THE ENGLISH IPA EAR TRAINER

BY

SEBASTIAN ROCHELEAU

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Teaching English as a Second Language
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Advisers:

Associate Professor Randall Sadler
Assistant Professor Suzanne Franks

ABSTRACT

The affordances of the internet, such as social networking sites, wikis, and various communication tools have become integral to daily life. Educators are increasingly dedicating more time to developing applications to take advantage of the Web 2.0 age for second language learning purposes. However, many applications lack a solid pedagogical foundation or evidence-based reasoning in the application's design. This study focused on the creation of a browser-based language learning application to aid language learners' perception of a selection of English sounds. Using the training method known as High Variability Phonetic (HVPT), this thesis will present the creation of a learning application: the English IPA Ear Trainer. This paper will also detail the creation method and timeline in order to provide instructors with a model process to aid the creation of their own language learning application.

Keywords: E-learning, Phonological Training, High Variability Phonetic Training, L2 Acquisition

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisers Dr. Randall Sadler and Dr. Suzanne Franks of the Department of Linguistics at the University of Illinois at Urbana Champaign. You both consistently allowed this paper to be my own work but steered me in the right the direction whenever I needed it. Without feedback and discussion from you both over past years I would not have been able to complete this program or become the teacher I am today.

I would also like to thank the volunteer participants who let me record them for use in the program. Without the time you committed to allowing me to record you making many sounds this project could not have been completed.

Finally, I would like to thank my wife Jennifer, without your love and support I would not have been able to complete this program. You are my rock and without your unfailing support and continuous encouragement throughout the years, none of this could have been accomplished. Thank you.

Sebastian Rocheleau

TABLE OF CONTENTS

CHAPTER 1—INTRODUCTION

While using technology to enhance language learning is not a new concept, recent advancements in computer technology and the internet provide new opportunities for language instructors and learners. The proliferation of personal computers (PCs) and the increasing accessibility of the internet from the 1990s onward has taken teaching outside of the classroom and into people's homes (Bax, 2003). The study and application of computers as a language learning and teaching tool, commonly referred to as Computer Assisted Language Learning (CALL) (Levy, 1997), is an integrated part of modern teaching pedagogy. Teaching techniques such as project-based telecollaborative language learning and self-regulated language learning are increasingly reliant on advancements such as the internet, PCs, and smart devices to facilitate more engaging and effective language learning. CALL software such as Rosetta Stone launched in 1992 and Duolingo launched in 2011 (Robertson, 2011) are commonly used by educational institutions around the globe.

High Variability Phonetic Training (HVPT) is returning to popularity due to advancements in computer technology. HVPT is a form of phonological perceptual training using natural stimuli from multiple speakers. The stimuli is presented to learners in a variety of phonetic contexts and includes immediate feedback on the learner’s response to the stimuli (Wong 2014). Initially proposed by Goto (1971), HVPT requires access to a training regime that includes audio recordings in very specific phonological contexts such as focused /CV/ audio stimuli for enhancing the perception of /l/ and /r/ for Japanese speakers (Goto, 1971; Lively, Logan, & Pisoni, 1993; Pisoni, & Lively, 1995; Iverson, Hazan, & Bannister 2005)**.** Due to these requirements, use of HVPT has been predominantly limited to within language labs from the

1970s onward. However, with advances in technology, accessibility to HVPT has finally become convenient for instructors and language learners.

With the mass proliferation of PCs and the internet, HVPT has become a practical teaching tool for language teachers and learners. However, there is a paucity of HVPT tools available online. Currently, there is no free and accessible HVPT training program that can be used by students to improve their perception or by teachers as a tool to add to their repertoire of easily accessible learning resources. Furthermore, a design overview of existing HVPT-based teaching tools is not available to instructors or students to facilitate the use these tools effectively. Many existing tools also use outdated code such as the ipatrainer.com or proprietary software such as Flash, which limits the service lives of these tools. Some tools like the English Accent Coach and UCL Vowel Trainer also restrict access by erecting a paywall. The proposed HVPT tool will enable both students and teachers to easily access free training software, while the accompanying website will offer browser and mobile compatibility and clear design methodology and background to promote the use of HVPT.

CHAPTER 2—LANGUAGE AND THE BRAIN

Language acquisition research has broken into two distinct camps. First language (L1) acquisition studies examine the acquisition of one’s native language from birth until the end of puberty, while second language acquisition (L2) research focuses on language learning after puberty. Though the cognitive processes used to acquire one’s L1 and L2 differ, L1 and L2 acquisition share many key perceptual features such as auditory and visual perception.

The perceptual foundation of language acquisition can allow us to imply some inherent universal elements concerning the perception of language. This foundation differs from the language acquisition device proposed by Chomsky (1967) and is rather an agreed upon set of phonological concepts among FLA and SLA researchers. The concept of phonological categories in the brain is a foundational element in both the Speech Learning Model (SLM) and the Perceptual Assimilation Model (PAM) (Best, 1994; Flege, 1995). Phonological categories are also an integral aspect of an infant's ability to acquire their L1 which later impacts L2 acquisition.

The creation of the L1 phonological template also has a direct impact on L2 acquisition. Born with the ability to perceive all sounds, infants’ discrimination of non L1 phonological input over time leads to the loss of perceptual categories that are no longer required for one's native language(s) (Goodman, Walley, & Michela, 1996). This degradation of perception leads to the phonological perceptual boundaries that can enable or inhibit SLA.

First Language Acquisition (FLA)

The impact of FLA on SLA is felt in two distinct ways. First, FLA creates the perceptual categories one uses for all language processing. Phonological categorical perception is a critical component to our ability to perceive language and is a critical aspect of current language perception theory. Secondly, infant L1 language preference leads to the loss of perceptual categories throughout the first year of childhood. Our L1 acts as the phonological filter for our perception of spoken language. In both PAM and SLM, all categorical actions are made in reference to learners' L1 (Best, 1994, Flege, 1995). Based on the perception of a new sound, it can be categorized into a pre-existing category or a new one.

Phonological Categories

The ability to categorize phonological input into specific categories is a critical aspect of language development. Without our ability to categorize these sounds into phonemes, a person would not be able to process the linguistic audio information of their L1 and must rely on other means to acquire language. The retention of the phonological foundation for language acquisition has some innate characteristics (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Eimas, 1975; Strange, & Jenkins, 1978), however, infants do not retain this full phonological palette of human language indefinitely (Goodman, et al., 1996). Infants' ability to perceive language changes over their development with the eventual loss of all phonological categories that are not required for their day to day use. Furthermore, this loss of general phonological categories occurs regardless of input as several famous cases of feral children and child neglect have shown: Without adequate audio stimuli in the first few years of life, one's ability to acquire language becomes deeply curtailed (Goodman et al., 1996).

Phonological categorization is an innate aspect of human language learning from the earliest stages of development. Infants as young as 6 months are able to perceive and distinguish between distinct sounds (Goodman et al., 1996; Ohala, 2008) and need these skills to develop their L1 competency. Eimas, Siqueland, Jusczyk, & Vigorito (1971) observed that infants were able to differentiate between /p/ and /b/ at the age of 1 to 4 months, demonstrating the existence of perceptual categorization. As mentioned above, this phonological foundation allows infants to perceive the audio stimuli of their surroundings and categorize them for their L1 (Eimas et al., 1971). Infants' ability to perceive the audio stimuli of their surroundings is more sensitive than adults. Goodman et al. (1996) observed that English L1 infants aged 6 to 8 months were better able to perceive contrasts between Hindi and English than adults with English as their L1. This perceptual foundation changes in depth and detail over time as the infants' discrimination to sounds of their L1 will cause profound changes to their perceptual ability and influence their L2 acquisition.

First Language Discrimination

From birth, a newborn will show recognition and preference for their mother's voice (DeCasper, & Spence, 1986). Infants will pay more attention to voices and sounds they hear often and over time. As this focus is honed in on the sounds in their environment, an L1 preference is formed and permanently impacts their language perception.

Parents provide constant and rich audio stimulus for the infant with adults and adolescents using "baby talk" to interact with the infant. Over time, the infant will pay more and more attention to phonological stimuli that is regularly repeated. The stimuli allows the infant to create the required categories to perceive their L1 with pre-existing categories adapting to fit the requirements of the L1. This heightened focus on repeated stimuli leads to the degradation and eventual loss of sounds that the infant does not encounter (Tees, & Werker, 1984).

Over time, the infants' focus on particular sounds critical to their L1 will lead to the loss of unneeded phonological categories. By 10 -12 months, infants already demonstrate losing the ability to perceive non L1 sounds. This perceptual change leads to infants completely losing the ability to discriminate between L1 and L2 sounds and L2 to L2 sounds. Sounds they had previously been able to distinguish such as the /l/ and /r/ could become one or disappear completely. This loss of categorical perception has been described as the first critical period in the language learning process (Tees, & Werker, 1984).

This change of the perceptual ability of infants and adolescents is responsible for creating the parameters of SLA theory. The perceptual boundaries created by L1 phonological acquisition creates the void in phonological perception that SLA theory inhabits. Both the SLM and PAM use the categories created in FLA as the key perceptual filter that must adapt to new input. By altering the boundaries of existing L1 categories or creating new perceptual categories through techniques like HVPT, L2 learners' change their initial L1 limitations as they learn new languages.

Second Language Acquisition

Though FLA and SLA are connected they are different processes. A learner's L1 can have a dramatic impact on their ability to perceive their L2. The phonological foundation of their L1 will transfer positively or negatively on the learners' perception of their L2 depending on the relationship between the L1 and the L2. These interactions do not exist when an infant is learning language for the first time. In SLA, learners can be both hindered or aided by their pre-existing linguistic knowledge. Though past the critical period to inherently acquire language, with proper input and output it is possible for learners to perceive and produce sounds foreign to their L1.

Both major theories in phonological SLA, Best's Perceptual Assimilation Model (PAM) and Flege's Speech Learning Model (SLM), agree that the categories from a learners' L1 play a vital role in L2 acquisition (Best, 1994; Flege, 1995; Best, et al., 2001; Major, 2001; Flege, 2005; Shafer, 2008.). These models differ on how new sounds are categorized by the brain once perceived. Though both models agree that distance from pre-existing perceptual categories is critical to the categorization process, neither model provides an explicit set of rules for L2 phonological perception and acquisition. However currently they offer the best explanation as to how the brain categorizes new L2 phonological input.

Perceptual Assimilation Model (PAM)

The Perceptual Assimilation Model (PAM) was originally created by Best, McRoberts, and Sithole (1988) to help describe native English speaker's affinity to Zulu clicks (Best, McRoberts, Goodell, 2001). In PAM, when a learner perceives a sound it can be processed in three distinct manners. First, if the sound is close to a pre-existing L1 category, the L2 sound will be assimilated into the L1 category. Second, if the sound is perceived as language and is too far from the learners' L1 categories to be assimilated, it is perceived as an L2 sound and will form a new phonological category. Finally, if the sound is not perceived as language, it is perceived as unintelligible or nonsense (Best 1995; Best et al., 2001; Shafer, 2008). Best hypothesized that the click sounds were so easily perceived by English speakers due to their distance from any phonemes in English (Best et al., 1988). Best believed that these clicks were perceived as so alien to native English speakers that they were not categorizing clicks as part of language, but rather as part of the unintelligible or nonsense category. If the clicks had been more similar to a preexisting L1 category, they would have been assimilated into the L1 category and the ease of their assimilation would be much faster than the time required to create a new L2 phonological category. Since its initial proposal in 1994 by Best, PAM has been evaluated by many SLA researchers (Major, 2001; Shafer, 2008.). Overall, it is a model with some limitations but works well to explain how L2 language learners perceive L2 phonological input in the very early stages of L2 learning.

Speech Learning Model (SLM)

Flege's Speech Learning Model (SLM) was initially proposed to account for variations in learners' ability to produce and perceive phonetic segments from L2 languages (Flege, 2005). The SLM contains four hypotheses relating directly to L2 perception (Major, 2001). First, the relation between L2 and L1 sounds exists perceptually at the allophonic level rather than the phonemic level. For example, Flege, Takagi, & Mann (1995) found that Japanese L1 speakers perceive and produce /l/ and /r/ differently depending on whether the sound is in the word initial or final position. Second, it is possible for learners to create new independent categories for L2s. Third, the larger the dissimilarity between the L1 sound and the L2 sound, there is a higher chance that the L2 will be discerned. Fourth, the acuity of one's ability to discern sounds decreases with age, however, one's ability to create new categories remains a constant throughout one's life (Flege 2005). Flege's SLM also provides some suggestions to improve phonological acquisition in learners. Flege recommends phonological training regimes that focus on raising learners' awareness to the differences between two phonological categories. Learners, especially after the initial stage of L2 acquisition, will become aware of the subtle differences between categories and either adjust pre-existing categories or create new categories.

Similarities

Both the SLM and the PAM share several key features. Some of the features, such as phonological categories within the brain, are considered universals of phonological language acquisition (Best, 1994; Flege, 1995; Flege, 2005; Best et al., 2001; Major, 2001; Shafer, 2008). L1 categories impact L2 acquisition and act as the phonological perception foundation and main filter used to analyze L2 phonological input. Both models also rely on contrastive analysis to describe how learners analyze new phonological input. By comparing the similarities between preexisting L1 and L2 categories, learners categorize new sounds.

Phonological Categories

Since the 1970s, SLA and FLA researchers have agreed that the sounds of language are divided into distinct perceptual categories (Goto, 1971, Goodman, Walley, & Michela, 1996). Both PAM and SLM use phonological categories as the unit or measurement of L2 acquisition (Best, 1994; Flege, 1995). For example, both PAM and SLM researchers have used the much studied /l/ vs /r/ category distinction of Japanese speakers as an acquisition environment to test each model (Best, & Strange, 1992; Flege et al., 1995; Best et al., 2001). Both models detail acquisition as different processes of phonological categorizations. In the PAM, new L2 phonological input is categorized into pre-existing phonemic categories or new phonemic categories (independent L2 or nonsense/unintelligible). In the SLM, new phonological input is categorized into pre-existing allophonic categories or new allophonic categories (independent L2 or nonsense/unintelligible). Though the manner in which the SLM and the PAM describe categorization is different, without phonological categories neither model would exist.

L1 impact on L2 acquisition

When describing the process of phonological categorization in PAM and SLM, both models use the learner's L1 as the main perceptual reference. Based on the learners' L1, sounds will be organized based on perceived phonological distance from pre-existing categories and the new sound. The concept of distance is based on contrastive analysis (CA). In CA, the similarities and differences between the L1 and L2 (James, 1980) impact the categorization of sounds. Based on the similarities and differences between the L1 and L2, instructors can try to predict which aspects of the L2 will be challenging to the learner. PAM and SLM both rely on the degree of difference as a main predictor of L2 phonological categorization (Best, 1994; Flege, 1995;

Shafer, 2008). If L2 sounds are similar or cross over phonological features with a pre-existing L1, it will most likely be assimilated into a pre-existing category. If the L2 sound is perceived as language but is distinct enough in phonological features to avoid assimilation, a new phonological category can be created. Finally, the sound could be perceived as so distinct that is does not belong to language and is simply nonsense (Major, 2001; Shafer, 2008).

Differences

Though both models agree on several universals in SLA and have foundations in CA, there remain several differences between these two models. Both models were products of their SLA research environment. PAM was originally conceptualized to explain learners' initial proficiency and perceptualization of L2 sounds with little to no exposure (Best, 1995), while the SLM was used to describe the categorization process of language learners who have already been exposed to L2 audio input (Flege, 1995; Shafer, 2008). This difference in original purpose has led to differences in rationale for facts such as the level of phonological categorization: PAM focuses on the phonetic level, while SLM focuses on the allophonic level description of the learners' phonological foundation and the process of input categorization (Best et al., 2001; Major, 2001; Flege, 2005).

Input Categorization

Once a learner receives L2 phonological input, both the SLM and PAM agree that the learner will categorize the sound in their brain. In both models, if the sound is near an existing L1 phonological category, it will be recognized as a sound in that pre-existing category. In both models, if the sound is distinct enough to be perceived as different from any sound in the learner's L1, a new category is created for the L2 sound (Best, 1994; Fledge, 1995). However, the PAM adds a nonsense category, which is not present in the SLM (Best, 1994). This category is a useful way to describe sounds that might be part of an L2 but have yet to be perceived as part of the L2 repertoire.

The SLM and PAM rely on two distinct levels (phonemic and allophonic) of the learner's L1 phonological system to explain phonological acquisition. In the SLM, the interaction and categorization of phonological input between the L1 and L2 happens at the allophonic level. Fledge, Takagi, & Mann (1995) found that Japanese L1 speakers could perceive and produce /l/ and /r/ at different levels of proficiency if the sounds where in word initial versus word final position. In the PAM, the interaction between L1 and L2 phonology occurs at the phonemic level. Returning to the /l/ and /r/ distinction in Japanese, L1 Japanese speakers will perceive and produce both /l/ and /r/ as a single Japanese phoneme (Best et al., 2001).

Phonological Foundation

The perceptual system used to perceive and categorize L2 sounds differs between the PAM and SLM. In FLM, both L1 and L2 acquisition use the same perceptual system (Flege, 2005). Though the categories used in the perceptual system will change over the development of a learner's L1, it is still the same system that the brain uses to categorize sound. In the PAM the categorization system transcends the FLA and SLA and includes the ability to perceive and categorize non-linguistic input (Best, 2001). The system used for comprehending sounds is not a unique ability to perceive language but "general-purpose auditory processes employed for perception of non-speech sounds (p.778)".

Criticisms

Though the PAM and SLM provide a rationale for how L2 sounds are perceived and categorized in the brain, both models lack a detailed set of rules that govern perception and acquisition. When describing the assimilation process both models fail to discuss the conditions required to use the learners’ categorization tool (Shafer, 2008). Is the categorization tool constantly on or is it only used in specific circumstances? Are there some environments where the categorization tool is more effective? Answering these questions will allow instructors to better create the appropriate conditions for learners to activate their categorization tool.

Both models also lack sufficient details on the cognitive parameters of the learners’ categorization tool. If an L2 sound is between two L1 categories, which L1 category will assimilate the sound? Will the L2 sound remain as one category or be broken in two? Is it possible to extract two L2 categories from a larger L1 category? Answering these questions will allow instructors to design more focused training to meet the need of specific L1 and L2 interactions. Phonological training programs using HVPT allow researchers to test the limits of these theoretical models along with providing learners with an evidence-based training tool ready to integrate into their language learning regimen.

CHAPTER 3—HIGH VARIABILITY PHONETIC TRAINING

When learning a second language (L2), a key feature that many L2 instructors emphasize is the importance of intelligibility. Many learners may possess sounds in their L1 that are similar to the L2 target sounds, or may not have the sounds at all in their linguistic repertoire. Both situations provide different difficulties for L2 learners. The problems of how to teach the perception and accurate production of L2 sounds, especially those not present in the learner's L1, have received considerable attention in SLA and classroom research. The practice of high variability phonetic training (HVPT), the exposure and training with multiple speakers producing target sounds in varied phonetic contexts, was first introduced in the 1980s. With the increasing affordability and availability of technological resources, it has recently become a topic of great interest and relevance in SLA research. This section will examine the theoretical background of HVPT, summarize key prior research, and critically evaluate current research with the perspective of the practical implementation of HVPT research findings and the language classroom setting in mind.

HVPT is the use of multiple, varied voices to provide pronunciation models of target sounds in varied phonetic environments to improve language learners' L2 perception and production (Thomson, 2011). The groundwork for HVPT as a teaching technique relies on two distinct foundations: the development of the study of categorical perception of phonological features and the study of the effectiveness of varied stimuli as a learning tool. Stimuli variation was initially studied by Keele and Posner using visual stimuli in the 1960s. Keele and Posner (1968) compared the performance of two groups of participants who were given two types of novel visual stimuli during the training procedure of the experiment. One group received high variability novel stimuli and the other group received low variability novel stimuli. The group that received the high variability stimuli significantly outperformed the low variability group in performance tests. The effectiveness of using varied stimuli as a learning tool would become a topic that would be sporadically investigated over the next five decades, gradually increasing in interest and applicability with the increasing accessibility and availability of technology and mediums with which to provide the varied stimuli.

Previous Research Into HVPT

Within a phonological training context, the first study to address the importance of variation for perception and production was carried out by Goto in 1971. Goto described the impact of a single input source of phonological information on Japanese L1 English learners' ability to perceive target L2 sounds. Goto noted that learners required more training if they relied on a single source of stimuli compared to those that received multiple sources of stimuli. Goto stated that a single source of stimuli seemed to have a negative impact on learner gains. He posited that learners would adapt to their stimuli source and require significant training to break down the preferences for the single source once the source was changed. Goto (1971) reported that using multiple sources improves learners' ability to identify specific phonological features of sounds in a language. The phonological features can then be identified and categorized within the mind.

During the 1970s and 1980s, researchers explored the specific phonological categories for each unique L1 sound within learners' minds. Goto (1971) found that as learners age they establish specific categories for phonological sounds in the L1 (e.g., /a/, /v/, /b/). These categories provide the framework for future phonological identification after hearing a sound. Samuel's (1982) research on category construction described mental representation of phonetic categories as having 'foci' (p.2), which serve as the representative member of a category, and other phonetic segments are judged according to their distance from the focus of a particular category. Kuhl's (1991) research argued that the foci of categories are assumed to be more accurately perceived and more stable in long-term memory than outlying members of categories.

SLA researchers expanded the concept of categories to L2s and detailed the impact of categorical organization. Two features were found to be the most important for changing the categories in their L2: negative transfer and perceptual salience. Extensive research has been done on the acquisition of /l/ and /r/ by Japanese L1 learners of English and shown that these sounds, which belong to the same category in Japanese, are extremely difficult for learners to accurately perceive and produce in English where the /l/ and /r/ are contrastive (Strange, & Dittmann, 1984; Logan et al., 1991 & 1993; Iverson et al., 2005; Uther et al., 2007). This conflict between categorization in a learner's L1 and L2 was used as an example of negative transfer from the learner's deep L1 structure (Samuel, 1982).

When target sounds were completely different (or not present) in the learner's L2 repertoire, learners had little conflict in creating new categories. In these categories, the impact of the learner's L1 was non-existent as there were no similarities in productive features (manner and place of articulation) within existing L1 categories. Learners were easily able to perceive how different the new L2 category was from categories within their L1 repertoire.

The categorization of phonological sounds also provided a framework for researchers to focus on HVPT. By confirming that humans create specific categories in their L1, researchers could compare the similarities and differences between categories between L1 and L2. Logan, Lively, & Pisoni (1991) found that through HVPT learners could see improvements in creating L2 perceptual and productive categories and also found in their 1993 study that Japanese L1 learners could achieve substantial gains in perceptual differences between /l/ and /r/. This was done by bringing learners attention to the specific areas of concern through the use of minimal pairs. Gains achieved through HVPT were not equal in rate however as more success was found in word final position than word initial (Logan, Lively, & Pisoni 1991). This is most likely due to the difference in the features of the /l/ and /r/ in word final position in English (the dark /l/) compared to the /l/ in word initial position.

Studies in the late 1980s and early 1990s confirmed the benefits of variation in stimuli and provided a framework for assessment. The AX comparison system developed by Pisoni (1973) allowed researchers to assess learners on changes in their L2 phonological categories. AX comparisons assess learners through a series of perceptual judgement tasks where learners determine if the stimuli presented to them in two instances is the same or different. Exposure to two instances of audio stimuli can happen consecutively or with a delay between exposures. The use of AX discrimination tasks has become a standard in recent HVPT studies. This allows for a uniformity in assessment over time, allowing research to expand the applicability of HVPT.

The modern concept of HVPT finally came together with a series of four papers on /l/ and /r/ training of Japanese learners by Logan, Lively and Pisoni (1991, 1993), Pisoni and Lively (1995) and Iverson, Hazan, and Bannister (2005). Through the use of minimal pairs, participants were exposed to various minimal pair stimuli with the target /l/ and /r/ sound in word initial, middle, and final positions. Logan, Lively, and Pisoni (1991) created stimuli from ten L1 speakers of English. Results showed that Japanese speakers could improve their categorical perception of the /l/ & /r/ distinction but that the gains were not equal in all positions. They argue that the reason for the gains was HVPT’s ability to draw learner attention to the specific target language. By varying the stimuli, learners were made more aware of the differences in the stimuli and were forced to create or adjust their categories to create a more ideal version of the phonological category.

Iverson, Hazan, and Bannister (2005) also compared the results of HVPT stimuli to an individual source. Iverson, Hazan, and Bannister divided participants into two groups. The control group used a variety of stimuli of /l/ and /r/ minimal pairs in all three word positions while the test group experienced stimuli from a single speaker. Both groups experienced stimuli within the language lab setting over a 2-3 week period. Both groups performed an AX comparison test with alternative stimuli for the stimuli presented in the training sessions. The results of the AX comparisons showed significant gains in the HVPT group compared to the low variation group. This supported the earlier observations by Goto (1971) that both groups experienced gains in their L2 phonological categorization but that the HVPT group outperformed the low variability group by a significant margin. These results were also confirmed by Wong (2012) when comparing HVPT and low variability phonetic training (LVPT) of vowels with Chinese L1 speakers at the University of Hong Kong.

Current Research

Current research in HVPT has branched in several different directions. With the popularization of mobile devices and smartphones, HVPT research has focused on extending the scope of HVPT applicability. In Uther et al.'s 2007 study, MALL-based HVPT was found to be an effective learning modality. Though the study found no difference between the two forms of HVPT being tested, it found that MALL-based HVPT was an effective modality of facilitating phonological acquisition of L2 sounds. Mobile modality allows HVPT to discard its dependence on language labs and move more into the sphere of self-regulated language learning. Instructors can create HVPT content that students can access through their mobile phones at their convenience.

Originally confined to the /l/ and /r/ perception of L1 Japanese learners, researchers are currently expanding the use of HVPT into a variety of L2 speaker background and categorical environments. Kondaurova and Francis (2010) examined HVPT training in L1 Spanish learners of English for vowels and stops and found that HVPT provided measurable improvement in the voice onset times (VOT) and vowel accuracy of learner production. Alia-Garcia and Mora (2009) also found increased levels of perception of these speakers in these categories. Thomson (2011) studied the effect of HVPT on Chinese Cantonese and Mandarin while Wong (2012, 2014) assessed HVPTs ability to aid in geminate acquisition. In both cases, HVPT was found to be an effective training tool at improving learners’ perception and production, however, all of the studies mentioned above used HVPT in a language lab based setting.

PAM, SLM, and High Variability Phonetic Training (HVPT)

The Speech Learning Model (SLM) and Perceptual Assimilation Model (PAM) provide a theoretical rationale for the effectiveness of HVPT in improving the categorical perception of phonological features (Best 1994, Flege 1995, Thomson, 2011). The PAM focuses on non-native speech perception in general while SLM focuses on L2 perception and production by active learners (Flege 1995, Best, & Tyler 2007). Both the SLM and PAM propose that listeners perceive and categorize L2 phonological categories into learners' pre-existing L1 categories, however, the two models posit different processes for how this assimilation is carried out. In the SLM, contrastive features between auditory stimuli and pre-existing phonological categories raises learners’ awareness to new phonological features (Flege 1995). In the PAM there is a focus on the perceptual warping of phonological categories due biases of the listeners’ L1 (Best

1994, Best, & Tyler 2007). Both models agree that phonetically distinct L2 sounds can be assimilated into L1 phonological categories that have similar phonological properties. As the SLM is also a model for learners to actively expand their L2 phonological categories, Flege (1995, 2003, 2005) suggests a method to help learners overcome their L1 perceptual filter. In the SLM, Flege (1995) stresses the importance of "accurate perceptual targets" to raise learners' awareness of the target L2 phonological categories. This is especially important for L2 phonological categories that are similar to pre-existing L1 categories, such as /l/ and /r/ distinction in L1 speakers of Japanese. By creating stimuli that focus on the phonological salience between similar L1 and L2 categories through repeated yet varied stimuli, HVPT can raise learners' awareness to the differences between sounds.

Relationship between HVPT and CALL

Since the inception of using HVPT to increase the perception and production of target L2 sounds, this training method has relied on the integration of technology in language learning. Bax (2003) and Warschauer and Healey (1998) partition Computer Assisted Language Learning (CALL) into three distinct phases: Structural, communicative, and integrative. Each of these stages contain specific characteristics due to the technological limitations of each phase. As computer use progressed from the structural to the integrative phases of computer integration, learner accessibility to HVPT increased dramatically.

The initial phase of Warschauer, and Healey's (1998) model, the structural phase, corresponds to computer use in teaching pedagogy in the 1970s. This phase relied on the audio-lingual methodology (Salaberry, 2001), cognitive methods (Alexander, 2007), and language labs as the main point of interaction between the user and the teaching tool. Rote learning drills paired with audio recording devices allowed instructors to employ HVPT within the limited confines of the language lab. Unfortunately, many teaching environments lacked a dedicated language lab to enable the use of HVPT, as labs are limited to schools, colleges, and universities (Alexander, 2007). Furthermore, the use of language labs for language learning is a trend that fell out of favor due to little faculty involvement, a paucity of programs for advanced students, low quality commercial recordings, and lack of meaning in rote drills (Salaberry, 2001). Popular teaching pedagogy also shifted from audio lingual and cognitive approaches to a more communicative approach to language teaching that eschewed the language lab for more interpersonal interaction. The technological requirements to enable HVPT limited its availability to language learners outside of a rigid learning environment. Until the proliferation of the home computer, HVPT lacked a practical device that language learners could access on an impactful scale.

Warschauer and Healey's (1998) second phase of CALL, the communicative phase, corresponds to the 1980s and early 1990s. During this phase, the personal computer became the main foci of CALL based instruction. The English-teaching paradigm shift to communicative language teaching led to language labs falling into disfavor among educators. This had a major impact on the proliferation of HVPT, as it was still dependent on the language lab as the major point of instruction.

A hallmark of the communicative phase was the introduction of the personal computer (PC) as an instructional tool. However, several limitations of PCs at that time also attributed to the lack of HVPT adoption as a mainstream teaching technique during this phase. PCs were prohibitively expensive for the majority of language learners, which restricted access to language learning software. Though several language learning programs were developed to aid language learners and instructor, such as Rosetta Stone (Greenfield, 1993) and the PLATO system (Hart, 1981), HVPT was not integrated into these programs and systems. Finally, PCs still lacked the ability to effectively connect on a large scale. Predecessors of the modern version of the internet were limited to major educational institutions. Though PLATO offered a home based educational platform, PLATO was still not recognized by mainstream language learning and instructional communities.

With the mainstream adoption of the internet and PCs, CALL has transitioned into the final phase described by Bax (2003) and Warschauer and Healey (1998): the integrative phase.

From the late 1990s onward, the personal computer has become integrated into our lives to such an extent that their use has become normalized (Rogers, 1995, Bax, 2003). Personal computers and the internet have become an essential part of our everyday lives. Current research has still been limited to the language lab in an effort to create controlled environments. However, in this stage HVPT has the potential to become an effective and widely accepted teaching practice.

The normalization of CALL has eliminated many of the practical limitations that previously existed for HVPT. PCs allow instructors and teachers to access the same resources at home that were once available only in language labs. The proliferation of language learning software enables software developers to include HVPT as part of a language learning regimen. Computer and mobile programs such as Rosetta Stone and Duolingo are widely used as independent learning tools or part of a structured curriculum at educational institutions around the world.

The internet has also enabled the diffusion of HVPT tools into mainstream language learning. Internet browser based programs such as *The English Accent Coach* (Thomson 2012), *IPA Trainer* (n.d), and *UCL Vowel Trainer* (1999) have been released over the internet for students and instructors to use worldwide.

CHAPTER 4—EDUCATIONAL GAME DESIGN

While there is no wide interdisciplinary consensus on what features constitute a game, there are common themes in definitions of games amongst educational game design research (Juul, 2003). Dempsey, Haynes, Lucassen, and Casey (2002) define ‘game’ as a set of activities involving one or more players. It has goals, constraints, payoffs, consequences, and is rule-guided and artificial in some respects. Finally, a game involves some aspect of competition, even if that competition is with oneself (p.159). Grendler (1996) states that games consist of, “... rules that describe allowable player moves, game constraints and privileges (such as ways of earning extra turns) and penalties for illegal (non-permissible) actions’’ (p.523). A general overview is best summed up by Webster's dictionary: “A game is a structured form of play, usually undertaken for enjoyment and sometimes used as an educational tool” (n.d.). Whether used for simple enjoyment or with a focused learning goal, games are typically enjoyable activities; games can be physical or mental; played individually or in a group; they included specific goals and methods to achieve the goals through a form of action or “moves/turns”; and games are constrained through rules (Hainey, Connolly, Stansfield, & Boyle, 2011).

All games, even those that may appear to be random, are rule-based. For example, the game *Calvinball* (Watterson, 2005), whose gameplay is in a constant state of flux and chaotic to the outside observer, still follows a core set of rules;

1. The players make up their own rules as they go along (see Figure 1.).
2. Rules cannot be used twice.
3. Any plays made in one game may not be made again in any future games.

Figure 1: A game of Calvinball

Rules provide the structure needed for players to experience the game. Rules can break down into two distinct types: emergence types rules and progression type rules (Juul, 2005). Emergence type rules use a limited number of rules and are used to maximize the number of possible outcomes and play options, such as the game *Calvinball*. Progression type rules, frequently used in adventure games, have all potential outcomes predetermined by the game's creators. Rules provide the framework for games; however, they are not what differentiates an educational game from a drill. While both drills and games have rules, the addition of an aspect of perceived randomness is critical in games (Juul, 2005; Koster, 2013).

A game's aspect of randomness is what differentiates it from a drill and provides a critical part of a players enjoyment. Game designers must find an appropriate balance between the game's governing rules and the amount of randomness in the game to maximize their players enjoyment in completing the game. If a game is too rigid, it becomes boring. If a game is perceived as being completely random, players might will lose a sense of their agency within the game. Koster (2013) discusses the need for players to find new experience within the confines of a game as the main motivator for players to return. Perceived randomness of games helps players perceive that each time they play something new can happen. The random aspect of a game also allows games to adapt to their players and become more challenging. In *Tetris* (Pajitnov, &

Gerasimov, 1984) a shape is randomly selected for the user to place and progressively increases in gameplay difficulty. As the player reaches a new competency threshold they are continually challenged until they successfully pass the next threshold or fail and must try again. Games have the ability to scaffold to the user's current ability in a similar way to scaffolding in educational theory.

Games tap into the concept of play and can be harnessed for educational means. While Piaget (1962) pioneered play research with studies of children, more modern approaches to play research have expanded their focus to all ages of life. Play can impact the mental and social development of the players and acts as an excellent medium for learning. Games can increase motivation through user engagement. Game can improve the internal motivation of L2 learners (Ryan, Rigby, & Przybylski 2006; Przybylski, Rigby, & Ryan, 2010). Games can be focused on a specific topic, stimulate metacognition, and provide a variety of feedback (Malone, 1981; Malone, Lepper, Snow, & Far, 1987; Ke, 2008). These positive traits have led to many researchers studying the “gamification of language learning” (Burgers et al., 2015; Su, & Cheng, 2015)**.** The swell in research on this topic in recent years can be partially attributed to the acceptance of gaming in popular culture. Apps like *Duolingo*, *Memrise* (2009) and others aim to use games and game design concepts to enhance language learning.

Why Games?

Games can provide us with a sense of fun that drills may not necessarily provide. A drill can be turned into a game by the inclusion of scores and ranking performance. This small change helps build the motivation of learners. Depending on the game content, games can satisfy several psychological needs such as competence, autonomy, and relatedness (Przybylski et al., 2010; Blumberg, & Blumberg, 2014). By adding a score to a drill, a learner must now use their own skill to demonstrate their level of competence. They can relate their performance level to that of their peers in a form of competition such as a featuring a high score. Though both a drill and a game can make learners repeat the same task repeatedly, games can make the repetition engaging through randomness, whereas drills do not. Games can use a variety of feedback methods to show learners that the game's outcomes are dependent on their actions (Blumberg, & Blumberg, 2014) and transform players into active learners (Rossano, Roselli, & Calvano, 2017). Games require learners to analyze and access problems that need to be solved to complete their assigned task (Uskov, Bakken, Penumatsa, Heinemann, & Rachakonda, 2017). The results of forcing learners through this process is increased learner autonomy and the reward of solving the problem. The complexity of the task can impact the complexity of the game design: A complex task requires a sufficiently complex game.

Relatedness

Games promote the use of reflective processes such as metacognition. When a user is challenged with a task in a game they must think strategically and weigh their strengths and weaknesses in order to formulate a method to complete the task (Blumberg, & Blumberg, 2014). In some games, it might take a user several hours of forced repetition to complete a task. For example, the *Dark Souls* series action role-playing video game is famous for the unforgiving and repetitive gameplay required to beat the game. But with time, reflection, and practice, users can develop strategic skills along with the operational skills required for specific topics (e.g. city taxation in *SimCity*) (Bogost, 2010). Language learning games can help learners relate the current task target to their current skills through metacognition (Snow, & Far, 1987; Huffaker, & Calvert, 2003; Ke, 2008). By reflecting on how their actions have either failed or succeeded in completing their current task, users must alter their actions must to ensure a positive outcome in subsequent game play. Problem solving skills such as analysis, classification, evaluation, and the formulation of new plans (Koster 2013) can be applied to specific problems or topics (Bogost, 2010). This differs from a drill in that development of strategic competencies stagnates over time due to non-random repetition.

Games can use drill-like repetition to limit the cognitive load on the learner. By recycling operational knowledge and adapting to the learner's current skill level, games can focus the learner's attention on the content of the activity (Koster, 2013). For example, *Duolingo* has a limited number of activity types. However, by replacing the target vocabulary as users improve their lexical knowledge, the same task is completed with new and more challenging input data. This allows education learning tools to limit the cognitive load of the game as users focus on new input instead of learning new rules and systems.

Research suggests that games can effectively objectify concepts and create learner narratives. In a learner's narrative, they must use skills including classifying, quantifying, and analyzing in order to succeed in their game tasks (Koster, 2013). In Tetris, the user must classify the shape of the current piece, analyze the current rows and stacks of blocks, and finally choose where the current piece should fall. Though the acts of classification, quantification, and analysis can be part of a drill activity, games allow learners to create player-based narratives created by the learners' own actions (Koster, 2013). Users can compare the progress of their ability to complete the task or the time it took for them to complete the task for example. It allows for users to claim ownership of their learning through the narrative with themselves as the protagonist of their own learning adventure (Koster, 2013). Learning tasks or learner-to-learner competition in games can be seen as an obstacle that needs to be overcome through the growth of the learners' skills.

Games facilitate learner engagement in a variety of ways. Instead of relying on lectures from an instructor, games rely on visceral, behavioral, and reflective engagement (Blumberg, & Blumberg, 2014). Completing a set of a language learning drill rarely results in celebration. However, games rely on a sense of accomplishment and desire to keep people playing. The pleasure derived from games is vital for the survival of the games, for if a game was not pleasurable, in most cases people would stop playing the game. Games can also appeal to our senses as a manner to increase our engagement. Games such as broken telephone rely on listening while video games such as *Rock Band* (2007) rely on players' sense of rhythm, pitch, and sound duration. Others, such as the 1970s classic *Simon Says* (Helwig, 2012), rely on visual and audio stimuli. The *Rubik's Cube*® (1977) uses the player's visual spatial sense along with touch to align the colors on each side. Games are not required to provoke strong emotions such as happiness, sorrow or anger to be successful, however, a successful game can illuminate aspects of ourselves that we do not fully understand (Koster, 2013). Active reflection is a critical aspect of any game, as a player who does not reflect on their interactions with the game will never improve their performance. The drive to be successful at the game promotes internal motivation, which facilitates successes and leads to more motivation. This positive feedback loop has led to games becoming a multi-billion dollar industry and a popular topic for educational researchers.

Gamification and Serious Games

Though users can use a game designed for enjoyment to learn, for example, city taxation in *SimCity* (Bogost, 2010) or conversation practice in *Second Life*, games can be purpose-built with an educational goal in mind. Serious games are games that use entertainment to focus players towards an educational goal (Zyda, 2005). Serious games have been used to train pilots, teach surgery, simulate combat for soldiers, practice computer coding, along with traditional educational subjects like mathematics, language learning, science and history (Sykes, Reinhardt, & Thorne, 2010). Gamification, as stated by Cerrato, Ferrara Ponticorvo, Sica, Di Ferdinando, & Miglino (2017) is the concept of "using games and game-like activities in a non-leisure settings". In serious games, designers implement concept from entertainment focused games "to increase user interaction and participation" Cerrato, et al., (2017). For language learning, the primary goal of a serious game is to improve some aspect of the user's communicative competence. The target language goal could take the form of a socio-pragmatic exercise in a virtual online world or improving one's reading skills through a flashcard game. With the increase in sophistication of computers and the internet, there has been a renewed interest in how digital games can be used in education (Juul, 2001; Rice, 2007).

CALL Programs

The relationship between computers and games is long and rich in its history. Though various computer games were created in the 1950s to help with academic research and demonstrations of computer performance, *Spacewar!*, the first popular computer game designed for fun, was released in 1962 (Russell, Graetz, Witaenem, 1962). However, due to the restrictions on computing technology at the time, games such as *Spacewar!* were mainly played by other programmers on large mainframe computers (Rutter, 2006). The growth in game development accelerated dramatically after the invention of the microchip and distribution of the PC into the household. PC proliferation led to computer games being released with more frequency and complexity. The gamification of language learning and CALL have progressed in tandem. The development of the PLATO network pioneered many technological advances such as the computer screen, chat programs, touch screens, and curriculum integration. However, widespread adoption of these innovation would take several decades (Dear, 2017). As computers became increasingly powerful and complex so have the CALL programs that run on them, mirroring the software development of non-serious games.

Modern CALL programs like *Duolingo*, *Rosetta Stone*, *Babbel*, *Memrise*, *Lingua.ly* combine features common in both gaming and language learning such as negative feedback, variety of mini-tasks, user generated content, and cross platform compatibility along with the advantages of modern computers including instant feedback, analytics, cloud based personal progress, and online community integration. Modern CALL programs aim to return self-determination to language learning through play.

CALL programs take the internal motivating characteristics of a game and combine them with the ability of computers to deliver immediate feedback (Burgers et al., 2015; Lavolette et al., 2015). If a task is quantifiable, computers can measure the performance of a task: For example, a question by question performance report of a multiple-choice quiz like the English IPA Ear Trainer. Score is a common tool used in a variety of CALL programs and many games monitor and award user accuracy and speed with higher scores or other incentives. The instantaneous nature of computer feedback promotes learner experimentation and helps learners immediately identify areas they need to improve to complete the task at hand.

One area where internet based computer programs hold a significant advantage over a textbook-based curriculum is the speed and ease of updates. Through the internet, applications can be quickly updated with new features and content. Developers can react to user feedback to make improvements to the program to better fit the needs of the users in real-time. These changes can be made incrementally allowing a computer program to grow in complexity over time. This allows learners to ensure that they are using the most up to date content and materials and improve interaction with the training regimen as possible task issues are corrected over time.

CALL programs can also take advantage of saving learner progress in a variety of competencies to make sure that when a task is commenced learners are exposed to material that is at their learning level. Computers games can save a learners' progress and allow them to restart their language learning training regimen at the same place they previously stopped. The convenience of this feature cannot be overstated. Learners do not need to be reassessed or find materials if they paused learning for any period of time: for example, resuming an activity, or returning to learning after a long absence. The ability to save progress allows learners to jump back into their tasks or quickly access their current level based on previous achievements.

The widespread use of computers and the internet allows CALL games to reach a larger audience than any other medium in human history. As stated by Bax (2003) and Warschauer et al. (1998), we have entered the integrative period of computer use in most societies. As of 2017, more than 3.7 billion people were connected through the internet (Kemp, 2017). Desktop and browser-based CALL programs take advantage of the large potential user base and can be within fingertips of billions of possible language learners.

Current Gap

Though programs like *Duolingo*, *Rosetta Stone*, and *Lingua.ly* focus on the acquisition of several different and distinct aspects of language (vocabulary, grammar, listening, speaking, reading, and writing), they do not specialize in any one area. HVPT training software is one area that is highly specialized and generally not included in larger language programs. When programs do include some kind of phonological perception tool, there is little to no information on the design of the application. Without this information, researchers and instructors cannot assess the applications' strengths, weaknesses, nor the pedagogical foundations of the applications' design. Currently the program that best applies the concept of HVPT is the *English Accent Coach* (EAC) created by Ron Thomson in 2011 at Brock University.

Since 2011, the EAC has gone through various versions and updates that have allowed the EAC to grow in functionally as a tool. The current version allows users to choose English vowels, consonants, or select specific sounds to practice. Users can also choose the lengths of their sessions, and there appears to be a large variety of audio sources. However, due the lack of information of stimuli and program design, the total amount is not known. Though the program takes advantage of the customizability of CALL programs, some design choices lead to limitations of the EAC.

Limited use of context words, distractor presentation, and negative feedback design all limit the effectiveness of the EAC as a learning tool. The presentation of distractors might be overwhelming to some students. The EAC include icons for all the sounds chosen for the training session and the learner must select from every option for every question. This could lead to a high cognitive load as learners have to remember the characteristics of every sound for every question. Furthermore, there is no context given to audio stimuli when presented but rather after the learners have answered the questions. Learners could use the context as a tool to help them acquire the sound, but by presenting the context after the learner has answered, the impact of the context word is severely curtailed. Learners are only allowed a single attempt at each question, limiting the impact of negative feedback. If a learner chooses a distractor, the EAC moves onto the next question and states that their answer was incorrect. This choice limits the effectiveness of negative feedback and fails to take advantage of task repetition. Learners are not forced to reflect on their L2 perception and, by simply moving on, a chance to engage the learners in a brief moment of metacognition is lost. The EAC has been demonstrated as a useful tool that learners have demonstrated gains in vowel acquisition (Thomson, 2011), but is limited by aspects of game design that have not been fully incorporated or taken into account.

Project Goals

The current goal for this project is to create the English IPA Ear Trainer. This program will use the concept of HVPT to create a series of training regimens to help L2 learners perceive the consonants of North American English (NAE). To address some of the limitations of the EAC, the English IPA Ear Trainer will display only three distractors in the form of a multiple-choice quiz along with the correct answer. For the correct answer and distractors, the IPA symbol of the consonant and a context word will be provided. The syllable containing the IPA consonant represented will be bolded and highlighted to raise learners' awareness of the context of the sounds' use. The English IPA Ear Trainer will also not require the user to create an account to gain full access to the program to ensure the open accessibility of the program. The English IPA Ear Trainer will also take advantage of the accessibility of CALL programs by being a browser-based program readable by desktops and mobile devices along with backwards and forwards compatibility. It is the hope that the initial design experience of this tool will create a platform to design future HVPT tools to target the IPA as a whole and more language specific regimens.

CHAPTER 5— ENGLISH IPA EAR TRAINER DESIGN METHODOLOGY

Design Rationale

The English IPA Ear Trainer is designed as an easy to access serious game based on SLA pedagogy and theory. In-program aspects such as accessibility and feedback, audio stimuli, distractors, score, and accessibility were designed with the goal of creating a tool that learners are able to access and use anywhere.

Accessibility

The English IPA Ear Trainer was created using a combination of JavaScript, HTML5, and Flash. These coding languages were chosen to enable cross platform compatibility between computers and smartphones along with forward and backwards browser compatibility. By incorporating Adobe Flash, browsers from older computers running Windows XP forwards, Mac OS 10.7 or Linux 5.6 can read the English IPA Ear Trainer. Many of these older computers are not compatible with HTML5 and would not be able to access this tool without this backwards compatibility. The use of HTML5 enables the English IPA Ear Trainer to be compatible with modern computer operating systems and internet browsers. HTML5 is the latest and current major version of the HTML standard (Juntunen, Jalonen, & Luukkaine, 2013). The use of Javascript enables online mobile compatibility of the English IPA Ear Trainer. Many mobile based browsers currently lack Flash and HTML compatibility. By using Javascript, mobile users can view the English IPA Ear Trainer in an optimal manner on their mobile browser.

A critical aspect of this serious game design is free access for both teachers and students. Providing a transparent design methodology will enable students, teachers, and future software developers to understand the limitations and benefits of HVPT and incorporate HVPT into an individual's learning regime, a class syllabus, or an integrated portion of more complex language learning computer programs.

Development Timeline

Creating a research driven learning tool takes a significant amount of time and effort. This section will detail the development process in an effort to help other teaching instructor streamline the process for creating their own computer program and serious game. In general, the development of the English IPA Ear Trainer can be broken into several distinct phases: the initial game idea, idea refinement and concept development, developing a proof of concept, game design, and implementation. It is impossible to create a quantitative timeline applicable to all games, but this timeline is meant to detail the process required to create a browser-assessable free serious game for instructors with limited to no coding experience.

Stage	Features	Timeline
Game Idea	• Conceptualization of program • Establish overall characteristics	Spring 2014 - Spring 2016
Idea Refinement and Concept Development	• Define specific goals for the program • Development of key elements such as technology needed, user interactions, settings, and game mechanics	Summer 2016 – Fall 2017
Developing a proof of concept	• Test and select building technologies required. • Practical assessment of what aspects of the program are possible. • Develop a basic program demonstrating key features of the program.	Fall 2017 - April 2018
Game Design	• Utilize building technologies to finish program construction.	April 2018 - June 2018
Implementation	• Upload program to the selected hosting service and open access to the program.	July 2018

Table 1: Software Development Timeline

Game Idea

The initial idea for the English IPA Ear Trainer was developed after my own experiences learning the international phonetic alphabet (IPA) at the University of Victoria. Required to learn all consonants, I searched for an ear training tool that focused on quizzing users on the perception of various IPA symbols. Having found little in the way of digital aids, I was struck with the idea to make one. Serious game ideas can come from actively searching for a need or by realization through the creator's experience. The English IPA Ear Trainer's creation stems from my own need for a tool to help my studies of the IPA. Research into free language learning software led me to realize there is a dearth of research-based tools with data demonstrating their effectiveness. Therefore, my serious game must be rooted in proven science. It was during this period that I discovered HVPT as a training method and researched the work of modern tools using this method. I chose to create a browser-based free tool to allow learners across a variety of platforms to access the English IPA Ear Trainer. The initial idea was conceptualized in the Spring of 2014 with the final research into HVPT completed in Spring 2016 (see Table 1).

Idea Refinement and Concept Development

In this stage of the process, detailed aspects on the game's mechanics, technology and interactions are defined. Upon completing background research into phonological training programs, specific goals for the program were developed

1. The program would be accessible on smart devices and PCs.
2. The program would take the form of a serious quiz game.
3. The program would be free.
4. The program will use HVPT as the training method.
5. The program can be replayed as many times as possible.

The goals help define the aforementioned game mechanics, the technology needed to use the program, and how learners will interact with the program. To better conceptualize how learners would interact with the program, I developed a flowchart to visualize the workflow of the program and the building blocks required to complete the program (see Figure 2). Overall this stage commenced in Summer of 2016 and was completed in the Fall of 2017 (see Table 1).

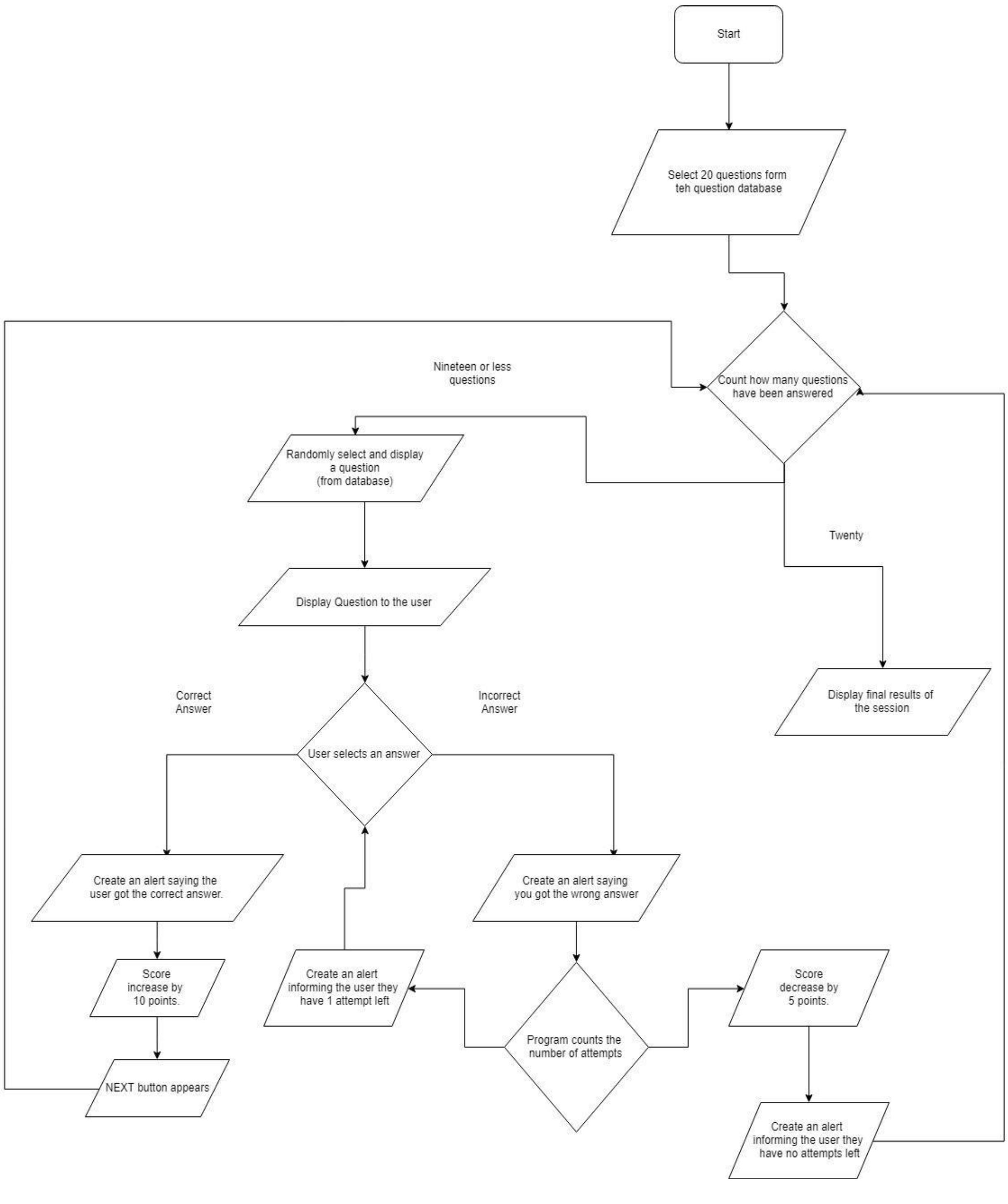

Figure 2: Program Operation Flow Chart.

Developing the Proof Concept

During this stage, developers test the feasibility of a design concept. This can be done through experimenting with technologies (development software, hosting services, media software) used to create and implement the program. It is also common to develop a pilot program to test the features of the final program at this time. This stage of the process proved the most challenging. Without prior coding experience, I initially learned Javascript and HTML to aid me in writing my program. Learning to code takes a significant amount of time and may not be practical for the majority of language instructors who wish to create a simple serious game or learning tool for their students. A far more time efficient method is to use pre-existing technologies to create the program. There are several options available such as the *MIT APP Designer*, *Adobe Captivate*, and *iSpring Quizmaker* to make program designer tools accessible to language learning instructors that cannot code. For the English IPA Ear Trainer, *iSpring Quizmaker 9.0* was the main program used.

In order to make the English IPA Ear Trainer available online, I had to find a hosting service for my software. There are limited options to host software without having to pay monthly fees. Some websites like GitHub have limited hosting services, however, if you wish to maintain full ownership of all aspects of your program, you will have to pay for a monthly hosting service. If an instructor does not mind losing some control over access to their language learning game, *itch.io* is a free game based website and hosting service that is available. Once both a hosting and software development method have been chosen, a small proof of the concept was created in the form a simple pilot program. In this stage I devoted a significant amount of my working day to developing my proof of concept. Overall, this stage lasted from Fall 2017 until April of 2018, with this project taking the majority of my time from January 2018 onwards (see Table 1).

Game Design

During this phase the English IPA Ear Trainer was fully developed. Audio stimuli were recorded, tested, and edited in various sessions. The input questions were generated along with the distractor items and scoring system. Several minor changes to scoring and distractor choices were undertaken after testing the limitations of the iSpring Quizmaker. This stage was the most labor intensive taking several weeks of full time commitment to complete. Layout and final tests of the program were completed in June 2018 (see Table 1).

Implementation

During the developing proof of concept phase the hosting service Itch.io was selected for the program. It was selected due to the free and openly available platform Itch.io provides. The final version of the program was uploaded to itch.io hosting service in June 2018 (see Table 1).

User Interaction

When the user starts the training session in their internet browser, a question will be displayed to the user (see Figure 2). For each question prompt, the program will randomly select a question with an audio file of an English consonant (m,n, ŋ,p,t,t͡ʃ, k,b,d, d͡ʒ,g,f, θ, s, ʃ, v, ð, z, ʒ, h, l, r, j, w) (Ladefoged & Disner, 2012) (see Figure 3).

Consonant phonemes

	Labial	Dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Nasal	m		n			ŋ	
Stop	p b		t d			k g	
Affricate				tʃ dʒ			
Fricative	f v	θ ð	s z	ʃ ʒ			h
Approximant			l	ɹ	j	(ʍ) w	

Figure 3: Consonant Phonemes of North American English.

The learner will press the play icon and hear an audio recording of an English consonant in /CV/ structure and select the corresponding consonant IPA transcription from four options (Figure 4).

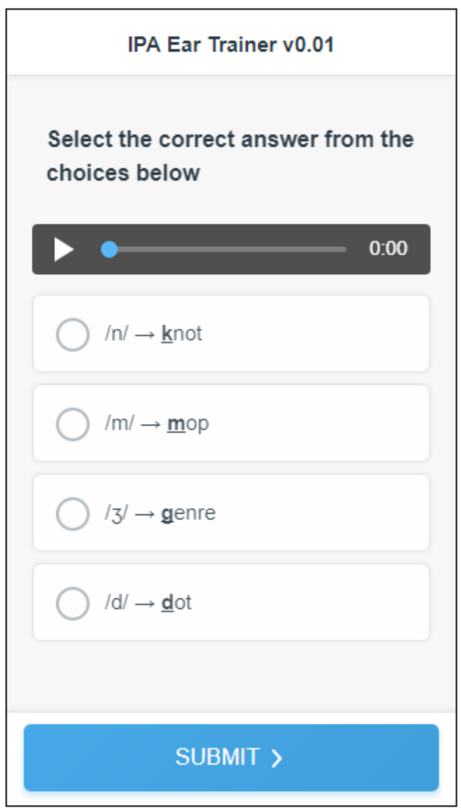

Figure 4: Question Example (Mobile)

The learner can also replay the audio file as many times as they like to receive more input to help them perceive the consonant in the audio file. Once the learner is ready to answer, they can select from four options (one correct answer and three distractors). If the learner chooses the correct answer, an alert notifies the learner that the choice is correct and allows them to move the next question. At this point, the application will randomly select another question from within the question bank and repeat the question process. If the learner chooses an incorrect answer, an alert notifies them that they have chosen incorrectly and that they will need to try again. Both forms of feedback are provided instantaneously and allow the learner to repeat the identification task several times to practice identifying the correct audio stimuli. Using instantaneous feedback maximizes the benefits of the English IPA Ear Trainer as a computer based tool (Blumberg, & Blumberg, 2014). By limiting their progress, the English IPA Ear Trainer engages the learner by showing the connection between the learners' outcomes and actions within the training sessions. This form of negative feedback also promotes autonomy in learners as they must rely on their own skills to proceed (Przybylski, et al., 2010).

Audio Stimuli

The audio stimuli used in the English IPA Ear Trainer was carefully designed to maximize the learners' ability to perceive the target English sound. Audio stimuli is provided to the learner in /CV/ structure. This structure was chosen due to the simplicity of the CV syllable and the widespread use of the CV syllable across the majority of the world's languages (Ladefoged & Disner, 2012). The /CV/ syllable is the most basic syllable containing a consonant. The simplicity of this syllable structure and lack of distractor consonants within the syllable, which would be prevalent in other syllables structures (e.g. /CVC/, /CVCC/ etc.), allow the learner to focus on the target language. /VC/ structure also provides the same level of potential focus as /CV/ structure.

To focus the learners' attention on the target consonants, all audio stimuli in the English IPA Ear Trainer uses the same vowel. Limiting the sound files to /CV/ structure allows for consonant exposure with the least amount of phonological complexity and increased phonological saliency of the target consonant. For the stimuli, the low front unrounded cardinal vowel /a/ will be used in all audio samples due to its prevalence in a large variety of the world's languages (Ladefoged, & Disner, 2012). The majority of learners from most language backgrounds will be able to perceive the /a/ vowel and will be able to use their perception tool to analyze the differences between the new consonant and a previously categorized sound.

For this tool, four native speakers of American English were recorded producing the full complement of North American English (NAE) (m,n, ŋ,p,t,t͡ʃ, k,b,d, d͡ʒ,g,f, θ, s, ʃ, v, ð, z, ʒ, h, l, r, j, w), for a total of 96 audio recordings. To ensure the comprehensibility of the recordings, 40 of the 96 recordings were randomly selected for an identification task and played back to the four native speakers (Ladefoged, & Disner, 2012). The four native speakers scored 116/120 for a total accuracy of 95%. With a comprehension rate of 95%, it can be stated that the item sounds accurately represent the specific English consonants that are targeted in the IPA Ear Trainer tool.

HVPT connection syllable structure Variation

Two critical aspects of audio stimuli define the IPA Ear Trainer as an HVPT tool. HVPT requires variation in the source of stimuli. By using several different speakers as sources of audio stimuli, a variety of models for each sound are used to perceive and create the parameters for categorization. This process is accelerated by the second cardinal feature of HVPT: simple and repeated syllable structure. By repeating the same basic syllable structure, learners can focus on the phonological variation by lowering the cognitive load.

Question Design

For each audio stimuli, ten questions were created for a total question bank of 960 questions. Each question contains three distractor items along with the correct answer. In order to provide varying levels of difficulty to the learner, different types of distractor items were selected (see Table 2).

Consonant: /h/	1-2 Questions	Example Distractors: m, n, ŋ, p, t, t͡ʃ, k, b, d, d͡ʒ, g, l, r, j, w
Consonant: /h/	6-8 Questions	Example Distractor: m, n, ŋ, p, t, t͡ʃ, k, b, d, d͡ʒ, g, l, r, j, w
Consonant: /b/	1-2 Questions	Example Distractors m, n, ŋ, t͡ʃ, d͡ʒ, f, θ, s, ʃ, v, ð, z, ʒ, h, l, r, j, w
Consonant: /b/	3 Questions	Example Distractors: (t, d, k, g) + m, n, ŋ, t͡ʃ, d͡ʒ, f, θ, s, ʃ, v, ð, z, ʒ, h, l, r, j, w
Consonant: /b/	3 Questions	Example Distractors: (p) + m, n, ŋ, t͡ʃ, d͡ʒ, f, θ, s, ʃ, v, ð, z, ʒ, h, l, r, j, w

Table 2: Various question difficulty levels for /C/ & /C/

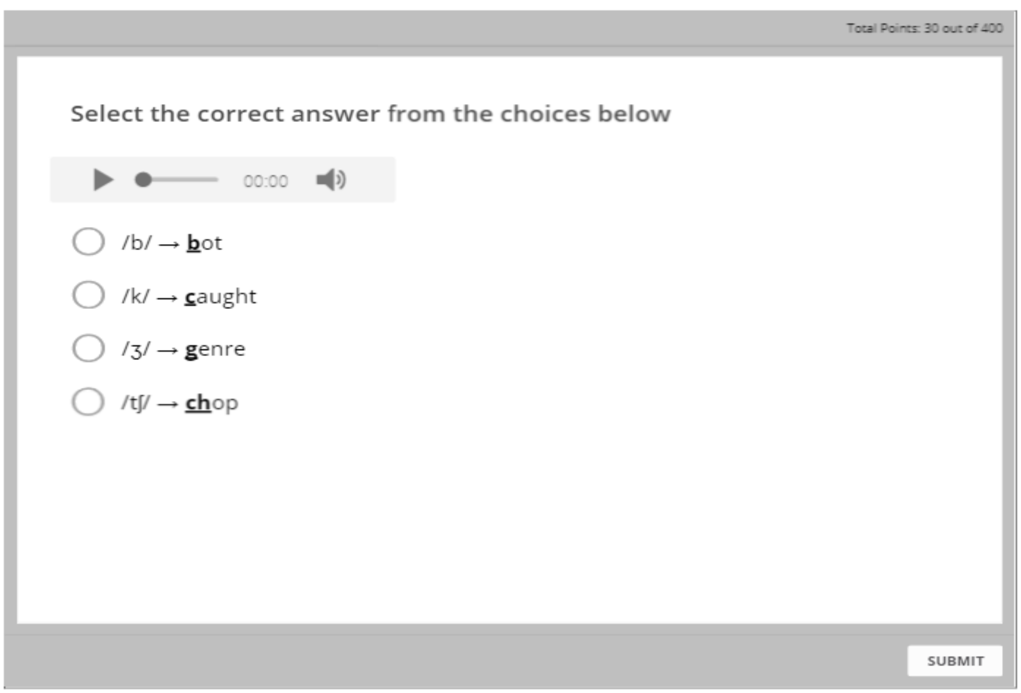

Figure 5: Question with distractors differ in manner of production (correct answer: /b/)

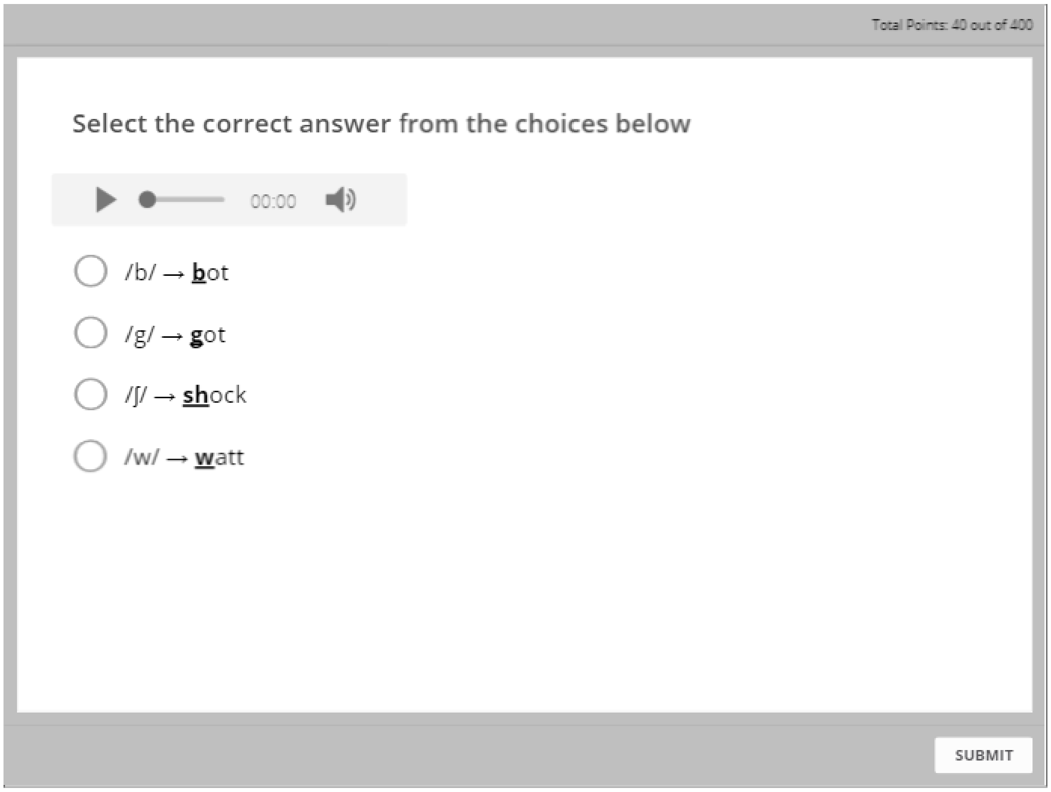

Figure 6: Question with distractor that shares a manner of production (correct answer)

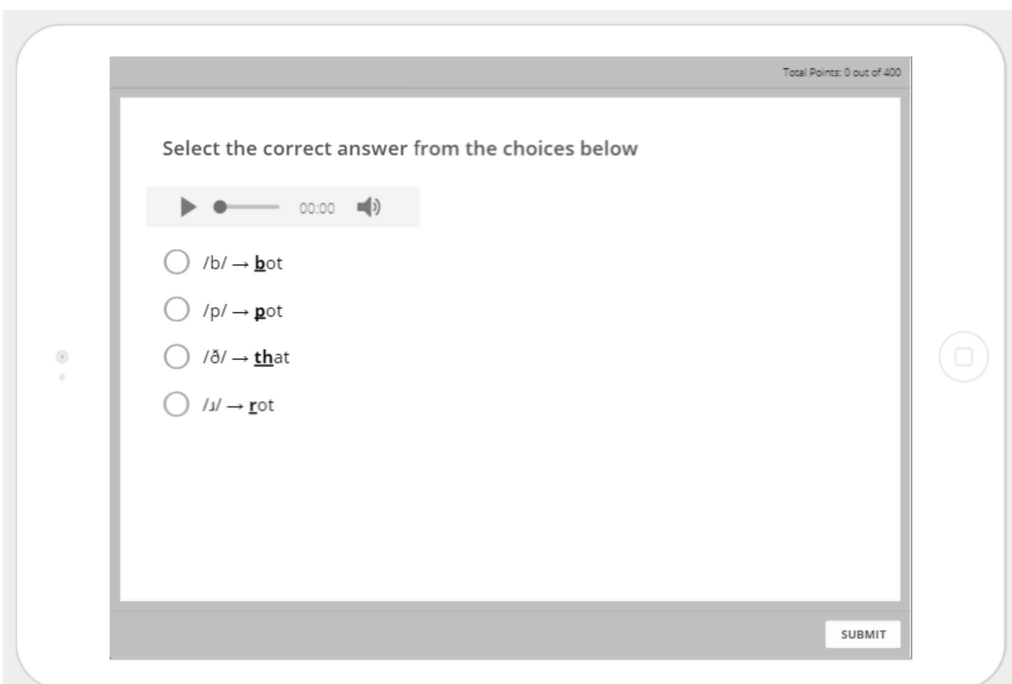

Figure 7: Question with minimal pair as a distractor (correct answer: /b/)

For each audio stimuli, two questions contain distractors from different manners of production (see Table 2 and Figure 5). In these questions, the correct answer is the most salient to the user of all the questions created for each consonant. In eight of the ten questions, one distractor item shares the same manner of production with the correct answer (see Table 2 and Figure 6). This will cause learners to focus on the manner, production, and place of articulation. If the consonant has a voiced or voiceless pair, for example /b/ and /p/, three of the eight manner questions will include the pair sound to draw learns attention to the manner and place of articulation along the voice or voiceless characteristics of the stimuli (see Table 2 and Figure 7). Due to the lack of affricates outside of the post-alveolar place of pronunciation, all the affricate questions contain the voiced or voiceless counterpart (/dʒ/ and /tʃ/) in the set of manner distractors. (/dʒ/ and /tʃ/).

For each question, a context word is provided for the answer and distractor. Including a contextual cue for each consonant allows the user the better conceptualize how the sound is represented orthographically in English and contextualize the sound in a real word. Exemplar words were chosen based on their syllabic uniformity, simplicity, and vowel usage. Exemplar words were chosen based on their inclusion of the low front unrounded cardinal vowel /a/ in a simple syllabic structure. Due to the lack of /CV/structure words in English using the vowel /a/, the majority of the exemplar words chosen follow /CVC/ structure. To mirror the /CV/ structure of the stimuli the exemplar word contained the target consonant in word initial position with the /a/ vowel (eg. pot.). The velar nasal /ŋ/ is not found in word initial position in English, therefore, the exemplar word chosen for /ŋ/ contains the consonant in its natural environment.

Feedback

Feedback within the tool will focus on two main features: speed to feedback delivery (instant feedback) and the results of input (negative feedback). After the audio input has been played, the user will select the appropriate IPA symbol from a group of potential answer icons. If the user selects the correct answer, they will move to the next questions. If they choose a distractor, they will be immediately informed of their incorrect choice through an alert (see Figure 6).

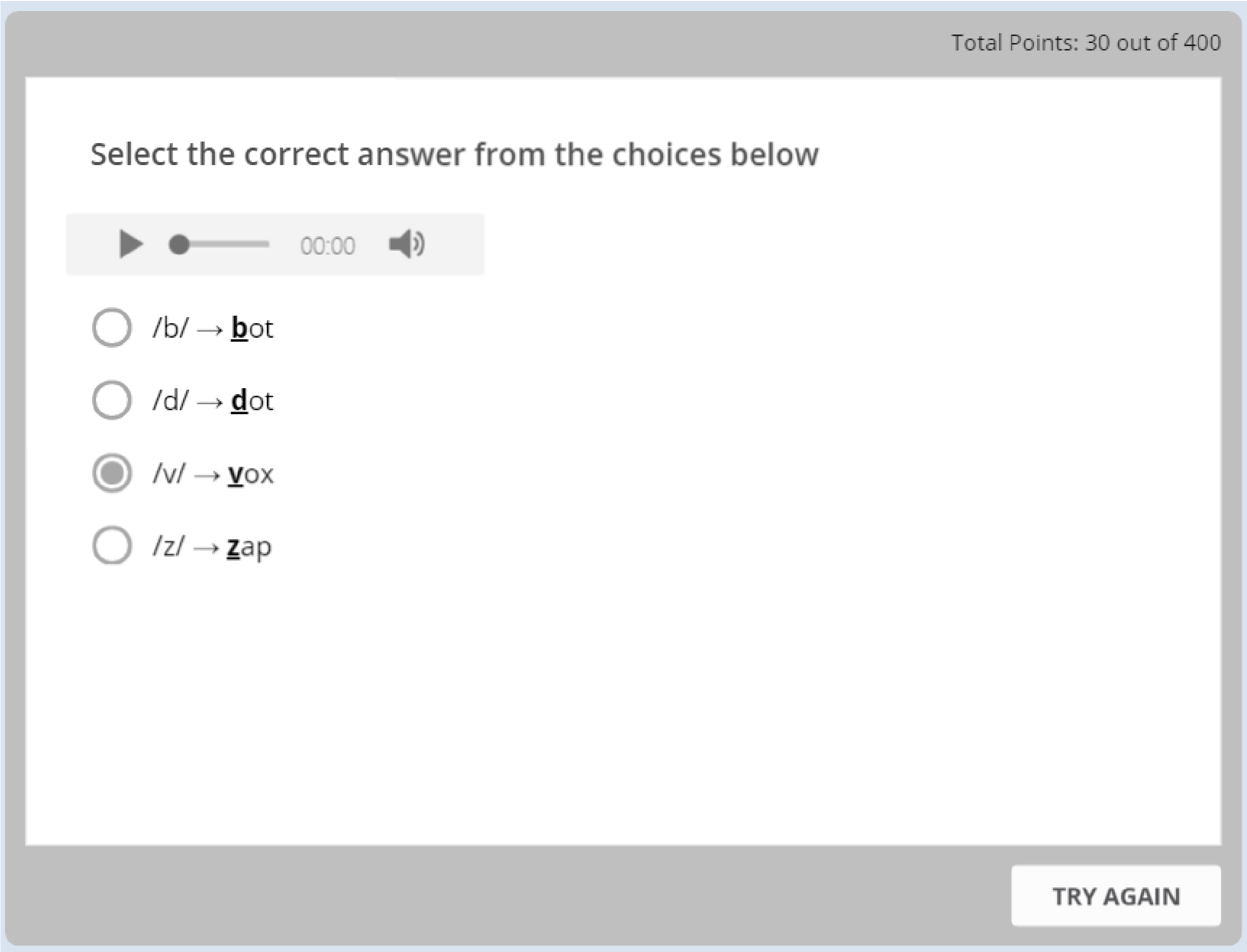

Figure 8: Incorrect answer to a question within a training session

Based on the research of Burgers, Eden, van Engelenburg, and Buningh, (2015), this form of negative feedback will motivate the user to immediately correct their perceptual error. This will be beneficial for simulating short term tool use behavior. The immediacy of the feedback is also critical as immediate error correction in low cognitive load tasks can accelerate the speed of acquisition (Hattie & Timperley, 2007). The instantaneous nature of this feedback is critical to improve learners' perception as any delay in feedback will negatively affect the effectiveness of user intake (Lavolette, Polio, Kahng, 2015), and immediate feedback is more likely to activate a learner's monitor to the target language (Sanz, 2004; Lavolette, Polio, & Kahng, 2015). Repeated exposure to input and practice will allow users' phonological perception awareness to become routine and eventually automatic (Dekeyser, 2007). This move from declarative knowledge to implicit knowledge is a critical step in improving users' final level of perception.

Motivation

Through the use of negative feedback and end of session scores, the IPA trainer aims to foster internal motivation in learners. As an assessment tool, the end of session score has yet to be analyzed for their predictive validity, however, the score acts as a benchmark for the learner to improve upon in their next session. The end of session score focuses the learners' motivation on a specific goal.

If the learner selects the wrong answer to the audio stimuli, they will be unable move on to the next question. The need to complete the session to receive a score motivates learners to find the solution to the question and move on. In order to maximize learners' opportunity to answer the questions correctly, learners have unlimited attempts to listen to the audio stimuli before answering the question. This manner of negative feedback promotes learner autonomy by encouraging the learner realize their impact on their learning and can motivate them learn more.

Variation in stimuli also limits the effects of task habitation. If learners are constantly exposed to the same stimuli either in manner, articulation, or source audio, the arousal they receive from the stimuli decreases over time (Koster, 2013; Blumberg, & Blumberg, 2014). By adding variety to the training regimen, learners will regularly be exposed to new stimuli. In this way the base concept of varying audio sources in HVPT regimens helps maintain learner motivation.

CHAPTER 6— LIMITATIONS AND FUTURE CHANGES

Tool Limitations

There are many potential benefits of the English IPA Ear Trainer tool for English language learners, however, the scope of the current tool, level of customization, and off-line accessibility are limited. These limitations are caused by the programing skill of the tool's creator and the design structure of HVPT. Currently, students have little control of the audio stimuli selection and the syllabic structure of stimuli items. Because this tool focuses on the phonemic level of English phonology, a variety of allophones used in English were omitted from the final design (e.g. dark l). Students must also have an active internet connection for the tool to load each new question. This required connection to the internet could limit the use of the tool in areas where internet connectivity is a finite or unpredictable resource.

Phonological Restrictions

In this version of the tool, the IPA trainer will be limited to American Standard English. Containing only the 24 consonant sounds of NAE, the tool does not contain English phonemes from other dialects of English (e.g. Received Pronunciation, Australian English, Standard Singapore English, etc.), vowels, or sounds outside of English. Therefore, the use of this tool is limited to American Standard English L2 acquisition.

This tool also focuses on the phonemic level of a learners' phonological foundation. Allophones of a phoneme such as alveolar flap and "dark" or velar /l/, which are common in English, are not currently covered in the English IPA Ear Trainer. Furthermore, by representing all sounds in /CV/ structure, some sounds are presented in environments that do not naturally exist in English. The velar nasal /ŋ/ is found in word final position and syllable coda (eg.

singing) in English. By choosing to follow the guidelines of HVPT audio stimuli, /ŋ/ will be presented in an unauthentic environment.

Though the use of four different native speakers to create recordings allows for a sufficient level of variation in audio stimuli for current purposes, it is still a limited pool of sources. When interacting in the real world, we speak with and listen to a large variation of people. The variation currently provided does not reflect the level of variation encountered in the real world and thus limits the ecological validity of the audio stimuli.

Customization

In other tools, such as the English Accent Coach, it is possible to select specific NAE sounds for perception training. Due to the current programming limitations of the IPA Ear Trainer, learners are able to select manners of production to study through premade training regimens but cannot build their own regimen. Adding this feature would greatly improve the individualization of the IPA Ear Trainer and increase the possible uses of this tool.

Browser Dependency

Finally, due to this trainer being a browser-based tool, the preferred mode of interaction will be through an internet browser such as Google Chrome, Firefox, Safari, or Internet Explorer. Though a JavaScript based program is accessible as a native code for mobile operating systems, an app based version with all sounds installed on the mobile device would increase the accessibility of this tool.

Future Changes

In the future, several improvements can be made to the English IPA Ear Trainer to enhance the tools accessibility, depth of content, and customization. These improvements are based on the current limitations of the tool and the experiences gained in the tools' construction.

Accessibility & Customization

In order to improve accessibility to the English IPA Ear Trainer, a standalone downloadable app would allow smart device users to use the tool offline. This version of the tool would have several distinct advantages. By downloading the tool directly to the users' smart device, learners could access the tool in any location. This could help promote the use of the tool in places where internet accessibility is not a guarantee such as a subway commute, a rural area, or an underdeveloped area. Downloading the tool will also eliminate a learners' monthly data consumption plan if the learners are using the tool over a wireless network. Mobile smart devices have become a staple of the day to day lives of people across the world with 4.92 billion global mobile users in 2017 (Kemp, 2017). Creating a version of the tool specifically designed for smartphones and tablets will grant access to the largest growing section of the world's digital savvy population.

Following in the footsteps of the English Accent Coach, the ability to select specific sounds into a training list would be very beneficial. By creating training lists, the learners could focus on the specific sounds that they struggle to categorize and create lists to compare a pre-existing sound to a new one. If the content of the trainer were expanded to include sounds from outside English, the creation of specific lists of sounds could be used as a training template for a variety of languages.

Content

Expanding the variety and depth of content in the English IPA Ear Trainer will allow the tool to add more variation, context, and scope for L2 learners. Recording more L1 English speakers will increase the variation of input. Adding vowels would cover the total L2 phonological repertoire. Adding more exemplar words for each sound would provide the learners with more examples of the target sound in context. Adding phonological categories outside of English would allow the tool to target multiple languages, dramatically expanding the scope of the tool. These changes would enable the English IPA Ear Trainer to grow into a multilingual tool with varied contextual references and a full repertoire of all phonemic categories.

Variation

Increasing the total number of native speaker sources will strengthen the English IPA Ear Trainers' input variation and authenticity. By increasing the variation of audio stimuli sources, learners will be exposed to a more detailed representation of North American English's (NAE) phonological palette. Learners would be exposed to the same target stimuli in breathy and creaky voices, different pitches, and level of rhoticity based on the variation in each individuals' voice. Improving the variation in exposure will more adequately represent the learners' possible interactions in English in the real world. When interacting with speakers of English, an L2 learner will be exposed to distinct characteristics of each individuals' pronunciation.

Vowel Inclusion

By only including the low front unrounded cardinal vowel /a/, the English IPA Ear Trainer restricts the ability of the tool to improve learners' perception of vowels. Recording more audio stimuli in /CV/ structure using the entire vowel and diphthong repertoire would allow learners to practice both the vowel and consonants. The customization of training sets would allow learners to choose the specific vowels they would like to study. The accompaniment of training customizability and the inclusion of vowels will increase the personalization of the tool and will make the English IPA Ear Trainer into a more complete tool.

Cross Linguistic Expansion

The current version of the tool was designed to fulfill the requirements of the Master of Arts degree in Teaching English as a Second Language. However, the principles applied in the creation of this language specific tool could be expanded to any language. Recording the remainder of IPA sounds would allow users to be exposed to input from any language. Paired with the ability to create a custom practice list, learners could choose preset lists for specific languages or create lists focusing on L2 phonological characteristics of their choice. Cross-linguistics expansion will allow the tool to grow beyond an L2 perceptual tool and could be used in speech pathology, language revitalization, and phonetics training.

Increasing the variety and target language of context words would allow learners to better understand the orthographic context of the sound’s representation in the language. Currently, the same context word exemplar is used in the answer/distractor section of the question each time a consonant is represented. Providing the variety of context word exemplars for each consonant and vowel in several languages would allow for better contextualization of the target sounds. By allowing learners to select the language of the context words they can create a more authentic learning experience when using the tool allowing the English IPA Ear Trainer to evolve into the L2 IPA Ear Trainer. Of these changes, the increase of content by adding vowels and more recordings would greatly improve the usefulness of this language learning game. As active learners constantly improve their skills with new learning opportunities, I hope that this project will continue to upgrade over time and grow into a more complete learning tool to help any L2 learner become a better listener.

REFERENCES

Alexander, C. (2007). Language labs: An overview of the trends. *Teaching English with Technology*, *7*(3).

Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. *Recent research in second language phonetics/phonology: Perception and production*, *231*

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese/r/and English/l/and/r. *Journal of Phonetics*, *32*(2), 233-250.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological bulletin*, *128*(4), 612.

Bax, S. (2003). CALL—past, present and future. *System*, 31(1), 13-28.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of experimental psychology: human perception and performance*, *14*(3), 345.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of phonetics*, *20*(3), 305-330.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words*, *167*(224), 233-277

Best, C. T. (1995). A direct realist view of cross-language speech perception. In. W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research (pp. 171–204).

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775-794.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception:

Commonalities and complementarities. *Language experience in second language speech learning: In honor of James Emil Flege*, *1334*, 1-47.

Boersma, P., Weenink, D. (n.d). Praat: doing phonetics by computer. Retrieved from: http://www.fon.hum.uva.nl/praat/

Bogost, I. (2010). *Persuasive games: The expressive power of videogames*. Mit Press.

Blumberg, F. C., & Blumberg, F. (Eds.). (2014). *Learning by playing: Video gaming in education*. Oxford University Press, USA.

Burgers, C., Eden, A., van Engelenburg, M. D., & Buningh, S. (2015). How feedback boosts motivation and play in a brain-training game. *Computers in Human Behavior*, *48*, 94-103.

Butler-Pascoe, M. E. (2011). The history of CALL: The intertwining paths of technology and second/foreign language teaching. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, *1*(1), 16-32.

Cerrato, A., Ferrara, F., Ponticorvo, M., Sica, L. S., Di Ferdinando, A., & Miglino, O. (2017, June). DILIGO Assessment Tool: A Smart and Gamified Approach for Preschool Children Assessment. In *International Conference on Smart Education and Smart E-Learning* (pp. 235-244). Springer, Cham.

Chomsky, N. (1967). Recent contributions to the theory of innate ideas. *Synthese*, *17*(1), 2-11.

Clark, E. V. (2009). *First language acquisition*. Cambridge University Press.

Dear, B. (2017). *The Friendly Orange Glow: The Untold Story of the PLATO System and the Dawn of Cyberculture*. Pantheon.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant behavior and Development*, *9*(2), 133-150.

DeKeyser, R. (2007). Skill acquisition theory. *Theories in second language acquisition: An introduction*, *97113*.

Dempsey, J. V., Haynes, L. L., Lucassen, B. A., & Casey, M. S. (2002). Forty simple computer games and what they could mean to educators. *Simulation & Gaming*, *33*(2), 157-168.

Duolingo [Computer software]. (2011). Retrieved from https://www.rosettastone.com/lp/about/

Kemp, S. (2017) *Digital in 2017: Global Overview,* Retrieved from: https://wearesocial.com/special-reports/digital-in-2017-global-overview

Ellis, N. (2012). Frequency-based accounts of second language acquisition. *The Routledge handbook of second language acquisition*, 193-210.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303-306.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [rl] distinction by young infants. *Perception & Psychophysics*, *18*(5), 341-347.

Flege, JE (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience* , 233-277.

Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English/I/and/l/accurately. *Language and Speech*, *38*(1), 25-55.

Flege, J. E. (2005). Origins and development of the Speech Learning Model. *Retrieved December*, *13*, 2005.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R”. *Neuropsychologia*, *9*(3), 317-323.

Goodman, J. C., Walley, A. C., & Michela, V. L. (1996). The Development of Speech Perception.
Psyccritiques, *41*(1), 47-48.

Gredler, M. E. (1996). 17. Educational games and simulations: A technology in search of a (research) paradigm. *Technology*, *39*, 521-540.

Greenfield, E. (1993). Foreign language products: Increasing communication in a global community. *The Journal (Technological Horizons In Education)*, 21(2), 12.

Hart, R. (1981). Language study and the PLATO system. *Studies in Language Learning*, 3(1), 1-24.Chicago

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.

Hainey, T., Connolly, T., Stansfield, M., & Boyle, L. (2011). The use of computer games in education: A review of the literature. In *Handbook of research on improving learning and motivation through educational games: Multidisciplinary approaches* (pp. 29-50). IGI Global.

Helwig, D. (2012). *Simon says*. Ottawa: Oberon Press.

Huffaker, D. A., & Calvert, S. L. (2003). The new science of learning: Active learning, metacognition, and transfer of knowledge in e-learning applications. *Journal of Educational Computing Research*, 29(3), 325-334.

IPA Trainer. (n.d) Retrieved from: http://www.ipatrainer.com/user/site/?language=

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267-3278.

James, C. (1980). Contrastive Analysis.

Juntunen, A., Jalonen, E., & Luukkainen, S. (2013). HTML 5 in Mobile Devices--Drivers and Restraints. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp. 1053-1062). IEEE.

Juul, J. (2001). The repeatedly lost art of studying games. [Review of the book The Study of Games]. Game Studies,1(1). Retrieved September 5, 2018, from http://www.gamestudies.org/0101/juul-review/

Juul, J. (2003). The game, the player, the world: Looking for a heart of gameness. In Copier, M., & Raessens, J. (Eds.), *Proceedings of Level-Up: Digital games research Conference* (pp. 30–45). University of Utrecht.

Juul, J. (2005). *Half-real: Video games between real rules and fictional worlds*. Cambridge, MA: MIT Press.

Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay?. *Computers & education*, 51(4), 1609-1620.

Keele, S. W., & Posner, M. I. (1968). Processing of visual feedback in rapid movements. *Journal of experimental psychology*, *77*(1), 155.

Kondaurova, M. V., & Francis, A. L. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: comparison of three training methods. *Journal of phonetics*, *38*(4), 569-587.

Koster, R. (2013). *Theory of fun for game design*. " O'Reilly Media, Inc.".

Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & psychophysics*, *50*(2), 93-107.

Ladefoged, P., & Disner, S. F. (2012). *Vowels and consonants*. John Wiley & Sons.

Lavolette, E., Polio, C., & Kahng, J. (2015). The accuracy of computer-assisted feedback and students’ responses to it.

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

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242-1255.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/l: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874-886.

Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Routledge.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive science,* 5(4), 333-369.

Malone, T. W., Lepper, M. R., Snow, R. E., & Farr, M. J. (1987). Aptitude, learning and instruction III: Cognitive and affective process analysis.

Memrise [Computer software]. (1992). Retrieved from https://www.rosettastone.com/lp/about/

Merriam-Webster (n.d.) *Game*. Retrieved from: https://www.merriam-webster.com/dictionary/game

Ohala, D. K. (2008). Phonological acquisition in first language. In Edwards, J. G. H., & Zampini, M. L. (Eds.). *Phonology and second language acquisition* (Vol. 36). John Benjamins Publishing.

Pajitnov, A., & Gerasimov, V. (1984). Tetris [videogame]. *A. Pajitnov*.

Peterson, M. (2010). Computerized games and simulations in computer-assisted language learning: A meta-analysis of research. *Simulation & Gaming, 41*(1), 72-93.

Piaget, J. (1962). Play, Dreams, and Imitation in Childhood.(translation by C. Gattengno & FM Hodgson; original work published in 1951).

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13*(2), 253-260.

Pisoni, D. B., & Lively, S. E. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. *Speech perception and linguistic experience: Issues in cross-language speech research*, 433-459.

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of general psychology, 14*(2), 154.

Rock Band. (2007). [Computer Software]. MTV Games.

Rogers, E., (1995). Diffusion of Innovations, fourth ed. Free Press, *New York.*

Robertson, A. (2011). Duolingo will translate the internet while teaching languages. *The Verge*. Retrieved from https://www.theverge.com/2011/12/16/2639262/duolingo-translate-internet-teach-languages

Rossano, V., Roselli, T., & Calvano, G. (2017, June). A Serious Game to Promote Environmental Attitude. In *International Conference on Smart Education and Smart E-Learning* (pp. 48-55). Springer, Cham.

Rosetta Stone [Computer software]. (1992). Retrieved from https://www.rosettastone.com/lp/about/

Rice, J. (2007). Assessing higher order thinking in video games. *Journal of Technology and Teacher Education*, *15*(1), 87-100.

Rubik, E. (1977). Rubik's Cube, [Combination Puzzle] Rubik's Brand Ltd.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and emotion*, *30*(4), 344-360.

Russell, S., Graetz, M., & Witaenem, W. (1962). Spacewar. *Computer software*.

Rutter, J., & Bryce, J. (Eds.). (2006). *Understanding digital games*. Sage.

Salaberry, M. R. (2001). The use of technology for second language learning and teaching: A retrospective. *The modern language journal*, *85*(1), 39-56.

Sanz, C. (2004). Computer delivered implicit vs. explicit feedback in processing instruction. *Processing instruction: Theory, research, and commentary*, *241*.

Shafer, V.L. (2008). Speech Perception in Second Language learners; the re-education of selective perception. In Edwards, J. G. H., & Zampini, M. L. (Eds.). *Phonology and second language acquisition* (Vol. 36). John Benjamins Publishing.

Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, *31*(4), 307-314.

Strange, W., & Jenkins, J. J. (1978). Role of linguistic experience in the perception of speech. In *Perception and experience* (pp. 125-169). Springer, Boston, MA.

Su, C. H., & Cheng, C. H. (2015). A mobile gamification learning system for improving the learning motivation and achievements. *Journal of Computer Assisted Learning*, *31*(3), 268-286.

Sykes, J. M., Reinhardt, J., & Thorne, S. L. (2010). Multiuser digital games as sites for research and practice. In *Directions and prospects for educational linguistics* (pp. 117-135). Springer, Dordrecht.

UCL Vowel Trainer. (1999). Retrieved from: https://www.phon.ucl.ac.uk/shop/voweltrainer.php

Uskov, V. L., Bakken, J. P., Penumatsa, A., Heinemann, C., & Rachakonda, R. (2017, June). Smart pedagogy for smart universities. In *International Conference on Smart Education* *and Smart E-Learning* (pp. 3-16). Springer, Cham.

Uther, M., Uther, J., Athanasopoulos, P., Singh, P., & Akahane-Yamada, R. (2007). Mobile Adaptive CALL (MAC): A lightweight speech-based intervention for mobile language learners. In *Eighth Annual Conference of the International Speech Communication Association*.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *38*(4), 579.

Thomson, R. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, Vol. 28, No. 3 (2011), pp.744-765.

Thomson, R. (2012). English Accent Coach. Retrieved from: https://www.englishaccentcoach.com/

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language teaching*, *31*(2), 57-71.

Watterson, B. (2005). *The Complete Calvin and Hobbes*. Andrews McMeel Publishing.

Wong, J. (2012). Training the perception and production of English /e/ and /ae/ of Cantonese ESL Learners: A Comparison of Low vs. High Variability Phonetic Training. In *Proceedings of the 14th Australasian International Conference on Speech Science and Technology. Sydney, Australia* (pp. 3-6).

Wong, J. (2014). The effects of high and low variability phonetic training on the perception and production of English vowels /e/-/æ/ by Cantonese ESL learners with high and low L2 proficiency levels. In *Fifteenth Annual Conference of the International Speech Communication Association*.

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer*, *38*(9), 25-32.