3D Talking-Head Mobile App: A Conceptual Framework for English Pronunciation Learning among Non-Native Speakers

Ahmad Zamzuri Mohamad Ali[1] & Kogilathah Segaran[1]

[1] Faculty of Art, Computing and Creative Industry, Universiti Pendidikan Sultan Idris, Tanjong Malim, 35900, Perak, Malaysia

Correspondence: Ahmad Zamzuri Mohamad Ali, Faculty of Art, Computing and Creative Industry, Universiti Pendidikan Sultan Idris, Tanjong Malim, 35900, Perak, Malaysia. Tel: 6012-4777-395. E-mail: zamzuri@fskik.upsi.edu.my

Abstract

One of the critical issues pertaining learning English as second language successfully is pronunciation, which consequently contributes to learners' poor communicative power. This situation is moreover crucial among non-native speakers. Therefore, various initiatives have been taken in order to promote effective language learning, which includes 3d talking-head on mobile technology. 3D talking-head appeared to be a sufficient instructional material in supporting language learning, mostly in pronunciation aspects among non-native speakers. On this regard, this paper proposed a conceptual framework for 3D talking-head Mobile Assisted Language Learning (MALL), specifically for pronunciation learning. The framework was developed based on theories, principles and literature overview done. The paper also suggests potential studies to further affirm the framework developed.

Keywords: animation, m-learning, pronunciation, talking-head

1. Introduction

It was globally known that multimedia had played an important role in linguistic learning, predominantly the animation applications. It has significant contribution in the language-learning process among various age groups of learners (Tamburini & Paci, 2002); particularly, the 3D talking-head animation, which appeared to be the virtual teacher in many computer-assisted language-learning applications (Wik, 2011; Wik & Hjalmarsson, 2009; Voce & Hamel, 2001). 3D talking-head seems to be an essential instructional material in supporting language learning, mostly in pronunciation aspects among non-native speakers (Badin, Tarabalka, Elisei, & Bailly, 2010).

This condition is true, since it is a norm for non-native speakers facing difficulty in using English as second language due to poor pronunciation skills (Fraser, 2000). The difficulty arises, for the most part, among those who only pay serious attention in learning English after the school years (Gilakjani & Mohammad Reza, 2011). This problem seems moreover 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. namely, Kelly Global Workforce Index, revealed that communication skills as one of the top five most desired skills required by the corporate sectors (Azizan & Mun, 2011b). A survey done by Malaysian Employers Federation (MEF) also revealed similar findings (Figure 1). This clearly specifies the importance of the skills, in particular, for career development purpose.
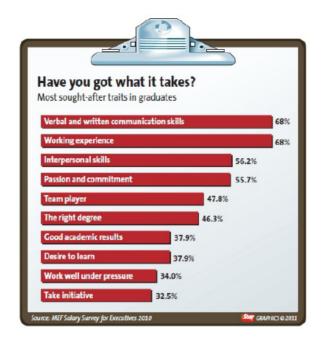
Figure 1. Most Sought-After Traits in Graduates (Azizan & Mun, 2011b)

The standard of English among Malaysian youngsters is actually declining; the communication skill among school leavers is viewed poor by employers. Study revealed that an average of six out of ten Malaysian graduates could not convey the massage effectively in interviews due to poor English communicative skills (Azizan & Mun, 2011b). Many efforts have been taken to address this issue in education aspects, which includes collaborating language learning and multimedia, collaborating language learning and technology and so forth. Today, with rapidly evolving mobile technology, m-Learning appears to be a new approach in aiding education field as well. Academia and mobile world have paved a new path for language learning, which has led in the introduction of Mobile Assisted Language Learning (MALL) (Kukulska-Hulme & Shield, 2008). Throughout the years, much research, studies and forum discussions about the topic of MALL have been carried out to identify the effective way of implementing and benefiting it for successful language learning.

Therefore, this paper is principally aimed to propose a conceptual framework for an effective way of incorporating 3D talking-head animation in MALL, specifically in assisting English language learning as second language among tertiary level students. The focus in the paper will mainly be on pronunciation skills, whereby in acquiring good English communication skills, pronunciation plays an important role in shaping the language (Wei, 2006; Derwing, 2003)

2. Animation and MALL

Debate on animation typically relates to movies, cartoons or special effects. Though, there were many positive results showing that animation has significant contribution within the education field too (Balasubramanyam, 2012, McMenemy & Ferguson, 2009; Doyle, 2001). Studies show that implementation of animation in learning has led to promising outcome since decades ago (Williamson & Abraham, 1995). Animation plays a potential role in improving the human learning process, particularly in promoting profound understanding of the subject matter (Mayer & Moreno, 2002). Nowadays, animations are incorporated as part of computer-based multimedia learning aid in many subject matters, including in language learning (Cheng Lin & Fang Tseng, 2012; Kayaoğlu, Dağ 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 in teaching pronunciation, vocabulary, articulation and so forth (Engwall & Bälter, 2007; Wik, 2004; Wik & Hjalmarsso, 2009). However, similar applications seem infrequent in m-Learning initiative; particularly focusing on teaching pronunciation.

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

Mohammad Reza, 2011). Learners who start to learn English after school years usually have severe difficulties in mastering lucid pronunciation and the difficulty increases as they age (Gilakjani & Mohammad Reza, 2011). There are a number of factors that might influence the acquisition of proper pronunciation such as speaking the native language which involves the accent, the age of the learners, motivation the learners receive to learn, 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 need to be considered in order to facilitate pronunciation skills development among different levels of learners, specifically on the mobile platform, which appears to be a must have tool among tertiary level students nowadays.

As pointed out earlier, Mobile Assisted Language Learning (MALL) appears to be a new teaching aid in the educational field during recent years (Chinnery, 2006). Numerous researchers believe that just in-class activities are not sufficient for effective language learning. Learners should be given opportunities to learn the language beyond the classroom activities (Saran, Seferoglu & Cagiltay, 2009). Advancement in technology, particularly mobile technologies, actually has paved 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, anywhere or well 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 opened up a huge array of possibilities in the domain of language teaching too (Joseph & Uther, 2009). Researches and studies have been carried out continually to prove that mobile learning could enhance and ease the learning process for different groups of learners, especially in MALL (Idrus, 2011; Sood, 2010; Fotouhi-Ghazvini, Earnshaw, Haji-Esmaeili, 2009).

3. 3D Talking-Head

Multimedia in point of fact has significant contribution in MALL. Several methods are practicable in MALL, such as usage of multimedia messages in improving pronunciation and so forth (Saran, Seferoghu & Cagilitay, 2009). In recent studies, researchers explored the usage of multimedia messages (MMS) via mobile phones for improving pronunciation of words. For instance, study by Saran, Seferoglu and Cagiltay (2009), which looked at 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-head in assisting language learning is mostly computer based. There are research done on the 3D talking-head on mobile phones as voice interactive services but there is scarce research done on the use of the 3D talking-head as pronunciation learning aid. Therefore, study filling the gap with concentration on the effectiveness of 3D talking-head mobile application in aiding pronunciation learning seems vital.

Basically, the term talking-head refers to a computer generated animated character, lifelike video character or just a person on a website that can 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 expression combined with audio-visual speech processing (Lun, n.d). 3D talking-heads are widely being used nowadays for web services or as an agent to replace real human (Lun, n.d). Besides, these talking-heads will appear as virtual tutor or teacher in learners' computers and eventually contributes in various aspects of the language-learning process such as from reading to pronunciation and to conversation and practice (Busa, 2008). Talking-heads are also becoming a tool for children learning their first language and disabled people such as the deaf to be specific (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, 2006a; 2006b).

On the other hand, character which is designed with non realism appearance has its own advantages. 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 has been illustrated through a graph produced by a Japanese roboticist, Masahiro Mori in 1970 namely the Uncanny Valley.
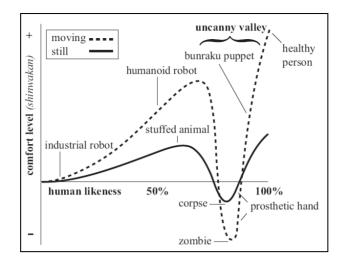
Figure 2. Masahari Mori's Graph of The Uncanny Valley (MacDorman, Green, Ho, & Koch, 2009)

Figure 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 deepened when there are movements added to the character (MacDorman, Green, Ho, & Koch, 2009). Adapting to these findings, the 3D talking-head model developed with non-realistic facial proportion yet with proper modeled features of human to show the movement of the lip is believed to promote comfortable experience and realistic environment for the user of the application.

The main feature of a talking-head would be the lip synchronization (Lun, n.d.). Lip synchronization or well known as lip sync is one of the important arts in any animated character in order to make the artificial character speak (Arkinson, n.d.). Numerous research and experiments have been carried out in order to improve the art of lip synching (Lewis, 1991). There are several methods of creating the lip sync such as 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 done 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. The researcher believes that to ease and to meet the accuracy of lip syncing, usage of the automated lip-sync technique combining with recorded human voice might be a better choice.

There are two things needed in order to create lip syncs, 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 need to be looked into as pronunciation needs the movement of the mouth and also how a word sounds like when we pronounce it (Osipa, 2010).

VISEME	DESCRIPTION	SCHEMATIC
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 3. The Visemes' 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; Benoıˆ & 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), which engages 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 the facial expressions, emotions and gestures produced by a speaker (Massaro, 1998). In current usage of the talking-head, especially in CALL, facial expression is also among the concern to make the language learning more efficient and robust (Wik & Hjalmarsson, 2009). Previous studies within the field of neuroscience, cognitive science, and psychology specify that emotions have a significant role in attention, planning, reasoning, learning, memory, and decision making (Picard, 1997). Emotions also play the role as motivator that influences perception, cognition, coping, and creativity (Johnson, Rickel & Lester, 2000; Picard, 1997).

4. Theoretical Framework

The theoretical framework of this study is grounded on Mayer's Cognitive Theory of Multimedia Learning and Constructivist Learning Theory. According to Mayer's cognitive theory of multimedia learning, human processes information through the dual channels; which is the visual channel that processes visually represented materials and the verbal channel that process audio and text materials (Mayer, 2001). Mayer (2001) believes that human understanding occurs when learners are able to mentally integrate visual and verbal representations of a subject matter as both channels being activated simultaneously. Figure 4 outlines Mayer's cognitive theory of multimedia learning on how the information processed in human memory. This illustration actually draws on Paivio's (1986; Clark & 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.
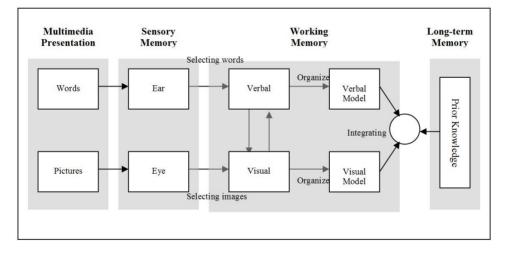

Figure 4. A framework for cognitive theory of multimedia learning drawn from Mayer (2001)

There are three assumptions that support Mayer's (2001) cognitive theory of multimedia learning. First assumption states that multimedia learning is *dual channel activities*, which are visual-pictorial channel and auditory-verbal channel. For example, 3D talking-head animation will be processed in the visual-pictorial channel and the pronounced word will be processed in the auditory-verbal channel. Second assumption is *limited capacity*, where each channel in the human cognitive system has limited capacity in processing information. The third assumption is the *active processing*, where learners are involved in active processing in the channels, which includes media selection (words and pictures), organizing the media into the verbal and pictorial mental model 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 important in 3D talking-head learning condition. The application must have the capability in assisting learners to integrate the visual form of the 3D talking-head with facial expression and lip movement with the audio. Likewise, they are able to store the knowledge acquired from the 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.

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

Meanwhile, the constructivist learning theory approach claims that learning is an active, creative, and socially interactive process in which learners constructs new ideas based upon their current and past knowledge (Bruner, 1990). This theory also applies for language learning in which successful language learning is achieved through learners' exposure level towards the language and also their genuine interaction with the language (Pemberton, Fallahkhair & Masthoff, 2004). In a constructivist-inspired program, learners would be required to perform tasks and solve problems involving listening, reading, writing and speaking in the foreign language that ensures a high level of communication (Pemberton, Fallahkhair & Masthoff, 2005). The constructivist viewpoint is often strongly associated with communicative teaching approaches, and this has been the basis for many initiatives in interactive computer-assisted language learning (Pemberton, Fallahkhair & Masthoff, 2005). Even though constructivist learning theory is one of the traditional learning theories, thus far it still has the fine impact on new innovative teaching and learning method such as mobile learning (Craig & van Lom, 2009). It enables the mobile technology to focus on students' ability to be self-directed and to draw conclusions on the knowledge acquired on their own (Karagiorgi & Symeou, 2005). In this context, learners are given the luxury to learn at their own pace and to construct a new way of learning pronunciation by watching, listening and pronouncing the word using 3D talking-head mobile applications. This theory also will direct the study by allowing the targeted learners to focus on their ability to learn without the guidance of an actual teacher at their own learning pace, anytime and anywhere (Pemberton, Fallahkhair & Masthoff, 2004). As 3D talking-head is a computer simulation of an actual human, it can be associated with the constructivist theory since constructivist approach allows the students to relate the knowledge acquired from the simulation with a similar situation in the real world (Hoek, 2009). For example, the pronunciation exercise experience that the students acquire from the 3D talking-head mobile application can be applied in the real world when they come across in pronouncing the same words when needed.

5. Conceptual Framework

In sum, by adapting to the theories, principles and the literature overview, the paper proposes a conceptual framework of 3D talking-head mobile phone application for the purpose of pronunciation learning as depicted in Figure 5.
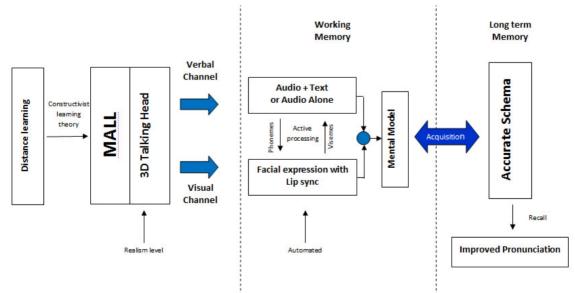
Figure 5. A conceptual framework for 3D talking-head mobile phone application for pronunciation learning

Based upon the conceptual framework developed, it is believed that the constructivist learning theory approach assists MALL by making the whole process as an active, creative, and socially interactive. Whereby, learners construct new ideas based on their past and current knowledge attained throughout the pronunciation practice. Learners are given the comfort to learn on their own pace and construct a new way of learning pronunciation by watching, listening and pronouncing words referring to the 3D talking-head on mobile devices. This approach also directs the study in allowing the targeted learners to focus on their ability to learn without the guidance of a real teacher on their own learning pace, anytime and anywhere. In meantime, the aspect of non-realistic yet proper modeled features of human like 3D talking-head also plays an important role in eliminating creepy feeling and promoting comfortable condition in learning.

When the learners utilize 3D talking-head as their pronunciation assistant, they perceive the practice in working memory structure via two channels namely visual channel and verbal channel. Facial expression with automated lip-sync strategy of 3D talking-head animation will be processed in the visual channel, while the pronounced word and text will be processed in the verbal channel concurrently. By utilizing both channels, the issue of limited capacity could be minimized. However, question arises on the usage of text on the application. Even so, text is categorized as verbal information; it is represented in a visual form. The viewers need additional effort in transforming it from visual representation to verbal, whereby extra processing is needed in the working memory. Nevertheless, excluding text in pronunciation learning might cause difficulties among learners in identifying the syllable brakes for proper pronunciation. Thus, studies identifying the solution to this issue are important.

Besides addressing limited capacity issue, dual-channel activation has potential in promoting active processing in the working memory. When corresponding visual and verbal representations are processed simultaneously in the working memory, referential or activation of verbal system by visual system or vice versa will happen. The learner at this point might develop a more accurate mental model of the pronounced word. This will help the learner to process the knowledge acquired from the sensory memory through listening and watching the 3D talking-head in the working memory through integrating facial expression with lip sync (visemes), pronounced word (phonemes) and prior knowledge for adequate schema acquisition. Adequate schema acquisition is important for accurate schema formation to be stored on more-or-less permanent basis in the long-term memory. As a result, in the future, when learners came across the same word, they may have improved pronunciation level of the respective word.

6. Conclusion

Animation has significant contribution in education industry among various subject matters for the past decades. It also plays an important role in second language learning, which was determined through findings from recent studies. Animation, specifically 3D talking-head has been developed in many studies as a virtual teacher in aiding second language learning. This can be seen through the emergence of CALL and MALL, which includes Multimedia elements, particularly animation in their existing system or application. However, it has scarce research done on mobile phone based application, specifically on the 3D talking-head. As what can be seen, mobile learning is becoming trendier in aiding education field nowadays. Furthermore, Mobile Assisted Language Learning shows growing bodies of development in assisting language learning.

One of the issues in learning second language effectively is pronunciation, whereby it is one of the reasons why learners do not acquire good verbal communication skill. This usually happens in non native English countries and consequently, contributes to low employability rate among graduates. Highlighting to that, researches identifying the effects of 3D talking-head mobile phone application in improving English pronunciation skills among non-native speakers is important. Referring to the conceptual framework developed, the modality effects and realism level of the 3D talking-head character appeared to be among the potential issues to look into.

References

Arkinson, D. (n.d.). *Animation notes*. Retrieved September 30, 2011, from http://minyos.its.rmit.edu.au/aim/a_notes/anim_contents.html

Azizan, H., & Mun, L. (2011a, April 10). Minding our language. *The Star Online.* Retrieved from http://thestar.com.my/news/story.asp?file=/2011/4/10/nation/8357917&sec=nation

Azizan, H., & Mun, L. (2011b, April 10). Top jobs only for those who knows the language well. *The Star Online.* Retrieved from http://thestar.com.my/news/story.asp?file=/2011/4/10/nation/8357917&sec=nation

Badin, P., Tarabalka, Y., Elisei, F., & Bailly, G. (2010). Can you 'read' tongue movements? Evaluation of the contribution of tongue display to speech understanding. *Speech Communication*, *52*, 493-503. http://dx.doi.org/10.1016/j.specom.2010.03.002

Balasubramanyam, V. (2012). Animations in medical education. *Medical Journal of Dr. D.Y. Patil University, 5*(1), 22. Retrieved February 22, 2013, from http://www.mjdrdypu.org/article.asp?issn=0975-2870;year=2012;volume=5;issue=1;spage=22;epage=22;au last=Balasubramanyam

Baddeley, A. (1992). Working memory. *Science*, *255*, 556-559. http://dx.doi.org/10.1126/science.1736359

Baxter, B. (1993). *Studying tips: Speaking tips*. Retrieved December 20, 2011, from http://www.musicalenglishlessons.org/tips-speaking.htm

Benoıˆt, C., & Le Goff, B. (1998). Audio–visual speech synthesis from French text: Eight years of models, designs and evaluation at the ICP. *Speech Communication, 26*, 117-129. http://dx.doi.org/10.1016/S0167-6393(98)00045-4

Bruner, J. (1990). *Act of Meaning*. Cambridge, MA: Harvard University Press.

Bruner, J. (1966). *Toward a Theory of Instruction*. Cambridge, MA: Harvard University Press.

Busa, M. G. (2008). New perspectives in teaching pronunciation. *Didactas to Ecolingua: An Ongoing Research Project on Translation and Corpus Linguistics*, *8*, 165-182.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332. http://dx.doi.org/10.1207/s1532690xci0804_2

Cheng Lin, C., & Fang Tseng, Y. (2012). Videos and animations for vocabulary learning: A study on difficult words. *Turkish Online Journal of Educational Technology, 11*(4), 346-355.

Chinnery, G. M. (2006). Emerging technologies going to the mall: mobile assisted language learning. *Language Learning & Technology, 10*(1), 9-16.

Chong, E. (n.d). *International Phonetic Alphabets*. Kuala Lumpur, Malaysia: Erican Language Centre.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*, 149-210. http://dx.doi.org/10.1007/BF01320076

Craig, T., & van Lom, M. (2009). *Impact constructivist learning theory and mobile technology integration*. Retrieved December 18, 2011, from https://sites.google.com/a/boisestate.edu/edtechtheories/craig_and_vanlom

Derwing, T. (2003). What do ESL students say about their accents? *Canadian Modern Language Review, 59,* 545-564.

Doyle, A. (2001). Web animation: Learning in motion. *Tech & Learning, 22*(2), 30-42. Retrieved January 26, 2012, from http://search.proquest.com/docview/212089737?accountid=48462

Engwall, O., & Bälter, O. (2007). Pronunciation feedback from real and virtual language teachers. *Journal of Computer Assisted Language Learning, 20*(3), 235-262. http://dx.doi.org/10.1080/09588220701489507

Fotouhi-Ghazvini, F., Earnshaw R. A., & Haji-Esmaeili L. (2009). Mobile assisted language learning in a developing country context. *Proceedings of the 2009 International Conference on CyberWorlds,* 391-397.

Frank, T., Hoch, M., & Trogemann, G. (1997). Automated Lip-Sync for 3D-Character Animation. *Proceedings of the 15th IMACS World Congress on Scientific Computation, Modelling and Applied Mathematics, 6*(1), 721-726.

Fraser, H. (2000). Coordinating improvements in pronunciation teaching for adult learners of English as a second language. Canberra: Department of Education, Training and Youth Affairs.

Geller, T. (2008). Overcoming the uncanny valley. *IEEE Computer Graphics and Applications, 28*(4), 11-17.

Gilakjani, A. P. (2012). The significance of pronunciation in English language teaching. *English language Teaching, 5*(4), 96-107.

Gilakjani, A. P., & Mohammad Reza, A. (2011). Why is pronunciation so difficult? *Canadian Center of Science and Education*, *4*(3), 74-83.

Godwin-Jones, R. (2005). Messaging, gaming, peer-to-peer sharing: Language learning strategies and tools for the millennial generation. *Language Learning & Technology, 9*(1), 17-22. Retrieved November 14, 2011,

from http://llt.msu.edu/vol9num1/emerging/default.html

Gueorguiev, V., & Velcheva, D. (2005). Bulgarian language speech and lip-sync animation. *Paper presented at the International Conference on Computer Systems and Technologies - CompSysTech' 2005.*

Hoek, P. (2009). *Constructivist learning theory and the use of computer simulations.* Retrieved January 10, 2011, from https://sites.google.com/a/boisestate.edu/edtech504/phoeck

Idrus, M. (2011). Impact of Wireless Technology on Language Learning. *Paper presented at the WVTESOL 14th Annual Spring Conference, Ripley, West Virginia.*

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, *11*, 47-78.

Joseph, S., & Uther, M. (2009). Mobile devices for language learning: Multimedia approaches. *Research and Practice in Technology Enhanced Learning*, *4*(1), 7-32. http://dx.doi.org/10.1142/S179320680900060X

Kadyte, V. (2004). Learning can happen anywhere: A mobile system for language learning. In J. Attewell, & C. Savill-Smith (Eds.), *Learning with mobile devices* (pp. 73-78). London: Learning and Skills Development Agency. Retrieved January 20, 2012, from http://www.lsda.org.uk/pubs/

Karagiorgi, Y., & Symeou, L. (2005). Translating constructivism into instructional design: Potential and limitations. *Educational Technology & Society*, *8*(1), 17-27.

Kayaoǧlu, M. N., Daǧ Akbaş, R., & Öztürk, Z. (2011). A small scale experimental study: Using animations to learn vocabulary. *Turkish Online Journal of Educational Technology, 10*(2), 24-30.

Kukulska-Hulme, A. (2010). Mobile learning as a catalyst for change. *Open Learning: The Journal of Open and Distance Learning*, *25*(3), 181-185. http://dx.doi.org/10.1080/02680513.2010.511945

Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL*, *20*(3), 271-289.

Kukulska-Hulme, A., Traxler, J., & Pettit, J. (2007). Designed and user-generated activity in the mobile age. *Journal of Learning Design*, *2*(1), 52-65.

Kukulska-Hulme, A. (2006). Mobile language learning now and in the future. In S. Patrik (Ed.), *Fr˚an vision till praktik: Spr˚akutbildning och Informationsteknik [From vision to practice: language learning and IT]* (pp. 295-310). Sweden: Swedish Net University.

Kukulska-Hulme, A. (2005). Mobile usability and user experience. In A. Kukulska-Hulme, & J. Traxler, J. (Eds.), *Mobile Learning: A handbook for educators and trainers* (pp. 45-56). London: Routledge.

Lewis, J. (1991). Automated lip-sync: Background and techniques. *Journal of Vizualisation and Computer Animation*, *2*(4), 118-122. http://dx.doi.org/10.1002/vis.4340020404

Lun, E. V. (n.d.). *Talking head*. Retrieved December 18, 2011, from http://www.chatbots.org/talking_head/

MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, *25*, 695-710.

Massaro, D. W. (2006a). Embodied agents in language learning for children with language challenges. In K. Miesenberger, J. Klaus, W. Zagler, & A. Karshmer (Eds.), *Proceedings of the 10th International Conference on Computers Helping People with Special Needs, ICCHP 2006* (pp. 809-816). Berlin: Springer.

Massaro, D. W. (2006b). The psychology and technology of talking heads: Applications in Language Learning. In O. Bernsen, L. Dybkjaer, & J. van Kuppevelt (Eds.), *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems* (pp. 183-214). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

Mayer, R. E. (2002). Cognitive theory and the design of multimedia instruction: An example of the two way street between cognition and instruction. *New Directions for Teaching and Learning, 89,* 55-71. http://dx.doi.org/10.1002/tl.47

Mayer, R., & Moreno, R. (2002). Animation as an Aid to Multimedia Learning. *Educational Psychology Review, 14*(1), 87-99.

Mayer, R. E. (2001). *Multimedia learning.* Cambridge, England: Cambridge University Press.

Mayer, R., & Moreno, R. (1998, April). A cognitive theory of multimedia learning: Implications for design

principles. *Paper presented at the annual meeting of the ACM SIGCHI Conference on Human Factors in Computing Systems. Los Angeles, CA.*

Mayer, R., & Moreno, R. (1998). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358-368.

Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review, 8*, 357-371. http://dx.doi.org/10.1007/BF01463939

McMenemy, K., & Ferguson, S. (2009). Enhancing the teaching of professional practice and key skills in engineering through the use of computer animation. *International Journal of Electrical Engineering Education, 46*(2), 164-174. Retrieved January 26, 2012, from http://search.proquest.com/docview/219870653?accountid=48462

Osipa, J. (2010). *Stop Staring: Facial Modeling and Animation Done Right* (3rd ed.). Indiana, Indianapolis: Wiley Publishing.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press.

Pemberton, L., Fallahkhair, S., & Masthoff, J. (2005). Learner Centred Development of a Mobile and iTV Language Learning Support System. *Educational Technology & Society, 8*(4), 52-63.

Pemberton, L., & Fallahkhair, S., & Masthoff, J. (2004, December). Toward a theoretical framework for informal language learning via interactive television. *Paper presented at the IADIS International Conference, Lisbon, Portugal.*

Picard, R. W. (1997). *Affective Computing*. Cambridge, MA: MIT Press.

Pollick F. E. (2009). In search of the Uncanny Valley. In P. Daras, & O. M. Ibarra (Eds.), *User Centric Media*: *Lecture notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering* (pp. 69-78). Venice: Springer.

Rodgers, T. (2001). Language teaching methodology. Retrieved September 30, 2011, from www.cal.org/resources/Digest/rodgers.html

Ruttkay, Z., & Noot, H. (2000). Animated chartoon faces. *Proceedings of First International Symposium on Non Photorealistic Animation and Rendering, Annecy, France,* 91-100. http://dx.doi.org/10.1145/340916.340928

Saran, M., Seferoglu, G., & Cagiltay, K. (2009). Mobile Assisted Language Learning: English Pronunciation at Learners' Fingertips. *Eurasian Journal of Educational Research*, *34*, 97-114.

Sood, K. (2010). Mobile Media Archive Using Significant Tools For Mobile Assisted Language Learning (Bachelor Degree Thesis, San Diego State University). Retrieved November 11, 2011, from http://hdl.handle.net/10211.10/602

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*(2), 212-215.

Sundberg, P. A. (1998, July). Animation in CALL: Learning to think in the fourth dimension. *Paper presented at CALICO '98 Symposium San Diego, California.*

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structure of technical material. *Journal of Experimental Psychology: General, 119*, 176-192. http://dx.doi.org/10.1037/0096-3445.119.2.176

Tamburini, F., & Paci, S. (2002, November). Web based language learning: Authoring and assessment technologies. *International Conference on Information and Communication Technology In Education, Badajoz, Spain.*

Voce, J., & Hamel, M. J. (2001). Towards a compromise between talking heads & interface agents: A web-based "mentor" for computer assisted language learning (CALL). In C. Montgomerie, & J. Viteli (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications* (pp. 1964-1965). Chesapeake, VA: AACE.

Wei, M. (2006). *A literature review on strategies for teaching pronunciation.* Online Submission. (ERIC Document Reproduction Service No. ED491566).

Wik, P. (2011). The Virtual Language Teacher: Models and Applications For Language Learning Using Embodied Conversational Agents. Unpublished doctoral dissertation, KTH Royal Institute of Technology,

Sweden.

Wik, P., & Hjalmarsson, A. (2009). Embodied conversational agents in computer assisted language learning. *Speech Communication*, *51*(10), 1024-1037. http://dx.doi.org/10.1016/j.specom.2009.05.006

Wik, P. (2004). Designing a virtual language teacher. *Proceedings of The XVIIth Swedish Phonetics Conference FONETIK 2004,* 136-588.

Williamson, V. M., & Abraham, M. R. (1995). The effects of computer animation on the particulate mental models of college chemistry students. *Journal of Research in Science Teaching, 32*(5), 521-534. http://dx.doi.org/10.1002/tea.3660320508

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist, 24*, 345-376. http://dx.doi.org/10.1207/s15326985ep2404_2