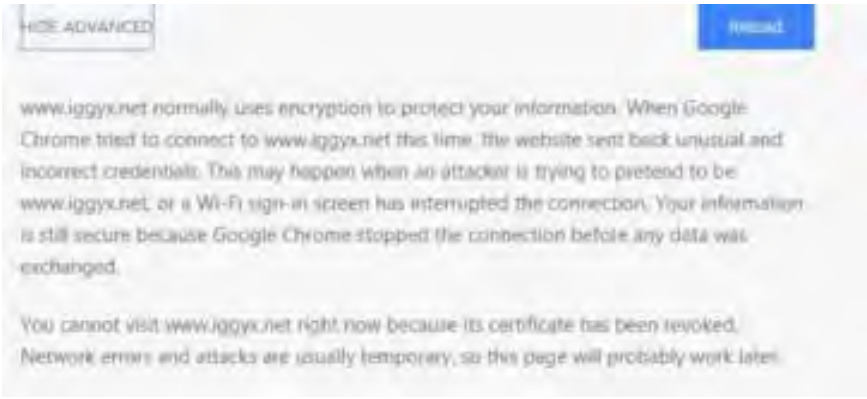

Figure 5.2 Website's credentials warning

From these reported issues, it can be concluded that all required software to work support the game's functionality must be provided to the players to ensure that the game can reach its audience at minimum effort. The potential problems can be identified by testing the game on multiple platforms and environment before launching it to the public, and the solutions to the potential problems need to be prepared.

The players' emotion analysis shows that all children felt interested and engaged by the game except for one child. This is because the child was in the situation of the microphone issue, and he became impatient while waiting for the game to set up. However, he became less frustrated after he played with the game, although he did not show emotion of being interested. These findings show that RPG game can motivate the children to speak more often by attracting their attention and engage them to speak clearly and with better grammar. A possible explanation is that the game gives the children an impression that their speech is powerful and can make the character moves, but only if they speak correctly. Another possible reason is that with an adult playing with them, they feel more courageous and motivated to speak as to fulfil their self-esteem. However, since the game is not tested for children without adult's supervision, the claim requires further research to support.

Most of the adult participants also felt that the game appears interesting to them. The three adults who did not have special feeling toward the game feel that the game is too simple and only suits for children. One of the three adults points out that the prototype is not attractive and not interesting because the character movement is very much like Pokemon, and the storyline is too short. She suggested that the game should include speech guide to teach the children how to pronounce the sentence, which is also suggested by others. Another suggestion from other adults is that more animations, such as character jumping when it is selected, should be added to make the game look more attractive. Nevertheless, this game can be an interesting self-paced practising tools for the non-Chinese-speaking people to learn

basic Mandarin, as the sentences they utter return easily comprehensible actions with English annotations.

5.2 Theoretical Implication

When applying the RPG to Mandarin speaking practice, interesting outcomes are observed. For the children accompanied by adults during the prototype testing, most of them gave positive emotions and showed learning engagement. A general assumption can hence be drawn, that is, with interesting animation and the encouragement from adults, the children are more motivated and interested to speak. Therefore, theoretically, the children need to be encouraged and their efforts need to be recognized in order for them to engage in learning. In addition to that, multimedia aids the adult's encouragement by attracting and sustaining the children's attention, allowing them to force themselves into speaking the sentences being display on screen. If we apply this theory in other research, such as for other languages, we can expect to see similar outcomes on the children's performance and emotion. Thus, further research is needed to test for the theory.

5.3 Practical Implication

In practical, this research can be applied to the mildly impaired down syndrome children to improve their spoken grammar by encourage them to practice more often. From the results, it shows that not only for children, but simple and relaxing RPG can be enjoyable to the adults. Unlike other courseware for language learning, educational RPG provides a different approach to learning by giving contextual meaning about the sentence, rather than individual words. Therefore, the practical implication is that, RPG is not only beneficial to the down syndrome children, but also helpful to adults who wish to learn a new language by modifying the contents and visualisation.

CHAPTER 6

CONCLUSION

6.0 Introduction

The section presents the flaws and limitations of this research from the analysis phase to the evaluation and testing phase from the aspect of instructional design models and SDLC model followed. With the gaps of this research be identified, the future research suggestions are proposed, together with the contributions of this research. Finally, this section briefly sums up for the entire work including methods, results and major recommendations arising.

6.1 Limitation of Research

Analysis: The major limitation of this research on the analysis phase is the lack of actual needs survey on the down syndrome children, such as the game's visualisation that they prefer. The graphic design and text styling are based solely on the recommendation for visually and auditorily impaired learners. The analysation of the down syndrome children's prior knowledge on the language aspect is also insufficient. The teaching material selected is based solely on the material suitable for all children below eight years old. Meanwhile, the analysis of the speech recognition API on children's voice is not performed and their speaking style is not evaluated. The hypothesis that speech recognition is suitable to be used for down syndrome children is based completely on the word error rate of current speech recognition technology under controlled condition, experimented by their researchers, which is not replicated in this study.

Specification of content: As mentioned above, the teaching material selected is based on the syllabus suitable for the general children below eight years old. The selected contents are not reviewed by the experts nor the down syndrome children to identify whether it is commensurate with their comprehensive levels, before being employed in the game. The scope of the game is only limited to the daily activities of general children, and it is not reviewed by the parents of the down syndrome children whether the situations in the game is relatable to their children as well.

Determination of strategy: The use of RPG to improve the grammar quotient of the down syndrome children is hypothesized on English language. There is no detailed research studied the Chinese-speaking down syndrome children's grammar performance. This research also does not compare the learners' performance on RPG with other approaches, such as conventional courseware or games of other genres. Consequently, it is not known whether RPG is the best solution for the down syndrome children to improve their grammar ability.

Selection and utilisation of media: The accessibility of the game is rigid, wherein no scaling of the game is allowed, and the platform of the game is only computer based. There is also no option to turn off the sound effects. This game is also not fashioned to work with screen readers, which is usually used by the severe visually impaired players. Another limitation of this research is that only two browsers tested are functioning with the speech recognition API.

Design and development for game: The physical design for the game is by using raw programming, which means that no gaming tools were involved. This is possible to develop a shorter storyline prototype, but impractical for the full-length game and future extension. This is because raw programming is extremely time consuming and many errors can exist due to negligence and typing errors. In addition, most of the time the developer must

remember all the functions related to one scene. In this research, a prototype is managed to be produced by developing the scenes, path, and speech comparison as different units, however, since the game is programmed using raw programming, debugging is very difficult and unit testing is impossible without affecting the overall game.

Testing and evaluation: The manual evaluation on the performance of down syndrome children is not performed in this research, but based solely on the game's feedback system instead. The sufficiency of the feedback system is not evaluated, and it is not known whether more feedbacks should be given by the game for each falsely detected utterance. The game is also too pronunciation demanding, that is, the pronunciation must be clear and unambiguous, which might be challenging for children with developing tongues.

6.2 Recommendation for Future Research

Based on the limitations stated in section 6.1, some recommendations for future research are described in this section.

Analysis: A thorough needs survey on the down syndrome children is recommended to be performed in future research, including but not limited to the things and colour that draw their attention, font size and text styling that are more appealing to them. The target group is recommended to be analysed in detail so understand their cognitive preferences. The children's knowledge on the language used in the game is also recommended to be surveyed in the future, which is to identify whether they can play the game with least supervision from their family members. Their prior knowledge to operating a computer should also be surveyed based on the children's geographical location and severity of their intelligence impairment. Before the development process, experiment on the speech recognition API on the targeted children is recommended to be performed to determine whether the system can correctly interpret their speaking style and their voices, so that when

they play with the game, they are not frustrated by the inability of the game to recognize their speech, suppose they uttered the correct sentence. In this research, the analysis is done by comparing the differences between typical kindergarten with down syndrome rehabilitation centre's teaching methods and the teaching syllabus, and the analysis shown that the children with mild down syndrome are capable to learn languages through playing games, similar to the typically developing children. This analysis helps to hypothesize that learning through computer games may achieve similar results.

Specification of content: For future research, the teaching materials are recommended to be reviewed by the contents expert after their selection to ensure that the content is suitable and effective for the targeted children. That is because for the down syndrome children, their cognitive preferences might be largely varied due to different level of intelligence impairment, the contents are recommended match with the target groups cognitive preferences based on the previous analysis outcomes. The scope of the game can be extended to more complex daily tasks that the children with down syndrome are more likely to encounter, so that they can relate the game more to themselves and make use of the game more effectively. In this research, the contents suitable for general children not only allow mildly impaired down syndrome children to improve their spoken language and grammar, but enables the typically developing children to play with the game for entertainment purpose. For the children who learn Chinese as a second language, they can practice their speaking since it is related to their daily routine, rather than bewildered by the irrelevant tasks.

Determination of strategy: Future research can focus on the comparison between the effects of RPG and other approaches on the degree of grammar improvement achieved by the down syndrome children. The identification of which game genre works best for these children can help to draw research attention to that specific area, and a standard to using

game for grammar practices could be formed. The contribution of this research regarding to strategy determination is that it opens a new perspective for research, emerges a research question of whether RPG is effective to bring visual context to down syndrome children.

Selection and utilisation of media: In the future, the media selected and created is recommended to allow scaling or magnified to cater for different display monitor's dimensions and special needs players who are more severely impaired visually. The platform for the game can be extended to mobile devices to increase the portability of the game. With smaller screen display of the mobile devices, the zooming ability of the selected media becomes essential. It is also recommended that more effort to be put into overall accessibility of the game such as volume control and screen reader friendly.

Design and development for game: Future researchers are advised to programme the game with an engine that supports speech recognition API or extension so that the maintenance of the game become easier, which is required after the game is implemented and distributed. The future research can also focus on the method to transcribe the recognized speech to the corresponding sentence on the instruction, based on a defined confident value. For example, if the instruction required the players to say "xiao ming" but "xia min" is detected instead, the system can intelligently correct it without returning error. With this ability, the children are less likely to become frustrated when their grammar is correct but pronunciation is unclear due to certain personal limitations.

Testing and evaluation: Future research is recommended to evaluate the down syndrome children's performance manually to identify the sufficiency of the feedback system using text highlighted with different colours, or if better approaches for the feedback system exist. This study focus on the children with mild down syndrome, who are known to have the learning capability as the typically developing children. Although not tested with down syndrome children, similar results is expected since children with mild down

syndrome can speak like normal children, and also attracted to things that attract normal children. The results from the evaluation of the participant's emotion can be used as an approximation for future research.

6.3 Conclusion

In conclusion, this study produces a game prototype for the Chinese-speaking children with mild down syndrome to practice the language with speaking. Although the game is intended for these children, it may be played by the typically developing children for entertainment or educational purpose since the contents are relatable with their daily routines. The educational contents and the methods to convey knowledge were planned by using the hybrid ASSURE and Gerlach-Ely ID model, whereas the game prototype was developed by conforming to the Waterfall SDLC model. This study has employed the standards for developing multimedia courseware for down syndrome children into RPG, and the feedbacks from participants are evaluated. The results support the claim that RPG is suitable for speak motivation as the participant's impression towards the game are mostly positive. Despite many limitations exist throughout the entire research, this study has successfully achieved the intended objectives. The research gaps are identified and improvements are recommended so that this research area can draw more attention for future investigation and experiments.

REFERENCES

Abbott, S., 2016. Student Engagement [online]. *The glossary of education reform*. Available at: http://edglossary.org/student-engagement/.

Adel, H., Vu, N.T., Kraus, F., Schlippe, T., Li, H., and Schultz, T., 2013. Recurrent Neural Network Language Modeling For Code Switching Conversational Speech Cognitive Systems Lab. *Computer*, pp.8411–8415.

AECT, 1977. *Educational Technology: Definition and glossary of Terms*. Association for Educational Communications and Technology.

Allison, F., Carter, M., and Gibbs, M., 2016. A History of Voice Interaction in Games. *Proceedings of 1st International Joint Conference of DiGRA and FDG*, pp.1–2.

Alonso, F., López, G., Manrique, D., and Viñes, J.M., 2005. with a blended learning process approach. *British Journal of Educational Technology*, 36 (2), pp.217–235.

Arısoy, E., Sainath, T.N., Kingsbury, B., Ramabhadran, B., Arisoy, E., Sainath, T.N., Kingsbury, B., and Ramabhadran, B., 2012. Deep Neural Network Language Models. *NAACL-HLT 2012 Workshop: Will We Ever Really Replace the N-gram Model? On the Future of Language Modeling for HLT*, pp.20–28.

Arora, R. and Arora, N., 2016. Analysis of SDLC Models. *International Journal of Current Engineering and Technology*, 6 (1), pp.2277–4106.

Aziz, N., Mohamad Roseli, N.H.M., and Abdul Mutalib, A., 2011. Visually Impaired Children's Acceptances on Assistive Courseware. *American Journal of Applied Sciences*, 8 (10), pp.1019–1026.

Baddeley, A., 2003. Working memory and language: An overview. *Journal of Communication Disorders*, 36 (3), pp.189–208.

Balaji, S., 2012. Waterfall vs v-model vs agile : A comparative study on SDLC. *International Journal of Information Technology and Business Management*, 2 (1), pp.26–30.

Blanco, J., Gabau, E., Gómez, D., Baena, N., Guitart, M., Egozcue, J., and Vidal, F., 1998. Chromosome 21 Disomy in the Spermatozoa of the Fathers of Children with Trisomy 21, in a Population with a High Prevalence of Down Syndrome: Increased Incidence in Cases of Paternal Origin. *The American Journal of Human Genetics*, 63 (4), pp.1067–1072.

Buckley, S., 1993. *Language development in children with Down syndrome - Reasons for optimism*.

Buckley, S., 2002. *Can children with Down syndrome learn more than one language?*

Bull, M.J., 2011. Clinical Report — Health Supervision for Children With Down Syndrome ABBREVIATIONS. *Journal of the American Academy of Pediatrics*, pp.393–406.

Burgoyne, K., Duff, F.J., Nielsen, D., Ulicheva, A., and Snowling, M.J., 2016. Bilingualism and Biliteracy in Down Syndrome: Insights From a Case Study. *Language Learning, A Journal of Research in Language Studies*, 66 (4), pp.945–971.

Callison, D., 2002. Instructional Models: (Part 3). *School Library Media Activities Monthly. Baltomore*, 19 (13), pp.36.

Chapman, R. and Hesketh, L., 2001. Language, cognition, and short-term memory in individuals with Down syndrome. *Down Syndrome Research and Practice*, 7 (1), pp.1–7.

Cleave, P.L., Bird, E.K., Trudeau, N., and Sutton, A., 2014. Disorders Syntactic bootstrapping in children with Down syndrome : The impact of bilingualism. *Journal of Communication Disorders*, 49, pp.42–54.

Cockburn, A., 2004. Revisiting 2D vs 3D Implications on Spatial Memory. *In Proceedings of the Fifth Conference on Australasian User Interface*, 28, pp.25–31.

Coggins, T., Carpenter, R.L., and Owings, N.O., 1993. Examining early intentional communication in Down syndrome children and non-retarded children. *British Journal of Disorders of Communication*, 18 (2), pp.98–106.

Cornillie, F., Clarebout, G., and Desmet, P., 2012. The role of feedback in foreign language learning through digital role playing games. In: *Procedia - Social and Behavioral Sciences*. pp.49–53.

Dehghan, M., Yadegari, F., Shirazi, S., Kazemnejad, A., and Sciences, R., 2014. Effect of Teaching Reading and Traditional Methods of Language Therapy on Grammatical Quotient of Children with Down Syndrome. *Zahedan Journal of Research in Medical Sciences*, 16 (5), pp.54–58.

Faryadi, Q., 2007. Instructional Design Models: What a Revolution.

Felix, V.G., Mena, L.J., Ostos, R., and Maestre, G.E., 2016. A pilot study of the use of emerging computer technologies to improve the effectiveness of reading and writing therapies in children with Down syndrome. *British Journal of Educational Technology*, (April).

Feltmate, K. and Kay-raining Bird, E., 2008. Language Learning in Four Bilingual Children with Down Syndrome : A Detailed Analysis of Vocabulary and, 32 (1), pp.6–20.

Forest, E., 2016. The Gerlach and Ely Design Model [online]. *Educational Technology*. Available at: https://educationaltechnology.net/gerlach-ely-design-model/ [Accessed 17 Aug 2017].

Frenkel, S. and Bourdin, B., 2009. Verbal, visual, and spatio-sequential short-term memory: Assessment of the storage capacities of children and teenagers with Down's syndrome. *Journal of Intellectual Disability Research*, 53 (2), pp.152–160.

Fung, P. and Schultz, T., 2008. TECHNOLOGIES AND APPLICATIONS. *IEEE SIGNAL PROCESSING MAGAZINE*, pp.89–97.

Galeote, M., Soto, P., Sebastián, E., Checa, E., and Sánchez-palacios, C., 2013. Early grammatical development in Spanish children with Down syndrome.pdf. *Journal of Child Language*, pp.1–21.

Gathercole, S.E. and Baddeley, A.D., 2014. *Working Memory and Language*. Psychology Press.

Grabowski, S., 2003. Teaching & Media: A Systematic Approach The Gerlach & Ely Model.

Graves, A., Geoffrey, H., and Abdel-rahman, M., 2013. SPEECH RECOGNITION WITH DEEP RECURRENT NEURAL NETWORKS. *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference*, (3).

Gustafson, K.L. and Branch, R.M., 2002. What is instructional design. In: *Trends and issues in instructional design and technology*. pp.16–25.

Hassan, N.W.M., Tahar, M.M., and Yasin, M.H.M., 2017. Teaching And Learning Using Software 'let's Reading' For Malay Language Subjects For Students With Learning Disabilities. *Jurnal Penelitian Dan Pengembangan Pendidikan Luar Biasa*.

Hong Kong Lutheran Social Service, 2014. 龍情訓練坊2 [online]. *Google Play*. Available at: https://play.google.com/store/apps/details?id=hk.hklss.android.lodac [Accessed 21 May 2017].

Huang, D.W., Diefes-Dux, H., Imbrie, P.K., Daku, B., and Kallimani, J.G., 2004. Learning motivation evaluation for a computer-based instructional tutorial using ARCS model of motivational design. *34th Annual Frontiers in Education*, pp.30–36.

Huang, J.T., Li, J., Yu, D., Deng, L., and Gong, Y., 2013. Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers. In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. pp.7304–7308.

Hughey, D., 2009. The Traditional Waterfall Approach [online].

Husni, E., 2013. Mobile Applications BIUTIS : Let ' s Study Vocabulary Learning as a Media for Children with Autism. In: *Procedia Technology*. Elsevier B.V., pp.1147–1155.

Isman, A., Caglar, M., Dabaj, F., and Ersozlu, H., 2005. A New Model for the World of Instructional Design : a New. *Turkish Online Journal of Educational Technology*, 4 (3).

Juang, B.H. and Rabiner, L.R., 2005. Automatic Speech Recognition – A Brief History of the Technology Development. *Elsevier Encyclopedia of Language and Linguistics*, pp.1–24.

Karakis, H., Karamete, A., and Okcu, A., 2016. The Effects of a Computer-Assisted Teaching Material, Designed According to the ASSURE Instructional Design and the ARCS Model of Motivation, on Students' Achievement Levels in a Mathematics Lesson and Their Resulting Attitudes. *European Journal of Contemporary Education*, 15 (1), pp.105–113.

Kay-Raining Bird, E., Cleave, P., Trudeau, N., Thordardottir, E., Sutton, A., and Thorpe, A., 2005. The Language Abilities of Bilingual Children With Down Syndrome. *American Journal of Speech-Language Pathology*, 14, pp.187–199.

Kennedy, J., Montassier, C., Lavalade, P., Pierre, U., Senft, E., and Belpaeme, T., 2017. Child Speech Recognition in Human-Robot Interaction : Evaluations and Recommendations. In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, pp.82–90.

Kenneth E., H.I., Andrew, Z., Ming, L., and Dinesh, M., 2001. Fast and simple 2D geometric proximity queries using graphics hardware. *Proceedings of the 2001 symposium on Interactive 3D graphics*, pp.145–148.

Khan, M.M. and Bayoumi, S., 2015. Multimedia as a Help for Children with Special Learning Needs.

Krause, K.-L. and Coates, H., 2008. Students' engagement in first-year university. *Assessment & Evaluation in Higher Education*, 33 (5), pp.493–505.

Kuh, G.D., Kinzie, J., Buckley, J.A., Bridges, B.K., and Hayek, J.C., 2007. Piecing Together the Student Success Puzzle: Research, Propositions, and Recommendations. *ASHE Higher Education Report*, 32 (5).

Kyritsis, M., Stevens, R., Kyritsis, M., and Gulliver, S.R., 2013. Issues and Benefits of Using 3D Intefaces: Visual and Verbal Tasks Issues and Benefits of Using 3D Intefaces: Visual and Verbal Tasks. *Proceedings of the Fifth International Conference on Management of Emergent Digital EcoSystems*, pp.241–245.

Lanfranchi, S., Carretti, B., Spanò, G., and Cornoldi, C., 2009. A specific deficit in visuospatial simultaneous working memory in Down syndrome. *Journal of Intellectual Disability Research*, 53 (5), pp.474–483.

Larochelle, H., Bengio, Y., Louradour, J., and Lamblin, P., 2009. Exploring Strategies for Training Deep Neural Networks. *Journal of Machine Learning Research*, 1, pp.1–40.

Lau, K.H., 2015. Self-Paced Speech Rehabilitation Among Children With Down Syndrome Using Malay Automatic Voice Recognition (AVR). Universiti Teknikal Malaysia Melaka.

Laws, G. and Hall, A., 2014. Early hearing loss and language abilities in children with Down syndrome. *International Journal of Language & Communication Disorders*, 49 (3), pp.333–342.

Lazar, J., 2007. *Universal usability: Designing computer interfaces for diverse user populations.* John Wiley & Sons.

Lee, T.T., 2011. Effectiveness of interactive multimedia module with pedagogical agent (IMMPA) in the learning of electrochemistry : A preliminary investigation. *Asia-Pasific Forum on Science Learning and Teaching.*, 12 (2), pp.1–24.

Liao, H., Pundak, G., Siohan, O., Carroll, M.K., Coccaro, N., Jiang, Q., Sainath, T.N., Senior, A., and Bacchiani, M., 2015. Large Vocabulary Automatic Speech Recognition for Children. *INTERSPEECH 2015*, pp.1611–1615.

Mahalakshmi, M. and Sundararajan, M., 2013. Traditional SDLC Vs Scrum Methodology – A Comparative Study. *International Journal of Emerging Technology and Advanced Engineering*, 3 (6), pp.2–6.

Mammarella, I.C., Pazzaglia, F., and Cornoldi, C., 2008. Evidence for different components in children's visuospatial working memory. *British Journal of Developmental Psychology*, 26 (3), pp.337–355.

Megaw, A.E., 2001. Deconstructing the Heinich, Moldena, Russell, and Smaldino Instructional Design Model.

Mohid, S.Z. and Mat Zin, N.A., 2010. Courseware Accessibility for Hearing Impaired. In: *2010 International Symposium on Information Technology*. IEEE, pp.26–30.

Næss, K.B., Lervåg, A., Lyster, S.H., and Hulme, C., 2015. Longitudinal relationships between language and verbal short-term memory skills in children with Down syndrome. *Journal of Experimental Child Psychology*, 135, pp.43–55.

Navarro-newball, A.A., Loaiza, D., Oviedo, C., Castillo, A., Portilla, A., Linares, D., and Álvarez, G., 2014. Talking to Teo : Video game supported speech therapy q. *Entertainment Computing*, 5 (4), pp.401–412.

Ng, K.H., Aryati, B., and Azizah, A.R., 2014. A Review on Courseware for Down Syndrome Children. *Journal of Information Systems Research and Innovation*, (8), pp.56–65.

Osman, K. and Lee, T.T., 2014. Impact of Interactive Multimedia Module With Pedagogical Agents on Students' Understanding and Motivation in the Learning of Electrochemistry. *International Journal of Science and Mathematics Education*, 12 (2), pp.395–421.

Peskett, R. and Wootton, A.J., 1985. Turn-taking and overlap in the speech of young Down syndrome children. *Journal of Mental Deficiency Research*, 29, pp.263–273.

Phillips, B.D., 1994. Roleplaying Games in the English As a Foreign Language Classroom. In: *Tenth National Conference on English Teaching and Learning in the Republic of China*. pp.625–648.

Prena, K., 2014. Down Syndrome Video Game Preferences. Michigan State University.

Protalinski, E., 2017. Google's speech recognition technology now has a 4.9% word error rate [online]. *VentureBeat*. Available at: https://venturebeat.com/2017/05/17/googles-speech-recognition-technology-now-has-a-4-9-word-error-rate/ [Accessed 16 Aug 2017].

Pulina, F., Carretti, B., Lanfranchi, S., and Mammarella, I.C., 2015. Improving spatial-simultaneous working memory in Down syndrome: Effect of a training program led by parents instead of an expert. *Frontiers in Psychology*, 6 (August), pp.1–9.

Ragunath, P., Velmourougan, S., Davachelvan, P., Kayalvizhi, S., and Ravimohan, R., 2010. Evolving A New Model (SDLC Model-2010) For Software Development Life Cycle (SDLC). *IJCSNS International Journal of Computer Science and Network Security*, 10 (1), pp.112–119.

Ramli, R. and Zaman, H.B., 2011. Designing usability evaluation methodology framework of Augmented Reality basic reading courseware (AR BACA SindD) for Down Syndrome learner. *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, ICEEI 2011*, (July).

Reale, M.J., Member, S., Liu, P., Member, S., Yin, L., and Member, S., 2013. Art Critic : Multisignal Vision and Speech Interaction System in a Gaming Context, 43 (6), pp.1546–1559.

Reiser, R.A., 2001. A history of instructional design and technology: Part I: A history of instructional media. *Educational Technology Research and Development*, 49 (53), pp.17–34.

Reiser, R.A. and Dempsey, J. V., 2012. *Trends and issues in instructional design and technology*. Boston: MA: Pearson.

Richards, D. and Taylor, M., 2015. A Comparison of learning gains when using a 2D simulation tool versus a 3D virtual world: An experiment to find the right representation involving the Marginal Value Theorem. *Computers & Education*, 86, pp.157–171.

Saon, G., Kurata, G., Sercu, T., Audhkhasi, K., Thomas, S., Dimitriadis, D., Cui, X., Ramabhadran, B., Picheny, M., Lim, L., Roomi, B., Hall, P., and Mar, C.L., 2017. English Conversational Telephone Speech Recognition by Humans and Machines.

Sarikaya, R., Hinton, G.E., and Deoras, A., 2014. Application of deep belief networks for natural language understanding. *IEEE Transactions on Audio, Speech and Language Processing*, 22 (4), pp.778–784.

Shima, N., Aziz, A., Fatimah, W., and Ahmad, W., 2014. User Experience Study on Mobile Numerical Application for Children with Mental Disabilities. In: *2014 4th World Congress on Information and Communication Technologies (WICT 2014)*. IEEE, pp.118–122.

South, J.B., 2008. Views and Use of Theory by Practicing Instructional Designers. Brigham Young University.

Tactus Therapy Solutions Ltd., 2017. Conversation Therapy [online]. *Google Play*.

Wei, Q., Li, H., and Zhou, X., 2016. The appropriate hidden layers of deep belief networks for speech recognition. In: *Proceedings - The 2015 10th International Conference on Intelligent Systems and Knowledge Engineering, ISKE 2015*. pp.397–402.

Witecy, B. and Penke, M., 2017. Language comprehension in children, adolescents, and adults with Down syndrome. *Research in Developmental Disabilities*, 62, pp.184–196.

Xiong, W., Droppo, J., Huang, X., Seide, F., Seltzer, M., Stolcke, A., Yu, D., and Zweig, G., 2016. Achieving Human Parity in Conversational Speech Recognition. *arXiv preprint*.

Xue, S., Jiang, H., Dai, L., and Liu, Q., 2016. Speaker Adaptation of Hybrid NN/HMM Model for Speech Recognition Based on Singular Value Decomposition. *Journal of Signal Processing Systems*, 82 (2), pp.175–185.

Ye, G., Liu, C., and Gong, Y., 2016. Geo-location dependent deep neural network acoustic model for speech recognition. In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pp.5870–5874.

Yussof, R.L., Shima, T.N., and Zaman, H.B., 2010. An Interaction Design for MEL-SindD : In: *2010 International Conference on User Science Engineering (i-USEr)*. IEEE Conference Publications, pp.116–121.

Zhang, Y. and Ng, A., 2013. Speech Recognition Using Deep Learning Algorithms, pp.1–5.

APPENDIX A

List of Media Contributors

1. Ayene (ayene-chan.deviantart.com)
2. bbcpersian7.com
3. freedesignfile.com
4. freepik.com
5. gen8.deviantart.com (gen8.deviantart.com)
6. myfreetextures.com
7. painhurt (painhurt.deviantart.com)
8. SchwarzeNacht (schwarzenacht.deviantart.com)
9. soundbible.com (sound effects)
10. Sprite Creator 3
11. vecteezy.com