2022

Methodology of Augmented Reality Chinese Language Articulatory Pronunciation Practice: Game and Study Design

Daria Sinyagovskaya
University of Central Florida

METHODOLOGY OF AUGMENTED REALITY CHINESE LANGUAGE ARTICULATORY PRONUNCIATION PRACTICE: GAME AND STUDY DESIGN

by

DARIA SINYAGOVSKAYA
M.A. Russian Foreign Trade Academy, 2015

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Texts and Technology
in the College of Arts and Humanities
at the University of Central Florida
Orlando, Florida

Fall Term
2022

Major Professor: John Murray

ABSTRACT

Learning a language can be hard. Learning a language that contains tones to convey meaning is even harder. This dissertation presents a novel methodology for creating a language practice using augmented reality that has never been used before. The design of a new app in AR and non-AR versions can evaluate the same practice methodology. This methodology was applied to new software and was examined in regard to the importance of this software. Although the study results are inconclusive, progress has been made in answering research questions on the effectiveness of AR versus non-AR and the reliability of peer assessment. This study is essential for developing future language applications using design and methodologies in AR and peer evaluation.

To my family

ACKNOWLEDGEMENTS

I express my deepest gratitude to my committee chair, John Murray, for his guidance, support, and encouragement throughout my journey in the Texts and Technology Ph.D. program. I also thank my co-chair Rudy McDaniel for his invaluable feedback and support throughout my dissertation journey. In addition, I extend my thanks to my dissertation committee members, Emily Johnson, Amy Giroux, and Rustam Shadiev, for their instrumental advice and feedback throughout the dissertation process.

I want to thank Zijun Shen for serving as a second language expert in my study. In addition, I would like to thank my dearest friend, Yawen (Penny) Huang, for recording the Chinese phonemes for PinyinGuo. Finally, I would like to thank my family for always providing me security and love in achieving my dreams.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

Introduction

Augmented reality (AR) is a technology that uses computers and smartphones to overlay the virtual world onto the physical surroundings of the user. The affordances of the AR technology provide accessibility for learning. AR has a unique ability to situate an activity within a physical context, making it valuable for training and education. Moreover, AR can be used on smartphones anywhere anytime. These affordances create numerous possibilities for language instruction.

The contribution of this dissertation is developing a methodology for language practice with digital media. I was interested in combining AR and mobile technology to make language learning a more independent, interactive, and effective process. AR avatar would be more effective than a 3D tutor in an app, since AR technology replicates the presence of a teacher in the physical space. In addition, the learner can view the avatar mouth and lip movements from any angle in AR.

These interactive graphics with sound overlayed onto user's physical space are an "effective medium for information display" (Biocca et al., 2007, p. 164). My hypothesis is that an AR avatar will draw learner's attention to the main aspect of what they need to learn—articulatory movements and sounds. The 3D avatar placed in the user's spatial surroundings will be primary, while the 2D labels, like text, buttons, and timer would be secondary in navigating the system. Hence, AR affordances will allow user to focus on *doing* the task more than on reading the instructions.

Another contribution of this dissertation is developing a methodology for online peer-assessment practice. This project is focused on limited, anonymous, asynchronous communication that is possible with mobile technology. Mobile phone affordances allow peer feedback in the remote settings. I was interested in investigating whether online peer assessment can replace an expert feedback.

The contribution of this dissertation will add to the knowledge about designing an AR CAPT practice with peer assessment capabilities. The results presented in this dissertation will shed light on the role of AR for remote language learning and speech training.

Learning a new language can be hard. It is even harder when the language you are learning contains features like tones that aren't in your native language. Pitch is important in all languages to express intonation and inflection. But only tonal languages, like Chinese, use pitch to "distinguish words" (Duanmu, 2007, p. 225). Tones make learning Chinese especially difficult since the learner has to use an accurate pitch to convey the meanings of the words. For example, if the syllable *ma* is pronounced in a level tone (first tone), it means "mother"; if the syllable *ma* is pronounced in rising tone (second), it means “hemp”. If *ma* is pronounced in a falling-rising (third tone), it means "a horse." At last, *ma* in falling tone (fourth) means “to scold”, “to abuse”, “to curse”. Therefore, accurate pronunciation is needed to convey meaning of what the speaker wants to say. The context can help in the interpretation of the syllable if the syllable is mispronounced. For example, the word dumplings is 水餃 （shuǐjiǎo. The verb “to sleep” is 睡覺 (shuìjiào). The syllables look the same but the tones are different. If a speaker mispronounces the tones in a Chinese restaurant and says instead of “I want dumplings”, “I want to sleep,”, the listener might understand the meaning based on the context. However, without a context examples like this will create ambiguity in the interpretation.

This dissertation will advance knowledge in second language learning through both creating a new application and evaluating the role AR plays using a controlled study. The insights would help advance second language education and improve our understanding about how AR technology can be used in foreign language pronunciation practice.

The research questions of this dissertation are as follows:

RQ1: What effect on learner's motivation will articulatory demonstration in AR have on Chinese language pronunciation versus 3D virtual demonstration?

RQ2: How do AR conditions affect learning outcomes in pronunciation versus non-AR conditions?

RQ3: How accurate is anonymous peer assessment of phoneme pronunciation recordings?

Having a goal to answer these research questions, I have developed a methodology in order to address them. This methodology is part of the overall set of steps necessary to in the future prove things that answer the research questions. I have developed a mobile app *PinyinGuo* and designed an experiment to test learning effectiveness and motivation of AR in pronunciation practice and peer-feedback. PinyinGuo is available in AR and non-AR versions. The design of the software features peer assessment, drills, in-app assessment, surveys, and a posttest. The project is designed for a remote experiment without having participants come to the lab. The pedagogical requirements for the app are to facilitate Chinese pronunciation practice while ensuring the learners enjoy the learning process. The proposed system architecture, UI/UX design, and teaching methodologies will fill the gap in existing research on technology-enhanced pronunciation practice. This dissertation uses the affordances of AR for providing visualizations of abstract phonological concepts, interactivity for users, and environment for collaborative

language feedback. There have been limitations during the design of the experiment that have not produced the answers to the research questions.

This dissertation is organized in the following order:

Chapter One: Introduction

I outline my dissertation and present my research questions. This chapter offers a theoretical background for the motivation behind this dissertation project. One of the contributions of this dissertation is a new pronunciation software, PinyinGuo. PinyinGuo is a language learning app available in AR and non-AR versions. The iterative design process reflected both linguistic aspects of the learning content and usability elements of the mobile software. This dissertation draws on linguistic theory and usability theory.

The development was driven by theories in the humanities and computer science corresponds to the interdisciplinary nature of Texts and Technology methods. The design and development of the app was defined by the methods in second language pedagogy and the usability principles of mobile-assisted language learning (MALL) applications.

This chapter provides a literature review on mobile-assisted language learning. Two research questions of this study are focused on the effect of AR in Chinese language pronunciation practice regarding learning performance and motivation. The literature review addresses the use of AR in learning and language pedagogy. I hypothesize that visualizations of mouth movements of a 3D avatar in AR will provide have a positive effect of detail of mouthparts on the ability to replicate certain sounds. Another hypothesis is that the level of comfort with virtual avatars will provide a more motivating learning environment in AR than in non-AR.

Part of the literature review in this chapter is focused on peer assessment and corrective feedback in language learning. The peer assessment element in PinyinGuo is grounded in second language pedagogy and previously developed CAPT applications. This chapter provides an overview of how CAPT has been used to provide corrective feedback to improve language learning

Chapter Two: Methodology

This chapter discusses the method and the procedure used in the dissertation to answer the research questions. This chapter describes results from an informal usability test. The usability was conducted with two participants of the target group (age 28 and 29). Based on the usability testing results, the chapter explains the changes that were made to the app. A usability testing is currently being conducted with a secondary Chinese language expert to gather their feedback about the language pedagogy and usability of the system design.

The method for this study is an experiment. The experiment is conducted entirely online without having participants come to the lab. Such research design benefits from the affordances of mobile learning but also ensures safety for the participants during the time of COVID-19.

The procedure includes recruitment of the participants (age 18-45) via Facebook, Twitter, and UCF Discord channel. The participants had to download the app and go through a seven-day practice. The app was available in AR and the non-AR versions. The AR or non-AR versions were randomly assigned to the participants. At the end of the practice, the participants had to do a post-test. During the practice, the participants had to record phonemes that they've learned and submit for peer recording. They also had to assess two other peers randomly assigned each day.

At the end of each day of practice, the participants were answering three informal questions about how they liked the practice. This informal feedback refers to the motivational

aspect of the practice in AR and non-AR. This may also give us data on the continuous motivation that could be used for future research.

The usability testing followed by the expert evaluation. The expert reviewed the app in AR version. The expert wrote a feedback summary about the app contribution in the field of Chinese language learning. The expert highlighted the importance of a language app for pronunciation training. The expert shared suggestions about linguistic improvements for future research.

This experiment compares two different interventions – AR and non-AR. The data to look at the variance between the two groups (AR and non-AR) will be measured with a t-test. The informal feedback will be a descriptive piece. The motivation will be measured via intrinsic motivation inventory (IMI) in Qualtrics. The daily practice metadata will be saved on the server. This session metadata will give us an understanding of how much time the participants spend on the app and what type errors they make during the practice. This data will be visualized in a table and analyzed as a descriptive piece. The expert evaluations will be measured with Inter-rater reliability with Cohen's Kappa in SPSS.

Chapter Three: System Organization/Design and Development

This chapter explains the motivation behind the system organization of PinyinGuo. The design of the app reflects the complexity of pedagogical tasks. The client interface and a backend logic draw on pedagogical goals of peer assessment and feedback. In addition, this study investigates the effect of AR on motivation. For that purpose, PinyinGuo includes gamification elements (progress bar and timer) that encourage users to finish the session faster. This chapter includes the system architecture, description of front- and back end design features, pedagogical content, and backend logic for data collection and peer assessment.

Chapter Four: Overview of Existing Language Apps

This chapter provides a state-of the art review of the existing commercial language apps. The selection of the apps was driven by the user statistics and historical usage data. The analysis was grounded in the literature review on language theory, game theory, goal theory, and usability testing. This analysis precedes the development of the novice methodology for language learning in AR. The results of this review are recommendations for design and development of future MALL applications.

Chapter Five: Findings and Conclusion

This chapter presents recommendations for future MALL (Mobile-assisted language learning) applications based on the analysis in Chapter 3. The second part of Chapter 4 are the results of the experiment and the two rounds of usability testing. The data analysis from the server database and the Qualtrics will shed light on the learning outcomes and motivation from the two interventions. Hopefully, the results will prove the hypothesis about a more effective learning environment for pronunciation practice with AR versus non-AR in regards to learning outcomes and motivation. The peer assessment data will illustrate whether online peer assessment is effective without a teacher. In this part of the dissertation, I explain how I used the data and what type of data was eliminated from the study.

This chapter discusses the contributions of this dissertation. I provide answers to the research questions based on the results of the experiment. Further, I provide the implications of this study. The implications could be virtual language learning programs, hybrid classrooms, and computer-based pronunciation training systems. This chapter also includes limitations of this study and future research.

Literature Review

Context, Situational Learning, and Familiar Context in MALL

Context plays an important role in learning a language. Hwang, Shadiev, & Huang, 2011 provide evidence from an experiment to demonstrate that learning in familiar contexts is effective in terms of motivation and learning outcomes (Hwang et al., 2014). Hwang et al argue that affordances of mobile phones allow learners to learn new concepts in their natural environment. Such situational learning allows users to learn about objects, words, and scenarios anytime anywhere. The key results of this study are that the experiment group using mobile devices within familiar contexts has significantly better learning outcomes than the control group.

Augmented Reality in Language Learning

AR technology has been studied as a tool that can place virtual objects onto real life surroundings. The game by Thorne et al about the environment has shown positive effect on learners interest. Sykes at al have developed an AR game Mentira to teach Spanish. The spatial context played a vital role in the game, i.e. the users had to do a quest in the original town where the story took place (Holden & Sykes, 2011a) .

Technological innovation provides opportunities for creating and distributing instructional content (Golonka et al., 2014, p. 70). For example, AR overlays synthetic imagery onto the real world in such a way as to make it appear as if it were situated in the environment. The Reality-Virtuality continuum was proposed by Milgram et. al (Milgram et al., 1995) to characterize how technology can vary between completely replacing the world (Virtuality) to the world as observed unmediated (reality). AR falls toward the reality side of the continuum, where

virtual objects are placed in an otherwise real environment. The game that most popularized the term “AR” was Pokemon GO (Paavilainen et al., 2017), which was initially GPS-based but which eventually included more AR features as smartphones adopted camera-based tracking capabilities natively through libraries such as ARKit and ARCore. Other uses include altering facial appearances in real time (SnapChat “lenses”) and placing objects in an environment, as the Niantic sequel Harry Potter: Wizards Unite incorporates into its core gameplay (Baker, 2019). The core technology relies on efficient processing of camera data on modern consumer smartphones. The core technology provides information about objects visible in the frame or the device position in an environment through simultaneous location and mapping (SLAM).

AR has several affordances that can benefit language learning, including contextualize linguistic concepts into real situations and concrete actions for language learners. Smartphones can render content using realistic graphics, allowing learners to see the meaning in context and improve engagement through interaction with the materials. Studies have revealed that AR-based learning with gamification increases motivation and self-efficacy among L2 learners (Taskiran, 2019).

Popular commercial mobile apps like Duolingo offer gamified learning content which was curated through collaborative translation learning model (von Ahn, 2013). These apps present content as textual prompts and include simple graphics, but do not immerse learners in authentic contexts or incorporate active learning (Lotherington, 2018, p. 211). AR affordances, combined with the right content, can address this problem by combining multimodal visualizations of the context of the words and phrases used while offering improved interactivity. In addition to analyzing current applications in the market and the related literature, this paper

presents several possible approaches for how AR can be used in language learning and which design principles should be applied for language learning applications.

AR renders virtual objects, places and people onto the real world, it "incorporates digital information such as images, video, and audio into real-world spaces" (Taskiran, 2019, p. 122). Billingshurst et al. expand this definition by adding that AR is interactive and is "registered in 3D". This extended reality supports new modes of interaction with the world around us, and has been explored by researchers for how it can improve traditional human-computer interaction tasks (Xu et al., 2011) and offer new ways to interact with other people present (*Augmented Versus Virtual Reality in Education: An Exploratory Study Examining Science Knowledge Retention When Using Augmented Reality/Virtual Reality Mobile Applications | Cyberpsychology, Behavior, and Social Networking*, n.d.). Examples include social collaboration as present in Minecraft Earth[1], the ever-popular Zombie AR[2], and the previously mentioned Pokémon GO[3].

Smartphones can offer a fully immersive language learning environment which does not require other media or tools. This includes practicing a language though real-time interaction with virtual partners or incorporating virtual content in the user's environment. One area that is particularly promising is incorporating different modalities of corrective feedback (haptic, aural, visual) into the learning activity, which we discuss further in Section 0.

The affordances of AR allow learners visualize new vocabulary by placing the virtual objects into the physical space (Ibrahim et al., 2018). Therefore, AR provides opportunities for

[1] https://www.minecraft.net/en-us/article/new-game--minecraft-earth
[2] https://www.maysalward.com/zombiear/
[3] https://pokemongolive.com/en/

spatial learning and contextualization of concepts in learner's natural environment (Santos et al., 2016).

The affordances of AR include multimodal visualizations and auditory support which is effective for language learning (Santos et al., 2016) . The disciplines of phonetics and phonology in applied linguistics are concerned with how sounds are produced and articulated. Based on these promising results, we hypothesize that AR would benefit pronunciation training since the control of different muscles is often associated with mimicry in other practices. In this regard, the multimodal representations can serve as a tool to visualize articulatory movements. Tönnis et al. refer, AR allows "*Dimensionality*: the number of features (dimensions) that virtual and physical objects possess as well as methods to visualize and render them." (Tönnis et al., 2013, p. 998) Therefore, the way information is presented in AR allows to manipulate and adjust the 3D avatar for better learning experience. Dimensionality might have an effect on the ability to replicate certain sounds (X. Peng et al., 2018). In addition, AR provides an interactive element to the practice by allowing learners to manipulate an avatar at any angle.

Studies have revealed that due to uncanny effect, virtual embodied avatars provide a higher level of comfort among users than 2D avatars (Freeman et al., 2020; X. Peng et al., 2018). The hypothesis is that AR could provide a level of comfort for leaners that will make the learning process more motivating.

Another part of this dissertation is devoted to online peer assessment. Pronunciation is not only about how the sounds are produced but also how accurate the sounds are perceived. The question of input and output is one of the most crucial in phonological practice. Therefore, effective feedback helps learner understand their errors and improve pronunciation.

Theory

This dissertation draws on situated learning theory. Situated language learning refers to "physical, social, and cultural aspects" of the learner (Santos et al., p. 4). The situated learning theory drives the methodology to a novel practice app to be created in order to evaluate the effects of AR on pronunciation training. The learning content in the system is based on combining linguistic theory and pedagogy to teach Standard Chinese pronunciation.

Various CAPT tools have been developed and studied for decades (O'Brien et al., 2018). The development of technology broadened the range of teaching methodologies used in CAPT tools. The language pedagogies include three main methods: (1) the Audiolingual Method (ALM); (2) Communicative Language Teaching (CLT); (3) Task-based Language Teaching (TBLT) (Stern, 1983).

The ALM method was developed as one of the earliest methods in second language acquisition but is still widely used in some language programs. The teacher provides the students learning materials with the content to be memorized. The techniques in this method include drills and memorization exercises. ALM primarily helps students develop linguistic habits and avoid making errors (Ellis, 2014, p. 33). At the same time, ALM practice often lacks spontaneous dialogues and situational assignments where the students can contextualize language expressions and linguistic instances.

In comparison to ALM, Communicative Language Teaching (CLT) aims to develop sociolinguistic competency among L2 learners. This method is focused on communication through collaboration and work in pairs or groups. In CLT practice, the teacher provides a topic or a semi-authentic situation with conversations that serve as an example for in-classroom

discussion. Thus, the learning happens based on a situation rather than a practice of a particular linguistic concept.

In Task-based Language Teaching (TBLT), students are not provided with any prior linguistic content for the practice. Instead, the students learn a new language in communicating while trying to conceive a meaning during the course while completing meaningful tasks. There are three stages in TBLT: a pre-task stage, a main-task stage, a post-task stage (Ellis, 2014, p. 33). If there are any communication difficulties, the teacher can provide corrective feedback and explain some linguistic content.

These methods provide different levels of attention to linguistic form in teaching L2. In ALM, the linguistic form stands in the center of the learning activity. The CLT method merely provides pre-defined linguistic Content, but the method is focused on the topic and how certain linguistic elements fit into that topic. TBLT has three stages of the learning practice: pre-task, task, and post-task phase. In this method, the obligatory step is only the task phase. However, the teacher can introduce some linguistic elements in the pre-task or post-task stage if needed. The main goal of the TBLT method is to encourage students to focus on meaning rather than a linguistic form.

Apart from the linguistic pedagogy, these methods differ in the syllabus design. In ALM, the teacher focuses on linguistic aspects of L2. In CLT, the primary focus shifts towards conversational topics like weather, travel, etc. Lastly, the TBLT method provides communication activities to exchange information or share an opinion. These three language pedagogies use different approaches and techniques to achieve seamless communication in L2 with native speakers.

I have considered the three methods for the development of the learning activity. Because the linguistic aspect plays such a vital role in the practice, I have used some elements from ALT. These elements are drills and pre-defined tasks to help learners develop listening habits in the beginning of the session.

In addition, I used CLT method in the peer assessment part of practice. The participants received recordings from their peers daily. They have to assess the recordings of the others and send it back. Each participant assesses two recordings per day. The assessment criteria is whether the participants can recognize the sound (phoneme identification), how well they pronounced it (performance rating), and how confident the evaluator is about what phoneme it is (confidence rating). PinyinGuo allows participants to collaborate in the remote settings by providing feedback to peer recordings without having to directly interact with each other. Moreover, the learners are given a one-day timeframe to assess the recorded phonemes. The advantages of such setup are that not only the learners can provide feedback to multiple people remotely, but also they can do it at their own convenience. The system is set to send push notifications when new peer recordings are available and reminders to complete the practice.

Phonetics

In linguistic theory, phonetics is closely related to the single sound production and articulation. Sound is an abstract concept challenging to express in linguistics and L2 pedagogy (Kiparsky, 1982; Vaux & Samuels, 2018). In this regard, phonological practice (or generally known as pronunciation practice) uses visual representations of how the sound is produced. Historically, these representations have been images of articulatory apparatus with a text describing how the sound is produced through the vocal tract (Bickford & Floyd, 2006, pp. 4–5). In the age of computer and mobile technology, practitioners in language pedagogy have

extensively used CAPT tools to make learning more effective and engaging. Levis refers to the computer as almost "ideal" tools for pronunciation practice (Levis, 2007, p. 184). Computers can be set up for automated feedback without the presence of a teacher (G.-Z. Liu et al., 2016). These affordances save time and cost for learners.

Learners can use CAPT tools at the time and place of their convenience. These affordances are important in regards to personalization of the learning process (Bu et al., 2021). The learners can develop and adjust their learning style. The capacity of CAPT tools allow developers to include a choice of various tasks including audiovisual, audio only, and interactive activities (Hazan et al., 2005). Learners can choose from these tasks the ones that suit their learning style better. Such customization provides opportunities for further development of CAPT technology for autonomous learning (McCrocklin, 2019).

Research on CAPT has revealed that phonological practice with multimodal representations is more effective than without any visual representations (audio only) (Massaro et al., 2008). Indeed, illustrating articulation is important in teaching pronunciation. For example, the study on articulatory representations for pronunciation of hearing-impaired children showed a positive effect on learning outcomes from demonstrating Chinese sounds with 3D virtual tutor head (X. Liu et al., 2013). Indeed, visual representations have been images of articulatory apparatus with text, describing how the sound is produced through the vocal tract (Bickford & Floyd, 2006). In the age of computer and mobile technology, practitioners in language pedagogy have extensively used CAPT tools to make learning more effective and engaging. Levis refers to the computer as an almost "ideal" tool for pronunciation practice (Levis, 2007, p. 184) since the affordances of the computer allow automated visualizations of the sound input. Such visual aid will enable learners to compare their recordings to the correct model (Pennington & Rogerson-

Revell, 2019). This literature review elements directly relate to my research question about motivation and learning outcomes in mobile assisted pronunciation practice. The design and development of PinyinGuo is grounded in the body of literature on CAPT tools in language learning.

In relation to the 3D avatar representation in AR, it is vital to review various modes of representations in CAPT in applied linguistics research. For example, Peng et al. conducted an experimental study with three audio-visual representations for Chinese pronunciation training—audio-only, human video presentation, and 3D virtual model. The research questions included learners' perception of the 3D virtual head for pronunciation training, learners' impression from the three modes, and learning outcomes from three presentation conditions (X. Peng et al., 2018, p. 28). The study results revealed that the participants' attitudes towards a 3D head model representation were more favorable than the other presentation modes. The second experiment in this study aimed to answer the research question about the learning outcomes. The results revealed that 3D pronunciation training is more effective than the audio-only and the human video presentation modes (X. Peng et al., 2018).

Che Dalim et al. conducted two experiments testing the effectiveness of AR in combination with speech recognition technology on learning outcomes and user's enjoyment from the practice (Che Dalim et al., 2020). The experiment group with the AR system has shown better results in terms of learning gains and enjoyment compared to the control group.

Studies have revealed that AR is beneficial for learning new vocabulary since the affordances allow learners to visualize new words and see the words around them (Ibrahim et al., 2018; Santos et al., 2016). Santos et al. argue that AR is beneficial for vocabulary practice, since it provides opportunities for spatial learning and contextualization of concepts in learner’s

natural environment (Santos et al., 2016). Based on these promising results, it we hypothesize that AR would benefit pronunciation training since the control of different muscles is often associated with mimicry in other practices. AR further can provide an engaging interactive visual articulatory representations, allowing learners to manipulate and explore them at any angle.

Usability Theory

Following the principles of usability theory of "how people interact with interfaces"(Lazar et al., 2017, p. 6), we have designed a working prototype of a Chinese language pronunciation app (Lazar et al., 2017). HCI research has pointed out the importance of making software design and development a *process* of ensuring that the system architecture is user-friendly, easy to operate, and response to the learning goals (Gould & Lewis, 1985, p. 302).

At the same time, the app corresponds to the three principles of design developed by Gould and Lewis: (1) early focus on users and tasks; (2) empirical measurement of usage; (3) iterative design (Gould & Lewis, 1985).

The user group for this system will be adult learners from 18-45 years old. The choice of this age-category is determined by the results on a recent study on correlation of age and linguistic competence. The study included neuropsychological assessment of 30 young adults (18-45 years old) and 30 (65-85 years) old adults during a Chinese language practice. The study revealed an age impairment factor in language learning abilities among the latter group (Baker et al., 2020). In the scope of this dissertation project, a usability study for an older group is not possible to conduct knowing that there are certain challenges associated with the age factor. A younger group (minors) will also not be included to the sampling procedure due to challenges associated with parents' consent required by IRB.

The method of participatory design is also used in this project. The user profile in this study is a young adult learner with no prior Chinese language experience. Another category of stakeholders are teachers, since this CAPT can be integrated into classroom settings. The contribution of the teacher feedback to the design process will be invaluable. Such participatory method will be conducted remotely and will include interviews and survey.

Empirical measurement was followed by iterative design (Gould & Lewis, 1985, p. 300). The design of the system is aimed to facilitate learning content and increase learners' motivation. Based on the suggestions of the stakeholders, the app prototype will undergo another round of design cycle and usability testing before the deployment.

A round of usability testing with inconclusive results was conducted for this study. A convenience sampling of two participants of 28 and 29 years old was used to test the system's functionality. The participants used a think-aloud method while completing the first day of practice. The participants completed a SUS based on three customized questions about the app. The participants found the system useful for someone who wants to learn Chinese and overall easy to use.

The participants also provided informal feedback about ways to improve the UI. For example, both participants pointed out that the tone marks need to be somehow highlighted; otherwise, it is hard to understand that the users need to pay attention to tone marks. In addition, both participants found that a timer should be in the recording panel interface is confusing since it seems that the recording starts after the timer finishes. One participant suggested having a progress bar for a recording instead. Both participants noted that the font in the instructions should be bigger. The participants also pointed out that the buttons in the phoneme choosing

practice should change color depending on correct or wrong selections. These suggestions were noted and implemented into the system design.

Peer Assessment

Corrective feedback is an effective teaching method for foreign language learning (Lightbown & Spada, 1990; Sheen, 2011). Studies in applied linguistics have revealed the effectiveness of various types of corrective feedback on learning outcomes (Bitchener, 2008). Corrective feedback can be direct and indirect. The direct corrective feedback occurs when the teacher points out an error and provides the correct version. The indirect corrective feedback is when the teacher alludes that there may be an error but doesn't directly points at it. Indirect corrective feedback requires student to think about where they might have made an error and find a solution by themselves.

Language pedagogy suggest that pronunciation practice needs to be supported by corrective feedback (Tseng & Tsai, 2007). Research has shown that corrective feedback is more effective in collaborative settings rather than individual feedback (Mujtaba et al., 2021). In this regard, peer assessment has been used as a pedagogical method to improve learning outcomes (Dochy et al., 1999; Falchikov & Goldfinch, 2000).

Online peer assessment is included into learning practices. Traditional in-class language instruction includes collaboration as one of the main aspects for effective learning (Chung et al., 2019). Based on socio-cultural approach, learning is a social process. This process requires interaction with peers. The results of studies on collaborative language practices demonstrated positive learning outcomes from peer interaction and peer feedback (Evers & Chen, 2020; Hazan et al., 2005).

Indeed, one of the advantages of digital technology is that it can automate the process of providing corrective feedback to foreign language learners. Studies have revealed that CAPT tools, on the contrary, can address the issue of input-output in the listening and speaking practice by providing automated feedback (Tsai, 2019). In this regard, CAPT tools can also allow collaborative learning and gathering peer feedback in remote settings (Tseng & Tsai, 2007; Yan et al., 2018). Yan et al. have developed a web platform for pronunciation practice and weekly peer feedback. The experiment results have revealed positive improvements in learner's pronunciation and enjoyment (Yan et al., 2018, p. 82).

Computer assisted language learning (CALL) is a discipline in applied linguistics that uses computers as a tool to learn a language (Beatty, 2010, p. 7). The pedagogies in CALL have evolved along the development of computer technology. For example, first CALL tools in 1970 were based on audiolingual method that was widely used in that time (Levy, 1997). Mobile phones allow learning to happen anytime anywhere, i.e. in a comfort at one's home, in a classroom, on campus etc. Mobile assisted language learning (MALL) has become a subdiscipline of CALL. The affordances of mobile devices provide opportunities for spatial and contextual learning (Hwang et al., 2014).

MALL apps that offer entire courses for learning a foreign language are available on AppStore and Google Play. These commercial apps have a full structure of how to learn a language on your own without having to hire a teacher. Some of the most popular commercial language apps are Duolingo, Memrise, Busuu, and Mondly.

The most popular language app today is Duolingo. Duolingo has more than 500 million users (von Ahn, n.d.). Duolinguo offers a game-like language courses for thirty-nine languages. Duolinguo has collaborative translation method where users provide translations of phrases.

Then other users provide ratings of the translations. The app has machine learning algorithms embedded in the system by gathering scores from users of the translation and improving the translation (Monroe, n.d.).

However, along with the design features of these apps, there is a general critique about the lack of learning pedagogy in existing mobile language learning apps (Levis, 2007; Lotherington, 2018). The review refers to the popular commercial mobile-assisted language learning (MALL) apps like Duolingo, Memrise, and Busuu that use outdated learning techniques from the 1950s-1960s like drills and grammar-translation (Lotherington, 2018). One of the first examples of computer-assisted language software, PLATO, was developed in the 1950s and had the same audiolingual L2 method as Duolingo that came out in 2016. Lotherington refers to this phenomenon as design-driven app development that does not put language pedagogy first.

CALL and MALL research have established principles of language pedagogy with digital devices (Farr & Murray, 2016). In these disciplines, technology and pedagogy are intertwined that requires a different approach to developing learning content compared to the one in traditional in-class instruction. Contextual learning available with mobile phones can be achieved with AR and VR (Dünser et al., 2006). Another example of mobile phones as a medium to provide a link between learning environment and learning content is the project *Mentira* by Holden and Sykes (Holden & Sykes, 2011b, pp. 10–11). Mentira is a place-based mobile game, where the students went to a real location described in the game and play the game at the site (Holden & Sykes, 2011b, pp. 10–11). An interesting part of the study design was that the research activities took place in the classroom—for a classroom discussion as well as outside of the classroom. Such mixed mode of location and activities was possible due to the exploration nature of the method of design-based research (DBR) used in this study. DBR is a process of in-

situ research where students and teachers are collaborators on creating a better learning environment based on user experience, curriculum, suggestions, and solutions. Holder and Sykes describe the process as following: "[participants'] goal was to imagine themselves as the designers of this game for other students and revisit the existing design as a controlled experience to unravel its purposes and motives." (Holden & Sykes, 2011b, p. 10).

Holden and Sykes' project is an example of how language pedagogy can be adjusted for learning with technology. The authors of this project have developed a mobile game with the mind of language curriculum as well as cultural education crucial to the language learning process. The fact that the researchers added authentic location as a virtual environment to the game, allowed students to immediately create context about the history of the place, its traditions and surroundings. The linguistic aspect included some socio-cultural implications based on family values and cultural implications (Sykes & Holden, 20147, p. 167). The method used in this study provides an opportunity for user-centered research in foreign language taken into account the importance of situational learning and learner-centered sociocultural theory.

However, pronunciation practice with MALL and CALL is often overlooked and marginalized due to the lack of formal training among language instructors. In addition, the existing foreign language textbooks do not provide learning content for pronunciation practice (Dunleavy et al., 2009). In languages Indo-European like English, pronunciation practice is not as crucial as in Chinese, since even with poor pronunciation, speakers can be understood (G. Peng et al., 2008). In Chinese if the pitch is incorrect, the word might have another meaning. The tonal aspect might bring confusion and frustration (Wen, 1997; Yu & Watkins, 2008).

Indeed, tones are used to define meaning of words. Each tone is assigned to a vowel in a syllable. Each syllable equals one Chinese character. A character can be a single word or part of

a word (Honorof & Feldman, 2006). Hence, learning correct pronunciation of Chinese tones and sounds is crucial for communication in the language.

CHAPTER TWO: OVERVIEW OF EXISTING LANGUAGE APPS

Analysis of Seven Most Commercially Popular Apps

This chapter reviews key features and implicit pedagogical approaches of commercial gamified mobile educational apps in second language acquisition. This analysis was crucial for designing and developing PinyinGuo.

This chapter discusses the design and learning content of seven market-leader commercial gamified language apps. These apps are following apps: Duolingo (*Duolingo*, n.d.), Mondly (*Mondly*, n.d.), Busuu (*Busuu*, n.d.), Babbel(*Babbel*, n.d.), HelloTalk (*HelloTalk*, n.d.), Memrise (*Memrise*, n.d.), and MondlyAR (*MondlyAR*, n.d.). Smartphones have made learning a second language accessible to a wider audience. Mobile language apps have gained popularity among millions of language learners (Karjo & Andreani, 2018). These mobile language apps are worthy of analysis since they use language theory and gamified design. In comparison to other language apps, these seven apps merge language theory, peer-communication, and gamification elements. Such learning technology is related to the main arguments of the dissertation.

The statistics provided in this overview is from Sensor Tower[4]. I requested two datasets to determine the language app usage. This data (Fig. 1) represents the popularity of language apps based on iOS installs. The data for Android installs is very similar. In this review, I am analizing a general dynamics of app distribution. Since both iOS and Android graphs are very similar, I will use visualizations from iOS installs.

[4] https://sensortower.com/

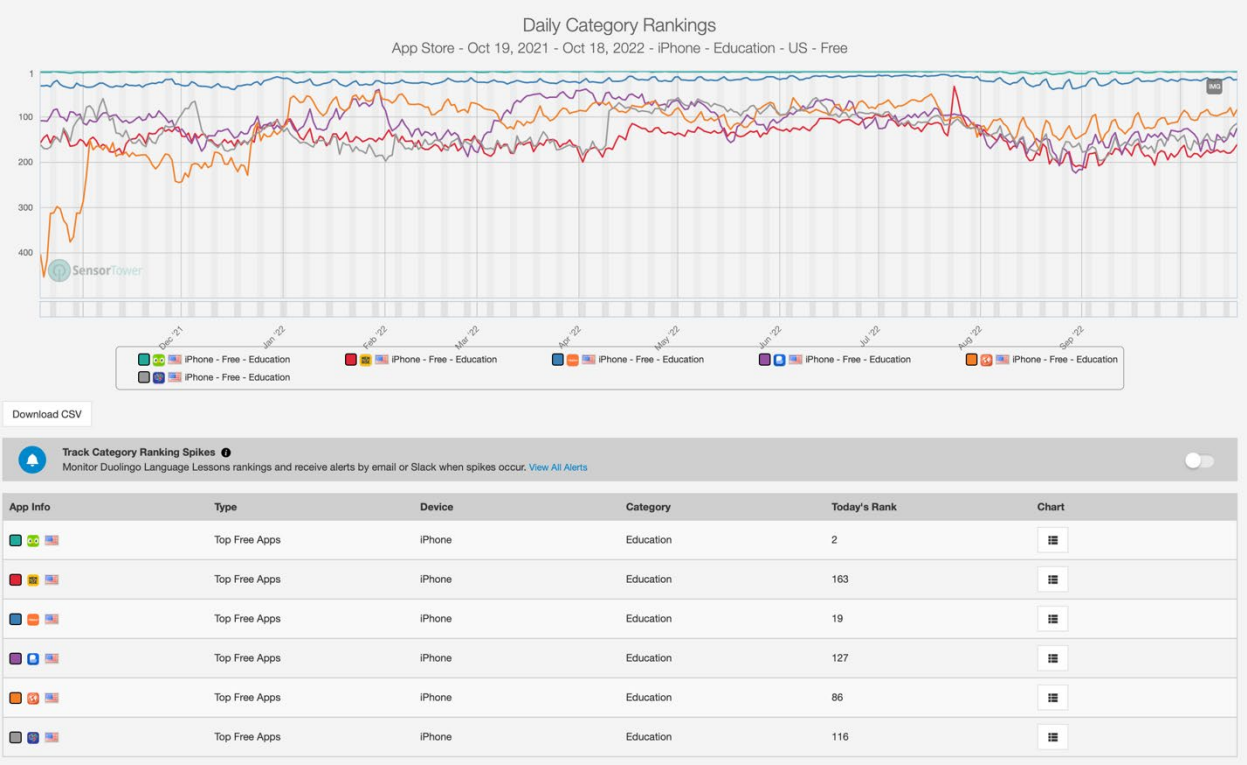

Figure 1: Free Language App Daily Category Ranking. App Store - Oct 19, 2021 - Oct 18, 2022 - iPhone - Educational

This analysis is grounded in the historical app usage data to determine usage data of these apps in the commercial app market. I focused on two types of data, i.e., longitudinal dataset since 2014 and the one-year dataset. The longitudinal data (Fig. 2) were referred to as “of all time.” 2014 was year when all the six apps were available, including Mondly that was founded in 2014 in Brașov, Romania (n.d.). Therefore the “all time” data start from 2014.

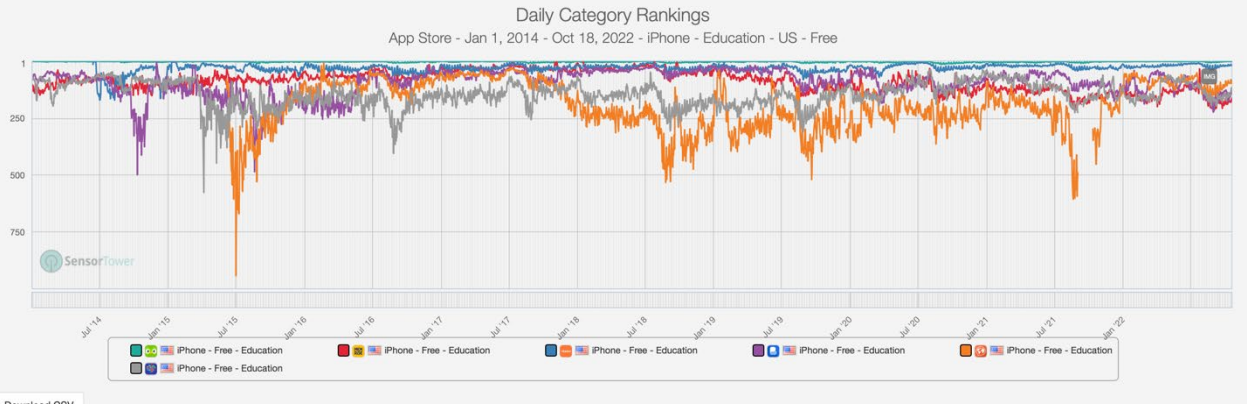

Figure 2: Daily Category Rankings App Store - Jan 1, 2014 - Oct 18, 2022 - iPhone - Education - US – Free. Retrieved from Sensortower

Figure 2 illustrates the overall app download statistics on the US App store over the last year. Duolingo is the number one app for language learning today. It has been the industry leader since 2014 (Fig 2). Historical data of all times shows almost unchanged Duolingo place in the language app market, while Mondly and HelloTalk apps show unstable development. Although Memrise and Babbel are not as popular as Duolingo, both apps have stable places in the language app market as well. The companies' founding dates might have something to do with taking up the largest share of the market.

Fig. 2 shows that Mondly AR has the most rapid falls and rises throughout the history of the apps. Mondly also has an AR version, which will be discussed later in this chapter in Mondly AR sectionI. However, Sensor Tower does not provide separate data for Mondly AR version. Both figures have the two versions combined as Mondly.

Duolingo is the most popular language app today (see Figs. 1 and 2). Duolingo started in 2009 as a free language app. Since then the app's user base has reached more than 500 million users (*About Us Research*, n.d.). Today Duolingo has the largest collection of language learning

user data in the world (*About Us Research*, n.d.). Duolingo's website says it provides content for more than 30 languages. This is the largest number of app installs from all other language apps. For a comparison, second and third popular apps Memrise and Babbel have 10 million installs.

Duolingo Method

Duolingo uses gamified translation methodology for language learning (Munday, 2017; *Translating with Duolingo for Language Learning*, 2013; von Ahn, 2013). Grammar translation method (GTM) is one of the oldest teaching methods in language acquisition (Kong, 2011). This method was used in the 18th and 19th century for translating texts into the native language (Kong, 2011, p. 76). Diane Larsen-Freeman describes the role of GTM as an outdated approach that was used in early twentieth century to teach students by translating foreign texts, understand their native language better (Larsen-Freeman, 2010). GTM was a mental exercise with the purpose similar to learning Latin and Greek (Stern, 1983, p. 455). Heinrich Stern refers to the translation method as follows: "Translation is to be used as a last resort. The IPA [International Phonetic Association], then, did not recommend a direct method at all costs" (Stern, 1983, p. 92). The idea behind GTM approach is keeping the native language "as the reference system in the acquisition of the second language" (Stern, 1983, p. 455). By constantly referencing the native language, GTM prevents the students from "dominance" of their first language (Stern, 1983, p. 456). The dominance of the first language can slow down fluency acquisition (Kaharuddin, 2018, p. 235).

GTM is used as a core approach in Duolingo. Lotherington refers to Duolingo as "driven principally by grammar translation" (Lotherington, 2018, p. 206). GTM approach makes the learning connected to the first language. Lotherington describes the method employed by Duolingo as "rigidly inflexible" (Lotherington, 2018, p. 208). Another issue Lotherington points out is that the translation happens into English. The app assumes that their users are fluent in

English. This attitude adds extra challenge to information processing if English is not user's native language. Moreover, Duolingo becomes inaccessible for individuals who do not speak English. There is a Duolingo version for Spanish speakers but it does not have working pages for other languages. (Nushi & Hosein Eqbali, n.d.)There is a tab on the right corner of the page "I speak…" where the user can select their native language. I chose Chinese and it showed me a mix of Chinese and English text. When I chose Russian, it took me to the language menu page where the user can select the menu. I have selected German. The translation of the Russian version of the website is very poor. I could hardly understand the text with the instructions.

Language accessibility is not considered in the design of the app. Especially taking into the account the large user base of more than half a billion people.

The gamification approach allows users to stay motivated throughout the learning process. Duolingo has the largest user database to conduct language pedagogy and research (Portnoff et al., 2021). Portnoff et al. have developed a method for assessment based on data mining. The data mining approach allows assessment via module quiz, *Checkpoint Quiz*, and skill quiz, called *Review Exercise* (Portnoff et al., 2021). Portnoff et al. describe the difference between these two assessment types is that Checkpoint Quiz is a quiz after the learning module, while the Review Exercise checks each skill early in the course. To complete the Checkpoint Quiz, the learner must go through the entire unit. The Review Exercise does not require to finish the unit. The learners test their learning on every skill level of the course (Portnoff et al., 2021, p. 867).

Levels take important part of the Duolingo design. Duolingo is a game-based language app. There is a "tree" of levels that get unlocked when the user goes through the learning content. The learner chooses the language they want to learn. After the leaner choses the

language course, they chose the skill level. The app asks the learner a series of questions to determine their skill level. There are two options to select from – a complete beginner or if the learner already has some knowledge in that language. The app offers a level assessment test to determine the language proficiency of the learner.

For this analysis, I have chosen the course in French. The practice consists of listening, choosing, and selecting sequences. The first practice starts with the word "cat". First, the user listens to the word and choose the matching picture. Then the user reads the words and find the word "le chat" – "cat" in French.

There are different voices for the phrases which makes it easier to get used to different sound of male and female voices. There are badges and rewards. When completing each step of the session, the user gets an encouraging message about their success. In addition, there are sound effects that make the experience more engaging. At the end of the session, there is a point checker with the bonus points for the session.

The outline of the whole course in the form of a tree helps user understand how long they need to go to achieve their learning goal. Such overview is similar to a progress bar. There are progress bars throughout the practice as well that keep the learner on track of their progress and motivated to complete it. The learning content is organized by themes, such as nature, shopping, weather, doctor etc. If the user wants to proceed to the next level, they have to take a skill assessment test in a Check Point.

Duolingo is a gamified translation-based app (Munday, 2017). The motivational game elements are badges, leaderboards, levels, rewards, progress bars, and challenges(Hamari et al., 2014, p. 3027). Gamification is an effective method to increase learner's motivation (Buckley & Doyle, 2016). Duolingo keeps the user motivated in translation exercises by having gamification

elements in the app design. However, studies point out the learning content includes repetitive activities and is too translation-focused (Loewen et al., 2019). In addition, Duolingo overemphasizes receptive language skills (listening and reading) instead of expressive skills (speaking and writing) (Loewen et al., 2019). Huynh et al. argues that the app might be boring for intermediate level learners (Huynh et al., 2016) since it is difficult to expand on their language skills beyond translation.

Mondly

On app store, Mondly advertises learn 33+ languages for free. However, the app is free for the first seven day trial period. The app's content layout is very similar to Rosetta Stone[5]. There are flashcards with images (like a cup of coffee) and words to match. There is audio support to listen to how the words sounds..

Another similarity is that the app claims to use a unique method to learn a language. Just like Rosetta Stone, Mondly claims to use special techniques for "quick memorizing": there is a disclaimer on the second screen after the user logs in: "Words and tests are interchanged in a special sequence that is optimal for memorizing." The app does not specify why this sequence is optimal for memorizing words. However, this statement refers to the teaching method in this app by memorization of words. This method is also known as audiolingual method (ALM) and it was discussed in Chapter 1. Similarly to the GTM, ALM is an outdated method for teaching foreign languages. ALM is based on memorization of words and drills. Memorizing single words takes communication out of learning context and spontaneous conversation with native speakers (Belasco, 1965).

[5] https://www.rosettastone.com

Mondly has a feature of selecting the language, topics of interest, and time intended to be spent on learning. Just like Duolingo, the learning content is presented in topics. The topics are listed before the learning session begins. In contrast, Duolingo provides topics in the learning content. There is a tree with various levels and greyed out topics. Mondly's topics include travel, business, romance, fun, school. The learner can choose topics thy are interested in from multiple selection. There is also a goal system, i.e. how many minutes per day the user plans on learning. This is less than a learning goal than a time management goal. However, the app emphasizes that a daily commitment is part of Mondly's learning method. The goal selection begins from 5 minutes per day to 20 minutes per day. Duolingo has the same time goal selection as Mondly. The app does not provide options for longer periods of time for learning. Such a time limit shows a supplementary nature of these two app instead of being the primary language courses. Mondly could become primary in their current point. To become primary Mondly would need to have peer collaboration option or some social outlet to expand language knowledge.

However, daily goals encourage learners to develop consistency in language learning . Second language literacy is developed through consistent daily practice. In case on Duolingo a qualitative study has indicated that short duration practice does not replace comprehensive language instruction (Leung, n.d.).

I selected the language Chinese with practice 20 minutes per day. The session begins with listening to a dialogue "at the restaurant." There are two speakers in this dialogue—a male speaker and a female speaker. The user sees dialogue bubbles, Chinese text, pinyin under the Chinese, and the English translation (I specified that I speak English while selecting the target language). This session begins with a Task-based Language Teaching (TBLT). The app does not

provide any explanations or background for what needs to be learned in the module. The conversation “in the restaurant” starts spontaneously without any prior context.

After listening to the conversation, the user goes to the “match the word with a picture” exercise. The words are used from the restaurant dialogue. After the lesson is finishes, there is a statistics tab on the app. The app tells the user how many words you have learned and how much was spent on learning.

Mondly has a leaderboard element just like Duolingo. The difference between the two apps is that Duolingo does not allow users to unlock Leaderboards before they have completed at least nine lessons. Leaderboards add a social aspect to the learning process when people compare themselves to others (Baldwin & Mussweiler, 2018). Comparison as a social mechanism might have an effect on user’s engagement (Schlömmer et al., 2021). In combination with other gamification elements in MALL apps, leaderboards might have both positive (Deterding et al., 2011) or negative effects on user’s experience (Zuckerman & Gal-Oz, 2014). Duolingo explains leaderboards as a “fun experience” to motivate users(“Duolingo Blog,” n.d.). There are ten leagues in the leaderboards: “Bronze, Silver, Gold, Sapphire, Ruby, Emerald, Amethyst, Pearl, Obsidian, and Diamond.”

Leaderboards combined with badges are aimed to motivate learners (Bai et al., 2020). Gamification elements like badges and leaderboards improve self-efficacy of learners (Gnauk et al., 2012). Therefore, these gamification elements in both apps are aimed to keep the learner engaged and motivated. Having read the research, gamification elements are essential for language app development. Self-efficacy is important for keeping the learner motivated to continue learning, This allows to create a learning environment where learners have high level of self-efficacy and sense of achievement. Leaderboards are vital to engage learner with the

community. Communicative aspect is important, since language learning is a social process even in the remote settings.

Busuu

On its website, Busuu presents it's courses as "award-winning.[6]" Just like in Duolingo and Mondly, the user answers questions about what language they want to learn. Next, it asks the user for their language proficiency: "I don't know any French" or "I already know some French." Then, there are different topics of interest that the user needs to choose before being directed to the course. Finally, the user is asked how long they plan on practicing the language. The time duration is the same as the other two apps. Just like Mondly, Busuu has also a seven-day free trial period. The yearly price for the app is $89.99.

Similar to Duolingo and Mondly, Busuu is also a vocabulary-oriented app. Instead of images, Busuu offers short videos with real people pronouncing words. This is helpful, since the learners can engage with multimodal learning. The speakers are located in authentic environments, e.g. for French, the speakers are located in front of a café or a historical building. Authentic environments and real people in the videos make the learning more engaging (Rosell-Aguilar, 2018). Autonomous language learning with MALL often lacks the interactive aspect (Beckman et al., 2014). Videos in the app provide a feeling of communicating with a native speaker. Such organization of layout allows students to have more context about speaking with native speakers in their natural environment.

The topics of each lesson are have similar layout to Duolingo. There is also a tree-like structure of the content. However, the content specifies some grammatical topics like "singular

[6] https://www.busuu.com

forms of être." From all commercial apps, Busuu app is the only app that has grammatical notes embedded in the course. The course also includes information of what each chapter (unit) contains. For example, instead of saying: "city" or "city 1" like it is in Duolingo, Busuu has more explicit information about module contents: "Saying your nationality," "Listening skills," "Grammar: subject pronouns" etc. This course logic makes it easier to understand what the course offers in each module.

Busuu provides a page to connect with native speakers. There is an option to choose the languages user speaks to connect with language exchange partners. In this regard, Busuu is different from the two other apps by using communicative language teaching (CLT) approach.

Babbel

Babbel app states on its homepage "Language lessons for every learner." The Babbel app has a rubric on the website describing its teaching method. All the lessons are based on real-life conversations.

The menu offers learners a selection of the language, the topics of interest, and the pre-determined language level. The app offers more options to choose the time for user's language practice: from five to sixty minutes per day. In my opinion, a longer duration provides the user with an opportunity to use Babbel as a primary language course.

Since Babbel mentions on the main website that its language courses are available to any type of learner, there are additional questions in the questionnaire about the age of the user. Babbel also asks users to provide information how they have heard from the app. Just like Mondly and Busuu, Babbel is not free, the yearly subscription is $215.88 for the first 12 months. With a special offer a twelve-month subscription is $74.99.

The Babbel method uses TBLT. The practice begins with a "guess the meaning" exercise. The user has to guess the meaning of basic words like "thank you" and "hello." After choosing the correct translation by guessing the meaning, the user has to match the words with it's translation. This exercise is similar to matching words sessions in Duolingo. Next session is to construct the words from different parts called "Can you build the words?" The user has to build a word from deconstructed parts of the word like "cias" and "gra" – "gracias." Each session has a voiceover pronouncing the words. Next, there is a dialog where the user has to complete sentences to answer the speaker. The dialogue has both male and female voices. The following exercise is a listening and translation task. The user listens to the word and has to choose the correct translation. The final session is to listen and build the sentences. The user has to listen and build sentences into their language. The user has to build sentence from random words. This is a listening practice when the user has to train their ear to catch the words in the sentence.

HelloTalk

HelloTalk is a social platform to learn 150 languages with 30 million native speakers for free. The app offers a VIP subscription that is $6.67 per month. HelloTalk is not a course-based MALL app. A new user has to fill out their personal information, i.e. profile picture, name, gender, and date of birth. Those fields are mandatory. The app explains this requirement to bring authenticity to the community. The app asks for user's location, native language, and the language they want to learn. Then, the user selects the level. The app collects data of where the users have learnt about HelloTalk.

The app interface looks like a social media platform. On the interface, there is a messaging, a discussion board, a "connect" button, live stream events, and the profile. This app

provides a feeling of a community of language enthusiasts. The user can message anyone on the app including during live stream sessions. The community rules are specified before you open a messaging board or live stream sessions. Some of the community rules prohibit using the app for : “attempted/committed fraud, send sexual content, used the app for dating purposes, abusive language.” Before proceeding to using the app, the user has to agree to those rules.

The native speaker’s profiles on HelloTalk include a pin of the word map of where the individual is located. In addition, there is a current time displayed in the location of that user. The personal information includes a profile picture, first name, languages that the user knows and wants to learn. There is also learning statistics with total learning points. The learning statistics include community engagement like used translations and submitted corrective feedback, number of language exchange sessions, favorites, transcriptions, and pronunciations.

Translations is a feature when the user taps and holds the message to see a translation. Corrections is available through a correction button that the users can use in the chat (Fig. 3).

Figure 3: HelloTalk Correct Sentence UI

The profile also displays user's engagement in the discussion board. The discussion board is called "moments." The statistics have an option to view user's posts and interactions in the discussion board.

The user profile also include a self-introduction section where the user can write about their background and interests. Following self-introduction, there are separate sections of interests and hobbies and desired places of travel.

Live stream events look like a Facebook or Instagram livestream, which is a live video with a live chat. Collaborative learning environment has been used in teaching a second language to make learning more effective and interactive (Lan et al., 2007).

The peer-assisted concept is used in Duolingo for correcting translations. Additionally, Duolingo has a collaborative translation feature (*Translating with Duolingo for Language Learning*, 2013; von Ahn, 2013). The users help translate the phrases into their native language. The algorithm collects the translation data and improves the translations in the app.

Memrise

Memrise mentions on the home page of the website: "learn a language that's really useful [7]". The app promises users that the learning content will be authentic and useful for communication with native speakers. After the user selects the language, they select a course. The course begins from basic words. The words are supplemented with short videos of native speakers pronouncing those words with the traditional background.

For example, I chose to learn Chinese language. The native speaker was standing in front of a Chinese temple saying the phrases in Chinese. This approach is very similar to the Busuu. After the listening practice, the user has to listen and choose correct words. I have selected a Chinese course, but there are no Chinese characters introduced in the practice. Unlike Duolingo, the Chinese characters are not displayed next to the Hanyu Pinyin and the English translation.

[7] https://www.memrise.com

Instead, there is only Hanyu Pinyin. The Chinese characters only appear in the following session, when the learners are already introduced to the sounds of the new words. When the user selects the correct word, the app displays the corresponding Chinese character. After the choosing state, there is a dialog to train listening skills. There is a short video of a conversation in the native language. There is a word on top of the screen that the user needs to catch while listening to the dialog. For example, the user has to hear the word “good” in Chinese. While listening to the conversation in Chinese, the user needs to press the word when they hear it in the conversation. In the following sessions, there is an emphasis on natural language usage. The videos show scenarios where the new vocabulary can be used in a conversation. The user has to be focused to hear new vocabulary in a spontaneous conversation like with the example of the word “good” in Chinese. The feature is called: “immerse.” The “immerse” button is displayed on the menu on the button of the screen that allows the user to navigate to the immersive language scenario at any time.

The app has a feature of AR image capturing (Fig 4.) The user can point at an object and the camera will scan and recognize the object like the keyboard in the example. There is a sound attached to the word. In addition, the caption button allows users to save the picture of the object with the translation. This allows users to contextualize their learning experience in their physical space. AR design has capabilities for spatial learning (Dünser et al., 2006). Specifically I will review the affordances of AR in the Recommendations section. Also, the learner can save images and text to create their own digital vocabulary. Such approach allows customized learning experience.

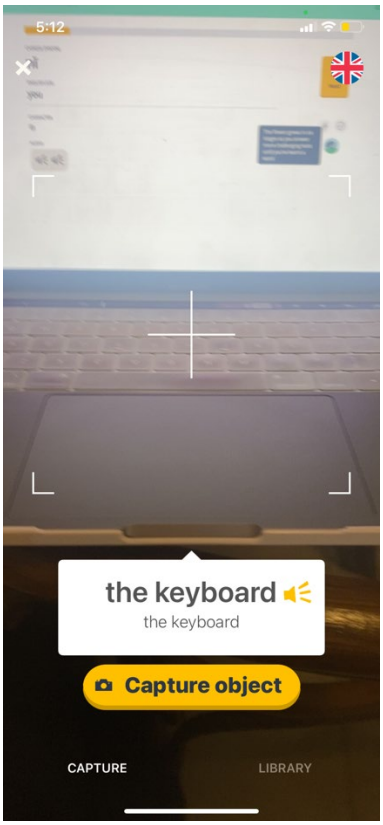

Figure 4: Memrise Image Caption Feature

At the end of the practice, there is a calendar with a list of words learnt each day. The app displays the status of user's daily goal, i.e. how many words have been reviewed, how many new words have been learned, and how long the user has spent on the language practice. The daily goal seems more clear than in the above-mentioned apps due to the specificity of the goal. For example, Duolingo has a fire icon for day streak with a number. However, the app does not specify what that number means.

A study on goal theories for language learning defined two main types of language learning goals: "goals as aims: goal setting theory" and "goals as purposes: achievement goal theory" (Lee & Bong, 2019). Goal setting theory refers to reaching a certain level of proficiency focusing on "goal difficulty, goal specificity" (Lee & Bong, 2019, p. 3). Achievement goal

theory is focused on “mastery-oriented goals” (Lee & Bong, 2019, p. 3). Mastery-oriented goals are aimed at the ability of learner to get good learning results from the learning content instead of competing with the others (Locke & Latham, 1990).

The goal interface in Memrise is an example of mastery-oriented goal setting practice. The leaderboard interface to compete with the other is not as easy to find as in Duolingo and Mondly. This shows that the emphasis is more on personal goals and meaningful learning practices. In addition, list of the words learnt along with the time input show the performance-oriented goals.

MondlyAR

Mondly has a VR and AR editions available for download on App Store and Google play. In this review, there will be no discussion about the VR version since the topic of VR goes beyond MALL. Mondly VR is available on Oculus and STREAM platforms[8]. MondlyAR (Figure 5) is a commercial AR language app that provides an animated 3D tutor and 3D models of content for new vocabulary.

[8] https://www.mondly.com/vr

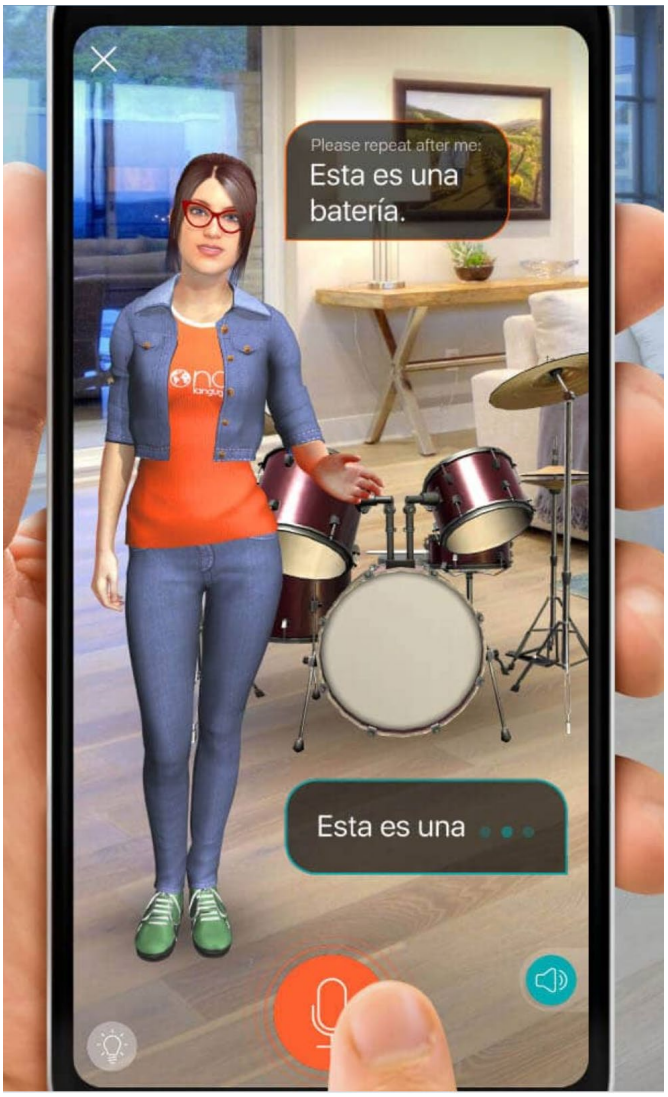

Figure 5: MondlyAR Marketing Materials. Retrieved from https://www.mondly.com/ar on Feb 1st, 2021

MondlyAR has a learning content laid out in the manner of flashcards. The cards are distributed based on various topics (see Fig. 2). The topics include ten topics, i.e. space, musical instruments, animals, fruits, vegetables 1, vegetables 2, food, eating utensils, and drinks.

Figure 6: MondlyAR Learning Topics

To begin the session, the user has to allow to access camera to initiate AR. After the camera access is granted, the user has to find the surface to place the AR objects onto the floor. The first object you see in AR in the first lesson is a full-body female instructor. The instructor says the welcome words in English following by content introduction. The first lesson is an introduction. The instructor explains that AR is a new technology that combines virtual world with the real one. Following the introduction, the instructor shows a lion next to her figure. The lion is a 3D model in AR as well. The lion roars and moves its head. The instructor pronounces lion in the selected language (I selected French). The user repeats the word in a microphone after the instructor. Next, the instructor shows a chicken. The instructor says a fun fact about the chicken that chicken is “the closest relative to t-Rex.” The instructor speaks English the entire time. No matter what the user records in the microphone, the avatar will “pass” them without any corrective feedback even for obvious errors. I recorded a word in English instead of French, and

it still showed me that I recoded the word correctly. The learning content is 3D models of animals with English description and facts about those animals. After single words, the instructor teaches learners simple phrases, i.e. “the elephant is big”, “the cat is small” etc. Finally, there is a review of the learned material that tests learner’s knowledge.

On the bottom of the screen in Figure 7, there is an section for “conversation”. There is a conversation button on the right bottom corner. The conversation is a dialogue that happens between the instructor and the learner (see Fig. 7). The app uses both speech recognition and a chatbot to provide feedback on correctness and prompted usage of vocabulary (CCSSP, 2020, p. 261). However, the speech recognition feature in MondlyAR is not currently accurate enough to provide feedback on pronunciation (Altinkaya & Smeulders, 2020, p. 3). The app uses a combination of ALM and TBLT methods (Stern, 1983). The ALM method in Mondly AR is used through single word memorization.

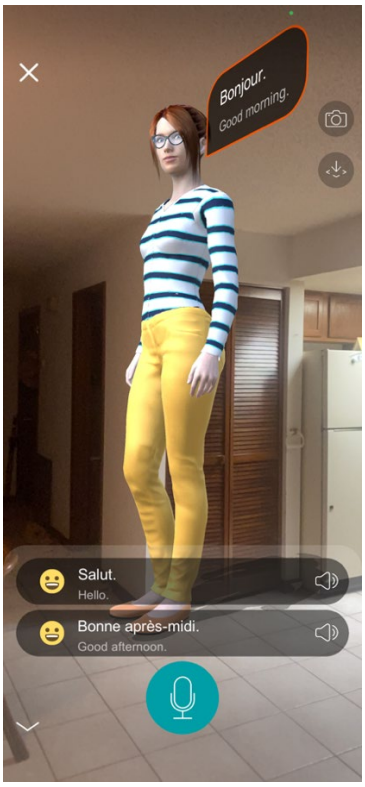

Figure 7: MondlyAR Conversation Stage

Similar to the lessons, the corrective feedback is not provided in the recording sessions. The app is more interactive in a way that it shows the ability of learning in AR. However, the learning content is very limited. In addition, the learning content is decontextualized. There is a possible solutions in the recommendations provided in this overview. The random display of a chicken or an elephant does not feel like a language practice. MondlyAR exhibits the affordances of AR. However, AR provides more opportunities for learning in terms of pedagogy, content, and interactivity. The section below describes the AR design for language learning.

Recommendations

It is always challenging for instructors in a traditional classroom environment to provide representations of some linguistic concepts without having to bring outside materials into class. For example, the topic "Food" is one of the most exciting subjects for L2 learners. Food is clearly a topic everyone can relate to and is interested in. However, teacher are constrained when creating an experience that involves talking about food in L2 language interactive and expecting students to engage. For instance, bringing food and beverage into the classroom is not an option for most educational settings due to the risk of allergy and food poisoning. While various digital and printed materials can allow students to explore the topic—images, video, text, there remain limitations to how much students can interact with the objects or how readily they can imagine their context of use. These hinder student learning of one of the most useful and engaging topics taught in L2, but could be addressed through richer presentation in sound and visuals.

L2 instructors can use the multimodal affordances of AR on smartphones to present new vocabularies using 3D models that behave realistically. These can help contextualize the meaning of the new words, while spatial audio can engage their auditory senses. AR, therefore, can add new senses and context in the teaching of vocabulary. BBC Civilisations AR app[9] is an example of an educational project that uses AR to engage learners with virtual artifacts. The AR context seeks to emulate some of the naturally curiosity evoked by encountering unusual exhibits in a museum. The app invites users to explore the content through a 3D model of the globe where historical artifacts are distributed. The user clicks on the artifact and is presented the history of the selected item and the region it is from.

[9] https://apps.apple.com/us/app/civilisations-ar/id1350792208

One recommendation for L2 would be to use the affordances of AR to offer a glossary of all nouns in a L2 course of one proficiency level, such as a beginner level course. One of the most challenging parts of teaching beginner L2 courses is the lack of learners' vocabulary knowledge. Therefore, it is especially important for instructors to select sets of primary vocabulary to be used during an introductory L2 course. Specific concepts such as *a book, a meal, a job, a plane* etc. could be presented as 3D models in AR in the learner's environment. Gamification can further keep learners engaged through common strategies such as points, timer, badges, etc. For a more complete review of gamification in education, see Hamari et al., who emphasize the context of the interactions over the gamified elements themselves (Hamari et al., 2014).

Content such as glossaries in AR apps can be represented as searchable databases. A glossary would include words and meanings, but might also include searching by related concept or situation. The annotated vocabulary can include 3D images, virtual text, and include external multimedia links. Apps that incorporate AR elements could maintain the same organization as a traditional L2 textbook, where the vocabulary section is placed in the lesson as well in end of the book.

Corrective Feedback

Besides the meaning of the words, learners require feedback to improve. The method of practice and the possible inclusion of a human avatar in the instruction are both opportunities for augmented reality to improve mobile language learning experience.

Corrective feedback (CF) is when a learner's usage is critiqued or errors in usage identified. This feedback can be delivered by simple scores of exams, but because language is often associated in a social and environmental context, the feedback on the performance can

mirror that of real-life interactions. A number of commercial apps have some version of solutions for corrective feedback. For instance, Mondly has a female tutor that provides introduction, feedback, and CF. This model attempts to duplicate some of the social dynamics of a human tutor in a traditional language learning context.

AR provides flexibility in designing a tool for teachers and learners. For example, Cuendet et al. assert that *flexibility* as a design principle "give[s] freedom to the teacher to access each part of the systems at any time" (Cuendet et al., 2013, p. 568). In this regard, if the teacher spontaneously decides that some practice doesn't need a virtual tutor, the students can easily switch off the tutor in the app. If the students are working with the app remotely from home or in small groups in class, they can turn on the virtual tutor option. Holden and Sykes have addressed the *flexibility* aspect in Mentira project by using design-based research (DBR) in developing the Mentira app (Holden & Sykes, 2011, p. 3). DBR is an iterative process that they used to receive feedback from students and stakeholders throughout the app development process (Holden & Sykes, 2011, p. 3). This study sets an example for future research involving teachers and students in the development process. AR can be one of the features to examine in research questions based on different learning settings and learning goals.

CF is an essential part of L2 learning, but incorporating indirect CF effectively has been a difficult challenge for gamified apps. Duolingo started out in 2009 as a game-based commercial language app with a concept of crowdsourcing translation feedback. In their early years, Duolingo's "immersion" feature allowed users to evaluate translation and contribute to the existing translations (*Translating with Duolingo for Language Learning*, 2013) while offering a form of indirect feedback in the sense of allowing learners to self-assess. Today the company focuses on using machine learning (ML) and artificial intelligence (AI) to improve CF and take

advantage of human language error recognition (*Research - Duolingo*, n.d.). Their transition from a manual-input collaborative translation approach to an ML, AI-driven platform, highlights the importance of technology-mediated CF and the opportunity to further improve. AR offers one such venue to increase learner immersion as well as deliver naturalistic corrective feedback.

AR and Immersive Learning Experiences

AR technologies make it easy to add sound and speech-recognition technologies to 3D content(Billinghurst et al., 2015, p. 145; Che Dalim et al., 2020). The spatial audio technologies allow creating an authentic experience for learners. For example, if the app has the tutor mode enabled, the sound source will come from the tutor. This could help maintain the illusion of authenticity (Wu et al., 2013, p. 45).

It is the author's opinion that AR combined with sound is a powerful tool for creating immersive and effective learning experiences. Peng et al. argue that "multimodal presentation conditions promote effective learning because humans process information through both visual and verbal channels" (Peng et al., 2018, p. 26). This suggests that combining audio and visual stimuli in necessary to create an authentic learning environment mimicking the real-world contexts. Likewise, it is important to develop sound-recognition systems that improves on existing speech recognition systems. These will create a bridge between the learner and the application context.

The tutor or the voice from the app (depending on the settings) will respond to what the learner vocalizes , providing immediate CF. The learner can have an option of recording each session to review it at later time by themselves or with the teacher in class.

Hence, AR provides opportunities for both aural and visual feedback. If the student has the AR tutor enabled, the avatar can provide CF with different voice tone, facial expressions, and

gestures. Billinghurst et al. state: “AR’s high level of interactivity enhances learning, particularly for students who learn through kinesthetic, visual, and other non-text-based methods” (Billinghurst et al., 2015, p. 211). In the above example, interactivity is achieved through the sound source and visual expressions.

Indeed, some people learn better through kinesthetic modalities. Santos et al. state that “vision-haptic visualization is the integration of both the sense of sight and the sense of touch in perceiving virtual information” (Santos et al., 2014, p. 51). The affordances of AR allow manipulation with the virtual objects by moving them around and rotating them. The users can also move away from the object or move towards the object (Santos et al., 2014, p. 51). Spatial manipulation of content opens new opportunities for CF using AR and ways to incorporate simulated interpersonal communication.

For example, a 3D tutor in AR could appear to listen to the learner response, appear to think about the answer while either an on-device or cloud-based process analyzes the audio, then provide indirect CF through embodied responses such as smiling, waving, or gesturing with a virtual object . These responses can be more animated or enticing in the case of correct answers. This section presented an example of using AR in vocabulary practice, illustrating how CF could work with three modalities (haptic, aural, visual).

Interpersonal Communication with AR

L2 acquisition is, however, more complex than just memorizing random words and phrases. The ultimate goal of learning a L2 is to communicate with native speakers in their natural environment. For that purpose, the L2 learner needs to acquire pronunciation, listening, writing, and reading skills. The learner should also gain knowledge about the culture, traditions,

and the people of the target language country. These skills and knowledge will help the learner achieve a high level of interpersonal communication with the native speakers.

Each of these L2 aspects takes an integral part in the language acquisition process. These aspects can be taught by applying the AR method, since AR provides all the necessary modalities to represent a natural L2 environment for the learner. In the authors' experience, when the learner is learning an L2, they are fully engaged into the learning process, not only with their mind, but with their body as well. For example, when learning a L2 language pronunciation, student's facial muscles and vocal apparatus are engaged in a new way. The student's ears and eyes are also engaged to process new information. Therefore, AR could represent these modalities and facilitate this complex cognitive process.

Visualization

AR can be used in pronunciation training, since AR can provide audiovisual representations of phonemes, sounds, and words. A 3D model could prove useful for presenting an example. Such a model should include animated facial features, vocal apparatus, and can be synchronized with sound production. One study used a 3D model for Chinese language articulation training, where the researchers examined the effects of visual representations of Chinese speech. The findings revealed the effectiveness of visual representations in 3D mode "to blade-alveolar, blade-palatal, lingua-palatal, open-mouth, open-mouth(-i) and round-mouth" (Peng et al., 2018, p. 26). Given the inherently 3D nature of AR, there is the potential to represent these 3D models even in a more contextualized and engaging way. For instance, students can imagine the speaker or themselves while being able to position the speaker and better visualize articulation.

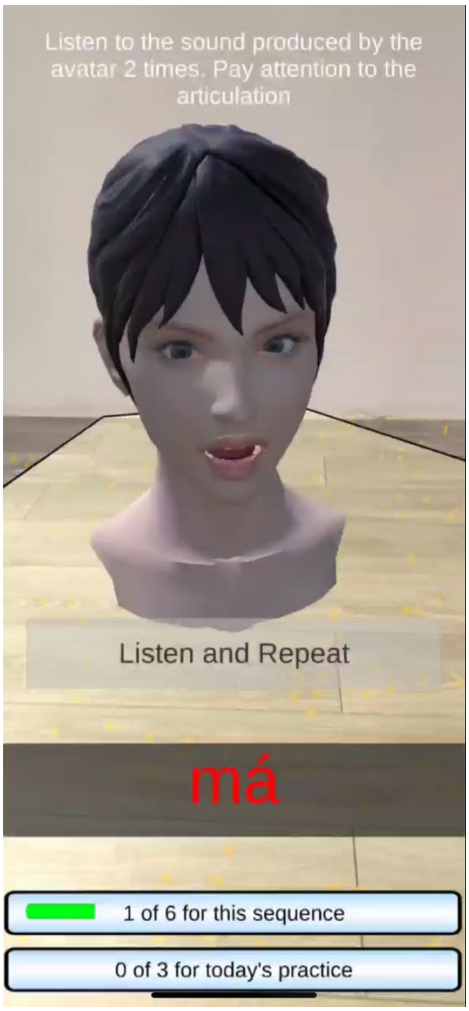

Figure 8: Example AR Facial Model for Articulatory Simulation for Chinese Sound ma in Second Tone (RISA HEAD v1 Low-Poly 3D Model)

Studies have revealed that AR can help learners visualize and understand abstract concepts [25]. In learning L2 it is vital to define types of practices where AR can be an effective method to convey complex information. For example, visual representations can assist learners in pronunciation practice. Peng et al.'s work was the falling-rising tone in the tone practice was recognized via aural-only practice better than with visual representation (Peng et al., 2018, p. 26). These results could support the argument that sound distribution are critical to the attention

of the learner. The falling-rising tone is usually a stepping stone for many new learners of Chinese as a second language. Chinese as a second language learners have difficulties recognizing the difference between the rising and the rising-falling tones [21]. This indicates the importance of flexibility of AR in L2. AR can be adjusted to pronunciation practice, where students can view visual representations of sound articulation or completely turn it off to train their ear for some listening practice.

Contextual Learning

Multiple studies on AR have focused on spatial learning and metacognitive awareness of linguistic development (Che Dalim et al., 2020; Dünser et al., 2006). Indeed, context is one of the central concepts in social cognitive theory, i.e. activity theory, on how people learn (Chung et al., 2019, p. 3). In the mobile-mediated activity theory, context and community are mutually replaceable concepts. Chung et al. define context (community) as "the relationship between the context of learning with the context of being. i.e. the environment of the learner" (Chung et al., 2019, p. 3). Thus, based on the taxonomy of activity theory, AR is a tool that mediates the learning context. The community is the environment of the learner that is also the context of the learner. AR should make an integral part in context (community) to make the learning process more effective.

The community in a L2 class are peers. Therefore, AR needs to be incorporated into collaborative practices to help learners create new knowledge and gain new L2 skills. In this regard, the recommendation is to incorporate peer-assisted feedback for L2 practices with AR app. For example, the learners can record the audio from the sessions in pronunciation with the tutor and share it with their peers. If the learners are not comfortable with sharing their identifiable audio recordings with others, they can share the recordings only with their teacher.

Context in the sense of activity theory can be achieved through gamification elements added to the AR app. For example, Thorne et al. conducted a study with an AR game for L2 students, based on a quest (Thorne et al., 2015). This suggests AR could be used to create similar quest-based activities to explore L2 topics outside of the classroom. “In the market” could be a topic for an AR game where players have to gather in teams to collect artifacts from an ancient Chinese market. While collecting the artifacts, the teams will have to solve a mystery of a Chinese pirate. The clues will be given to the team that can solve a series of communicative tasks with Chinese merchants, pirates, and government officials. This example interactive narrative scenario is grounded in the previous literature from the AR-mediated L2 (Sydorenko et al., 2019; Sykes & Holden, 20147). Sykes and Holden have introduced a project Mentira to learn Spanish as a second language. Mentira is a game-based AR app includes interaction with social cues and situational behavior. For example, if a player does not obey the family rules, they will be provided a feedback for future improvement. This corrective behavioral feedback is similar to the “Chimeria: Gatekeeper” project by D. Fox Harrel (Harrell, 2019). In this interactive narrative the user participates in a conversation with the gatekeeper to enter the keep. Depending on the user’s behavior and answers and importantly the cultural context of the interaction, the gatekeeper will open the gate or not. This conversational narrative and the responses involved provoke questions of what influences and is important for interpersonal interaction and self-representation.

Cultural Awareness

AR can be used to convey nuances of interpersonal communication and develop cultural awareness in L2 practice. This interpersonal practice can be a separated game app, or it can be incorporated into an existing AR app. One example is a scenario where students are provided

with an overview of traditions from the target language culture. The students then use the vocabulary from previous lessons and the knowledge about the tradition to communicate with a 3D avatar. For example, the user learns about the Chinese tradition of handing over a business card to a business partner with two hands and bow. The user will use the mobile phone and bow to the avatar, while saying a polite phrase. If the phrase or the gesture are incorrect, the avatar expresses discontent with a sad face or negative phrase. This exercise will serve as indirect metalinguistic CF as well. Such an approach using AR allows L2 learners to experience a truly immersive environment with haptic, aural, visual practice and feedback.

Billinghurst et al. argue that AR works better with images and kinesthetic activities rather than text (Billinghurst et al., 2015, p. 211). However, L2 writing and reading skills could also be taught in AR. In particular, writing skills that require some level of calligraphy, i.e., Chinese or Arabic writing. Chinese characters can be studied with a method of AR paper (Wen, 2018). Wen examines how AR can be a helpful tool for children to learn about Chinese characters and promote collaborative learning (Wen, 2018, p. 132).

Future Work

There are two broad categories of future work suggested by the analysis presented in this paper: specific lessons based on spatial knowledge and ecologically valid studies that evaluate whether AR affordances improve instruction in a full language course. Additionally, future work includes identifying and cataloguing design principles for AR in this domain. Comparing the effectiveness of AR in L2 learning with other alternative technologies such as VR and traditional casual game mechanics is another fruitful direction. We are specifically looking at the extent to which AR influences learner performance and motivation in our next planned study. We will compare both the performance on objective tests (as rated by an expert in the case of

pronunciation) as well as the subjective engagement of learners across these three conditions. Subjects will learn and practicing Pinyin syllables over the course of a week, an important topic in Chinese language acquisition. Our hypothesis is that the learner will attend more to content presented with a human that appears in the same space using AR compared with a non-human presentation or a context-free (non-AR) 3D model.

While specific topics are easier to address, ultimately effectiveness would need to be evaluated through a complete L2 course at a specific skill level. This includes evaluating whether AR environments are perceived as authentic and will require a combination of traditional skill assessment methodologies such as objective tests along with more qualitative instruments such as surveys or interviews What specific features of AR can or should be integrated into MALL by application developers? What are the effects on learner engagement or performance, if any?

Pronunciation

Pronunciation training is closely connected to auditory senses. It is our opinion that one of the greatest challenges for L2 instructors is to provide a real-time demonstration of what happens with the vocal tract in the instance of speaking. unique AR can demonstrate through an interactive visualization this movement. Research shows evidence that AR can be used in anatomy class as a tool for demonstrating abstract concepts in real time. AR visualizations “help [learners]… connect their understanding of human movements and muscles” (Santos et al., 2014, p. 43). Grounded in the findings of this research, the use of AR is promising for visualization of human vocal tract in pronunciation training, and is a critical part of the planned study.

Depending on complexity of the L2 pronunciation, practice can accompany lessons or be a dedicated activity. All languages are different and the L2 learners have different background as well. For example, Chinese has four tones and one neutral. The language is not alphabet-based,

so for a Western-language speaker it is particularly challenging to get used to the articulation and pronunciation. But to someone with Sino-Tibetan family background, Chinese pronunciation and tones might not be that difficult to pick up. Therefore, the amount of pronunciation practice required depends on the target audience of learners and the L2 they are trying to learn.

Customized Learning Styles

L2 AR app can assist the learner in understanding their learning type (visual, auditory, kinesthetic) (Vincent Annette & Ross Dianne, 2001). This division is important for L2, since many people have never been tested on how each of them personally learns L2 more effectively. Duolingo has a short survey of what type of learner you are to set your learning routine, but there is no kinesthetic aspect in that survey or any type of these learning intelligences (Vincent Annette & Ross Dianne, 2001). AR could offer a way to fill this gap in research on personalization of learning. Certain approaches to language learning only becomes only possible with AR, since AR allows to explore haptic, aural, and visual modalities and select the ones that are more enjoyable and effective. Needless to say, that in L2 traditional classroom settings there are limitation for personalized practices.

The high level of personalization available with AR technology has been noted in the literature. AR allows "creating new personalized learning opportunities in a physical space that is modeled, processed and labeled by automated machine learning (ML) classifiers, assisted by human users" (Ibrahim et al., 2018, p. 2867). Ibrahim et al. discuss how ML can be effectively incorporated into AR to increase learning outcomes in L2 acquisition (Ibrahim et al., 2018, p. 2867). Labeling the L2 vocabulary in the real time environment is one way AR can be used with text. This type of contextual vocabulary can be incorporated into AR app or game development. In the previous example of the lesson "Food", the suggestion was to present the virtual objects

onto the real world using augmentation. Ibrahim et al., examine a different approach of exploring the physical world around us. Similar studies have been conducted on AR to train spatial cognitive ability and spatial contexts (Ibrahim et al., 2018; Sydorenko et al., 2019). These studies have revealed an important role of AR in introducing multiple layers of contexts to the learning process.

Conclusion

This chapter analyses related literature that spans language learning, augmented reality and educational games. It presents a set of concrete suggestions based on a framework for incorporating AR into L2 instruction. The rationale provides a basis for L2 curriculum development using MALL. This paper focused on the spatial, personal and social affordances of situated avatars and embodied performance as particular for L2 curriculum design with use of AR for language instruction. This survey addresses future research questions and methods for use of AR in a full language course. This paper identified key gaps in research on AR features and affordances that could be used for improving linguistic pedagogy in the digital age. AR is a promising technology for language learning, offering opportunities for not only contextual learning, but personalization and immersion.

CHAPTER THREE: SYSTEM ORGANIZATION/DESIGN AND DEVELOPMENT

The smartphones are easy to carry around and use anywhere anytime because they are small. However, designing an app with a 3D tutor image in a way that would make the tutor's expression clear can be challenging (see Fig. 1). This was a motivation for adding as little text elements to the app UI as possible.

Another reason I did not add many text elements to the screen was the idea of task-based L2 pedagogy, where the learner is focused on the task rather than meaning (Ellis, 2014, p. 134). I wanted the learners to focus on the articulatory representation and the sound production rather than textual explanations in this practice.

The objective for the design of the PinyinGuo app was to create a mobile environment that incorporates AR into computer-assisted Chinese pronunciation practice. The original design goal to employ AR is grounded in the motivation aspect of learning with AR. Empirical research has provided evidence for using AR to increase motivation in learning (Chiang et al., 2014; Khan et al., 2019). Therefore, I have decided to create two versions for an experiment. To measure the motivational aspect of AR, I designed an experiment in AR and non-AR.

Even with AR, providing quality feedback is labor intensive. Studies have shown that computer-assisted corrective feedback can provide individualized remote feedback in writing or oral proficiency ((Bitchener, 2008; Mendonça et al., 1994)Cucchiarini et al., 2009). Regarding pronunciation aspect, automated computer-assisted feedback has been an understudied domain (Derwing et al., 2013). Topping argues: “peer assessment…makes available swifter feedback in greater quantity.” (Topping, 1998). Therefore, I incorporated a system for peer-assessment into the app.

One of the few examples in the literature review on corrective feedback for pronunciation training was Yan et al.'s study of 300 English as a Foreign Language (EFL) students (Yan et al., 2018). The research examined the role of computer-assisted platform on peer review and pronunciation practice. Yan et al. developed a software to evaluate automated peer feedback on EFL pronunciation. The study included 300 EFL students. The results of the study have revealed that computer-assisted peer feedback can complement teacher's feedback and automated speech algorithms (Yan et al., 2018).

My hypotheses is that the interactive visualizations of the mouthparts in AR have more effect on learning and intrinsic motivation on learners than non-AR, which would result in better learning outcomes. The second hypothesis was that the online peer assessment can replace the expert evaluation, which would be supported if the participants have accurate assessment results.

The language literature suggest that pronunciation practice needs to be supported by corrective feedback (Tseng & Tsai, 2007). Corrective feedback can be direct and indirect. However, corrective feedback is associated with a language expert, i.e. the teacher to provide corrections of errors. In the previous chapter, I have discussed the importance of correct output in pronunciation practice. In this regard, mobile phones can facilitate feedback in a more efficient way than in person.

Speech recognition and AI technology can provide automated feedback. There are a number of automated speech recognition software for language learning like Duolinguo and Rosetta Stone. The learner says a word or a sentence and the system provides feedback of the accuracy of the pronunciation. For pinyin practice, there are no current open source software for syllabus recognition and feedback. There are a number of apps and websites that provide pinyin

practice and tone training like Yabla.com for Chinese language (*Yabla.Com*, n.d.), Arch Chinese tone drill ,

However, in the pinyin practice, there is a challenge associated with syllable recognition. Therefore, the software developed in this dissertation project can be used to gather data that can be scaled up to apply machine learning to Chinese pronunciation training including different accents and voice tones. Therefore, while designing this system, I had in mind covering a set of Chinese phonemes with all four tones that will be recorded and assessed by the other participants.

This data will first shed light on whether online peer assessment could be as reliable as expert assessment. Second, this data will show where there are difficulties in pronouncing Chinese sounds and tones and in identifying those. The difficulty to assess Chinese syllables and tones remotely lays in lack of context of the sounds. There is also no way to clarify and confirm what phoneme and tone a participant assessing. As for the peer assessment, it is challenging for the participants to assess their peers since they are still do not have enough knowledge themselves, therefore, it may be difficult to recognize the sounds and tones of other speakers. In addition, PinyinGuo is a task-based practice that does not provide information about linguistic theory in the Chinese language. Participants do not know what initials and finals are and how to evaluate them. An expert would have an idea of what evaluation criteria to use to assess Hanyu Pinyin pronunciation.

Based on these differences, PinyinGuo peer assessment has the following grading criteria: Evaluator Assessment for Phoneme Identification (1-5); Evaluator Assessment Performance Rating (1-5); Evaluator Assessment Confidence Rating (1-5).

Confidence ratings have been used in a cross-linguistic study to identify Mandarin tones by French speakers (Hallé et al., 2004). This study focuses on confidence among French speaking participants while identifying Chinese tones. Confidence rating is used in Pinyin Guo peer assessment to ensure that the participants are confident in their assessment.

The participants have the option to record each phoneme three times, listen to the recorded audio files, and select the best version (see Fig. 1). This process of recording was used to give participants an option to listen to the recordings and adjust their pronunciation. Ding et al. refer: "…second-language learners are more likely to succeed when they imitate a speaker with a voice similar to their own, a so-called "golden speaker" (Ding et al., 2019, p. 52). Therefore, when the participants listen to the avatar and then do their own recording, they may have more success in learning pronunciation basics.

Figure 9: Screen of Day 1, Recording State. AR Version

The recorded sounds are sent to the server. They then are made available for assessment the following day two participants for assessment. the assessment data along with the recording ID is is stored in the database and web panel for the experts. The assessment data would only be collected when completed. The experts then assess the accuracy of the ratings given to the recorded phonemes. The experts will also have the assessment criteria of the recorded phonemes, i.e. the accuracy of original syllables overall and phoneme initials, finals, and tone (Ding et al., 2019, p. 52). The example of the expert evaluation is below.

Design Goals

The pedagogical goals in this study were to create a platform to introduce Chinese phonemes and tones in a gamified way and to facilitate online peer feedback. These goals have determined the design of the software. PinyinGuo was developed to have a client interface and a backend logic to receive, store, and transfer peer-assessment data. Per IRB requirements, this data will be stored on an encrypted hard drive for five years. PinyinGuo was designed to meet linguistic and pedagogical goals in the remote settings. Therefore, an alternative of in-person practice and peer assessment wasn't suitable due to the COVID-19 pandemic.

Motivation is vital for language learning (Dörnyei, 1998; Lasagabaster et al., 2014); Gamification elements improves motivation in learning and completing tasks (Mekler et al., 2013). The proposed system includes gamification elements to increase learners' motivation. These elements are progress bar, timer, and correct/wrong badge. Research has demonstrated that gamification elements provide psychological satisfaction (Sailer et al., 2017). In addition, gamification provides guidance and assistance in improving learning performance (Mekler et al., 2013). The progress bar and timer encourage users to finish the sessions. The badges reward the user when they are choosing the right option. The feedback is not provided via the app, but the correct/wrong badges are meant to show participant's success in the game. These gamification elements were chosen over the other ones like leaderboard and rewards due to the scope of the project. The leaderboard would require competition among participants. This activity was not approved by IRB, since the participants were not meant to communicate with each other directly or receive identifiable information about each other. Rewards were not included in the game mechanics either, since the gameplay is too short to accumulate rewards.

The timer for the entire daily practice is sixty minutes. The timer for each mini-sequence is set to three seconds. The decision on the amount of time for each session was based on a series of usability testing. As result, one sequence (listen to a single phoneme and repeat) would last three seconds.

Another important design decision was the Unity plugin LipSync Pro by Rogo Digital (*LipSync Pro*, n.d.). LipSync provides an editor that animates a model's lip movements. Such visual support enrichens the learning process by providing not only audio and images, but also movements of the mouthparts. My hypothesis is that AR will provide even more of the motivation for learners than non-AR interface since the learners will see the lip and mouth movements from any angle.

I configured a progress bar from Unity Asset store (*ProgressBar Pack*, n.d.) to fit the number of PinyinGuo learning sequences. The purpose of this gamification element is to keep the users informed on how much of the practice is left and how fast they are going through the sequences. This concept refers to the guidance of users to maintain their interest, give them satisfaction from completed tasks, and guide them through the learning sequence. Overall, the effect of these gamification features on user's perception and motivation can be measured in future studies.

The timer for the entire practice is one hour. The participants are allowed to leave the app for 15 min and return. If the participants leave the study for more than 15 min, they will be considered withdrawn. There is also a notification system for reminding the participants to continue the practice. The notifications appear twice a day at 6:00AM EST and at 6:00PM EST. If the participant does not complete the daily practice of that day, they are withdrawn automatically.

The UX design reflects the complexity of the pedagogical tasks and usability principles. The client-side should be interactive and engaging to keep learner's motivation while the server allows multiple users to be logged and the data to be collated remotely. PinyinGuo consists of a node express application (backend logic) that receives events from the Unity client (Fig. 10).

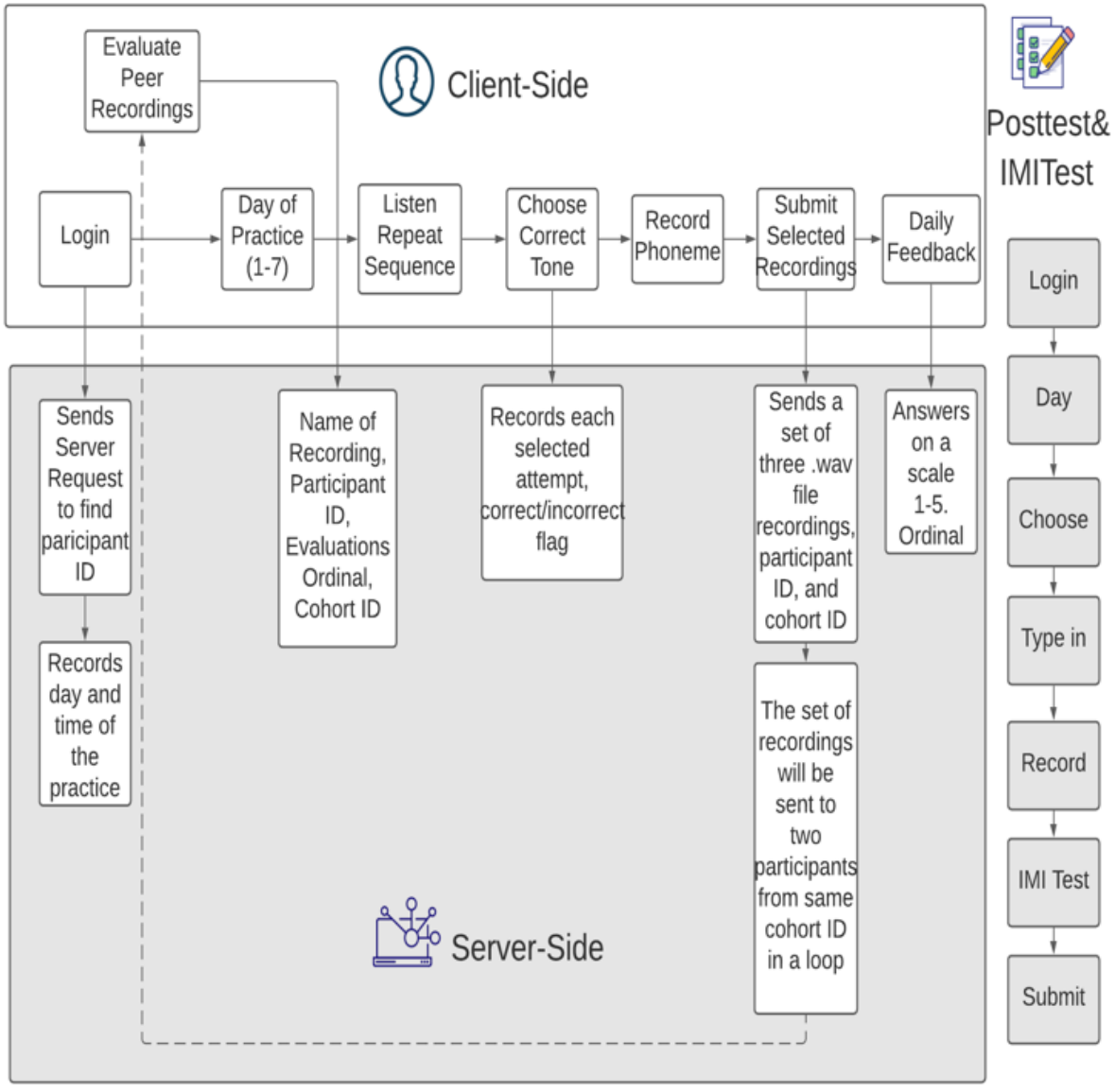

Figure 10: System Architecture Diagram

The proposed system serves pedagogical goals in terms of learners' collaboration, motivation, linguistic competency, and corrective feedback. The app was developed in two almost identical versions to do the experiment, i.e. mobile 3D version and mobile AR version. When I refer to the app or PinyinGuo development process, unless specified, I mean both versions.

Unity 3D is a cross-platform 3D engine that can produce both iOS and Android builds (Mekler et al., 2013) that are available to anyone with a mobile device. There are currently two main game engines for game development on the market— Unity 3D and Unreal Engine (*Unreal Engine*, n.d.). I chose Unity because of its s low barrier of entry to programming, i.e. C# in Unity vs C++ in Unreal. Second, the task management is easier in Unity as well. Unity provides access to the microphone for recordings, allows working with 3D imaging, and supports LipSync. In addition, Unity has a networking library that supports requests to the server .

I learned C# programming language to program the game functionality in Unity. In addition, I used Node.js for backend development. The server responds to the work with the database layer. MongoDB is the database where we store our data. The client makes requests to the server API, receives and sends data. The request methods are described in the Networking class while the server interacts with the database.

Mongo client for Node.js allows server code to requests to get data from collections and write data to collections. The system uses a Mongoose dependency connection, which is an ORM (Object-relational Mapper software) + mongo client. Mongoose is further used to connect the database:

```
const mongoose = require('mongoose')
```

To view the database, I used MongoDB Compass Isolated edition app for MacOS[10]. The collections of datasets appear in the compass edition. I have also used FileZilla to view the saved audio files from the recording practice [11].

MongoDB has a collection structure to store data. I have included the following collections to store the participants' data in this study:

{ collection: "feedback" }

In the feedback collection, I store the daily feedback data.

{ collection: "phonemes" }

Phonemes collection stores data about the audio files, i.e. file name, participant ID (who recorded the data), and day the audio was recorded.

{ collection: "users" }

Information about the users, the day they have begun the practice, the cohort number, and the activity log.

{ collection: "testData" }

Finally, in the collection testData, there is information about the participant, the day 7 posttest, and the test results (correct/incorrect choice).

The client side development flow is shown on Fig. 1. It includes the following screens: the login screen, each day of the practice, screens for various sequences, and daily informal feedback. In Fig. 1, the backend logic is highlighted in grey. This illustrates what information gets communicated to the server. The server data is available to admins on a web panel.

[10] https://www.mongodb.com/try/download/compass
[11] https://filezilla-project.org

I created an HTML landing page to advertise the project to the participants: . I used an album template from the Bootstrap library for markup languages to develop my study website[12]. The webpage was designed to advertise the study, inform potential participants about the requirements and the app. The webpage also has the link to the consent form on Qualtrics approved by IRB. I have also designed a poster in Adobe Express to advertise my study.

Prior to wireframing, I created a spreadsheet with the teaching content for each day of practice. Then I did wireframing of each screen in the app. The login screen has a field for the participant ID. The participants have to enter their user ID that had been sent to them when they signed up for the study. After the participants sign in, they see a Day screen. The Day screen shows the participants which day of the practice they are on. All the other days are locked, preventing the participants from accidently skipping or repeating practice. There is a withdrawal button on the Day screen as well. The withdrawal button allows the participants to withdraw from the study at any time. Once the participant clicks on the withdrawal button to mark the participant as withdrawn in the database.

The UI has two progress bars on the bottom of the screen to reflect how much of the single session has been completed and how much of the entire daily practice has been completed (see Fig. 11).

[12] https://getbootstrap.com

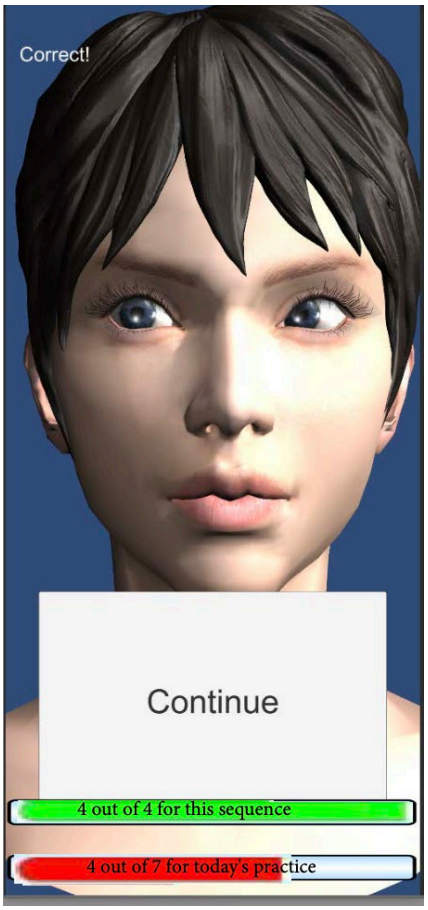

Figure 11: Progress Bar, Non-AR Version

The top bar marked green shows the current sequence and how much of the sequence is left. The full layout of the daily practices and its sequences is in Appendix A. The first three sequences include the following tasks:

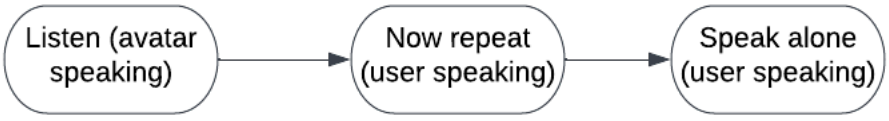

Figure 12: Flow of Linguistic Sequences in the Session

The fourth sequence is choosing the correct pinyin. The fifth sequence is designed to record sounds for peer assessment. Lastly, the sixth sequence is the peer assessment. Peer assessment is when each participant assesses two sets of recordings from two random participants.

In the beginning of the logic, I have added the participants that are working on the same day as the one who submitted the practice. Participants of the same day are listed as a “cohort”. If there is not enough participant on the same day, the scripts goes to the next day (the ones who are on the day before).

After each day of practice there is an informal feedback questionnaire. The participants answered the following questions on their perception about that day’s practice. The answers are recorded and sent to the server. The selection of the answers are on a 5-point Likert scale from strongly disagree to strongly agree. The participants had to choose an answer from the following statements: I want to continue practicing; Today's practice was very difficult; I had an excellent performance today. These feedback questions go up to day 7. On Day 7, the participants have to complete a post-test on the app and the IMI survey on Qualtrics (*Qualtrics*, n.d.). After the participants complete the post-test, they will follow an external link in the app to the Qualtrics survey. The survey records the participant’s ID and saves that data on Qualtrics.

I have created a web panel to reflect the saved server data with a more user friendly interface. In particular, the web panel reflects each user with a key (unique field) by email, the start date of the study, the active / completed / withdraw status, test results (phonemes and feedback), and recordings.

The UI elements in the daily practice are texts of the phonemes and the tasks, buttons, timer, correct/incorrect badges, and the progress bars.

In the recording practice, I have addressed issues that arose during the informal usability testing. For example, it was not clear to the users when to begin recording the sounds. The timer was there but both users thought that the timer is the countdown when they can start recording, not when the recording has started. Therefore, after rounds of consulting with my advisor, I have implemented a the following UX solution: I added a “be ready” button and a “recording…” button (Fig. 13).

Figure 13: Recording State in AR

The recording buttons also show the timer from 2 to 0. Then the button switches to "record" with a number of recording. The participants can record a phoneme three times. The button shows (record 1), for example. When the participant is ready to record, they click on the record button and the text changes to Be ready. Then there is a couple of seconds time to get ready (three seconds with a timer) before the recording begins. It says "recording". The timer to

record is set to three seconds. After all three attempts for one single phoneme are used, the participants can click on the buttons and listen to each recording. The participants are instructed to select the recordings for peer assessment that they think are the most successful ones. When the participants click on recordings, the button turns red. There is a “select recording” button below the recorded files. This button is used to make it clear which recording the participant is sending for peer assessment. The peer assessment panel appears on the screen before the next day of the practice. The assessment panel includes a play button, the assessment toggle group, and a submit button (Fig. 14).

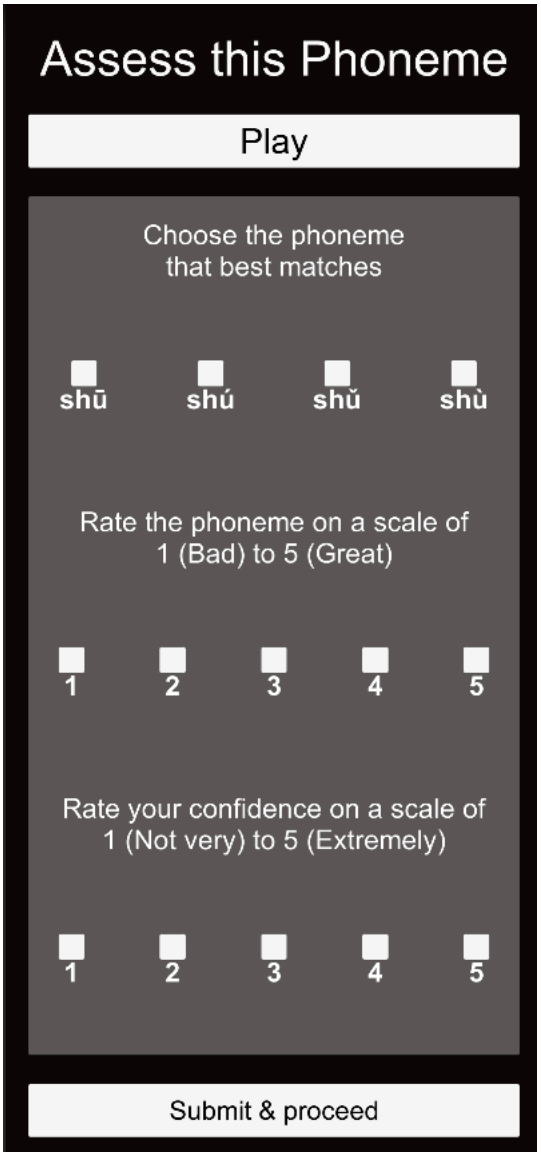

Figure 14: Phoneme Assessment Panel

The participants listens to the recordings, then they select the answers about the recordings, then they submit their answers. The statements are: 1. Choose the phoneme that best matches, i.e. identify the phoneme that you think is the one someone recorded. There is a selection from four phonemes. 2. The next statement is the rating: “rate the phoneme on a scale 1 to 4 from one being the worst and 4 being the best. 3. Finally, the participant has to rate the level

confidence about the phoneme that they selected (from 1 to 4). I have chosen the scale from 1 to 4 to visually have consistency with all the other buttons. The participants practiced recording four different phonemes on days 1-5. On day 6, the participants record five phonemes. See the table with the phonemes for each day in Appendix B. The first five days of the assessment structure were similar. However, the assessment panel on day 7 was different. Day 6 was a review day where the participants had to record all the phonemes they had practiced so far. Therefore, there were five recordings to assess. I have randomly placed the phonemes in the groups with one correct choice.

The UI elements include a 3D virtual tutor with audio support. To develop visual articulatory representations of how Chinese sounds are pronounced, I installed a plugin for high-quality facial animation and lip-syncing, Lip Sync Pro by Rogo Digital (*LipSync Pro*, n.d.). The plugin is available for purchase in Unity 3D Asset store. The decision behind selecting the virtual avatar was based on the literature review on virtual avatars. The research on users' perception of virtual avatars representation recommends a certain level of realism for the animation that should not be too realistic in order (Mohamad Ali & Hamdan, 2017). in order to avoid the uncanny valley effect, which is when a high quality human like avatar is so real that it creates a familiarity but shortly after has a terrifying effect (Mori et al., 2012). PinyinGuo was designed based on the research by Peng et al. on 3D talking head for pronunciation practice (Peng et al., 2018, p. 4). Based on these results, I have also selected a cartoon-like 3D animation of a woman's head as the virtual avatar[13]. I used to adjust the 3D image in the scene and to make sure the image can be viewed on any devices.

[13] cgtrader.com

The blend shapes were adjusted via the Lip Sync Pro Rogo Digital Plugin[14] to match Chinese sounds. Lip Sync Pro is an English alphabet-oriented Unity plugin. However, the plugin has a phoneme set feature that allows developers to adjust different phonemes or languages via blend shapes. I used this feature to ensure a proper Chinese articulation for every phoneme produced by the avatar. The audio recordings of the phonemes were made by a female Chinese native speaker to ensure the accuracy of input. The phonemes recorded by a Chinese native speaker were saved in .wav files and uploaded as a Lip Sync Data files to the Unity Assets folder. I have used the LipSync program to adjust the beginning of each syllable and the end of the syllable in the wave format to match the lip movements. The blend shapes on the animation were also adjusted to make eyes, lips, mouth, teeth sync with the consonants and vowels.

For example, the mouth when saying the phoneme [a] can be adjusted based on the proper articulation in Chinese. The phoneme [a] in every language is pronounced differently. To match the proper Chinese articulation, I have manually set the blend shapes and audio waves to make sure the mouth opens at the right timing and in the right width.

[14] https://lipsync.rogodigital.com

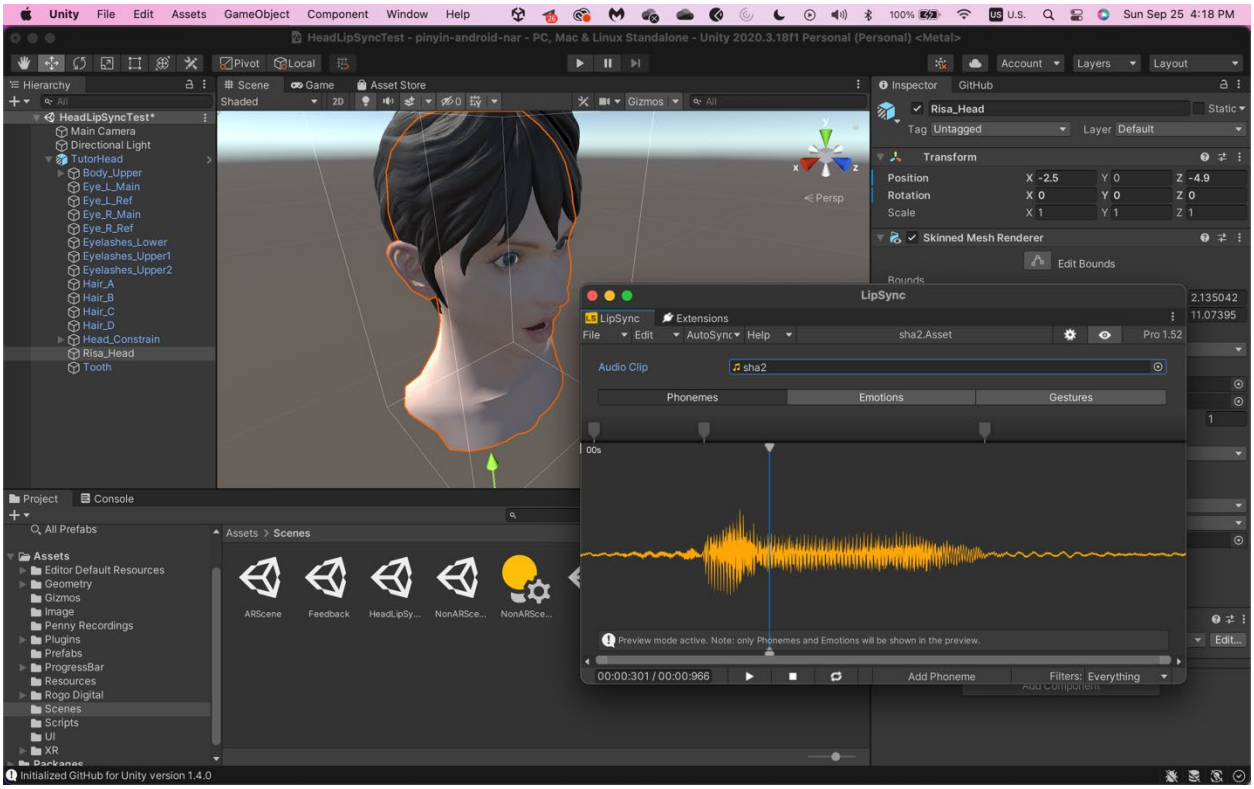

Figure 15: Rogo Digital Phoneme Right Width Adjustment for the Phoneme sha in Second Tone

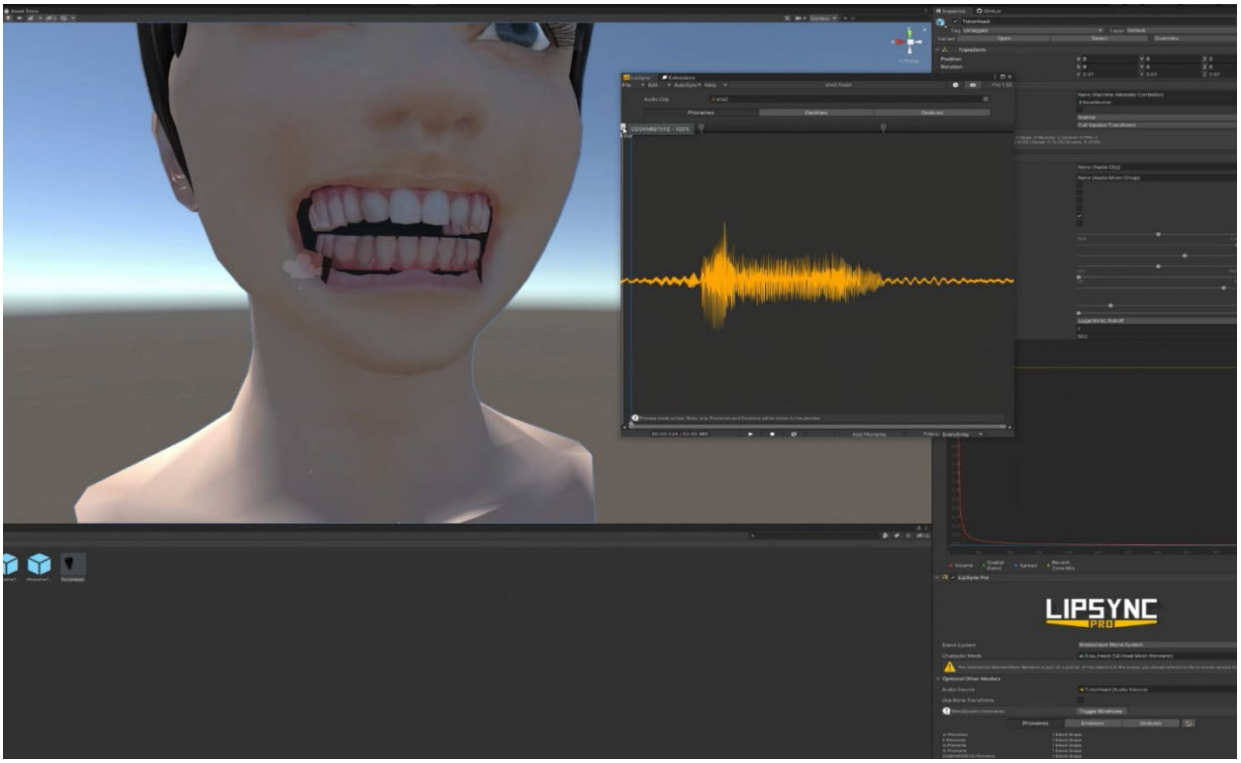

Figure 16: Wrong Width for the Phoneme sha in Second Tone

The learning content, was designed on the basis of linguistic theory and second language pedagogy. I have selected the first lesson from Zadoenko and Huang Chinese Language Beginner course to develop the practice (Zadoenko & Huang, 2019, pp. 15–20). The choice of this particular practice is that I was not able to find many academic textbooks for independent phonological course in Chinese. I also wanted to cover the four tones of Standard Chinese (level, rising, falling-rising, and falling) in a relatively short period of learning practice (Li & Thompson, 2009). I was also looking at including several basic-level initials and finals. A similar study to mine in terms of the research questions is by Peng et al. The research questions that are somewhat similar to mine are as follows: "RQ2: What are language learners' impressions of the three presentation conditions of the ALT?.... RQ3: How do language learners perform using the ALT in the three presentation conditions?" (Peng et al., 2018, p. 28).

The study included 60 Mandarin syllables for the experiment (Peng et al., 2018). The study limitations were the lack of tones in the syllables presented in the study. The study had only three second (rising) tones, two third (falling-rising) tones, and two fourth (falling) tones (Peng et al., 2018, p. 38). The study analysis was conducted on the user performance for the initials, finals, and the tones.

To address these limitations, I have ensured that the PinyinGuo practice includes all four tones in a balanced manner. The number of syllables in PinyinGuo practice is twenty. The design of the learning practice was primary to the app design process.

The app has undergone a number of changes based on how the objects are situated in the scene and the size of the UI elements. For example, the buttons and the text with instructions were made bigger after the user feedback during the usability testing.

In the AR version, the head can be adjusted in the space based on the view preferences of the user. The user can place the AR avatar onto any location in the space, reduce and enlarge the size of the head, and turn the head. These UX decisions were made after an inconclusive usability testing when I realized that the head could not be rotated or modified in space. This issue sometimes created a frustration with the avatar being too far from the viewer or too close. My intention was to benefit from the AR affordances where the user can manipulate the avatar in their physical space.

Indeed, the virtual 3D tutor head in the AR version is situated onto a natural environment of the learner; possibility for the learner to adjust the angle of the head in practice manually. Such interactivity allows the learner to see different angles of articulation apparatus and develop a personalized approach depending on their view preference. The head can be turned with the finger so that the user can see the head at different angles. This approach comes from the hypothesis that a virtual tutor in AR will provide better learning effects by viewing mouth parts from different angles in the physical environment as it was a real person tutor. The benefits of an avatar over a real person include less awkwardness and comfort in communication (Rogers et al., 2022). In addition, a virtual tutor allows the user to get closer to the avatar to view the details of the mouthpart compared to in-person interaction, while this might be inappropriate.

All other UX elements area the same in AR version as in the non-AR version. To develop the AR version, I worked in ARCore for Android and with ARKit for the iOS build. The iOS build requires adding participant devices to the Apple developer account , which makes it more challenging to add the participants to the study. Apple requires developers to register test users devices using a UDID (Unique Device Identifier).

This chapter documented the client and the backend specifications of the software PinyinGuo in both AR and non-AR versions. In this chapter, language pedagogy and usability theory have influenced the UX/UI design decisions and the full-stack development of the PinyinGuo app. Usability study has improved app design by changing the progress bar, test elements, buttons, button colors, and timer on recording. This study was driven by usability theory: usability theory determined how I incorporated user results into the design of the app.

The motivation behind creating this app was to answer the research questions about the role of AR affordances in enhancing Chinese language pronunciation practice. I have provided an overview of similar research projects to PinyinGuo that involve avatars and similar commercial apps. This chapter surveys computer-assisted language pronunciation training in the context learning outcomes and peer assessment. This chapter describes how the design process of PinyinGuo was determined by the literature review of the exiting apps, linguistic theory, and usability testing.

CHAPTER FOUR: METHODOLOGY

Usability Testing

The method used in this dissertation is design research followed by verification through usability testing of the experimental application. The key research questions of this my dissertation is related to the method of online experiment. In this study, I am evaluating two interventions – AR and non-AR. A similar research design to these methods was applied by Che Dalim et al. with an experiment with two interventions AR and non-AR (Che Dalim et al., 2020). The settings of an online experiment were previously used in research for large scale participation of 570 participants (Grimmelikhuijsen & Meijer, 2014). In the times of COVID-19 pandemic, an online experiment allows participants to participate in the study remotely without being exposed to potential risk associated with the pandemic.

Design research is an iterative process. Iterative design is an ever-changing process where "the initial problem formulation phase is not static, but reflects the current understanding of a problem (Adams & Atman, 1999, p. 2). During the design of PinyinGuo, I implemented changes into the software based on data that needed to be gathered or functionality of the app. The iterative design principles were applied on every stage of the app development. The usability testing included first consulting with Dr. John Murray about the features and functionality of the app interface.

After the app was in a low fidelity prototype, I have conducted an inconclusive usability testing with two participants of the target group for the study. The participants were 28 and 29 years old. They have never had any training or knowledge of the Chinese language. The participants used a concurrent think-aloud method to go through the procedure. The concurring

think-aloud method was adopted in studies on usability testing and technical communication (Cooke, 2010; Nørgaard & Hornbæk, 2006). The concurrent think-aloud method is used to understand how the participants are using the system and whether they are following your instructions. After the app usability testing, the participants completed an oral Intrinsic Motivation Inventory (IMI). IMI inventory was used to determine the motivational aspect provided by the system.

The participants were following my instructions. I was present in the room throughout both testing. I did not facilitate any information regarding the app usage. The instructions provided to the participants were the following: they needed to login into the app by entering a number one (1) into the login bar. In addition, the participants had to do the practice and speak their thoughts about the interactions out loud. The procedure was uniform for both participants. They completed the usability testing in different rooms one after another so they were not aware of each other's answers.

The voice memos with the feedback and the IMI test were saved along with the participant feedback in a document on an encrypted hard drive. The total time of the usability testing was 45 minutes. Both participants made comments and suggestions about the app during the usability testing. I recorded the interaction with the app on a video. Additionally, I did a voice recording of the feedback after the practice.

First, the participants had to enter a participant ID to log into the app. Then a panel with days showed up. At the beginning of the practice, one participant started randomly clicking on the buttons to proceed to the next sound. The participant mentioned that they could not see the text very clearly since the text with the instructions was too small. The participant had no difficulty understanding how to use the interface to accomplish the choosing phoneme task.. The

participant said that the selected state was evident and easy to navigate. However, the participant did not understand when the recording began. The recording started without any timer or instruction on when to start recording. The text in the feedback panel and the selection buttons were too small. Therefore, the participant did not understand what he needed to do in the feedback section. The first participant pointed out that the controls should be improved. He said that the UI is not very intuitive. He said the app UX should be comfortable to navigate, i.e., for the user not to be intimidated by the amount of text, the text font, and the absence of return buttons. He said the controls were too square, and he was afraid of missing a button. The buttons were too small, in his opinion. He said the UX is a bit outdated. In addition, he mentioned that the UI was a bit too busy and needed to be simplified. The participant mentioned that the avatar should be highlighted and brought to the user's attention. One option would be to make the head bigger and the buttons smaller. This suggestion makes an interesting point about the role of virtual avatars in AR. AR shifts the focus to the main element of practice—the avatar's mouth movements. Other UI elements are secondary.

The second participant found it easy to navigate through the practice up until the recording session. He did not understand why the sound in the practice was repeated by the avatar twice. While recording the sounds he thought he had to repeat each of them twice. The participant pointed out that the app should have clearer instructions about the importance of what the phonemes look like. He was under the impression that he was supposed to be pay attention to the listening not to the text. This was an issue with the the usability of the application where the text of the phonemes was too small. Therefore, the application did not reflect the pedagogical approach of paying attention to what phonemes look like. The phoneme text with the tone mark is important for the practice. The participant did not know thar they had to pay attention to the

tone marks. Also, the recording timer was taking too long. The timer counted down the recording for 5 seconds. The participant finished recording in less than 3 seconds but the count down was still happening, The recording button should have a timer to show how long the recording will take. The participant was confused when to start the recording and how much time is remaining to finish the recording. The instructions should be in bigger font since the participant did not see the text at all. The font size was changed to 65 font size which is based on the size of the screen. The recording buttons were big enough but the text with the instructions for the recording state was too small.

Based on the usability testing results and the feedback from users, the app has undergone a number of changes in the UI and the backend logic: text UI, button UI, timer, and progress bar.

After usability testing, I conducted app testing looking at the server and checked if the data was saving to the server. For testing purposes, I have created a button “advance day” to be able to proceed to the next day of the practice (Fig. 17).

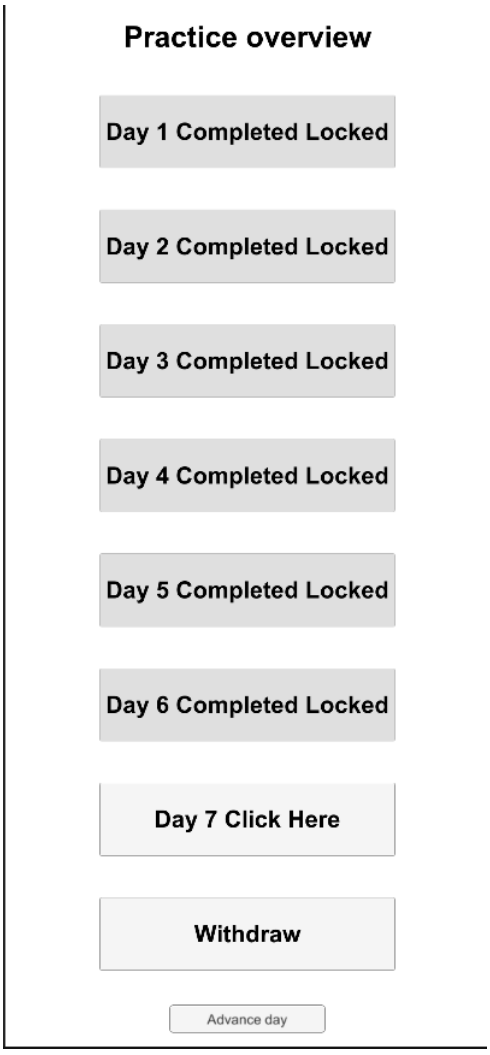

Figure 17: Testing UI

Additionally, I have changed the cohort timeframe to 15 minutes instead of one calendar day. I have used Postman [15] agent to generate test users. The app was ready for deployment after I tested the server and the client.

[15] https://www.postman.com/product/what-is-postman/

Heuristic Analysis

A Chinese expert approved for this study has reviewed the AR version of the app. The expert wrote a brief review of the app in the Chinese language. The review was saved in a Word document and is kept on a hard drive. The expert assessed the interface and the learning content. The expert stated that the app fills the gap of the lack of the tools that teach students correct pronunciation.

The expert noted that in recent years the echo method has become popular in the foreign language teaching field to teach pronunciation (Do et al., 2022). The echo method is in line with the method used in PinyinGuo allowing learners to listen to real voices and imitate them as much as possible, so that their pronunciation can be close to real voices (Chung, 2017). The expert stated that with the help of PinyinGuo an app, Chinese learners"will be able to fully master Chinese pronunciation, thereby increasing their confidence in speaking and expressing themselves more easily and without hindrance".

The expert pointed out that in the real world, however, native speakers rarely pronounce only one syllable. The expert argues that in future research, to help students become more familiar with the pronunciation of native speakers, some word pairs would be worth the effort. The expert stated that an improvement would be adding a long sentence practice after the monophony. The expert recommended to make the duration of sounds longer. Although the app is designed for the purpose of providing real voice, it may be difficult for beginners to grasp the difference between the four tones and lose the good intention of developing this app.

In addition, the expert suggested to add an option of longer sounds in addition to the shorter sounds. This approach can be useful for beginners since the learners can feel the

difference between the four sounds better. The expert believed having short and long sequences options would lay a solid foundation for pronunciation.

In the part where learners record their personal voices, if a comparison with real voices can be provided, it can also help learners to observe the gap between their own pronunciation and real voices, and they will be able to further correct their own pronunciation.

The expert also pointed out the importance of having the AR avatar placed in one spot and not move the camera. In their opinion, the image should be fixed in one spot to be will be more comfortable visually.

This user feedback highlights the importance of the PinyinGuo method and sheds lights on the future design research that can improve learner's results and experience.

Study Procedure

Acknowledgement About Retrieving Section From A Previously Published Paper (Sinyagovskaya & Murray, 2021)

The research method used in this study is usability testing and an experiment. This experiment was conducted entirely online and on personal devices (iPhone and Android). Since this is a multi-site experiment, we wanted to ensure the least amount of uncontrolled variables affecting the treatment. For example, we locked each day of the current intervention, so the participants cannot proceed to the next day without completing the current day.

The independent variable is the teaching content facilitated by the app. Another set of independent variables is peer assessment. Quantitative and qualitative methods measure the sets of dependent variables. The learning outcomes will be measured with a posttest. The tool to

measure the participants' answers and test scores will be the server. The posttest will be conducted via the app, and the results will be saved to the server.

IMI test will measure motivation among learners (Plant & Ryan, 1985). Participants' answers were recorded on Qualtrics and associated with the participant's ID. We also measured continuous motivation and enjoyment from the practice via the daily practice informal feedback. This information will shed light on how motivating and engaging our system design is as well. The feedback questions will provide us with more usability data. This data is recorded on the app and sent to the server. Lastly, the readability of peer assessment versus expert feedback was measured by the expert in terms of quality of initials, finals, and tones and the input the recordings have received from the participants. This data will be stored on a server.

This research design of this study is an online quasi experiment. Since the participants do not have to come in person, there was a goal to minimize the number of uncontrolled variables in the intervention. Therefore, participants could start at any time from the moment they received the link to download the app.

The cohort in PinyinGuo is a group of participants started on the same day. I developed a logic in the code for cohorts of participants to distribute the peer assessment between participants started on the same day. In peer assessment practice, participants have to assess each day the recordings from their peers. In order to distribute the recordings of the same day, I have connected the day count to the participants who started on that day. A minimum size for a cohort is two people, since this is the only way participants can assess each other. However, in practice, I waited for a cohort to be at least four people, since a lot of people who started did not finish the practice. So, if a cohort had five people, and four of them withdrew over the period of seven days, the assessment data was not gathered properly.

The logic of the backend counts the days in the app not the by the calendar days but the current day of the practice. This allowed me to group participants by cohorts of each day. So, if one participant withdraws from the study, the next participant will take their place in line. This backend logic was developed for the peer recording practice. The logic was designed so that the person who was on day two would receive two sets of recordings from two participants of the same day.

This is a multi-site experiment. Therefore, the study procedure had to go through the IRB approval prior to the intervention. The email addresses were replaced with a unique participant ID to ensure data is not linked to the identifiable information at the conclusion of the study. The identifier-email file will be kept for a minimum of 5 years separately in an encrypted file. The audio recordings will be associated with the participant ID and saved on a hard drive.

Recruitment

The requirements to participate in the study were the following: the participants needed to be 18-45 years old without any prior Chinese language training or knowledge. The choice behind this age category was determined based on the results from a recent study on the correlation between age and linguistic competence (Baker et al., 2020). The study included neuropsychological assessment of 30 young adults (18-45 years old) and 30 (65-85 years) older adults during a Chinese language practice. The study revealed an age impairment factor in language learning abilities among the latter group (Baker et al., 2020). I am referring to these results to determine the age range of my sample. Since the age group older than 65 years old has additional age-related challenges with learning a new language, this would require further research questions associated with this age group. The focus on the lower age group

acknowledges the limitations of my approach to address differences identified and helps to scope my work.

The study admits participants on a first-come, first-serve basis and not on the basis of any identifiable characteristics as long as prospective subjects meet the eligibility criteria. The advertisements were broadly distributed on social media without restricting access in any way for the duration of the study.

The participants could leave the research at any time. The compensation for participation in this study is a $40 Amazon gift card. If the participants withdraw before the end of the study, they received an amount proportional to the number of days/sessions that they contributed to ($30/number of days). If they complete all sessions, they received an additional $10, for $40.

I posted call for participants on my social media (Facebook, Twitter, Instagram, Reddit). I mentioned in my social media posts that the link can be reposted further. In the recruitment posts, I shared an email with the website URL link https://pinyin.mrl.ai/study to the study with UCF faculty and staff. Dr. John Murray and I posted an email to the faculty listserv for digital media and Text and Technologies (with their permission).

A recruitment flyer was shared by Dr. Amy Giroux with faculty of Modern Languages Department and the Chinese American Librarians Association's listserv and Facebook, the Orlando Chinese Professionals Association and the Chinese Language Teachers Association–Florida Chapter (CLTA-FL) 's WeChat groups.

The website has a description of the goals of the study and the requirements to participate. I uploaded the screenshots of low fidelity prototype to the website. The screenshots and the text on the website received an approval from IRB. The website has a clear interface with

two buttons: Information about Research and Requirements & Consent Form. The information about research page has a brief description about the study and the eligibility criteria.

When the prospective participants click on the Requirements and Consent form, they are directed to the Qualtrics survey. In the survey, there is a a consent document for download (approved by IRB). The prospective participants have to agree to the consent form in order to continue. The next page of the survey is only applicable to iPhone users where they should provide their UDID. UDID is a unique device identifier for Apple devices that is used to deploy test applications to Apple devices.

Intervention

I have created a spreadsheet where I manually assigned AR or non-AR builds to the participants. I also have included the information on whether it is an iPhone and an Android build. I have taken cohorts of participants of four to distribute the app build. The choice of this number was determined by the small number of participants who signed up for the study in the beginning. Part of the assignment is peer-assessment. Therefore, I needed a cohort of people who would assess each other's voice recordings each day.

The experiment compares two different interventions – a group of participants with AR and a group of participants with non-AR. The intended number of participants was forty people twenty people for AR and twenty for non-AR. The sample size was determined by the sample size in the previous studies in the area of applied linguistics (Che Dalim et al., 2020; Peng et al., 2018). A smaller sample size is used in the field of language learning.

The total number of prospective participants who signed up for this study in was seventy people. The total number of participants was thirteen people n=13. The number of participants

completed seven days of the study was two n=2. The explanation of research limitations is Chapter 5.

The participants were randomly assigned to either the AR or non-AR condition, represented by two different versions of the same app. The learning content in both apps was identical, the only difference was whether the tutor was displayed using AR or not. Appendix A illustrates each day of practice the participants had to complete.

On day 1 of practice, the participants had to go through six sequences. The first one was Listen and Repeat sound ma1 (first tone). The second one was to Listen and Repeat ma2 (second tone). The third one was to Listen and Repeat ma3 (third tone). The fourth sequence was Listen and Repeat ma3 (third tone). The fifth one was to Choose the correct Pinyin from the four previous sequences. The task was: "Listen to the sound produced by the avatar two times. Choose correct Pinyin with tone. Time limit 5 seconds." The sounds were: ma2, ma3, ma1, and ma4. The app would tell them the correct or incorrect choice. If the choice was incorrect, the app would not allow the participants to proceed to the following sequence. The last sequence for recording started on a three-second count. The text on the button said: "Be ready," and immediately after, the text said: "Recording…." This logic was written based on the usability testing feedback. The participants in the usability study had a hard time figuring out when they were supposed to be recording. Therefore, we implemented a logic with the timer that allowed the participants to get ready to record the sound. Furthermore, the recording sequence provided the participants with an option to record the same sound three times and select the recording they thought was the best.

After the recordings were submitted, they were anonymously distributed among the cohort. Each person had to assess two recordings. The recording data were saved on the server

with the participant ID and the date when the file was recorded. Day 1 did not have an option for assessment of the recording because this was the first day of the intervention. But on day two, the participants would see an assessment panel before the started the second day of the practice. The assessment panel would have a Play sound button that the participant would need to assess. The panel interface included toggle groups with the following assessment criteria: "Choose the phoneme that you think best matches the sound produced"; "Rate the phoneme on the scale 1 (Bad) to 5 (Great)."; "Rate your confidence of what phoneme it is on a scale 1 (Not very) to 5 (Extremely)." Each participant was supposed to receive two of those panels every day from day 2 to day 7 included. The assessment panels were programmed in such a way that the participants could listen to the recordings once. The timer was 0.5 settings to listen to the recordings. The rating scale was 1 to 5.

The first set of recordings (from Day 1) was: ma1, ma4, ma3, ma2.

After each day of practice, the participant had to complete an in-app informal survey that collected feedback about the practice. The participants had to click on the buttons to select answers to the following feedback questions. The numbers are on the Likert scale (5): from strongly agree to strongly disagree

1. I want to continue practicing

Likert scale (5): from strongly agree to strongly disagree

2. Today's practice was very difficult

Likert scale (5): from strongly agree to strongly disagree

3. I had an excellent performance today

Although Days 2—5 have different phonemes in the practice, the backend logic, UI, and learning tasks are the same as Day 1. However, day 6 is different since it has a review of

everything the participants have studied in the last 5 days. Practice 6 is a randomized practice of previously learned Pinyin. The first sequence is still Listen and Repeat after the avatar to review the phonemes. Then, Listen (avatar speaking) and choose the correct Pinyin. There are eighteen phonemes in this sequence. Following the Choosing state, there is a Recording state. The participants had to record five sounds instead of four in this practice. The number five is determined based on the number of unique syllables learned on each day. For example, on day 1, there was a syllable MA with different tones. On day 2, the phoneme was MU. On day 3, the phoneme was HU. On day 4, the phoneme was SHU. On day 5, the phoneme was SHA. Therefore, on day 6, we have randomized the five phonemes by assigning them with different tones: shu1; ma3; sha4; mu2; hu2. After the participants had to select the correct Pinyin for the randomized practice, they had to record the last set of phonemes for peer review. They also had to complete an informal feedback survey about the practice.

On day 7, the participants had to assess five phonemes from day 6. The phonemes are: shu1, ma3, sha4, mu2, hu2. After that, the participants had to complete a customized posttest. The posttest included the following tasks:

1. Listen to the Pinyin and choose the correct number of tone. 1 attempt 3 seconds.
2. Type in Pinyin and the number of the tone following after the Pinyin without a blank. Example: ma1
3. Record the following pinyin sounds

The full post-test is in Appendix A in the Day 7 sequence. The anticipated time of completion was five minutes.

After the participants finished the post test, they were redirected to the external link on Qualtrics. The link was embedded in the app. The participants had to complete two

questionnaires from the IMI. First, the participants had to complete the Activity Perception Questionnaire. Then, they would have to complete Task Evaluation Questionnaire. The anticipated time for the completion of both questionnaires was ten minutes.

Measuring Instruments

The learning outcomes will be measured with a post-test. The tool to measure the participants' answers and test scores will be the post test The posttest will be conducted via the app, and the results will be saved to the server. The list of variables is available in Table 1.

Table 1: List of Variables

Independent Variable	Dependent Variables	Measurement Instrument	Tool
Teaching Content: 3D avatar, pinyin practice, assessment test	learning outcomes (participants' test scores)	Posttest	App–server
	motivation	IMI test	Qualtrics linked with participant IDs
	Continuous motivation	Daily Practice Feedback	App-server
	Reliability of peer-assessment	Expert evaluation	App-server and database

Data from both interventions in the posttest will be measured with a t-test. This experiment does not include a pre-test since this is a beginner Chinese practice. Furthermore, the requirements to participate in the study were specifically stated: This is a beginner Chinese practice, so if you already have any level of Chinese language, you are not eligible to participate in this study." Therefore, a pre-test is unnecessary since the participants did not have any level of Chinese to compare to.

The informal feedback will be a descriptive piece. The feedback results will shed light on the continuous motivation aspect that could be used for future research. In addition, the feedback survey will provide us with more usability data. This data will be recorded on the app and sent to

the server. For example, when a learner selects items on the Linkert scale, each value will be recorded on the server.

The independent variable is the teaching content facilitated by the app. I am collecting peer assessments, but they won't be presented to participants. Instead, the accuracy of peer assessment versus expert feedback will be measured by the expert in terms of the quality of initials, finals, and tones and the input the recordings have received from the participants. This data will be stored on a server.

I use the Task Evaluation Questionnaire and Activity Perception Questionnaire from the Intrinsic Motivation Inventory (IMI) to answer RQ1 about motivation. I have selected the IMI as a measuring instrument since IMI is one of the most reliable and validated methods to measure intrinsic motivation ("Intrinsic Motivation Inventory," 1994). McAuley et al. describe the IMI questionnaire as a "flexible assessment tool" where the inclusion or exclusion of any dimensions does not affect other factors (McAuley et al., 1989). The instrument can also be adjusted based on the activity and topic. Studies on the perception of mobile-assisted language applications have used IMI to measure participants' attitudes towards an activity (O'Reilly, 2014). Based on the features of flexibility and reliability of the IMI questionnaire, I have applied this method. I have selected the interest/enjoyment subscale from the IMI test and added two minor modifications. I replaced the word "task" with "recording the assessment" and the word "activity" with "practice" in two questionnaires, respectively. The IMI test will measure motivation among learners (Plant & Ryan, 1985). Participants' answers will be recorded on Qualtrics and associated with participants' IDs.

The Task Evaluation Questionnaire consists of twenty-two items. The questionnaire with modified features is in Appendix D. The Activity Perception Questionnaire includes twenty-five questions. The Activity Perception Questionnaire is in Appendix E.

The daily practice metadata is saved on the server. This session metadata will give us an understanding of how much time the participants spend on the app and what type of errors they make during the practice. This data is visualized in a table and analyzed as a descriptive piece.

The readability of peer-assessment data was evaluated by two language experts. I serve as the primary language expert for this study. My credentials to serve as a Chinese language expert include a BA from Institute of Asian and African studies at Moscow State University. I have 36 credits in Chinese language including 1748 in-class hours. I spent one year (2010-2011) in Taiwan on a Ministry of Education (MOE) Scholarship. I took advanced Chinese classes during this time. I have published a Chinese language self-study guide, Easy Chinese, in 2017. I have taught Chinese in class and privately since 2012.

The second expert, Zijun Shen, is a lecturer in the Chinese program at the Department of East Asian Studies at Indiana University. Before coming to work in the United States, Zijun Shen had spent most of her time teaching in Taiwanese universities, mainly at National Taiwan University and National Chengchi University. The expert has more than ten years of experience teaching Chinese.

The recording data with original audio recordings and peer-assessment results are stored on the server. The server needed due to the nature of the online-experiment. The data collected from the participants' interaction with the app, feedback, and recordings should be stored on a server. The recordings are saved immediately on the server and stored for five years. To evaluate the peer assessment, I will use a 5-point likers scale for successful evaluation of the original

sound (from 5 successful to 1 fail). I will also evaluate the accuracy of original syllables overall and phoneme initials, finals, and tone (Peng et al., 2018).

Follow-Up User-Testing

The online experiment design in this dissertation had multiple limitations associated with the length of the experiment and the retention rate. I have also fixed some bugs in the server code after the experiment based on the data received from the participants during the study.

I conduced a round of usability testing using the logic of a one-hour uninterrupted intervention. I used convenience sampling by asking two adult individuals to download the app in AR and non-AR and do the entire seven-day intervention in one day. Both individuals claimed that they never had Chinese practice. As described in the Usability Testing section in this chapter, I changed the server script logic to allow to advance next day immediately after the day the first day is completed. The participants completed the practice with IMI questionnaires. The results of the follow-up usability testing are described in the Findings Chapter.

In this interdisciplinary work, I have designed and developed a novel peer-to-peer pronunciation system with augmented reality features. The research design is one of the significant contributions of this dissertation. In addition, asynchronous anonymous online peer feedback and AR in pronunciation are novel in this dissertation. The overall procedure goals were to develop a new pronunciation software in AR, test AR effects on motivation and learning outcomes, and evaluate reliability of anonymous asynchronous online peer assessment. Future work to refine it needs to be done in scaling up the online experiment for a larger number of participants, using more phonemes for the practice.

CHAPTER FIVE: FINDINGS

The results of the analysis of the existing apps are recommendations for developing new language applications with AR. The recommendations are based in close analysis of seven mobile learning apps Duolingo[16], Mondly[17], Busuu[18], Babbel[19], HelloTalk [20], Memrise [21], and MondlyAR[22]. Although MALL offers new avenues to improve L2 language instruction, existing commercial language apps do not implement pedagogical and linguistic theories in language acquisition. The reviewed apps are the subject of several studies examining the course structure and design elements. The results of this analysis show that these seven apps have a number similarities and differences.

Table 2 presents the design and the content of the most popular commercial language apps. The result of the analysis are the similarities and differences in the content and design of these apps. This analysis allows to make a claim about future development of MALL apps.

[16] https://www.duolingo.com
[17] https://www.mondly.com
[18] https://www.busuu.com
[19] https://www.babbel.com
[20] https://www.hellotalk.com
[21] https://www.memrise.com
[22] https://www.mondly.com/ar

Table 2: Breakdown of Content and Design in Commercial Language Learning Apps

App Name	Teaching Method	Content Presentation	Gamification Elements	Peer Collaboration
Duolingo	GTM	Game-like app/ lesson tree	Badges, leaderboards,	Peer translation, chat forum
Mondly	ALM/TBLT	List of lessons by themes	Badges, leaderboard	Interaction with friends (Facebook, US, Global)
Busuu	CLT	List of lessons	Progress bar	Community page for adding friends who learn languages
Babbel	TBLT	List of lessons, virtual cards	Achievement points	n/a
HelloTalk	CLT	Social media-like interface, live streams, chats, personal profiles	n/a	Language exchange platform
Memrise	GTM, CLT, TBLT	Short videos with phrases, AR-camera to capture translation in the real word	Points, leaderboard, progress bar	n/a
MondlyAR	TBLT	AR teacher	Sound effects	n/a

This data shows that a number of these apps is still using outdated language teaching methods like GTM. The most popular app Duolingo uses GTM. Despite the fact that Duolingo has been critically reviewed in the literature, there is still a large number of downloads of this app. The interesting part that there is not much peer collaboration going on in those apps. The community option in Busuu and the leaderboards in Memrise, Mondly, and Duolingo are not

exactly peer-collaboration. From all these apps, Duolingo is the only one that has an option to collaborate on translation of words. Duolingo gathers a lot of translation data from the numerous users that the app can use to create new algorithms (Garcia, 2013). This approach empowers users to get confident in their translation skills as well as provides opportunity to create more features in the future.

TBLT is also a popular method used in the apps in this analysis. TBLT works well with mobile technology that can contextualize the learning content immediately. Short sentences in the form of the flashcards with audio support replace lengthy explanations of grammar. TBLT is used to some degree in all of these apps.

In terms of design, commercial mobile apps have begun to explore how language learning can take place through gamified activities and collaborative curated content, contextualizing language usage and engaging learners remains a challenge. However, MALL in second language acquisition is one under-explored domain which has both great demand and unique challenges.

These apps incorporate gamification elements into language learning. Lotherington views such gamification approach is technology-driven and not pedagogically-driven (Lotherington, 2018, p. 211). A large body of literature reviewed in Chapter 3 has shown that there are opportunities to develop new tools for L2 instruction to make learning process more immersive, personalized, and effective.

Technology-enhanced language learning is a research area in language acquisition focused on enhancing learning with new technology focusing in "basic description of what technology can do and toward a mature theory of what it should do to support teachers and learners, and when and how different technologies can best be used to support learning"

(Golonka et al., 2014, p. 93). Education and technology have long had a relationship through educational games, including classics such as Math Blaster and Oregon Trail. In the language learning domain, there are fewer full examples and many examples of applications like Duolingo that adopt game-like features while not being themselves a game. Another problem in the technology-enhanced learning is that the current language apps do not deliver enough value and interactive features that AR could offer. This survey addresses the parallels with casual and educational games in the context of augmented reality and language acquisition.

The design of the apps reviewed in my analysis reflects the concept of learning anytime and anywhere, importantly fitting language learning into interstitial times and places in similar ways to casual game mechanics that reward short, regular sessions (Juul, 2009).

Combining interactivity and 3D visualization holds great potential for L2 instruction. In traditional settings, L2 teachers use images and audiovisual modalities to illustrate new vocabulary, concepts, and situations to provide learning context. This is accompanied by rote learning from static media, usually textbooks, along with practice with other students or the instructor.

The design of the novice app was intended to have interactive features to make learning more engaging and fun. However, the design of PinyinGuo was driven by pedagogical goals. Therefore, it was essential to conduct the state of the art analysis of each app to determine the design concept for the app.

Usability Testing Results

The usability testing before the intervention is a descriptive piece. The results of IMI questionnaires show that the users find the app interactive and useful. The activity (practice)

perception questionnaire showed that there is value in the practice. Both participants stated that the practice was “very fun”. Both stated that the practice was valuable to them.

The results of usability testing led to minor changes in the app design. The design changes were associated with learning content and visual accessibility. In regard to learning content, the UI text elements should emphasize the importance of the pinyin tone marks. It was not obvious to participants what As a results of this user feedback, the size of the tone marks was increased to be visible for learners. The texts size was increased as well.

These UI changes aimed to reflect the pedagogical method incorporated into the app design. The linguistic method used in PinyinGuo is TLBT. Highlighting the tones is crucial for giving learners an idea of what they need to pay attention in the learning content.

In addition, the buttons and text were modified based on the principles of accessibility. Users had difficulties with clicking the buttons due to the size of their fingers. As a result, the buttons were increased. The backend logic of the recording state was changed to eliminate confusion about where to start the recording.

Experiment Results

The results of the experiment have shown the effectiveness of PinyinGuo for learning Chinese pronunciation. The data gathered from all thirteen participants shows improvement of their pronunciation skills overtime. The results on the motivational aspect of learning with PinyinGuo in AR versus non-AR is reflected by the retention rate of the control (non-AR) and treatment groups (AR).

The total number of participants in the treatment group who finished six days was five. To compare, the number of people in control group who completed six days was four. The

number of participants who finished less than six days of practice in control and treatment group was two and three respectively. This data shows us that there was a slightly larger number of participants completing practice with AR rather than non-AR.

Therefore, the retention rate shows that there was slightly more motivation in completing the task in AR rather non-AR. The informal feedback data was not reflected on the server for these thirteen participants. There was an issue with the server script not saving the selections of the sever. However, this issue was fixed after the experiment was concluded. I was able to get data from the informal feedback during the user testing after the intervention. The follow-up user testing is described further in this chapter.

The recruitment data shows that there is a genuine interest in learning Chinese pronunciation. There were n=70 prospective participants who answered call for participants. Those individuals have read the study information on social media, went on the study website, agreed to the consent form, and entered their email information. These steps show interest of random sample to participate in this topic.

The recording data shed light on learner's progress while using the app. Since the sample size is not very large, the focus of this analysis will be on the recording data of the participants who finished six days of study. The six days of the study is enough to see the progress of the participants from day one. The sixth day is a review of everything that has been learnt from day one to day fifth. Therefore, a descriptive analysis of the recording data sheds light on the overall effectiveness of PinyinGuo in both treatments.

The recorded data has shown that the participants are trying to replicate the sounds that they have learned in the sessions. The participants tried their best to replicate the Chinese sounds. Each recording had phonemes with tones. It takes effort to pronounce the tones. The fact

that the participants tried to pronounce the tones with the phonemes, tells us that the participants took this practice seriously.

The .wav recordings in the data base show the participant ID, phoneme day, and phoneme name. The data shows that almost in all cases, the initials and finals were successfully pronounced. However, in some cases, the tones were confused by a number of participants. For example, a common error was a confusion between third and second tone. In addition, there was a confusion between fourth and second tone. These errors can be seen throughout all phonemes even in day 6.

The data of peer recordings was not reflected in the experiment due to different cohorts. Participants who were assigned the same cohort, did not begin on the same day. Some participants did not begin the study at all. This led to lack of data of peer assessment in the experiment. However, a follow-up user testing has peer-assessment data that shows some preliminary results for the effectiveness of peer-assessment.

The test phoneme data for each day shows that participants were making the similar errors in identifying phonemes as when they were recording the phonemes. For example, the participants had difficulties distinguishing between second and third tones. Some participants were confused with differences in the third tone and fourth tones.

Despite the fact that n=2 participants completed day 7 of the practice, the test data was not recorded on the server. The test data was not available due to a bug in the app. The bug has been fixed since the intervention. However, due to the sampling challenges, there were no new participants who completed the seven-day intervention.

The results of the experiment have shown the challenges associated with linguistic aspects in learning, motivation, and peer assessment. The seven-day intervention required

participants to be extremely motivated to launch the app and continue learning such complex linguistic aspects of the hardest language in the world. However, the retention rate shows us inconclusive results of AR being more motivating than non-AR.

Follow-up Usability Testing

Due to the limitations of a seven-day online experiment, a follow-up user testing was needed to determine the effectiveness of PinyinGuo to teaching Chinese pronunciation. The peer-assessment and the IMI data.

The informal feedback at the end of each session shows that both participants wanted to continue the practice after each session. Both participants answered to the question "I want to continue the practice" were "agree" or "strongly agree" throughout the intervention. The data shows that each day, the practice becomes less difficult for the participants (Fig. 18).

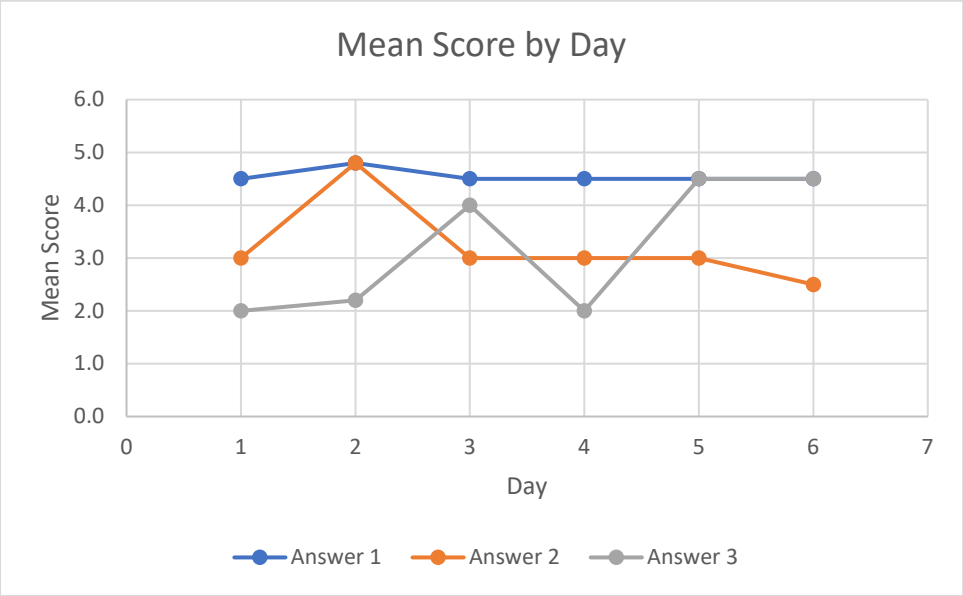

Figure 18: Feedback Results, Answer 1: I Want to Continue Practicing; Answer 2: Today's Practice Was Very Difficult For Me; Answer 3: I Had an Excellent Performance

This graph shows consistency in the continued interest in the practice. The line for Answer 1 shows that both users are motivated to continue. The line in the middle illustrates users' perception of difficulty. In the beginning of the practice, the users found the practice very difficult. However, by the end of day 6, they found it much easier. The bottom line shows user's performance trajectory. The users did not express linguistic confidence in the beginning of the practice, but at the end of the practice, they developed a linguistic confidence. This is an important dataset for contribution, since this data sheds light on continues motivation and linguistic confidence with PinyinGuo.

However, in regards of peer-assessment, overconfidence might be an issue. The peer-assessment dataset shows that there is a discrepancy between participant's confidence and their selection of the correct.

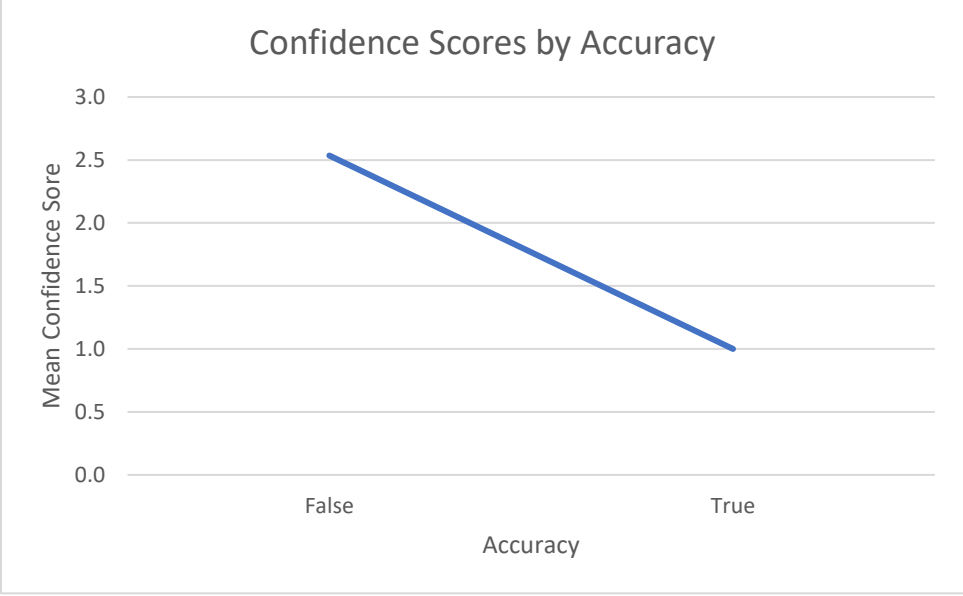

Figure 19: Confidence Scores by Accuracy Graph

This dataset illustrates users made more incorrect choices when they felt they were most confident. The confidence criteria is determined to show how confident the participants are in their choice of the phoneme. If they choose the wrong phoneme, it means that they were incorrect with their confidence rating. This two-person dataset is inconclusive in terms of results to answer a question for whether peer-assessment can replace expert feedback. However, this data can be used to make new hypothesis for scaling up this study.

The launch of experiment was challenging in a number of ways that have affected the readability of this data. However, based on the experiment results and two rounds of usability testing, there is a list of important results to answer the research questions of this dissertation. First, the experiment has shown that the retention rate in AR was higher than non-AR. This data shows that people were more interested in the learning content with AR than non-AR.

Second, the recording data has shown that users put a lot of efforts into pronouncing the sounds. This experiment was conducted with random sampling and the users were not enrolled into Chinese programs before. Therefore, the participants who tried the app even for a few days, followed the instructions and were motivated to learn Chinese pronunciation.

The users have completed the practice perception questionnaire. The interest/enjoyment scale was 7 points that is the highest score (Fig. 20).

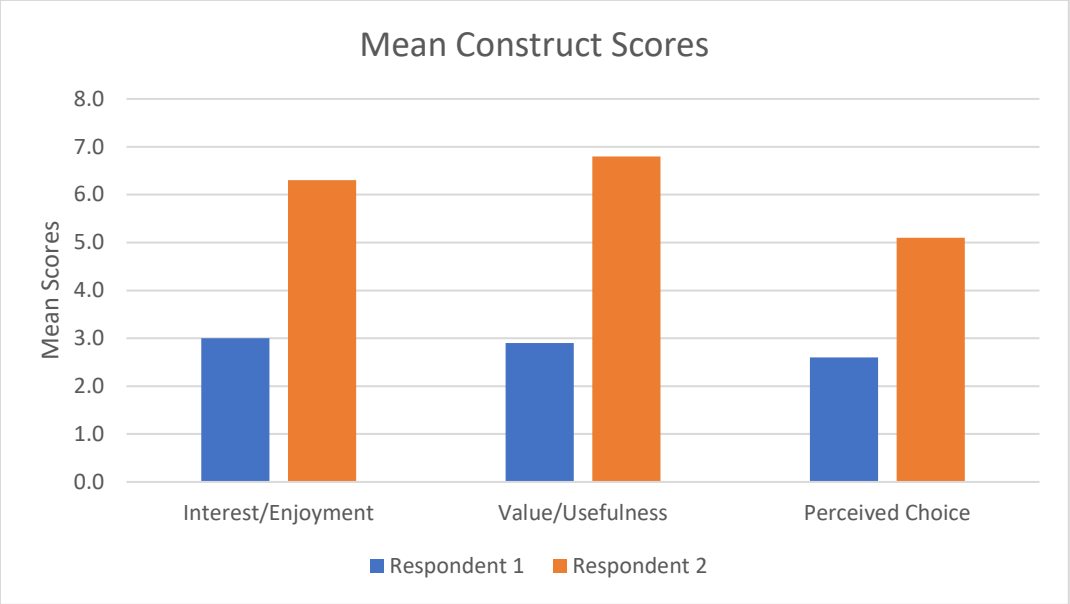

Figure 20: IMI Chart, by Respondents

The answer to the question from the enjoyment scale if the practice was boring was "somewhat true." Respondent 2 selected higher scores on interest/enjoyment scale. The "boring" aspect the participant selected was 4 on the scale 1-7 from "not at all true" to "very true". This answer shows that despite the fact that the user overall enjoyed the practice, they thought it was somewhat boring. However, the participant also selected "very true" for the practice to be fun. In general, the interest/enjoyment level in this practice was very high for respondent 2.

Respondent 1, however, did not enjoy the practice as much as respondent 2. Overall, the choice for interest/enjoyment in the middle. The participant did not find the practice fun or interesting.

Regarding the value/usefulness subscale, the respondent 1 found the practice very useful. The other participant found the practice not as useful. The choices the second participant made show that they did not find this practice useful.

Perceived choice in IMI expresses the motivation behind doing the activity. Since this was a convenience sampling for the purpose of user testing, the participants did not feel that they would do this activity if they did not have to. The perceived choice is a very valuable measuring instrument of intrinsic motivation for future research on MALL.

The recording for task perception questionnaire was not available. The participants have taken the perception questionnaire twice mistakenly thinking that they were taking different questionnaires. For the purpose of this analysis, I have used the first set of answers from the participants. Since this IMI questionnaire was what the original questionnaire. There was some variations in the data between two attempts, but the first set of questions was the one that was meant to be taken by the participants.

Regarding pronunciation of the participants: the initials and finals of the phonemes have no errors. The participants are having issues with pronouncing the second and the third tones. In terms of the AR and non-AR environment, the number of correctly pronounced phonemes was larger in AR than non-AR. These results are inconclusive and need to be addressed in a study of a larger scale.

Study Limitations

However, the small number of participants was a limitation for this research design. The cohorts of four participants does not guarantee that all of them start on the same day. This makes it harder to answer the research question if some participants get the audio recordings to assess and some not. Hence, I had planned a cohort of twenty participants who had to evaluate each other.

In the first round of intervention, I tracked each participant's progress on the database. Three participants experienced a technical issue with the app. Although I have tested the app multiple times before the deployment, the builds still had some bugs in the deployment stage. For example, various participants had issues with accessing day 6 and day 7. There were five participants discarded due to the issue. The issue was in the script. I had a day count for the current day. If the person left the session earlier, the day switches to Day 6th, but the day counter switches to 0. I fixed that issue and distributed the build to the new sets of participants. I could not update the app build that was already downloaded to the current participants. The participants were still compensated for participating in the seven days of the intervention. However, unfortunately, this issue arose too late in the first round of experiment, specifically on day 6. And the participants were not eligible to sign up for the practice again, since they have gone through six days of practice. This issue was noted for the second round of the study distribution.

In this part of the procedure, a lot of prospective participants were confused. Although I posted a web link with the instructions on how to retrieve the UDID, most people entered the wrong data. It was challenging to find a website with clear and easy instructions in the first place. Moreover, some of the websites were outdated, since they explained how to retrieve UDID's with iTunes. First, iTunes is no longer available on macOS. iTunes were replaced by Music. Second, not all iPhone users use Music. Moreover, one participant mentioned that she did not have a laptop or PC. Therefore, it is very challenging for iPhone users to retrieve a UDID to participate in the study.

The UDID part has negatively influenced the sampling procedure. First, the UDID requirement for iOS users created a delay in participant recruitment. The research team had to

email all the iOS prospective participants who have entered a wrong UDID with step-by-step instructions on how to retrieve UDID. Those instructions included a description of what a UDID is, what it is used for, and what websites they can click to learn more about UDID and how to retrieve it. I have provided an email with the website on how to retrieve UDID by yourself https://www.sourcefuse.com/blog/how-to-find-udid-in-the-new-iphone-xs-iphone-xr-and-iphone-xs-max/. Additionally, I have found an app that retrieves iOS users information: https://udid.tech. This website simplifies the process of obtaining UDID. However, it still requires participants to go to Settings, I previously provided another link to a similar website https://betafamily.com/superudid; however, the website was giving a server error for a while, so a number of iOS users in my study could not access it. To retrieve a UDID, the prospective participants needed to go through the following procedure: go to General Settings—VPN and Device Management—Downloaded Profile—UDID.tech Device Registration- Enter your passcode—Install. After they clicked Install, they are redirected to the browser. The page pulls up the information with UDID, IMEI, Product/OS Version, Serial Number.

I also did a screen recording to send to participants who had trouble following the written directions and the https://udid.tech directions. Some iOS users did not enter their UDID in the Qualtrics survey and they waited for a build. When they received an Android build, they had to email the research team stating that they cannot download the Android build. I had to send a reply to them that they need to provide their UDID. And provided the step-by-step instructions. At this point, a number of prospective participants lost interest in the study. I have followed up with twenty-five iPhone users from the first round of study advertisement in January 2022. I did not receive a single answer to this follow-up email.

I released a second round of call for participants in the Spring 2022. I received direct emails, text messages, and Facebook messages about the interest in the study. I have shared the screen recording of the procedure on https://udid.tech. I have also answered security concerns from users about what data the developers will get from their phone based on their UDID. Multiple iPhone users asked if the app is available on the AppStore.

Since the app is not available on AppStore, this has created another point of enrollment delay for iOS users. After all the steps with obtaining UDID's from the participants, the research team needed to update the provisioning profile on the Apple Developer portal, then build the archive and update the web portal. I had only partial access to his account to do the build. Therefore, I had to send a UDID to John Murray, wait until he adds the participants, and update my provisioning profile in XCode and recompile. After these steps, I sent emails with the iOS build to the participants with their participant IDs. These steps have affected the turnaround of the app distribution as well.

The participant ID is generated on the server when the prospective participants enter their email to the Qualtrics link while signing up for the study. I have created a spreadsheet to enter participants' progress. I found out that despite the fact that in the first round of recruitment there were sixty-eight people interested in the study, only five people who have completed the practice.

I followed up with thirty-eight potential participants who were Android users based on the Qualtrics survey and twenty-five iPhone users. I have received a response from twenty-nine Android users. However, only four participants completed all days of the study. However, they have not completed the Qualtrics surveys in the end of the study.

I did a second round of advertisement in May 2022. In this round I have received direct emails and messages from participants. Fourteen people were interested in the study. I have emailed them builds on the designated devices with the sign-in information. However, they did not complete the study.

Another issue was the time gap between when the users signed up for the study and when they received the email with the app download link and participant ID. At each day of the recruitment periods, I monitored how many people signed up for the study. Next day, I manually sent an email to each participant with the app build and participant ID. I have not asked for confirmation email acknowledging the receipt of the email with the credentials and the app. Not a single participant acknowledged the receipt of the email.

Such manual email distribution procedure is a serious limitation for prospective participants to begin the study. Especially if the email with the credentials got into a spam folder or got lost in the shuffle.

There was also an issue with some participants not providing their UDIDs. Based on the survey they completed to sign up for the study, if the UDID field was empty, I would assume that the users have Android phones. The question specified (for iOS users only). However, since UDID is a complicated concept for non-developers, some people just left the field blank.

When the iOS users received the Android build in their email, they would either think that the link is not working and give up the participation or the users would reach out to me via email and request a working link to download the app. However, this back and forth would take some additional time. These limitations have affected the participation as well.

Another reason for challenges during the intervention was the consistent commitment from the participants. There was a spike in interest in during the app advertising campaign.

However, an incentive for a seven-day consecutive practice was not enough to keep the commitment from the participants. Studies on commitment in online interventions have revealed lack of motivation of users to stick to the practice (Maeda et al., 2021). In Maeda et al. study people expressed interest in the first week but uninstalled the app in a second week. Similar retention rate decline was described in studies on health research (Dorsey et al., 2017; Yamaguchi et al., 2019).

In addition to the amount of compensation might not have been enough for a whole week of commitment. A $40 Amazon gift card for less than an hour of study participation may seem reasonable. However, research services like Focusgroup[23] or Userinterviews offer from $75 to $100 per hour respectively [24]. An commitment for a period at once is more feasible than a commitment for a 30 minutes throughout a week.

Contributions

The contributions of this dissertation contain the main five points:

(1) Survey of the apps and recommendations;

(2) PinyinGuo app development;

(3) Design and implementation of practice and methodology

(4) Usability test among students and the instructor;

(5) The experiment.

[23] https://www.focusgroup.com
[24] https://www.userinterviews.com/studies?sort=-id

Below I describe each of these contributions in more detail. I refer to the literature review related to each of these parts, relation to the research questions and hypotheses, and provide results and discussion.

First, I conducted a survey of language apps and recommendations. This analysis is grounded in the research questions on learning outcomes and motivation with AR, RQ1 and RQ2 respectively. The literature review on MALL, AR in learning, and game design provides support for the analysis of the existing apps. In addition, the systematic literature review focuses on methodological aspects for designing mobile apps for language learning with AR. The apps were searched based on the number of downloads on Google Play and App Store. I chose seven most downloaded apps: Duolingo, Mondly, Busuu, Babbel, HellowTalk, Memrise, and MondlyAR. I downloaded the data from the digital analytics website, SensorTower[25]. I looked at the teaching methods used in these apps, content presentation, gamification elements, and peer collaboration aspect. Based on these Finally, the survey of apps and recommendations provides a body of information for creating pronunciation training app PinyinGuo.

This leads to the second contribution of this dissertation is PinyinGuo app development. The hypothesis of this study was that articulatory demonstration of Chinese phonemes in AR will be more effective in terms of motivation and learning outcomes than in non-AR. The design and development of PinyinGuo was drawn on the body of previous research on MALL, learning theories, and usability theory. The design of PinyinGuo is a contribution in the field of Texts and Technology, MALL, and applied linguistics. The methodology used to develop this app merges theories from humanities and sciences. Such app has never existed before. The practices of backend and frontend development in this app can be used in future language applications.

[25] https://sensortower.com/

The third contribution is design and implementation of the practice and methodology to the new app. The practice is based on the tradition learning theory and supported by gamification elements from game theories. These foundational elements of the learning content contributed to the motivation and learning results in the practice. The backend logic is determined to receive and send data from measuring instruments and peer assessment. The peer assessment is part of the methodology developed in this study. The peer assessment aspect derived from the corrective feedback in language pedagogy and language theory. The peer assessment is part of the methodology and practice developed in this dissertation.

The fourth contribution is the usability test. The usability test from the students and the instructor. The usability test is grounded in usability theory. The usability testing in this study was conducted three times and led to changes in the design process of the app. Usability testing is part of the methodology for developing similar language apps. The first round of the usability testing was conducted with two participants from the learners' target group. The test has affected design decisions in the app development before the intervention. The second usability test was conducted with the language expert in the instructor role. The language expert provided feedback about the method used in this app and the use of this method in the classroom setting. Such heuristic testing is vital for the contribution of this dissertation in terms of reliability of the method. Finally, after the intervention, second round of student usability test was conduced in the remote settings. This follow-up usability testing was needed to determine the practice timeline and the experiment setup for future research.

The fifth contribution is the design of online experiment. The design of this experiment was driven by the COVID-19 pandemic-related risks of having to come to the lab in person. In addition, the design of PinyinGuo allowed to gather all the participants' data online. I designed

PinyinGuo around the concept of online experiment and data collection in remote settings. The design of the experiment was determined to answer the research questions on motivation, learning outcomes, and reliability of peer assessment. The measuring instruments were selected to answer those questions in the remote settings. The number of independent variables was reduced by backend logic and UI design of the app. The results of the experiment were inconclusive in terms of answering the research questions. However, progress was made in terms of the practice effectiveness in motivation and learning outcomes.

Conclusion

This dissertation makes a contribution to the second language learning by creating, evaluating, and testing a mobile assisted pronunciation software PinyinGuo. This dissertation explores the design and methodology of creating a new language app in AR.

The design and development of AR app for teaching Chinese pronunciation is a contribution to MALL. PinyinGuo is the first app that uses AR to teach Chinese pronunciation. The results of user testing have shown that AR can be used for teaching Chinese pronunciation.

The findings from two rounds of usability testing indicated that learners find this practice useful and fun. The heuristic analysis of the app has revealed that the app is pedagogically driven. The expert stated that PinyinGuo uses novice methods for pronunciation training, such as eco method. The expert pointed out the gap for pronunciation training apps.

The methodology of PinyinGuo includes pinyin learning content, AR features, lip-sync data, and a peer-assessment concept. Each of those components were added after conducting literature review, prototyping, and testing. The peer assessment aspect takes an vital part in the PinyinGuo method. Based on the state-of-the-art review of existing apps, it was determined that

the peer collaboration aspect is underrepresented. The method of online anonymous peer-assessment was developed after a thorough literature review.

The online experiment takes an essential part of the overall contribution of this dissertation. The UI functionality with the screens was created to reflect the research agenda. The approach to create a peer review criteria was developed based on literature review in language learning acquisition. This approach was determined to be scaled up for a study with a larger number of participants. The server-side scripting needed to deliver functionality for sending and saving user data to the server. In developing the peer-assessment practice, it was an great challenge to distribute the sets of recordings to different participants. The solution for this challenge was to ensure that the participants are assigned to someone to assess. Therefore, I created cohorts of participants based on the day they sign up for the study.

The design of PinyinGuo was driven by the research questions and by the method proposed in this study. The research questions determined the research method. The research method was an online experiment.

The design of this experiment is a contribution of this dissertation. The experiment procedure of recruiting participants online for a seven period intervention. The design of the app includes server logic that saves the experimental data. Designing the learning app interface with AR was challenging. The second challenge was to reflect requirements from the online experiment to save variables for future analysis.

The methodologies presented in this dissertation are driven from language theory and usability testing. Chapter 3 analyses related literature that spans commercial MALL apps. In this review, I compare different approaches, and point out the problems tackled by my design and

those are that out of scope. This chapter provides recommendations for future MALL design with AR.

RQ1: What effect on learner's motivation will articulatory demonstration in AR have on Chinese language pronunciation versus 3D virtual demonstration?

The online experiment presented in this dissertation was a seven-day continuous intervention. This is a long-term commitment for an online study. The seven-day period was chosen to investigate continuous motivation. The results of this experiment, though preliminary given the number of participants, suggest the retention rate of the participants with AR was longer than non-AR. The daily feedback results from the user testing did not provide statistical significance. However, the daily feedback results showed that the user with AR was more motivated than the user with non-AR. The results of IMI questionnaire have shown that the participant with AR app was more interested in the practice than the participant using the non-AR condition. Based on these three data sets from the experiment and the user testing, motivational aspect in AR is higher than non-AR.

RQ2: How do AR conditions affect learning outcomes in pronunciation versus non-AR conditions?

The learning outcomes from both AR and non-AR did not have much difference. The available dataset showed that there are less errors in the AR version. It is crucial to investigate this hypothesis further in future research with more data elements. Both conditions, however, showed an increased improvement of phoneme pronunciation. The previous findings have revealed the effectiveness of visual representations in 3D mode "to blade-alveolar, blade-palatal, lingua-palatal, open-mouth, open-mouth(-i) and round-mouth" (X. Peng et al., 2018b, p. 26). Given the inherently 3D nature of AR, there is the potential to represent these 3D models even in

a more contextualized and engaging way. For instance, students can imagine the speaker or themselves while being able to position the speaker and better visualize articulation. The findings from the online experiment and user testing show that the learning outcomes are very similar. Therefore, it is vital to conduct future research by scaling up the study.

RQ3: How accurate is anonymous peer assessment of phoneme pronunciation recordings?

The inconclusive user test results have revealed that peer-assessment may not be a reliable tool for correcting Chinese phoneme pronunciation. The data set (n=2) shows that the participants make errors in grading phonemes while they feel most confident. This claim requires further investigation due to the limited number of participants.

The main contribution of this dissertation is a new application that supports practice for Chinese pronunciation in AR. In addition, this dissertation contributes methodology for an online experiment and developing an AR app.

Future Research

The conclusions in this dissertation highlight the importance of measuring motivation and learning outcomes through the app data. The online experiment has a great potential for research on motivation and learning. Future research is needed in designing an online experiment. The online experiment is a promising method in the times of COVID-19. There are a number of app studies conducted online. However, the seven-day consecutive intervention should be changed to one or two days.

AR is a relatively novice technology in comparison to mobile phones. Therefore, it is challenging to recruit participants for a remote study for a week and use AR is challenging.

Additionally, Chinese language is one of the hardest languages. It takes a certain amount of intrinsic interest and motivation in learning Chinese for a week. This sampling procedure should be adjusted for future studies on Chinese learning in the remote settings.

Further research on sampling procedures for an online experiment is required. For example, recruiting students from Chinese language courses would be beneficial. Possibly, it would be helpful to remove the study requirement for no prior Chinese knowledge. Dropping this requirement would be useful for the peer-assessment aspect. Learners who have some knowledge of Chinese would enrichen the dataset gathered from the peer assessment. Additionally, Hanyu Pinyin practice is beneficial for students of higher language proficiency as well.

The outcome of this experiment has revealed the need to scale up the study to recruit more participants. Future research in designing an online experiment is needed. For example, a suggestion would be posting the app on the App Store and Google play. This step would eliminate confusion and difficulties associated with downloading the app from the email and entering participant ID. The UDID issue has been described in the Limitations section as well.

Marketing of the app would also valuable. Based on the results in the literature review, marketing is important to attract new users. Forthcoming analysis on MALL marketing would be needed. The session metadata and the user responses will be gathered and saved. This will replace the social media posts, faculty emails, and word of mouth recruitment strategies.

AI algorithms for automated corrective feedback can be done in the scope of a larger study. Automated online feedback on phonemes is an area of future work as well. The feedback for single sounds can be helpful for language learning and for speech training. Automated

corrective feedback in AR opens new possibilities for future research on accessibility in learning. In this dissertation, I reviewed the role of corrective feedback for language learning.

In future research, a larger phoneme set could be used to increase the number of phonemes for the practice. It is also important to provide different options of sound length and voice gender that pronounces the sounds. Finally, it is important to write learning content with dialogues, role play, and other communicative practices.

In regard to the app design, the app should have goal setting options, scores, leaderboards, and community. Google has a new feature on how to pronounce words correctly. For future design, it might be beneficial to make an AR face that would pronounce words correctly. The user would type in a word and get the AR avatar to demonstrate how to pronounce the word.

The field of Text and Technology merges methods from the Humanities and Sciences to research new media and culture. The contribution of this dissertation to the field of Texts and Technology is creating a new practice in AR that has not been previously created. I completed the design of a new app that can evaluate AR and non-AR editions of the same practice methodology. I situated this new methodology into the context of contemporary software and demonstrated why this app is important. I have made progress on several fronts of addressing the research questions, even though I cannot make claims that I have succeeded in answering the questions of whether AR or non-AR is better or if the peer-assessment practice is a reliable tool for corrective feedback in the online settings.

APPENDIX A: PRACTICE, RECORDING EXERCISES, AND POSTTESTS

Table 3 Details of the Practice

Day 1	Day 2		Day 3		Day 4	Day 5		Day 6
ma		mu	hu		shu	sha		Random
ma1		mu1	hu1		shu1	sha1		ma3
ma1		mu1	hu1		shu1	sha1		
								sha2
ma1		mu1	hu1		shu1	sha1		hu1
ma1		mu1	hu1		shu1	sha1		
ma1		mu1	hu1		shu1	sha1		mu4
ma1		mu1	hu1		shu1	sha1		
ma1		mu1	hu1		shu1	sha1		shu1
ma2		mu2	hu2		shu2	sha2		
ma2		mu2	hu2		shu2	sha2		hu2
								ma4
ma2		mu2	hu2		shu2	sha2		
ma2		mu2	hu2		shu2	sha2		sha3
ma2		mu2	hu2		shu2	sha2		
ma2		mu2	hu2		shu2	sha2		shu1
ma2		mu2	hu2		shu2	sha3		mu2
								shu2
								ma3
ma3		mu3	hu3		shu3	sha3		
ma3		mu3	hu3		shu3	sha3		sha2
	F			F			F	hu1
ma3	e	mu3	hu3	e	shu3	sha3	e	
ma3	e	mu3	hu3	e	shu3	sha3	e	ma1
ma3	d	mu3	hu3	d	shu3	sha3	d	
ma3	b	mu3	hu3	b	shu3	sha3	b	sha1
ma3	a	mu3	hu3	a	shu3	sha3	a	
	c			c			c	shu4
	k			k			k	
ma4		mu4	hu4		shu4	sha4		hu4
ma4		mu4	hu4		shu4	sha3		
								hu3
ma4		mu4	hu4		shu4	sha4		sha4
ma4		mu4	hu4		shu4	sha4		
ma4		mu4	hu4		shu4	sha4		ma2
ma4		mu4	hu4		shu4	sha4		
ma4		mu4	hu4		shu4	sha4		ma3
								shu3
ma2		mu4	hu1		shu3	sha1		
ma3		mu1	hu4		shu2	sha2		
ma1		mu3	hu2		shu1	sha4		
ma4		mu2	hu3		shu4	sha3		
ma1		mu1	hu2		shu3	sha4		shu1
ma4		mu4	hu1		shu4	sha2		ma3
ma3		mu3	hu4		shu2	sha1		sha4
ma2		mu2	hu3		shu1	sha3		mu2
								hu2
		Day 1	Day 2		Day 3	Day 4		Day 5
		ma1	mu1		hu2	shu3		sha4
		ma4	mu4		hu1	shu4		sha2
		ma3	mu3		hu4	shu2		sha1
		ma2	mu2		hu3	shu1		sha3

Day 7 Assessment			
	Day 7	Final Test	Anticipated time 5 min
	Assigment 1	Listen to the pinyin and choose the correct number of tone. 1 attempt 3 seconds	
Day 6			
shu1	Answer Keys	ma3	
ma3		hu2	
sha4		shu4	
mu2		sha1	
hu2		ma1	
		shu3	
Anticipated time 2 min		shu1	
		ma4	
		ma2	
		hu3	
		mu4	
		ma3	
		mu3	
		shu1	
		hu2	
		sha1	
		mu2	
		ma3	
		shu1	

After the Assesment Test	
Activity Perception Questionnaire	
Task Evaluation Questionnaire	

Color Codes		
	listening practice-repeat after the speaker	2x demo,
	listening practice-choose the correct pinyin	2x concurre
	assessment test	
	for users to record	
	assessment of peer recordings	2 peer recordings

APPENDIX B: SUPPLEMENTARY TABLE

Table 4: Recorded syllables ready for assessment by day

Day 2	mā	má	mǎ	mà
Day 3	mū	mú	mǔ	mù
Day 4	hū	hú	hǔ	hù
Day 5	shū	shú	shǔ	shù
Day 6	shā	shá	shǎ	shà
Day 7 1 Panel	mā	mú	shū	shú
Day 7 2 Panel	mǎ	mǔ	shā	shū
Day 7 3 Panel	shá	shà	shā	shù
Day 7 4 Panel	mú	mǔ	shú	mǎ
Day 7 5 Panel	shú	hú	hū	shā

APPENDIX C: QUESTIONNAIRES

concept described in the subscale name. Then calculate subscale scores by averaging the items scores for the items on each subscale. They are shown below. The (R) after an item number is just a reminder that the item score is the reverse of the participantÕs response on that item.

Interest/enjoyment: 1, 3(R), 5, 7, 9
Perceived competence: 4, 6,
Pressure/tension: 2(R), 8

* * * * * * * * * * * *

The next version of the questionnaire was used for a study of internalization with an uninteresting computer task (Deci et al., 1994).

ACTIVITY PERCEPTION QUESTIONNAIRE

The word “Activity” will be replaced with “Practice”

The following items concern your experience with the task. Please answer all items. For each item, please indicate how true the statement is for you, using the following scale as a guide:

1	2	3	4	5	6	7
not at all true			somewhat true			very true

1. I believe that doing this activity could be of some value for me.
2. I believe I had some choice about doing this activity.
3. While I was doing this activity, I was thinking about how much I enjoyed it.
4. I believe that doing this activity is useful for improved concentration.
5. This activity was fun to do.
6. I think this activity is important for my improvement.
7. I enjoyed doing this activity very much.
8. I really did not have a choice about doing this activity.

9. I did this activity because I wanted to.

10. I think this is an important activity.

11. I felt like I was enjoying the activity while I was doing it.

12. I thought this was a very boring activity.

13. It is possible that this activity could improve my studying habits.

14. I felt like I had no choice but to do this activity.

15. I thought this was a very interesting activity.

16. I am willing to do this activity again because I think it is somewhat useful.

17. I would describe this activity as very enjoyable.

18. I felt like I had to do this activity.

19. I believe doing this activity could be somewhat beneficial for me.

20. I did this activity because I had to.

21. I believe doing this activity could help me do better in school.

22. While doing this activity I felt like I had a choice.

23. I would describe this activity as very fun.

24. I felt like it was not my own choice to do this activity.

25. I would be willing to do this activity again because it has some value for me.

Scoring information. Begin by reverse scoring items # 8, 12, 14, 18, 20, and 24 by subtracting the item response from 8 and using the result as the item score for that item. Then calculate subscale scores by averaging the items scores for the items on each subscale. They are shown below. The (R) after an item number is just a reminder that the item score is the reverse of the participant's response on that item.

Interest/enjoyment: 3, 5, 7, 11, 12(R), 15, 17, 23
Value/usefulness: 1, 4, 6, 10, 13, 16, 19, 21, 25
Perceived choice: 2, 8(R), 9, 14(R), 18(R), 20(R), 22, 24(R)

better basketball player, or whatever, then fill in the blanks with that information. If you do not want to refer to a particular outcome, then just truncate the items with its being useful, helpful, or important.

Scoring information for the IMI. To score this instrument, you must first reverse score the items for which an (R) is shown after them. To do that, subtract the item response from 8, and use the resulting number as the item

score. Then, calculate subscale scores by averaging across all of the items on that subscale. The subscale scores are then used in the analyses of relevant questions.

* * * * * * * * * * * *

The following is a 22-item version of the scale that has been used in some lab studies on intrinsic motivation. It has four subscales: interest/enjoyment, perceived choice, perceived competence, and pressure/tension. The interest/enjoyment subscale is considered the self-report measure of intrinsic motivation; perceived choice and perceived competence are theorized to be positive predictors of both self-report and behavioral measures of intrinsic motivation. Pressure tension is theorized to be a negative predictor of intrinsic motivation. Scoring information is presented after the questionnaire itself.

TASK EVALUATION QUESTIONNAIRE

The word "task" will be changed to "recording the assessment"

For each of the following statements, please indicate how true it is for you, using the following scale:

1	2	3	4	5	6	7
not at all true			somewhat true			very true

1. While I was working on the task I was thinking about how much I enjoyed it.

2. I did not feel at all nervous about doing the task.

3. I felt that it was my choice to do the task.

4. I think I am pretty good at this task.

5. I found the task very interesting.

6. I felt tense while doing the task.

7. I think I did pretty well at this activity, compared to other students.

8. Doing the task was fun.

9. I felt relaxed while doing the task.

10. I enjoyed doing the task very much.

11. I didnÕt really have a choice about doing the task.

12. I am satisfied with my performance at this task.

13. I was anxious while doing the task.

14. I thought the task was very boring.

15. I felt like I was doing what I wanted to do while I was working on the task.

16. I felt pretty skilled at this task.

17. I thought the task was very interesting.

18. I felt pressured while doing the task.

19. I felt like I had to do the task.

20. I would describe the task as very enjoyable.

21. I did the task because I had no choice.
22. After working at this task for awhile, I felt pretty competent.

Scoring information. Begin by reverse scoring items # 2, 9, 11, 14, 19, 21. In other words, subtract the item response from 8, and use the result as the item score for that item. This way, a higher score will indicate more of the concept described in the subscale name. Thus, a higher score on pressure/tension means the person felt more pressured and tense; a higher score on perceived competence means the person felt more competent; and so on. Then calculate subscale scores by averaging the items scores for the items on each subscale. They are as follows. The (R) after an item number is just a reminder that the item score is the reverse of the participants response on that item.

Interest/enjoyment: 1, 5, 8, 10, 14(R), 17, 20
Perceived competence: 4,
7, 12, 16, 22 Perceived choice
3, 11(R),
15, 19(R), 21(R)
Pressure/tension: 2(R), 6,
9(R), 13, 18

APPENDIX D: IRB APPROVAL LETTER

UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board
FWA00000351
IRB00001138, IRB00012110
Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

APPROVAL

May 12, 2022

Dear Daria Sinyagovskaya:

On 5/12/2022, the IRB reviewed the following submission:

Type of Review:	Modification / Update
Title:	Using AR Articulatory Demonstration and Peer Assessment for Chinese Language Pronunciation Practice: Effects on Learning Performance and Motivation
Investigator:	Daria Sinyagovskaya
IRB ID:	MOD00002894
Funding:	None
Grant ID:	None
IND, IDE, or HDE:	None
Documents Reviewed:	• Recruitment Flyer, Category: Recruitment Materials; • Study Protocol, Category: IRB Protocol;

The IRB approved the minor modifications on 5/12/2022.

In conducting this protocol, you are required to follow the requirements listed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system. Guidance on submitting Modifications and a Continuing Review or Administrative Check-in are detailed in the manual. When you have completed your research, please submit a Study Closure request so that IRB records will be accurate.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Kamille C. Birkbeck

Kamille Birkbeck
Designated Reviewer

REFERENCES

About Us Research. (n.d.). https://research.duolingo.com

Adams, R. S., & Atman, C. J. (1999). Cognitive processes in iterative design behavior. *FIE'99 Frontiers in Education. 29th Annual Frontiers in Education Conference. Designing the Future of Science and Engineering Education. Conference Proceedings (IEEE Cat. No.99CH37011*, *1*, 11A6/13-11A6/18. https://doi.org/10.1109/FIE.1999.839114

Altinkaya, M., & Smeulders, A. W. M. (2020). Assisted Speech to Enable Second Language. *Proceedings of the 1st International Workshop on Multimodal Conversational AI*, 3–7. https://doi.org/10.1145/3423325.3423735

Augmented Versus Virtual Reality in Education: An Exploratory Study Examining Science Knowledge Retention When Using Augmented Reality/Virtual Reality Mobile Applications | Cyberpsychology, Behavior, and Social Networking. (n.d.). Retrieved April 2, 2020, from https://www.liebertpub.com/doi/full/10.1089/cyber.2018.0150?casa_token=Oe0qsk9aHxcAAAAA%3AIsmEVOdVYgp3sXbtDPx83Hje5aEnoGV361WjssAZMg50qQn-6NeHxgJyOBjy_sZFE6eCtn6Uag

Babbel. (n.d.). https://www.babbel.com

Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review*, *30*, 100322. https://doi.org/10.1016/j.edurev.2020.100322

Baker, H. (2019, June 17). *Harry Potter: Wizards Unite Proves Pokemon Lightning Doesn't Strike Twice*. UploadVR. https://uploadvr.com/harry-potter-wizards-unite-least-engaging-mobile-games/

Baker, J. E., Bruns Jr, L., Hassenstab, J., Masters, C. L., Maruff, P., & Lim, Y. Y. (2020). Use of an experimental language acquisition paradigm for standardized neuropsychological assessment of learning: A pilot study in young and older adults. *Journal of Clinical and Experimental Neuropsychology*, *42*(1), 55–65. https://doi.org/10.1080/13803395.2019.1665626

Baldwin, M., & Mussweiler, T. (2018). The culture of social comparison. *Proceedings of the National Academy of Sciences*, *115*(39). https://doi.org/10.1073/pnas.1721555115

Beatty, K. (2010). *Teaching and researching computer-assisted language learning* (2nd ed). Longman.

Beckman, K., Bennett, S., & Lockyer, L. (2014). Understanding students' use and value of technology for learning. *Learning, Media and Technology*, *39*(3), 346–367. https://doi.org/10.1080/17439884.2013.878353

Belasco, S. (1965). Nucleation and the Audio-Lingual Approach. *The Modern Language Journal*, *49*(8), 482–491. https://doi.org/10.1111/j.1540-4781.1965.tb02165.x

Bickford, A. C., & Floyd, R. (2006). *Articulatory phonetics: Tools for analyzing the world's languages* (4th ed). SIL International.

Billinghurst, M., Clark, A., & Lee, G. (2015). A Survey of Augmented Reality. *Foundations and Trends® in Human–Computer Interaction*, *8*(2–3), 73–272. https://doi.org/10.1561/1100000049

Biocca, F., Owen, C., Tang, A., & Bohil, C. (2007). Attention Issues in Spatial Information Systems: Directing Mobile Users' Visual Attention Using Augmented Reality. *Journal of Management Information Systems*, *23*(4), 163–184. https://doi.org/10.2753/MIS0742-1222230408

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, *17*(2), 102–118. https://doi.org/10.1016/j.jslw.2007.11.004

Bu, Y., Ma, T., Li, W., Zhou, H., Jia, J., Chen, S., Xu, K., Shi, D., Wu, H., Yang, Z., Li, K., Wu, Z., Shi, Y., Lu, X., & Liu, Z. (2021). PTeacher: A Computer-Aided Personalized Pronunciation Training System with Exaggerated Audio-Visual Corrective Feedback. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3411764.3445490

Buckley, P., & Doyle, E. (2016). Gamification and student motivation. *Interactive Learning Environments*, *24*(6), 1162–1175. https://doi.org/10.1080/10494820.2014.964263

Busuu. (n.d.). https://www.busuu.com

CCSSP. (2020). *CCSSP 2020 020 1st International Conference on Communications, Control Systems and Signal Processing: 16-17th March, 2020, El-Oued, Algeria*. https://ieeexplore.ieee.org/servlet/opac?punumber=9145688

Che Dalim, C. S., Sunar, M. S., Dey, A., & Billinghurst, M. (2020a). Using augmented reality with speech input for non-native children's language learning. *International Journal of Human-Computer Studies*, *134*, 44–64. https://doi.org/10.1016/j.ijhcs.2019.10.002

Che Dalim, C. S., Sunar, M. S., Dey, A., & Billinghurst, M. (2020b). Using augmented reality with speech input for non-native children's language learning. *International Journal of Human-Computer Studies*, *134*, 44–64. https://doi.org/10.1016/j.ijhcs.2019.10.002

Chiang, T. H. C., Yang, S. J. H., & Hwang, G.-J. (2014). An Augmented Reality-based Mobile Learning System to Improve Students' Learning Achievements and Motivations in Natural Science Inquiry Activities. *Educational Technology & Society*, *17*(4), 352-365.

Chung, C.-J., Hwang, G.-J., & Lai, C.-L. (2019a). A review of experimental mobile learning research in 2010–2016 based on the activity theory framework. *Computers & Education*, *129*, 1–13. https://doi.org/10.1016/j.compedu.2018.10.010

Chung, C.-J., Hwang, G.-J., & Lai, C.-L. (2019b). A review of experimental mobile learning research in 2010–2016 based on the activity theory framework. *Computers & Education*, *129*, 1–13. https://doi.org/10.1016/j.compedu.2018.10.010

Chung, K. S. (2017). Teaching pronunciation to adult learners of English. In H. P. Widodo, A. Wood, & D. Gupta (Eds.), Asian english language classrooms (pp. 131–149). Routledge. Https:// doi.org/10.4324/9781315755243-9. In H. P. Widodo, A. Wood, & D. Gupta (Eds.), *Asian English Language Classrooms: Where Theory and Practice Meet* (1st ed., pp. 131–149). Routledge. https://doi.org/10.4324/9781315755243

Cooke, L. (2010). Assessing Concurrent Think-Aloud Protocol as a Usability Test Method: A Technical Communication Approach. *IEEE Transactions on Professional Communication*, *53*(3), 202–215. https://doi.org/10.1109/TPC.2010.2052859

Cucchiarini, C., Neri, A., & Strik, H. (2009). Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback. *Speech Communication*, *51*(10), 853–863. https://doi.org/10.1016/j.specom.2009.03.003

Cuendet, S., Bonnard, Q., Do-Lenh, S., & Dillenbourg, P. (2013). Designing augmented reality for the classroom. *Computers & Education*, *68*, 557–569. https://doi.org/10.1016/j.compedu.2013.02.015

Derwing, T. M., Diepenbroek, L. G., & Foote, J. A. (2013). How Well do General-Skills ESL Textbooks Address Pronunciation? *TESL Canada Journal*, *30*(1), 22. https://doi.org/10.18806/tesl.v30i1.1124

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification." *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11*, 9. https://doi.org/10.1145/2181037.2181040

Ding, S., Liberatore, C., Sonsaat, S., Lučić, I., Silpachai, A., Zhao, G., Chukharev-Hudilainen, E., Levis, J., & Gutierrez-Osuna, R. (2019). Golden speaker builder – An interactive tool for pronunciation training. *Speech Communication*, *115*, 51–66. https://doi.org/10.1016/j.specom.2019.10.005

Do, H.-D., Wen, J.-M., & Huang, S. K. (2022). Impact of Humorous Chinese-Teaching Videos and Echo Method on Teaching Chinese Language: A Case Study of Vietnamese Students. *The Asia-Pacific Education Researcher*, *31*(6), 725–738. https://doi.org/10.1007/s40299-021-00622-5

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, *24*(3), 331–350. https://doi.org/10.1080/03075079912331379935

Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Language Teaching*, *31*(3), 117–135. https://doi.org/10.1017/S026144480001315X

Dorsey, E. R., "Yvonne Chan, Y.-F., McConnell, M. V., Shaw, S. Y., Trister, A. D., & Friend, S. H. (2017). The Use of Smartphones for Health Research: *Academic Medicine*, *92*(2), 157–160. https://doi.org/10.1097/ACM.0000000000001205

Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed). Oxford University Press.

Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and Limitations of Immersive Participatory Augmented Reality Simulations for Teaching and Learning. *Journal of Science Education and Technology*, *18*(1), 7–22. https://doi.org/10.1007/s10956-008-9119-1

Dünser, A., Steinbügl, K., Kaufmann, H., & Glück, J. (2006a). Virtual and augmented reality as spatial ability training tools. *Proceedings of the 6th ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction Design Centered HCI - CHINZ '06*, 125–132. https://doi.org/10.1145/1152760.1152776

Dünser, A., Steinbügl, K., Kaufmann, H., & Glück, J. (2006b). Virtual and augmented reality as spatial ability training tools. *Proceedings of the 6th ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction Design Centered HCI - CHINZ '06*, 125–132. https://doi.org/10.1145/1152760.1152776

Duolingo. (n.d.). https://www.duolingo.com

Duolingo Blog. (n.d.). *What Are Leaderboards?* https://support.duolingo.com/hc/en-us/articles/360035931732-What-are-Leaderboards-

Ellis, R. (Ed.). (2014). *Exploring language pedagogy through second language acquisition research*. Routledge/Taylor & Francis Group.

Evers, K., & Chen, S. (2020). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning*, 1–21. https://doi.org/10.1080/09588221.2020.1839504

Falchikov, N., & Goldfinch, J. (2000). Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research*, *70*(3), 287–322. https://doi.org/10.3102/00346543070003287

Farr, F., & Murray, L. (Eds.). (2016). *The Routledge Handbook of language learning and technology*. Routledge.

Freeman, G., Zamanifard, S., Maloney, D., & Adkins, A. (2020). My Body, My Avatar: How People Perceive Their Avatars in Social Virtual Reality. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–8. https://doi.org/10.1145/3334480.3382923

Garcia, I. (2013). Learning a Language for Free While Translating the Web. Does Duolingo Work? *International Journal of English Linguistics*, *3*(1), p19. https://doi.org/10.5539/ijel.v3n1p19

Gnauk, B., Dannecker, L., & Hahmann, M. (2012). Leveraging gamification in demand dispatch systems. *Proceedings of the 2012 Joint EDBT/ICDT Workshops on - EDBT-ICDT '12*, 103. https://doi.org/10.1145/2320765.2320799

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, *27*(1), 70–105. https://doi.org/10.1080/09588221.2012.700315

Gould, J. D., & Lewis, C. (1985). Designing for usability: Key principles and what designers think. *Communications of the ACM*, *28*(3), 300–311. https://doi.org/10.1145/3166.3170

Grimmelikhuijsen, S. G., & Meijer, A. J. (2014). Effects of Transparency on the Perceived Trustworthiness of a Government Organization: Evidence from an Online Experiment. *Journal of Public Administration Research and Theory*, *24*(1), 137–157. https://doi.org/10.1093/jopart/mus048

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421. https://doi.org/10.1016/S0095-4470(03)00016-0

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. *2014 47th Hawaii International Conference on System Sciences*, 3025–3034. https://doi.org/10.1109/HICSS.2014.377

Harrell, D. F. (2019, December 9). Chimeria: Gatekeeper. *D. Fox Harrell, P.h.D.* http://foxharrell.wpengine.com/chimeria-gatekeeper/

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, *47*(3), 360–378. https://doi.org/10.1016/j.specom.2005.04.007

HelloTalk. (n.d.). https://www.hellotalk.com

Holden, C. L., & Sykes, J. (2011a). Mentira: Prototyping language-based locative gameplay. In B. Coulter (Ed.), *Mobile media learning: Amazing uses of mobile devices for learning* (pp. 111–130). ETC Press.

Holden, C. L., & Sykes, J. M. (2011b). Leveraging Mobile Games for Place-Based Language Learning: *International Journal of Game-Based Learning*, *1*(2), 1–18. https://doi.org/10.4018/ijgbl.2011040101

Honorof, D. N., & Feldman, L. (2006). The Chinese character in psycholinguistic research: Form, structure, and the reader. In P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng (Eds.), *The Handbook of East Asian Psycholinguistics* (pp. 195–208). Cambridge University Press. https://doi.org/10.1017/CBO9780511550751.019

Huynh, D., Zuo, L., & Iida, H. (2016). Analyzing Gamification of "Duolingo" with Focus on Its Course Structure. In R. Bottino, J. Jeuring, & R. C. Veltkamp (Eds.), *Games and Learning Alliance* (Vol. 10056, pp. 268–277). Springer International Publishing. https://doi.org/10.1007/978-3-319-50182-6_24

Hwang, W.-Y., Chen, H. S. L., Shadiev, R., Huang, R. Y.-M., & Chen, C.-Y. (2014). Improving English as a foreign language writing in elementary schools using mobile devices in familiar situational contexts. *Computer Assisted Language Learning*, *27*(5), 359–378. https://doi.org/10.1080/09588221.2012.733711

Ibrahim, A., Huynh, B., Downey, J., Hollerer, T., Chun, D., & O'donovan, J. (2018a). ARbis Pictus: A Study of Vocabulary Learning with Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics*, *24*(11), 2867–2874. https://doi.org/10.1109/TVCG.2018.2868568

Ibrahim, A., Huynh, B., Downey, J., Hollerer, T., Chun, D., & O'donovan, J. (2018b). ARbis Pictus: A Study of Vocabulary Learning with Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics*, *24*(11), 2867–2874. https://doi.org/10.1109/TVCG.2018.2868568

Juul, J. (2009). A Casual Revolution: Reinventing Video Games and Their Players. In *Library*. MIT Press. http://www.jesperjuul.net/casualrevolution/

Kaharuddin, A. (2018). THE COMMUNICATIVE GRAMMAR TRANSLATION METHOD: A PRACTICAL METHOD TO TEACH COMMUNICATION SKILLS OF ENGLISH. *ETERNAL (English, Teaching, Learning, and Research Journal)*, *4*(2), 232. https://doi.org/10.24252/Eternal.V42.2018.A8

Karjo, C. H., & Andreani, W. (2018). Learning Foreign Languages With Duolingo and Memrise. *Proceedings of the 2018 International Conference on Distance Education and Learning - ICDEL '18*, 109–112. https://doi.org/10.1145/3231848.3231871

Khan, T., Johnston, K., & Ophoff, J. (2019). The Impact of an Augmented Reality Application on Learning Motivation of Students. *Advances in Human-Computer Interaction*, *2019*, 1–14. https://doi.org/10.1155/2019/7208494

Kiparsky, P. (1982). *Explanation in phonology*. Foris Publications.

Kong, N. (2011). Establishing a Comprehensive English Teaching Pattern Combining the Communicative Teaching Method and the Grammar-Translation Method. *English Language Teaching*, *v4*(n1), p76-78.

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. https://doi.org/10.1017/S0958344008000335

Lan, Y.-J., , Sung, Y.-T., & Chang, K.-E. (2007). A mobile-device-supported peer-assisted learning system for collaborative early EFL reading(Article). *Language Learning and Technology*, *Volume 11*(3), 130–151.

Larsen-Freeman, D. (2010). *Techniques and principles in language teaching* (2. ed., [Nachdr.]). Oxford Univ. Press.

Lasagabaster, D., Doiz, A., & Sierra, J. M. (Eds.). (2014). *Motivation and foreign language learning: From theory to practice*. John Benjamins Publishing Company.

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human computer interaction*.

Lee, M., & Bong, M. (2019). Relevance of goal theories to language learning research. *System*, *86*, 102122. https://doi.org/10.1016/j.system.2019.102122

Leung, R. C. Y. (n.d.). App-based Language Instruction for Developing L2 Oral Proficiency in English. *Jissen Women's University Faculty of Literature Bulletin*, *Vol.64*, 47–57.

Levis, J. (2007). COMPUTER TECHNOLOGY IN TEACHING AND RESEARCHING PRONUNCIATION. *Annual Review of Applied Linguistics*, *27*. https://doi.org/10.1017/S0267190508070098

Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Clarendon Press ; Oxford University Press.

Li, C. N., & Thompson, S. A. (2009). *Mandarin Chinese: A functional reference grammar* (1. paperback print, repr). Univ. of California Press.

Lightbown, P. M., & Spada, N. (1990). Focus-on-Form and Corrective Feedback in Communicative Language Teaching: Effects on Second Language Learning. *Studies in Second Language Acquisition*, *12*(4), 429–448. https://doi.org/10.1017/S0272263100009517

LipSync Pro. (n.d.). Rogo Digital. https://lipsync.rogodigital.com

Liu, G.-Z., Lu, H.-C., & Lai, C.-T. (2016). Towards the construction of a field: The developments and implications of mobile assisted language learning (MALL). *Digital Scholarship in the Humanities*, *31*(1), 164–180. https://doi.org/10.1093/llc/fqu070

Liu, X., Yan, N., Wang, L., Wu, X., & Ng, M. L. (2013). An interactive speech training system with virtual reality articulation for Mandarin-speaking hearing impaired children. *2013 IEEE International Conference on Information and Automation (ICIA)*, 191–196. https://doi.org/10.1109/ICInfA.2013.6720294

Locke, E. A., & Latham, G. P. (1990). A theory of goal setting & task performance. *Prentice-Hall, Inc*.

Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL*, *31*(3), 293–311. https://doi.org/10.1017/S0958344019000065

Lotherington, H. (2018). Mobile Language Learning: The Medium is ^not the Message. *L2 Journal*, *10*(2). https://doi.org/10.5070/L210235576

Maeda, N., Wales, L., & Ferguson-Pell, M. (2021). Verbal Augmented Reality Journaling App: Retention. *Archives of Physical Medicine and Rehabilitation*, *102*(10), e73. https://doi.org/10.1016/j.apmr.2021.07.689

Massaro, D. W., Bigler, S., Chen, T., Perlman, M., & Ouni, S. (2008). Pronunciation training: The role of eye and ear. *Interspeech 2008*, 2623–2626. https://doi.org/10.21437/Interspeech.2008-650

McCrocklin, S. (2019). ASR-based dictation practice for second language pronunciation improvement. *Journal of Second Language Pronunciation*, *5*(1), 98–118. https://doi.org/10.1075/jslp.16034.mcc

Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013). Do points, levels and leaderboards harm intrinsic motivation?: An empirical analysis of common gamification elements. *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, 66–73. https://doi.org/10.1145/2583008.2583017

Memrise. (n.d.). https://www.memrise.com

Mendonça, C. O., Johnson, K. E., & Mendonca, C. O. (1994). Peer Review Negotiations: Revision Activities in ESL Writing Instruction. *TESOL Quarterly*, *28*(4), 745. https://doi.org/10.2307/3587558

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). *Augmented reality: A class of displays on the reality-virtuality continuum* (H. Das, Ed.; pp. 282–292). https://doi.org/10.1117/12.197321

Mohamad Ali, A. Z., & Hamdan, M. N. (2017). The Effects of Talking-Head with Various Realism Levels on Students' Emotions in Learning. *Journal of Educational Computing Research*, *55*(3), 429–443. https://doi.org/10.1177/0735633116672057

Mondly. (n.d.). https://www.mondly.com

MondlyAR. (n.d.). https://www.mondly.com/ar

Monroe, W. (n.d.). *How Machine Learning Helps Duolingo Prioritize Course Improvements*. https://blog.duolingo.com/how-machine-learning-helps-duolingo-prioritize-course-improvements/

Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Mujtaba, S. M., Reynolds, B. L., Parkash, R., & Singh, M. K. M. (2021). Individual and collaborative processing of written corrective feedback affects second language writing accuracy and revision. *Assessing Writing*, *50*, 100566. https://doi.org/10.1016/j.asw.2021.100566

Munday, P. (2017). *Duolingo*. Gamified learning through translation. *Journal of Spanish Language Teaching*, *4*(2), 194–198. https://doi.org/10.1080/23247797.2017.1396071

Nørgaard, M., & Hornbæk, K. (2006). What do usability evaluators do in practice?: An explorative study of think-aloud testing. *Proceedings of the 6th ACM Conference on Designing Interactive Systems - DIS '06*, 209. https://doi.org/10.1145/1142405.1142439

Nushi, M., & Hosein Eqbali, M. (n.d.). *Teaching English with Technology*. *17*(1), 89–98.

O'Brien, M. G., Derwing, T. M., Cucchiarini, C., Hardison, D. M., Mixdorff, H., Thomson, R. I., Strik, H., Levis, J. M., Munro, M. J., Foote, J. A., & Levis, G. M. (2018). Directions for the future of technology in pronunciation research and teaching. *Journal of Second Language Pronunciation*, *4*(2), 182–207. https://doi.org/10.1075/jslp.17001.obr

Paavilainen, J., Korhonen, H., Alha, K., Stenros, J., Koskinen, E., & Mayra, F. (2017). The Pokémon GO Experience: A Location-Based Augmented Reality Mobile Game Goes Mainstream. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 2493–2498. https://doi.org/10.1145/3025453.3025871

Peng, G., Minett, J. W., & Wang, W. S.-Y. (2008). The networks of syllables and characters in Chinese∗. *Journal of Quantitative Linguistics*, *15*(3), 243–255. https://doi.org/10.1080/09296170802159488

Peng, X., Chen, H., Wang, L., & Wang, H. (2018a). Evaluating a 3-D virtual talking head on pronunciation learning. *International Journal of Human-Computer Studies*, *109*, 26–40. https://doi.org/10.1016/j.ijhcs.2017.08.001

Peng, X., Chen, H., Wang, L., & Wang, H. (2018b). Evaluating a 3-D virtual talking head on pronunciation learning. *International Journal of Human-Computer Studies*, *109*, 26–40. https://doi.org/10.1016/j.ijhcs.2017.08.001

Pennington, M. C. (1999). Computer-Aided Pronunciation Pedagogy: Promise, Limitations, Directions. *Computer Assisted Language Learning*, *12*(5), 427–440. https://doi.org/10.1076/call.12.5.427.5693

Pennington, M. C., & Rogerson-Revell, P. (2019). Using Technology for Pronunciation Teaching, Learning, and Assessment. In M. C. Pennington & P. Rogerson-Revell,

English Pronunciation Teaching and Research (pp. 235–286). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-47677-7_5

Plant, R. W., & Ryan, R. M. (1985). Intrinsic motivation and the effects of self-consciousness, self-awareness, and ego-involvement: An investigation of internally controlling styles. *Journal of Personality*, *53*(3), 435–449. https://doi.org/10.1111/j.1467-6494.1985.tb00375.x

Portnoff, L., Gustafson, E., & Rollinson, J. (2021). Methods for Language Learning Assessment at Scale: Duolingo Case Study. *Proceedings of The 14th International Conference on Educational Data Mining (EDM21)*, *EDM '21*, 865–871.

ProgressBar Pack. (n.d.). https://assetstore.unity.com/packages/tools/gui/progressbar-pack-120981

Qualtrics. (n.d.). https://www.qualtrics.com

Research—Duolingo. (n.d.). Retrieved December 1, 2020, from https://research.duolingo.com/

Rogers, S. L., Broadbent, R., Brown, J., Fraser, A., & Speelman, C. P. (2022). Realistic Motion Avatars are the Future for Social Interaction in Virtual Reality. *Frontiers in Virtual Reality*, *2*, 750729. https://doi.org/10.3389/frvir.2021.750729

Rosell-Aguilar, F. (2018). Autonomous language learning through a mobile application: A user evaluation of the *busuu* app. *Computer Assisted Language Learning*, *31*(8), 854–881. https://doi.org/10.1080/09588221.2018.1456465

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need

satisfaction. *Computers in Human Behavior*, *69*, 371–380. https://doi.org/10.1016/j.chb.2016.12.033

Santos, M. E. C., Chen, A., Taketomi, T., Yamamoto, G., Miyazaki, J., & Kato, H. (2014). Augmented Reality Learning Experiences: Survey of Prototype Design and Evaluation. *IEEE Transactions on Learning Technologies*, *7*(1), 38–56. https://doi.org/10.1109/TLT.2013.37

Santos, M. E. C., Lübke, A. in W., Taketomi, T., Yamamoto, G., Rodrigo, Ma. M. T., Sandor, C., & Kato, H. (2016). Augmented reality as multimedia: The case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning*, *11*(1), 4. https://doi.org/10.1186/s41039-016-0028-2

Schlömmer, M., Spieß, T., & Schlögl, S. (2021). Leaderboard Positions and Stress—Experimental Investigations into an Element of Gamification. *Sustainability*, *13*(12), 6608. https://doi.org/10.3390/su13126608

Sheen, Y. (2011). *Corrective feedback, individual differences and second language learning*. Springer. http://public.eblib.com/choice/PublicFullRecord.aspx?p=6350836

Sinyagovskaya, D., & Murray, J. T. (2021). Augmented Reality in Chinese Language Pronunciation Practice. *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 403–408. https://doi.org/10.1109/ISMAR-Adjunct54149.2021.00092

Stern, H. H. (1983). *Fundamental concepts of language teaching*. Oxford University Press.

Sydorenko, T., Hellermann, J., Thorne, S. L., & Howe, V. (2019). Mobile Augmented Reality and Language-Related Episodes. *TESOL Quarterly*, *53*(3), 712–740. https://doi.org/10.1002/tesq.507

Sykes, J. M., & Holden, C. L. (20147). Chapter 7 Complex L2 pragmatic feedback via place-based mobile games. In N. Taguchi (Ed.), *Technology in interlanguage pragmatics research and teaching* (pp. 155–184). John Benjamins Publishing Company.

Taskiran, A. (2019). The effect of augmented reality games on English as foreign language motivation. *E-Learning and Digital Media*, *16*(2), 122–135. https://doi.org/10.1177/2042753018817541

Technologies, U. (n.d.). *Unity Real-Time Development Platform | 3D, 2D VR & AR Engine*. Retrieved November 12, 2020, from https://unity.com/

Thomson, R. I. (2011). Computer Assisted Pronunciation Training: Targeting Second Language Vowel Perception Improves Pronunciation. *CALICO Journal*, *28*(3), 744–765. https://doi.org/10.11139/cj.28.3.744-765

Thorne, S. L., Hellermann, J., Jones, A., & Lester, D. (2015). Interactional practices and artifact orientation in mobile augmented reality game play. *PsychNology Journal*, *13*(2–3), 259–286.

Tönnis, M., Plecher, D. A., & Klinker, G. (2013). Representing information – Classifying the Augmented Reality presentation space. *Computers & Graphics*, *37*(8), 997–1011. https://doi.org/10.1016/j.cag.2013.09.002

Topping, K. (1998). Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research*, *68*(3), 249–276. https://doi.org/10.3102/00346543068003249

Translating with Duolingo for Language Learning. (2013, November 8). The FLTMAG. https://fltmag.com/duolingo/

Tseng, S.-C., & Tsai, C.-C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education*, *49*(4), 1161–1174. https://doi.org/10.1016/j.compedu.2006.01.007

Unreal Engine. (n.d.). Epic Games. https://www.unrealengine.com/en-US/

Vaux, B., & Samuels, B. (2018). Abstract Underlying Representations in Prosodic Structure. In D. Brentari & J. L. Lee (Eds.), *Shaping phonology* (pp. 141–181). The University of Chicago Press.

Vincent Annette & Ross Dianne. (2001). Personalize training: Determine learning styles, personality types and multiple intelligences online. *The Learning Organization*, *8*(1), 36–43. https://doi.org/10.1108/09696470110366525

von Ahn, L. (n.d.). *Language Learning Trends in a Year Like No Other*. https://blog.duolingo.com/global-language-report-2020/

von Ahn, L. (2013). Duolingo: Learn a language for free while helping to translate the web. *Proceedings of the 2013 International Conference on Intelligent User Interfaces - IUI '13*, 1. https://doi.org/10.1145/2449396.2449398

Wen, X. (1997). Motivation and Language Learning with Students of Chinese [1]. *Foreign Language Annals*, *30*(2), 235–251. https://doi.org/10.1111/j.1944-9720.1997.tb02345.x

Wen, Y. (2018). Chinese character composition game with the augment paper. *Educational Technology & Society*, *21*(3), 132–145.

Wu, H.-K., Lee, S. W.-Y., Chang, H.-Y., & Liang, J.-C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, *62*, 41–49. https://doi.org/10.1016/j.compedu.2012.10.024

Xu, Y., Barba, E., Radu, I., Gandy, M., Schrank, B., Macintyre, B., & Tseng, T. (2011). Pre-Patterns for Designing Embodied Interactions in Handheld Augmented Reality Games. *Information Systems Journal*, 19–28. https://doi.org/10.1109/ISMAR-AMH.2011.6093652

Yabla.com. (n.d.). https://chinese.yabla.com/chinese-tones-learn-the-right-way-with-tone-pairs.php

Yamaguchi, S., Waki, K., Nannya, Y., Nangaku, M., Kadowaki, T., & Ohe, K. (2019). Usage Patterns of GlucoNote, a Self-Management Smartphone App, Based on ResearchKit for Patients With Type 2 Diabetes and Prediabetes. *JMIR MHealth and UHealth*, *7*(4), e13204. https://doi.org/10.2196/13204

Yan, H., Lin, J., & Liu, Y. (2018). EFL Pronunciation Training With Computer-Assisted Adaptive Peer Review. *English Language Teaching*, *11*(11), 74. https://doi.org/10.5539/elt.v11n11p74

Yu, B., & Watkins, D. A. (2008). Motivational and cultural correlates of second language acquisition: An investigation of international students in the universities of the People's Republic of China. *Australian Review of Applied Linguistics*, *31*(2), 17.1-17.22. https://doi.org/10.2104/aral0817

Zadoenko, T. P., & Huang, S. (2019). *Beginner Course of Chinese Language* (5th ed., Vol. 1). Vostochnaya Literatura.

Zuckerman, O., & Gal-Oz, A. (2014). Deconstructing gamification: Evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. *Personal and Ubiquitous Computing*, *18*(7), 1705–1719. https://doi.org/10.1007/s00779-014-0783-2

(N.d.). *Mondly Timeline*. https://www.mondly.com/blog/2021/05/28/mondly-timeline/