THE EFFECTIVENESS OF MODALITY PRINCIPLE ON 3D TALKING HEAD LINGUISTIC LEARNING AID ON PRONUNCIATION AMONG TERTIARY LEVEL STUDENTS

KOGILATHAH SEGARAN

THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE
(MASTER BY RESEARCH)

ABSTRACT

This study aimed to investigate the effectiveness of various modality effects on 3D talking-head on pronunciation learning among tertiary level students. The research methodology is based on quasi-experimental study which investigates the effects of three different multimedia presentation strategies of 3D talking-head MALL on the learning achievement of students with low English pronunciation skills. The data is in the form of performance test scores of the pre-test and post-test. A number of 60 college students were selected as samples and they were divided into three groups of 20 students in each group. The scores obtained were analyzed statistically with one-way analysis of covariance (ANCOVA), with the pre-test scores as covariate. After adjusting for pre-test scores, there was a significant difference on post-test achievement of 3D talking-head with audio and text MALL in comparison with both 3D talking-head with audio alone MALL and audio and text alone MALL application. The adjusted mean scores indicate that students in the 3D talking-head with audio and text MALL application obtained a better mean score than students in the 3D talking-head with audio alone MALL application and audio and text alone MALL application. Therefore, it can be concluded that multiple sources of information to identify and decode language input are advantageous for effective pronunciation learning among non-native speakers.

TAJUK: Keberkesanan Modaliti dalam 3D "Talking-Head" terhadap Pembelajaran Sebutan dalam Kalangan Pelajar Peringkat Pengajian Tinggi.

ABSTRAK

Kajian ini bertujuan untuk melihat keberkesanan pelbagai kesan modaliti dalam 3D *talking-head* terhadap pembelajaran sebutan dalam kalangan pelajar peringkat pengajian tinggi. Metodologi kajian adalah berdasarkan pendekatan kuasi-eksperimen yang melihat kesan tiga strategi persembahan multimedia yang berbeza dalam 3D *talking-head* MALL terhadap pencapaian pelajar dengan kemahiran sebutan Bahasa Inggeris yang rendah. Data adalah dalam bentuk skor pencapaian ujian pra dan ujian pos. Seramai 60 pelajar kolej telah dipilih sebagai sampel kajian dan mereka telah dibahagikan kepada tiga kumpulan dengan 20 pelajar dalam setiap kumpulan. Skor yang diperolehi telah dianalisis dengan analisis sehala kovarians dengan skor ujian pra sebagai kovariat. Selepas pelarasan bagi skor ujian pra, terdapat perbezaan yang signifikan terhadap pencapaian ujian pos 3D *talking-head MALL* dengan audio dan teks berbanding dengan 3D *talking-head MALL* audio sahaja dan *MALL* dengan audio dan teks sahaja. Skor min yang diselaraskan menunjukkan bahawa pencapaian pelajar-pelajar 3D *talking-head MALL* dengan audio dan teks adalah lebih baik berbanding pelajar dalam 3D *talking-head MALL* dengan audio sahaja dan *MALL* dengan audio dan teks sahaja. Oleh demikian, kesimpulan boleh dilakukan bahawa pelbagai sumber maklumat untuk mengenal pasti dan *decode input* bahasa adalah berfaedah untuk pembelajaran sebutan yang berkesan dalam kalangan bukan penutur asli.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

PUBLICATIONS

Kogilathah, S., Ahmad Zamzuri, M. A., & TanW. H. (2013). Talking-head animation as pedagogical agent in language learning: A review on instructional strategy and media. *Malaysian Journal of Distance Education 15*(1), 55−71.

Ahmad Zamzuri, M. A., & Kogilathah, S. (2013). 3D talking-head mobile app: A conceptual framework for English pronunciation learning among non-native speakers. *English Language Teaching, 6*(8), 66-76.

Kogilathah, S., Ahmad Zamzuri, M. A., & TanW. H. (2014). Usability and user satisfaction of 3D talking-head Mobile Assisted Language Learning (MALL) for non-native speakers. *Procedia-Social and Behavioral Sciences, 131*(2014), 4-10.

Kogilathah, S., & Ahmad Zamzuri, M. A. (2012). Effects of 3D talking-head mobile phone application in improving English pronunciation among tertiary level students: A research proposal. *Proceedings of International Conference on Integrated Knowledge ICIK2012,* Malaysia, 78-89.

CHAPTER 1

INTRODUCTION

1.0 Introduction

It is widely acknowledged that multimedia plays an important role in linguistic learning, predominantly through animation applications. Such tools make significant contributions to the language learning process among various age groups of learners (Tamburini & Paci, 2002). This is particularly true of 3D talking-head animations, which serve as a virtual teacher or guide in many computer-assisted language learning applications (Wik, 2011; Wik & Hjalmarsson, 2009; Voce & Hamel, 2001). 3D talking-heads serve as an essential instructional material for supporting language learning, mostly in terms of pronunciation aspects among non-native speakers (Badin, Tarabalka, Elisei, & Bailly, 2010).

Unfortunately, non-native speakers face difficulty in using English as a second language due to poor pronunciation skills (Fraser, 2000). The difficulty arises mostly among those who only pay serious attention to learning English after the school years (Gilakjani & Mohammad Reza, 2011). The problem seems even more critical among Malaysian tertiary graduates who will face job searching challenges following the graduation, where effective English communication skill is crucial to determine their career success either locally or internationally (Azizan & Mun, 2011a). A survey conducted by Kelly Services (M) Sdn. Bhd., known as the Kelly Global Workforce Index, has revealed that communication skills as one of the top five most desired skills required by the corporate sectors (Azizan & Mun, 2011b). A survey from the Malaysian Employers Federation (MEF) has also revealed similar findings (Figure 1.1). This clearly emphasizes the importance of the skills, particularly for career development purpose.

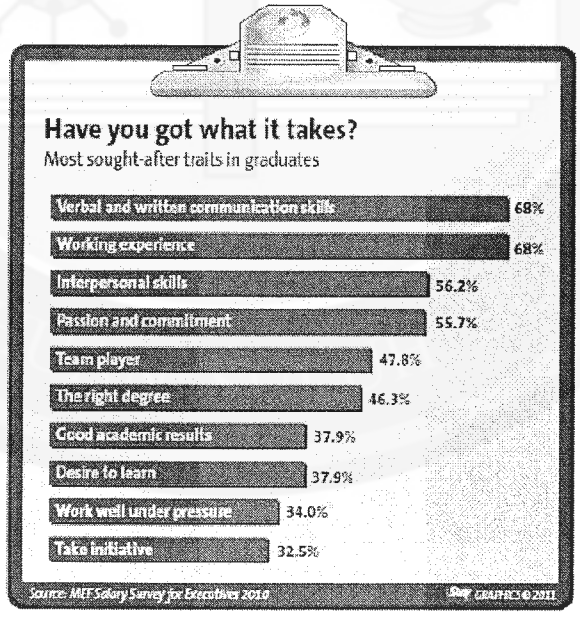


Figure 1.1. Most Sought-After Traits in Graduates (Azizan & Mun, 2011b)

The standard of English among Malaysian youngsters is actually declining; the communication skills among school leavers is viewed poor by employers. Study revealed that an average of six out of ten Malaysian graduates could not convey the message effectively in interviews due to poor English communicative skills (Azizan & Mun, 2011b). Many steps have been taken to address this issue in terms of education, including collaborating language learning and multimedia, collaborating language learning and technology, and so on. With today's rapidly-evolving mobile technology, m-Learning is a new approach in aiding education.

Academia and the mobile world have begun to forge a new path for language learning which has lead to the introduction of Mobile Assisted Language Learning (MALL) (Kukulska-Hulme & Shield, 2008). Throughout the years, a great deal of research, studies, and forum discussions on the topic of MALL have been performed to identify an effective way of implementing such benefits into for successful language-learning programs. Therefore, this research is principally aimed to identify the effective way of incorporating 3D talking-head animation in MALL, specifically in assisting English language learning as a second language among tertiary level students. The focus of the study will mainly be on pronunciation skills, as in acquiring strong English communication skills, pronunciation plays an important role in shaping the language (Wei, 2006; Derwing, 2003)

1.1 Background of Study

Debate on animation typically relates to movies, cartoons or special effects. There have also been many positive results showing that animation makes a significant contribution to the education field (Balasubramanyam, 2012; McMenemy & Ferguson, 2009; Doyle, 2001). Studies showed that implementation of animation in learning have led to promising outcome for decade (Williamson & Abraham, 1995). Animation has a potential role in improving human learning process, particularly in promoting a deeper understanding of the subject matter (Mayer & Moreno, 2002). Animations have become incorporated as part of computer-based multimedia learning aids in many subject matters, including language learning (Cheng Lin & Fang Tseng, 2012; Kayao□lu, Da□ Akbaş & Öztürk, 2011; Sundberg, 1998).

Notably, numerous initiatives were undertaken in establishing English as second language acquisition globally, which eventually resulted in introduction of Computer-assisted Language Learning (CALL). In CALL, 3D embodied agent or 3D talking-head animation becomes the prominent virtual aid for teaching pronunciation, vocabulary, articulation, and so forth (Engwall & Bälter, 2007; Wik, 2004; Wik & Hjalmarsso, 2009). However, similar applications seem lacking in m-Learning initiative, particularly those focusing on teaching pronunciation.

No doubt, good English communication starts with proper pronunciation (Saran, Seferoglu & Cagiltay, 2009). Improving pronunciation has positive effect on increasing the overall communicative power (Gilakjani & Mohammad Reza, 2011).

Learners who begin to learn English after their school years typically experience severe difficulties in mastering lucid pronunciation. This difficulty increases as they age (Gilakjani & Mohammad Reza, 2011). There are several factors that might influence the acquisition of proper pronunciation, such as speaking native language which involves an accent, the age of the learners, learners' motivation, the opportunity for the learners to learn, the learners attitude towards learning pronunciation, the time the learners have, and several other causes (Gilakjani & Mohammad Reza, 2011). For that rationale, various approaches must be considered to facilitate pronunciation skills development among various level of learners, specifically on mobile platforms, which appear to be a must-have tool among contemporary tertiary level students.

As stated earlier, Mobile Assisted Language Learning (MALL) has developed as a new teaching aid in the educational field in recent years (Chinnery, 2006). Numerous researchers have noted that in-class activities are not sufficient for effective language learning. Learners should be given opportunities to learn the language beyond simple classroom activities (Saran, Seferoglu & Cagiltay, 2009). Advancement in technology, particularly mobile technologies, has paved a new path for educational improvement in distance language learning via mobile technology. Gaining evidence from research and practice clearly indicates the potential of mobile technologies as effective learning and communication tools by a wide range of learners in a mixture of settings (Kukulska-Hulme, 2010).

Mobile learning has given the luxury of learning anytime and anywhere, a concept known as ubiquitous learning, for any learning style of students (Godwin Jones, 2005; Kadyte, 2004; Kukulska-Hulme, 2005). The development of mobile and wireless technologies had opened up a huge array of possibilities in the domain of language teaching too (Joseph & Uther, 2009).

Studies have been carried out to demonstrate that mobile learning can enhance and ease the learning process for different groups of learners, especially in MALL (Idrus, 2011; Sood, 2010; Fotouhi-Ghazvini, Earnshaw, Haji-Esmaeili, 2009). Several methods are practicable in MALL, including the use of multimedia messages in improving pronunciation (Saran, Seferoghu & Cagilitay, 2009). Multimedia has made a significant contribution in MALL. In recent studies, researchers explored the usage of multimedia messages (MMS) via mobile phones for improving pronunciation of words. For instance, a study by Saran, Seferoglu and Cagiltay (2009) looked on the effectiveness of three different mode of delivery, namely web base, handouts and MMS. They found that students in the MMS delivery mode outperformed students in handout and web base delivery mode. However, application of animation, particularly 3D talking-heads, is mostly computer based. There has been research done on 3D talking-heads using mobile phones as voice interactive services, but there has been little research done on the use of 3D talking-heads as pronunciation learning aids. This study will fill the gap by focusing on the effectiveness of 3D talking-head mobile applications to aid pronunciation learning.

Basically, the term talking-head refers to a computer generated animated character, lifelike video character, or just a simulated person on a website that is able to talk and hold a conversation with human users (Lun, n.d). The talking-head is also used by graphic designers working on facial animation and facial expressions combined with audio-visual speech processing (Lun, n.d). 3D talking heads are widely used for web services or as agents to replace real humans (Lun, n.d). Also, these talking-heads will appear as virtual tutor or teacher in learners' computers and eventually contributes in various aspects of language learning process such as from reading to pronunciation and to conversation and practice (Busa, 2008). Talking-heads have also become a tool for children learning their first language and for disabled people such as the deaf (Busa, 2008). It is due to the realistic speech and expressions, and the convincing emotions applied on the talking-heads which results in patient and exciting interactive tutors for learners to learn the second languages efficiently (Massaro, 2006b).

On the other hand, character which is designed with non realism appearance has its own advantage. A user usually would like to have an eye-catching, expressive face with easy to identify unique communicational, cognitive and emotional expressions such as paying attention, agreeing and joy (Ruttkay & Noot, 2000). Nevertheless, users tend to feel creepy when a virtual character looks too human like or too realistic (MacDorman, Green, Ho, & Koch, 2009). Nowadays, computer graphic characters are designed to look like real people and it has become less convincing, such as the computer graphic heroes in The Polar Express and Final Fantasy: The Spirits Within (Geller, 2008; Pollick, 2009).

This was illustrated by a famous graph produced in 1970 by a Japanese roboticist, Masahiro Mori, known as the Uncanny Valley.

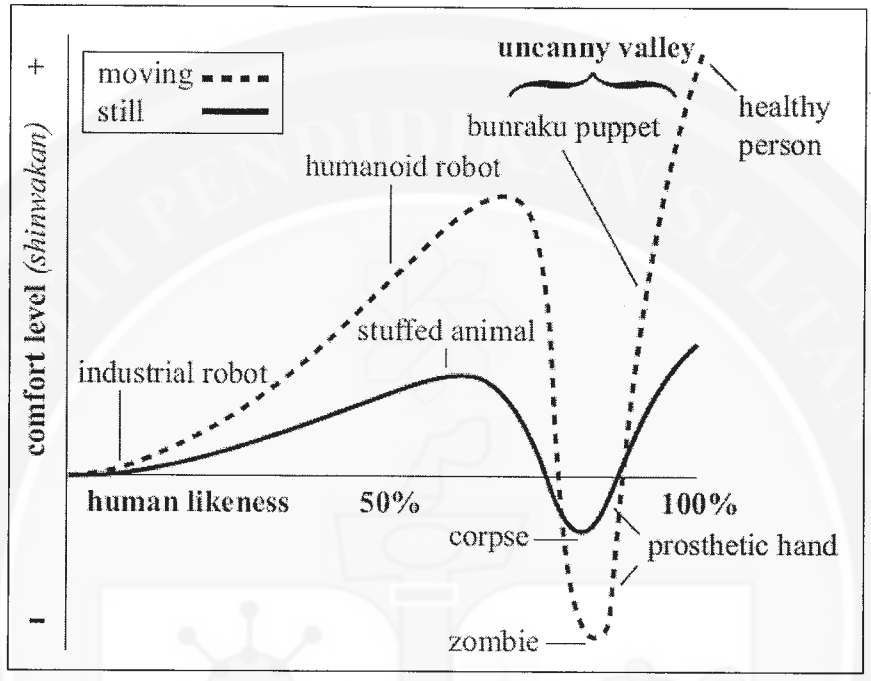


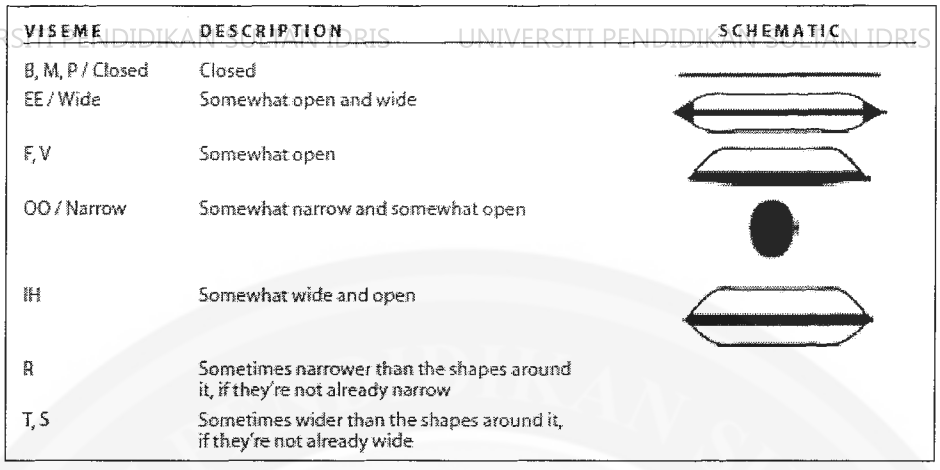
Figure 1.2. Masahari Mori's graph of The Uncanny Valley (MacDorman, Green, Ho, & Koch, 2009)

Figure 1.2 shows Masahiro Mori's proposed relation between human likeness and the comfort level in which it explains when robots look too human, even the slightest flaws will make it look creepy (MacDorman, Green, Ho, & Koch, 2009). Mori also added that this situation deepens when movements are added to the character (MacDorman, Green, Ho, & Koch, 2009). Adapting these findings, the 3D talking-head model for this research will be developed into non-realistic facial proportions yet with proper modeled features of human to show the movement of the lip.

It is believed to give a realistic environment and comfortable experience to the user of the application.

The main feature of a talking-head is lip synchronization (Lun, n.d). Lip synchronization, also known as lip sync, is one of the important parts of any animated character for speech (Arkinson, n.d.). Numerous research and experiments have been conducted to improve the art of lip synching (Lewis, 1991). There are several methods of creating lip sync, including metamorphoses, automated lip sync, rotoscoping through obtaining live-action footage of actors performing the desired motion, and adopting a canonical mapping from a subset of speech sounds onto corresponding mouth positions (Gueorguiev & Velcheva, 2005; Frank, Hoch, Trogemann, 1997; Lewis, 1991). Metamorphoses lip sync method is carried out by replacing one mouth with another mouth shape; meanwhile, automated lip sync uses special plug-in in 2D or 3D animation software to synchronize the lip movement with audio. This research believes that to ease implementation and increase the accuracy of lip syncing, usage of automated lip sync technique combined with recorded human voices would be the best choice.

There are two things needed to create lip sync, which are phonemes and visemes (Osipa, 2010). Visemes are visual phonemes or usually known as shapes that represent open/close/narrow wide mouth, meanwhile phonemes are sounds that are created when we speak (Osipa, 2010). For this research purpose, both visemes and phonemes must be examined, as pronunciation must consider both the movement of the mouth and also the sound of a word as it is pronounced (Osipa, 2010).



VISEME	DESCRIPTION
B, M, P / Closed	Closed
EE / Wide	Somewhat open and wide
F, V	Somewhat open
OO / Narrow	Somewhat narrow and somewhat open
IH	Somewhat wide and open
R	Sometimes narrower than the shapes around it, if they're not already narrow
T, S	Sometimes wider than the shapes around it, if they're not already wide

Figure 1.3. Viseme representation on an open/closed narrow/wide mouth (Osipa, 2010)

Lip synching has several contributions in language learning especially in improving pronunciation. English is a language which depends upon; airflow, lip shape, tongue position, teeth position and jaw movement (Baxter, 1993), where the process can be practiced through watching the lip syncing activities (Sumby & Pollack, 1954; Benoı & Le Goff, 1998). Apart from these, in recent studies on learning pronunciation, new methods and methodological prediction have been introduced such as using nonverbal features and full-frontal communicativity (Rodgers, 2001). Full-frontal communicativity is one of the 10 scenarios introduced by Rodgers (2001), engaging all aspects of human communicative capacities such as facial expression, gesture, tone and so forth to shape the teaching of second language (Rodgers, 2001). Moreover, speech is enriched by facial expressions, emotions, and gestures produced by a speaker (Massaro, 1998).

In current usage of talking-head especially in CALL, facial expressions are also necessary to make language learning more efficient and robust (Wik & Hjalmarsson, 2009). Previous studies in the field of neuroscience, cognitive science, and psychology have demonstrated that emotions have a significant role in attention, planning, reasoning, learning, memory, and decision making (Picard, 1997). Emotions also play a role as motivator that influences perception, cognition, coping, and creativity (Johnson, Rickel & Lester, 2000; Picard, 1997).

Considering these views, this research aims to investigate the effectiveness of 3D talking-head with facial expression, lip sync, audio, and text MALL in improving pronunciation learning among tertiary level learners.

1.2 Theoretical Framework

The theoretical framework of this study is grounded on Mayer's Cognitive Theory of Multimedia Learning. According to Mayer's cognitive theory of multimedia learning, human processes information through dual channels, namely the visual channel that processes visually represented materials and the verbal channel that processes audio and text materials (Mayer, 2001). Mayer (2001) stated that human understanding occurs when learners are able to mentally integrate visual and verbal representations of a subject matter, as both channels are activated simultaneously. Figure 1.3 outlines Mayer's cognitive theory of multimedia learning regarding how information is processed in human memory.

This illustration actually draws on Paivio's (1986), Clark and Paivio (1991) dual coding theory, Baddeley's (1992) model of working memory, Sweller's (Chandler & Sweller, 1991; Sweller, Chandler, Tierney & Cooper, 1990) cognitive load theory, Wittrock's (1989) generative theory, and Mayer's (1996) SOI model of meaningful learning.

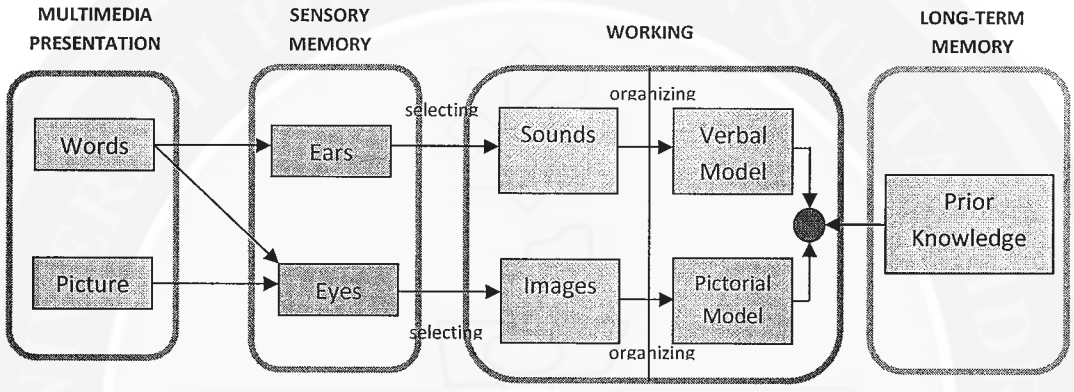


Figure 1.4. A framework for cognitive theory of multimedia learning drawn from Mayer (2001, p. 44)

There are three assumptions that support Mayer's (2001) cognitive theory of multimedia learning. The first assumption is that multimedia learning is a *dual channel activity*, consisting of a visual-pictorial channel and an auditory-verbal channel. For example, a 3D talking-head animation will be processed in the visual-pictorial channel and the pronounced word will be processed in the auditory-verbal channel. The second assumption is *limited capacity*, in which each channel in the human cognitive system has limited capacity for processing information.

The third assumption is *active processing*, in which learners are involved in active processing in the channels, including media selection (words and pictures), organizing media into the verbal and pictorial mental models, and finally integrating them with preexisting knowledge which results in meaningful schema acquisition. This happens when corresponding verbal and pictorial representations are in working memory at the same time (Mayer, 2002). The issue of integrating visual and audio information in order to retain it in the long term memory is an important 3D talking-head learning condition. The application must be able to assist learners to integrate the visual form of the 3D talking-head with facial expressions and lip movement with the audio. Likewise, they are able to store knowledge acquired from both sensory memory (listening and watching the 3D talking- head) and working memory (integrating 3D talking-head and the pronounced word) in the long term memory and apply it precisely.

Mayer's modality principle was also applied in this research. According to the modality principle as suggested by Mayer (2001), students are better able to acquire knowledge when a multimedia message is spoken, rather than simply displayed as on-screen text. Moreover, when pictures and text are both visually presented, the visual channel will be overloaded, consequently leaving the verbal channel unused (Mayer, 2001). Therefore, if text is presented in audio form, it can be processed by the verbal channel, while the visual channel processes visual information (Mayer, 2001). However, the question arises as to why this principle applies for 3D talking-head linguistic learning aids. Text might be helpful in assisting learners in determining the syllable break for correct pronunciation. Further study is needed to address this issue.

By adapting these theories and principles per the literature overview, the researcher proposes a conceptual framework of 3D talking-head mobile device application for the purpose of pronunciation learning, as depicted in Figure 1.5.