SPRING: Speech and Pronunciation Improvement through Games

For Hispanic children in US

Master's Thesis

Nitesh Goyal
Lehr-und Forschungsgebiet Informatik 9 „Computerunterstütztes Lernen" an der RWTH Aachen University
26. Dezember 2010

Lehr-und Forschungsgebiet Informatik 9
„Computerunterstütztes Lernen" an der
RWTH Aachen University

Contents

List of Figures

List of Tables

Abstract

Lack of understandable English pronunciations is a major problem for immigrant population in developed countries like U.S. This poses various problems, including a barrier to entry into mainstream society. This work presents a research study that explores the use of speech technologies merged with activity-based and arcade-based games to do pronunciation feedback for Hispanic children within the U.S.

A 4-month long study with immigrant population in California was used to investigate and analyze the effectiveness of computer aided pronunciation feedback through games to make the speech more understandable and intelligible. In addition to quantitative findings that point to statistically significant gains in pronunciation quality, the work also explores qualitative findings, interaction patterns and challenges faced by the researchers in dealing with this community. It also describes the issues involved in dealing with pronunciation as a competency.

Aim

Design and employ age-appropriate games similar to the ones that English as Second Language Learning (ESL/ELL) Hispanic High-School students in California already enjoy playing as extra aids to give feedback for improving English pronunciation to make their speech more intelligible and understandable.

Initial Proposal

Introduction

Integration of the immigrant families in the host country has been an issue of discontent across the world. It is felt more so in the developed world, which is the destination for most of the immigrants. According to the U.S. Bureau of the Census, USA gains one net immigrant person every 35 seconds and is witnessing an increasing immigrant population [1]. The foreign born population has increased from less than 5% in 1970 to more than 10% in 2000 [2]. Between 1970 and 2000 the share of Asian immigrants has risen from 9 to 25% and Latin American immigrants from 19% to 51% [2]. As indicated, the immigrant population sometimes does not share the L1 of their new countries in their original countries (Spanish and Chinese vs. English). Hence, this lack of communication skills is preventing the newcomers from completely integrating in the society, benefiting the society and gaining acceptance from the society as well. The newer data from Census 2010 is unavailable. However, there are reasons like economic growth, demand for people for certain jobs, and inertial momentum for me to believe that the figures are comparable.

Of particular interest, to me, are the children of these immigrant families who are facing problems in speaking English in USA. These children attend the same schools as the native population. However, L1 (Spanish, Chinese, Hindi etc) might still be spoken at home, amongst other resident members of the same community and at other special events. Besides, large immigrant community of the same origin might even reduce the motivation to learn English. Hence, it is a challenge for these children - the future of USA - to integrate better with the native population.

Mobile gaming, including games on mobile phones, game devices like PSP, Nintendo Game Cube, portable devices like iPod etc is becoming popular. Since the popular Nokia Snake game in 1970s, there seem to be a plethora of games available for afore mentioned devices, including over 17,000 games on Apple App store alone itself [3]. According to an article in SIGMOBILE in May, 2005: “Total global revenues from mobile games are forecast to increase from USD 2.6 billion this year to USD 11.2 billion by 2010, according to Mobile Games, a new strategic research report from Informa Telecoms and Media” [4]. While the mobile gaming has increased, the usage of mobile phones by kids has also increased. In 2004, 18 %(12 year old) up to 64 %(17 year old) kids owned mobile phones [5]. This number jumped to 51% and 84% in 2008, respectively [5]. Even, 62% of all the teens in households with less than $30k annual income have a mobile phone [5]. More than seven in ten (71%) teens ages 12-14 play games on a portable device or mobile phone, compared with just under half (49%) of teens ages 15-17 [6]. Hence, this ownership of mobile phones translates well into their use of mobile phones or other portable devices for playing games.

Hence, mobile games might be an interesting possibility for disseminating knowledge in a private captivating environment beyond traditional schooling. In my thesis, I intend to focus on designing, and creating a mobile game using the User-Centered Design Process that such kids can play to learn, and improve their English skills, with focus on correct pronunciation.

Who are the users?

School going children (12 – 18 year old) with English as a Second Language (ESL) facing pronunciation problems.

What do they want?

Improve their pronunciation in a non-traditional and captivating fashion.

Existing Work

Computer Assisted Language Learning (CALL) has existed for almost 70 years now. Several methods and systems have been proposed to help improve particular focus areas in language learning using computers.

[7, 8] present a system being used by the US Army to learn Arabic in Iraq. The system is based on Task-based-Learning paradigm for Culture and language Training. This serious interactive pedagogical drama based game is meant to be used by adults and involves several missions, game play, and feedback to learn Arabic language and culture. Due to nature of the intended use and lack of a particular pronunciation focus, this product is unsuitable for use by young children.

[9] describes Baldy, a virtual talking head on a screen with focus on helping users learn how to pronounce the phonemes properly as a virtual teacher. The open cross sections of the mouth are also displayed to help learners reciprocate the sounds by identifying the right internal and external mouth movements. [10, 11] are also similar systems and go a step further and act as Embodied Conversational Agents (ECA). These systems improve upon [9] by including other features like vocabulary learning etc. These systems mention encouraging results. However, they represent additional learning materials to existent school syllabus. Besides, they are computationally expensive and cannot be used in mobile phones.

[12] uses both the ECA and game based design in its system, called DEAL. the Agent in this game gets orders from the user to perform tasks, for example, in a virtual room like picking up things. The users hence, learn how to structure the sentences properly and learn appropriate word placements. This game offers little motivation to the young children and there seems to be no discussion about the methodology and process of game design.

Multimodality [13, 14, and 15] has also been briefly investigated for pedagogical benefits in ESL. [13] suggests that spoken language can be grounded sensory perceptions of the real world. It describes a learning interface that bridges a gap between the real world physical objects and the virtual interface.[14] describes a collaborative table top based simple matching to help develop the reading skills of young groups of children. [15] describes a system with 2 inter coupled-interfaces: TV as an audio visual aid, and mobile phone as a supporting aid to help learners learn the vocabulary. These systems also continue to focus on the writing, reading, and vocabulary parts of the language.

Proposed work

As evident from above, there seems to be a lack of usage of games, especially in mobile devices, to help in pronunciation feedback and correction. My aim is to design mobile phone based games which utilize the anytime, anywhere affordance offered by the mobile phones. These games should be user-centered and designed using the iterative DIA process. For the games to be successful: firstly, games should be motivating to ensure the frequent playability; secondly, they should be sufficiently enriching to ensure that certain measurable learning outcome is evident.

I begin by modeling and adapting three different games, which try to help improve the pronunciation of the children based upon popular existing phone games. Afterwards, I take these games to the target users and let them play with these games. I video record them and interview them to understand their requirements better. This would help me improve the design better. I will also try to find m-learning and relevant theories to create better designs. Finally, I will follow the process iteratively, as many times as I can within the time frame.

Project Schedule

DIA Cycle: Design, Implement, Analysis (5 months)

I will develop a few initial game prototypes. More than one game prototype is necessary to cover the differences and preferences of the various children, owing to their predisposed predilection for a certain genre of games, or gender based choices. These pro types will be deployed using a computer to ensure better user experience. The contextual enquiries and evaluations would suggest the changes necessary in the design of these games. These user tests would help inform me how to improve the motivation, the UI, and learning to develop the next game.

End of December 2nd week: Finish development of first game.

December 3rd Week: Deployment, and user studies.

End of January 2nd Week: Finish development of two more games.

January 3rd Week: Deployment, and user studies.

January 4th Week: Compilation of user studies.

February End: Next game design

End of Mid March: porting to Mobile phone

March 3-4 Week: Mobile Phone Deployment

April 1-2 Week: Game Design changes

April 3 Week: Deployment

April 4 Week: Final Compilation

References

[1] http://www.census.gov/population/www/popclock us.html

[2] http://www.census.gov/population/www/pop-profile /files/2000/chap17.pdf

[3] http://148apps.biz/app-store-metrics/

[4] http://www.sigmobile.org/newsletter-archive/May - 2005.html

[5] http://www.pewinternet.org/Reports/2009/14--Teens-and-Mobile-Phones-Data-Memo.aspx?r=1

[6] http://www.pewinternet.org/Reports/2008/Teens-Video-Games-and-Civics

[7] W. Lewis Johnson: Serious Use of a Serious Game for Language Learning, Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work, 2007

[8] W. Lewis Johnson, Hannes Vilhjalmsson, Stacy Marsella: Serious Games for Language Learning: How Much Game, How Much AI?, Conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology, 2005

[9]Massaro: A Computer-Animated Tutor for

Spoken and Written Language Learning, ICMI, 2003

[10] David Powers, Richard Leibbrandt: PETA – a Pedagogical Embodied Teaching Agent, PETRA, 2008

[11] Olle Bälter, Olov Engwall, et al: Wizard-of-Oz Test of ARTUR - a Computer-Based Speech Training System with Articulation Correction, ASSETS, 2005

[12] Anna Hjalmarsson, Wik: Dealing with DEAL: A dialogue system for conversation training, 8th SIGdial Workshop on Discourse and Dialogue, 2007

[13] Chen Yu, Dana H. Ballard: A Multimodal Learning Interface for Grounding Spoken Language in Sensory Perceptions, Applied Perceptions, 2004

[14] R.J.W. Sluis, I. Weevers, C.H.G.J. van Schijndel, L. Kolos-Mazuryk, S.Fitrianie, J.B.O.S. Martens: Read-It Five-to-seven-year-old children learn to read in a tabletop environment, IDC, 2004

[15] Sanaz Fallahkhair et al: Dual device Interface for Ubiquitous Language Learning: Mobile Phone and Interactive Television (iTV), WMTE, 2004

Acknowledgements

This thesis is a culmination of joint efforts of many, without whose help and support – this could not have been possible.

First, I would like to thank Prof. John Canny. Without your support, I could not have finished this work. I am grateful to you for the hours you took out of your schedule to give my thoughts a clearer direction, and ensure that the project was always on the track despite the multiple logistical and administrative challenges. Also, I am grateful that you hosted me at the BiD lab at University of California, Berkeley and gave me chance to observe, work with, and learn from the talented group of people.

Second, I would like to thank Prof. Ulrik Schroeder for encouraging me to pursue this thesis beyond the traditional realms of possibilities. You provided the most timely and effective help to work with the administrative requirements at RWTH University, Aachen. Had it not been the lessons I learnt from you in your classes, I would not have been sufficiently prepared to embark on this research. I am also grateful to you for your assistance beyond the official thesis advisor requirements to help further my future goals.

Third, I would like to thank my two student advisors: Anuj Tewari and Mostafa Akbari. Without your continued help, and previous preparations to me embarking on this journey, this project would not have been the same as it turned out to be. I was fortunate to have been advised by these two industrious fellows who made the trans-Atlantic endeavor much easier.

I would like to thank Ms. Beate Wassenberg, Dr. Hilde Akam and Dr. Heide Naderer at the International Office, RWTH University, Aachen for choosing me as the UROP RWTH Research Ambassador to US and providing the much-needed scholarship. Had, it not been for this graciousness, my stay at Berkeley would not have fructified.

At last many thanks go to my family whose unwavering support in me brought me so far. You allowed me to pursue my dreams and stood by me through thick and thin. Your encouragement and words of wisdom continue to guide me.

1. Introduction

“Inclusive Education” is a part of *Improving Education Quality*, one of the themes of UNESCO. Children belonging to indigenous groups and linguistic minorities are classified as vulnerable to exclusion from the benefits of the education system. Traditionally such minorities have been believed to exist only in the developing and the less developed world. Howe caver, statistics and experiences suggest that such minorities exist even in the developed world.

To establish this I would like to provide data about such minorities in USA.

According to the US Census Bureau’s American Community Survey in 2007, Latin America has been the largest source of immigrants to United States of America since 1990 (US Census Bureau, 2007). In 2007 alone, 53.7 percent of all the immigrants in USA emigrated from Latin America and about 80 percent from Latin America or Asia. Approximately, 6 percent of the total population reported themselves as Foreign-born Hispanics, mostly from Mexico. According to the Pew Hispanic Center, over half (55 percent) of all of these Mexican immigrants in the United States were unauthorized in 2008 (Pew hispanic Centre, 2010). Thus, around 6 percent of the total population in USA is of Mexican origin.

Migration Policy Institute (MPI), an independent, nonpartisan, nonprofit organization, which provides analysis, development, and evaluation of migration and refugee policies at the local, national, and international levels, corroborates this data. According to MPI’s Data Hub, available online, about 70 percent of these Mexican born immigrants live in close communities in just four of the fifty states in USA: California, Texas, Illinois, and Arizona (Migration Policy Institute, 2010). These communities are expected to not just live together for cultural and social benefits.

I believe that their similar economic and financial conditions also bring them closer. Of the foreign-born population in the United States in 2007, about 20 percent of noncitizens lived in poverty, with about 15 percent living below the poverty threshold (Migration Policy Institute, 2010). Also, 20 percent of persons who spoke Spanish at home lived in poverty, compared to around 12 percent of persons who spoke Asian or Pacific Island languages, and 11 percent of persons who spoke other Indo-European languages (Migration Policy Institute, 2010). It seems that Spanish-speaking immigrant population is almost twice likely to live in poverty.

International Bureau of Education (IBE), an international centre for the content of education, is an integral, yet autonomous part of UNESCO and International Academy of Education (IAE). IBE's Teaching Additional Languages booklet classifies "speaking" as an integral part of language learning for additional language learners (International Bureau of Eduation, 2001).

According to MPI's release in February 2010, about three-quarters of Mexican immigrants in 2008 were limited English proficient. Specifically, 73.8 percent Mexican immigrants reported speaking English less than "very well," much higher than the 52.1 percent reported among all foreign-born immigrants age 5 and older (Migration policy Institute, 2010). This might be attributed to the fact that in 2008, 61.5 percent of the 9.2 million Mexican-born adults in USA age 25 and older had no high school diploma or the equivalent general education diploma (GED), compared to 32.5 percent among all foreign-born adults (Migration policy Institute, 2010).

This highlights the plight of Hispanic, and specifically Mexican, immigrant cultural and linguistic minority living in USA, one of the most developed countries. Evidently, this community suffers from exclusion of benefits of the infrastructure and society available in the developed world. Moreover, their lack of knowledge of primary language of communication: English hampers prospects of improvement.

This work aims to draw attention of the research community to the marginalized linguistic minorities of the developed world and describes an application of Information and Communication Technologies (ICT) to address English language learning related challenges faced by such minorities as a contribution to the discourse.

Several factors contribute to the lack of English Language fluency amongst the immigrants, especially the immigrant children.

According to the 2000 Census (United States Census Bureau, 2000), Hispanic youth group has the highest High School dropout rate (about 28 percent) amongst immigrant, African American, and White youth groups. Almost half of all the Hispanic youth born outside USA, drop out of high school. According to the National Center for Education Statistics (NCES) Survey (National Center for Education statistics, 2010), Financial Problems and Incarceration were reported to be the biggest reasons that prevented immigrant children from 12 to 18 years age from finishing High School in USA. Other reasons included lost interest, behavior or academic problems.

Hispanic Children arriving in USA before the age of 12 resembled those born here in their level of English language usage. Both these groups performed better than Hispanic children arriving in the age group of 12-18 in all the three literacy tests: Prose, Document, and Quantitative Test. This is expected, since the lower aged children had an opportunity to grow in an English-speaking environment.

Thus, I focused the study on the age group of 12-18 year old immigrant Hispanic children who lacked an opportunity to gain English Language knowledge previously by trying to minimize the cause of “lack of interest” amongst the youth using games similar to the ones enjoyed playing the high-school students.

The project described in this work is a formative study to investigate the possibility, and effects of using computer-based games as a motivational tool to teach and improve pronunciation of immigrant children with limited exposure to spoken English language. Focus of the study is on the students at a Public High School located in California with a very high Hispanic immigrant population.

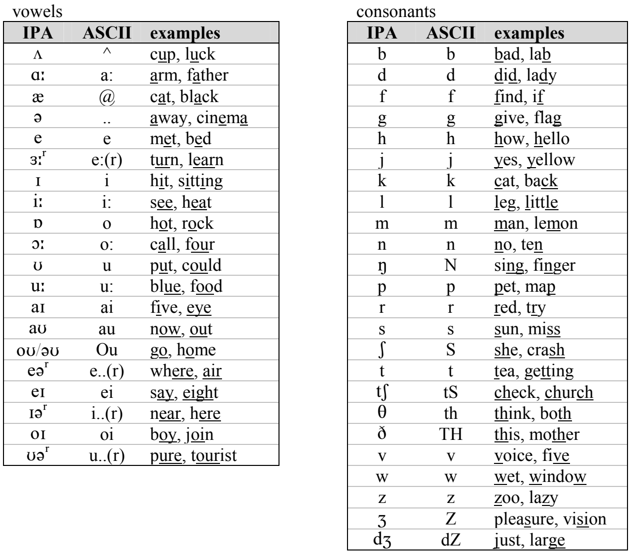
2. Background

Before proceeding further it is necessary to understand a few concepts. These include: the terms used by linguists to denote speech signals, and the basic knowledge of how speech recognition engine functions.

2.1 Phoneme

A Phoneme is the smallest discernible unit of speech sound in spoken language. (Fritzsche, 1997) Each language is hence constituted by a series of distinctive phonemes. American English consists of 16 vowels and 26 consonants, i.e. 32 characters. But, it consists of 40 phonemes. Thus, each written character is not singularly mapped to a single phoneme. In 1888, International Phonetic Alphabet (IPA) was developed to describe all the phonemes in the world language. (International Phoenetic Association, 1888) For example, Table 1 gives a brief overview of the vowel and consonant depictions using phonemes (Antimoon, 2010). For more details, please refer to Appendix A.

Table 1. Vowels and Consonants in English Language and their IPA Phonemic representations



vowels

IPA	ASCII	examples
ʌ	^	cup, luck
ɑː	a:	arm, father
æ	@	cat, black
ə	..	away, cinema
e	e	met, bed
ɜː^{r}	e:(r)	turn, learn
ɪ	i	hit, sitting
iː	i:	see, heat
ɒ	o	hot, rock
ɔː	o:	call, four
ʊ	u	put, could
uː	u:	blue, food
aɪ	ai	five, eye
aʊ	au	now, out
oʊ/əʊ	Ou	go, home
eə^{r}	e..(r)	where, air
eɪ	ei	say, eight
ɪə^{r}	i..(r)	near, here
oɪ	oi	boy, join
ʊə^{r}	u..(r)	pure, tourist

consonants

IPA	ASCII	examples
b	b	bad, lab
d	d	did, lady
f	f	find, if
g	g	give, flag
h	h	how, hello
j	j	yes, yellow
k	k	cat, back
l	l	leg, little
m	m	man, lemon
n	n	no, ten
ŋ	N	sing, finger
p	p	pet, map
r	r	red, try
s	s	sun, miss
ʃ	S	she, crash
t	t	tea, getting
tʃ	tS	check, church
θ	th	think, both
ð	TH	this, mother
v	v	voice, five
w	w	wet, window
z	z	zoo, lazy
ʒ	Z	pleasure, vision
dʒ	dZ	just, large

As can be seen, the English alphabets have not been used to represent the phonemes. These depictions might be easy for native English speakers to understand and mentally map between the written characters and associated phonemic sounds, but might prove difficult for non-native speakers to learn and remember. While a phoneme is a theoretical construct used by the linguists to differentiate between the different speech sounds, a Phone is the actual sound produced by the speaker. Throughout this work, they would be used interchangeably since the true meanings of the two words have now blurred through intermittent usage over time.

2.2 Spelled Out Pronunciation

Recently there has been a move to move away from the IPA style pronunciation to Spelled-out-Pronunciations. This is because a spelled out pronunciation uses the existing English alphabets or a combination of those to represent a sound. (Bloomfield, 1984) This enables the English language learners to read how sounds should be pronounced without learning a new language with different symbols. The Table 2 gives the spelled-out-pronunciations from Dictionary.com (Dictionary.com, 2010).

Table 2 Spelled Out Pronunciation examples

Word	Spelled-out Pronunciation by Dictionary.com
Circular	sur-kyuh-ler
Swirling	Swurl-ih-ng
Mothball	mawth-bawl
Overflow	oh-ver-floh
Wreckage	rek-ij
Aluminum	uh-loo-muh-nuh m
Menacing	men-is-ng
Equipment	ih-kwip-muh nt
Locomotive	loh-kuh-moh-tiv
Glowed	gloh d

2.3 Speech Recognition Engine

Speech is a continuous audio stream and is usually a mix of stable and dynamic states. To understand speech waveform, one needs to detect the end of a word and beginning of the next word. This means that a speech recognition engine should be able to detect the filling space between different words and then detect the word itself. The words are detected by matching the audio stream with known properties of each spoken word.

One way to categorize the spoken word is to divide it into frames of equal lengths and deducing feature vectors of each frame that represent unique features of the sounds in that frame. These are constituted using 39 parameters over a frame of length 10 milli-seconds usually (Carnegie Mellon University, 2010)

However, these feature vectors need to be matched with known a model, which should mathematically describe the most probable feature vector for that sound. Several models exist: Acoustic Model, Phonetic Dictionary, and a Language Model. Each of these has distinct advantages and disadvantages.

3. Related Work

Computer Assisted Language Learning (CALL) has existed for almost 70 years now. Several methods and systems have been proposed to help improve particular focus areas in language learning using computers. Most work in the CALL domain does not explore the ability of technology to teach English pronunciation using persuasive computer games to immigrant high school children.

Horowitz et al (J. E. Horowitz, 2006) describes an 8-week long study that promotes literacy in USA with participants from households below the poverty line. During the course of the study, cell phones were provided to preschool children and Sesame Street videos were streamed on these devices. The focus of this study was to improve literacy and teach the English alphabet. While the videos were persuasive, they lacked focus on improving the English pronunciation and were targeted at very young children.

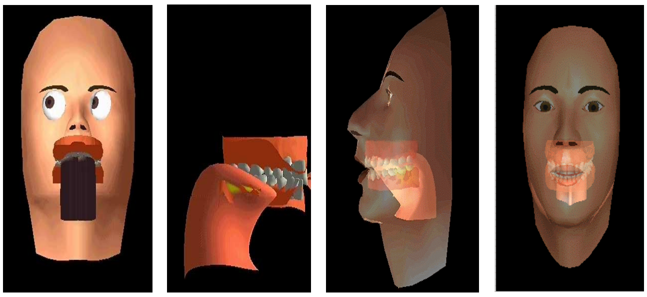


Figure 1 Massaro's Baldi showing a cross section of the anatomical movements while pronouncing

Massaro et al (Massaro, 2003) described Baldi as shown in Figure 1, a virtual talking head on a screen with focus on helping users learn how to pronounce the phonemes properly as a virtual teacher. This is bar the only study I know that focuses on teaching pronunciation and provided a visually detailed feedback and training. The open cross sections of the mouth are also displayed to help learners reciprocate the sounds by identifying the right internal and external mouth movements.

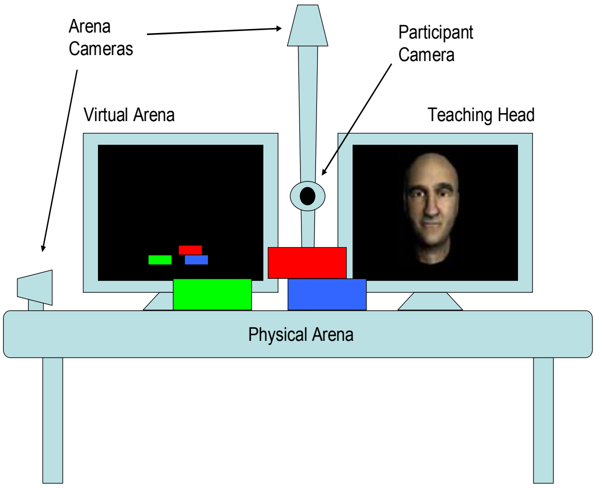


Figure 2 Powers et al's system installation showing the ECA connected to a camera

Powers et al 's pedagogical embodies teaching agent (PETA) (D. Powers, 2008) is also a similar system and goes a step further by acting as Embodied Conversational Agent (ECA). PETA improves upon Massaro by including other features like vocabulary learning etc. PETA also provides an opportunity to merge into the real world by capturing a part of the real word virtually and encouraging conversations about this part in the form of full sentences with participants. While this system mentions encouraging results, they lack information about how motivational these systems might be and are rather too artificial and obtrusive into the natural way of the daily behavior of the participants.

Multimodality has also been briefly investigated for pedagogical benefits in English Language Learning. Chen Yu et al suggests that spoken language can be grounded with the sensory perceptions of the real world. It describes a learning interface that bridges a gap between the real world physical objects and the virtual interface. (Chen Yu, 2004) While PETA involved conversations about a restricted and limited part of the physical world, the aim in this case is slightly different. The aim here is to use speech data and data from multiple other sources like visual data, perceptual data etc to create an intelligent machine capable of understanding the context of the user's actions, and objects under influence, leading to an intelligent conversation, as shown in Figure 3.

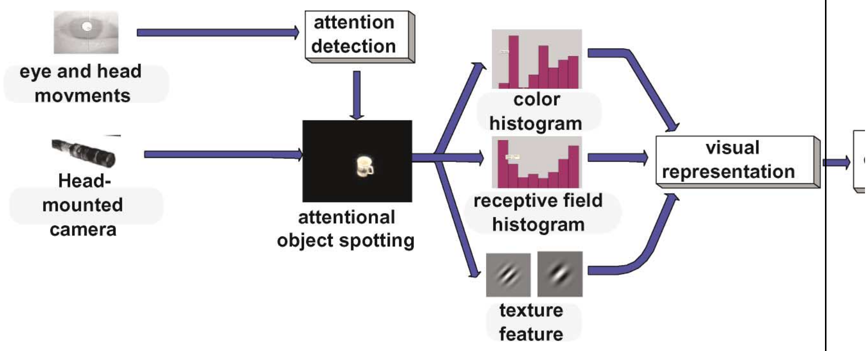


Figure 3 An overview of system implementation by Yu et al

Sluis et al's Read-It (R. J. W. Sluis, 2004) describes a collaborative tabletop based simple matching to help develop the reading skills of young groups of children. The aim of this system is to use a game as an learning tool to enable players of five to seven year old age to match words that begin with the same phonetic sound. While, this system does sound fun, it does not provide any opportunity to the players to perform pronunciation themselves or even hear it. So, there is no pronunciation and pronunciation feedback opportunity

Fallahkhair et al (Fallahkhair, 2004) describes an interactive TAMALLE application system with 2 inter coupled-interfaces: TV as an audio visual aid, and mobile phone as a supporting aid to help learners learn the vocabulary. This allows creating featured programming and allows pulling/pushing data on to a WAP enabled phone that can provide extra information about the content on the television screen. For example difficult words and terms can be explained on the mobile phone screen while the television screen is primarily used to stream the content and display the multiple options. This system focuses on the understanding of the scripted content, and improving the vocabulary of the users without focusing on the pronunciation related parts of the language.

However, recently there has been a growing interest in including computer-based tools that use automated speech recognition to provide a guided reading experience for the users. Mostow et al's Project LISTEN based Reading Tutor (G.Aist, 2001) has been used with a variety of audiences in improving the English reading ability of children by using storytelling techniques , with English as a first language and with English as a second language (ESL/ELL) in USA.

Later Mostow et al (J.Mostow, 2003) conducted a yearlong study with over a hundred second and third grade students to determine the influence, effectiveness, and differences between a human tutor and a reading tutor based on previously created Project LISTEN. The treatment groups exposed to the system outgained the control human-tutored groups in word comprehension and passage comprehension but not in word attack, building upon the previously shown results. The analysis of the video tapes showed that the experiment group using the reading tutor faced technical challenges due to the slower response times of the system, requested more help, and eventually picked easier stories to minimize the interaction with the slow system. However, the human tutors corrected more errors by focusing on particular error in the sounds, and hence improved the word attack. Hence, it is imperative to be able to find the exact error in speech and correct it by making the users sound it and hear it as well.

Recently Mostow et al (J.Mostow G. R., 2008) has shown that similar automated technologies can be also beneficial for developing communities, as in Africa in this particular work. Figure 4 shows the general setup of how one child would sit on a computer with the featured automated tutor. The child could see the story displayed on the screen. The story can be selected form multiple stories existing in the tutor system by the child according to self-interest. The headphones would read out aloud the stories and would give a chance for the readers to listen to the words while they appear on the screen.

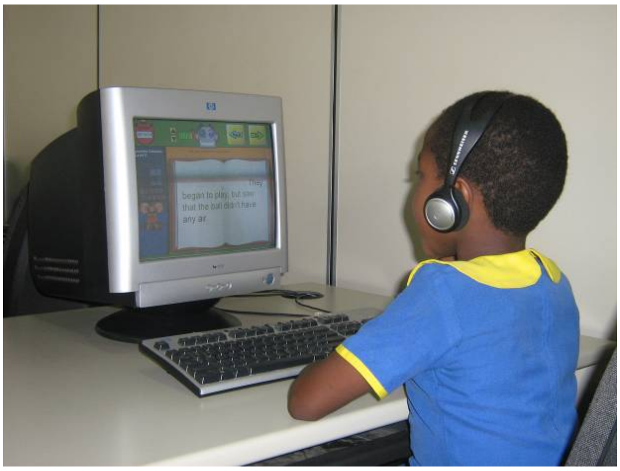
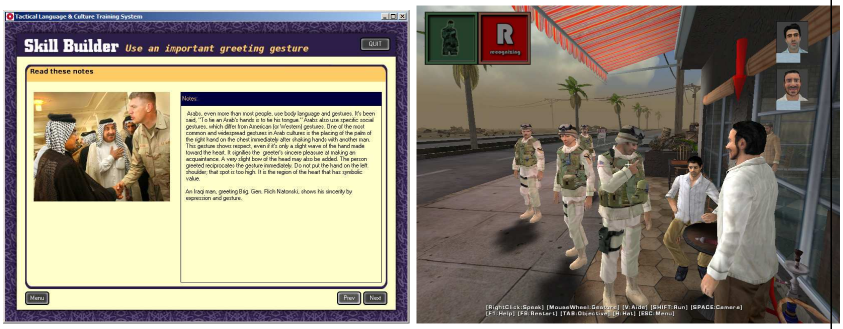


Figure 4 Mostow et al 's work in Ghana involving a child listening to an autonated tutor

Project LISTEN based Reading Tutors have been deployed later by Poulsen et al (R. Poulsen, 2007) and in Canada by Reeder et al (K.Reeder, 2005).

Use of games has been popular for education in other works. Johnson et al in (W. Lewis Johnson, 2005) and (Johnson, 2007) present a Tactical Language and Culture Training System (TLCTS) system being used by the US Army to learn Arabic in Iraq, as shown in Figure 5.



Tactical Language & Culture Training System
Skill Builder
Use an important greeting gesture
QUIT
Read these notes
Notes:
Menu
Prev
Next
R
recognizing

Arabs, even more than most people, use body language and gestures. It's been said, "To tie an Arab's hands is to tie his tongue." Arabs also use specific social gestures, which differ from American (or Western) gestures. One of the most common and widespread gestures in Arab cultures is the placing of the palm of the right hand on the chest immediately after shaking hands with another man. This gesture shows respect, even if it's only a slight wave of the hand made toward the heart. It signifies the greeter's sincere pleasure at making an acquaintance. A very slight bow of the head may also be added. The person greeted reciprocates the gesture immediately. Do not put the hand on the left shoulder; that spot is too high. It is the region of the heart that has symbolic value.

An Iraqi man, greeting Brig. Gen. Rich Natonski, shows his sincerity by expression and gesture.

[RightClick:Speak] [MouseWheel:Gesture] [V:Aide] [SHIFT:Run] [SPACE:Camera]
[F1:Help] [F8:Restart] [TAB:Objective] [H:Hat] [ESC:Menu]

Figure 5 Tactical Iraqi system created by Johnson et al showing a virtual environment

The system is based on Task-based-Learning paradigm for Culture and language Training. This serious interactive pedagogical drama based game is meant to be used by adults and involves several missions, game play, and feedback to learn Arabic language and culture. However, I feel that due to nature of the intended use and lack of a particular pronunciation focus, this product is unsuitable for use by young children.

Anna et al's DEAL (Anna Hjalmarsson, 2007) uses both the ECA and game based design in its system. The Agent in this game gets orders from the user to perform tasks, for example, in a virtual room like picking up things. The users hence, learn how to structure the sentences properly and learn appropriate word placements. This game offers little motivation to the children and there seems to be no discussion about the methodology and process of game design. Besides, this system is focused more on the grammar than spoken language.

MILLEE

This work is an extension of Mobile and Immersive Learning for Literacy in Emerging Economies (MILLEE). MILLEE involved Cell phone applications that enable children in the developing world to acquire language literacy in immersive, game-like environments. Over about 10+ rounds of field studies in the past 6 years, the aim has been to make localized language learning resources more accessible to underprivileged children, at times and places that are more convenient than schools.

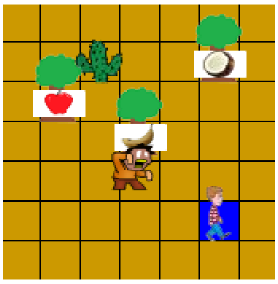


Figure 6 The digital game developed by Kam et al based on a popular traditional game

Kam et al (M.Kam, 2009) and (Matthew Kam, 2009) has successfully shown use of games (Figure 6) as persuasive tools for improving the English literacy of the illiterate English as Second Language (ESL/ELL) children in India. The work encourages the use of games with children to motivate them for learning English

Vocabulary while playing. This work highlights the gap between the existing digital games and the traditional games played by the rural children in Indian villages to analyze and inform the design of a new videogame that is more intuitive and intriguing.

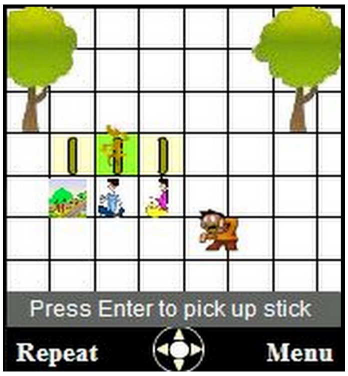


Figure 7 Kumar et al's mobile phone based educational game

Kumar et al (Anuj Kumar, 2010) emphasizes the need to explore possibilities of using mobile phones and mobile phone-based games to impart learning in out-of-school settings and improve the English vocabulary of the students at their geographical and temporal convenience while divulging details of social relationships and their impacts on such studies. One of the games played by the participants in the study is shown in Figure 7.

As explained above, the existing works successfully describe using games or speech recognition or both for literacy improvement. However, unlike the system, none of them employs usage of both games and speech recognition for pronunciation improvement amongst English as Language Learner (ELL) children.

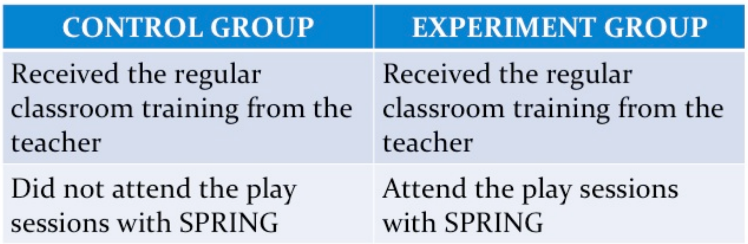
4. Pilot Study

4.1 Overview

The pilot study was carried out at a public high school located in a highly populated Hispanic immigrant location in California, USA for about three months from December 2009 to March 2010. The study took place within the school premises during the extended school timings and involved demographic study, the pre-test, the experiment, and the post-test within the school premises with permissions from the school authorities, the teachers involved, the students and/or their parents.

The goal was to investigate the learning impacts of pedagogy enriched games designed like the popular available commercial games. All the participants attended the same and equal amounts of ELL training by the same teacher in the same class. However, a randomly selected half of them (Experiment Group) were able to play with the games that were part of SPRING after their classes, while the other half (Control Group) continued with their normal schedule of work and play, as shown in the Table 3.

Table 3 Control Group vs. Experiment Group in the study



CONTROL GROUP	EXPERIMENT GROUP
Received the regular classroom training from the teacher	Received the regular classroom training from the teacher
Did not attend the play sessions with SPRING	Attend the play sessions with SPRING

Three sessions were held, on an average, per week for four weeks. Each session accommodated approximately three students, one after the other. So, each student played freely in seclusion from the other students for about ten minutes per week. There were two different games that each student was able to play. These games were alternated each week to keep up the interest level of the students. Hence, each student received a total of about forty minutes worth of play time with the two games of SPRING during the four week long endeavor.

4.2 Study locale and setup

This section describes the steps followed before beginning the study. This involved finding the user group.

To find the user group, I began by contacting the teachers and school authorities at several public middle and high schools located in the vicinity. The aim was to locate a school with a high immigrant population having a low level of spoken English fluency. Based upon the anonymous demographic and diversity data that were received from these schools, I shortlisted three schools where I submitted a request for conducting research with the students within the school premises. One of the public High Schools that accepted the request was perfect for the study.

The school district, where this school is located, conducted a data survey in 2007-08 that looked at the would-be seniors of that year from the beginning of their high school careers (2004-05) to that year.

Of all students who began 9th grade at a school in this district, 33 % were no longer attending school in the district.

Of the English Learners, the rate was double – 66% were no longer attending school in the district.

For this particular High School, those numbers are higher than for the district as a whole. Approximately 70-75% of the English Language Learners who start 9th grade in the school do not return for the beginning of their senior years.

This is tabulated in the Table 4 on the next page.

Table 4 Showing the amount of attrition in ESL or ELL classes

Organization	Attrition Rate (%)
District (general, including ESL/ELL)	33
District (ESL/ELL only)	66
Selected School (ESL/ELL only)	75

As previously stated, lack of interest is one of the major reasons for the high school – dropouts at this age group. So, the long-term aim of the study is to devise a motivational tool that would generate or increase interest in the ELL classes and possibly, in future, help reduce attrition.

I was guided to one of the English Language Learner's (ELL) classes at this school. The class consisted of 20 students, at ELL level 2. These students had been in USA for less than two years, and had over the time attended, and cleared ELL level 1. The class had a 100 percent immigrant population. In this class, 90 percent immigrants were from Latin America. Of the students in this class, 95% are labeled as SED (Socio-Economically Disadvantaged). That means one of two things (or both things) is true of all but one student. Either

(1) the students are living at an economic level qualifying them for the federal free/reduced lunch program or

(2) his/her parents did not graduate from high school, or both are true.

This situation at this school compares favorably with the previously quoted national data. So, it was decided to choose this particular ELL class at this school.

4.3 Data Collection

Solely the four researchers involved in this project managed the pilot study. However, due to the nature of the participants, a local member of the school volunteered to help translate between English and Spanish for children who could not understand the use of English language.

The children were interviewed to get details about their demographics, including age, sex, the duration of stay in USA so far, the education background, the nature and living situation of their stay in USA, the family background, and the nature of motivation and opportunities available at home for learning and speaking English. During the interviews, questions were also asked about their average daily schedule and the amount of hours they get each day to study at home, if any. Besides, I gathered information about their exposure to media and technology like TV, Computers, Internet, Cell phones and Games to understand the technical competence of the participants and their ability to learn how to play computer games.

The class had strength of 20 students. This class was divided, for the purpose of the study, into two groups of 10 students each. One of the groups was the CONTROL GROUP, which received the regular classroom training from the teacher and did not attend the play sessions with SPRING. The other half, EXPERIMENT GROUP, received exactly the same training in the classroom as the CONTROL GROUP. However, they also received the opportunity to attend play sessions with SPRING.

To reduce any bias due to pre existing knowledge between the two groups, I randomly picked and assigned the students to either of the two groups. Next, they were administered a simple qualifying test (all the 20 participants) to gather their existing level of knowledge. The test consisted of a slideshow of 30 words, one after the other, on a computer. The test taker was required to speak the word shown on the screen and a speech recognition engine (discussed later) recorded and scored the utterance.

The scores were not made visible to the students to reduce anxiety. The tests were done in private with each student to minimize any learning effects. The words selected for the test were kept constant for the entire pool of the participants and were selected from the syllabus and the recommended textbook for that class.

These sessions were also audio recorded. During the course of the study, I evaluated the participants using a similar test to prevent test anxiety and for consistent comparable results. These were administered as a series of pre-tests and post-tests. The words used in the Pre-Test and the Post-test are given in Table 5.

Table 5 List of words chosen for Pre-test and Post-test

Glowed	Wagon	Anchor
Mothball	Violence	Wreckage
Swirling	Carpenter	Soggy
Attic	Chimney	Haul
Circular	Eagle	Oar
Weapon	Barren	Menacing
Leap	Cupboard	Gnawed
Overflow	Bear	Locomotive
Crouch	Porcupine	Equipment
Peek	Curling	Aluminum

Each play session with SPRING was video taped to record the emotional state of the participants while playing. This was captured by facial and body expressions, exclamations, sighs, gasps and other auditory feedback. These recordings created the contextual data by providing us with more data about the playability of the different stages, elements and parts of the games. During the play sessions, the games also recorded the interactions and utterances and created logs of real time short-term gains or losses during the session itself. This gave the information about how pedagogically helpful and progressive the games were.

These video recordings were supplemented with silent observations and notes taking to summarize the general atmosphere of the game play. These notes later also proved successful in identifying and removing usability issues and learning obstacles and deduce interaction patterns of different players. Finally each session concluded with a quick interview of qualitative questioning about the game play and the session itself. The questionnaire is attached later in the Appendix B.

4.4 Participants

This pilot study was one of the first kinds to be established at the partner school, especially with the immigrant population. So, the participants were very new to this new arrangement and I benefitted from their enthusiasm to participate in “something new”. Figure 8 shows one of the participants.

Figure 8 One of the participants enthusiastically playing one of the SPRING games

Initially, in total I obtained consent from 20 children and/or their parents to participate in the study. They were all part of the same ELL Grade 2 class at the school and represented the total strength of the class, as well. I began the pilot study with all the 20 of them. However, during the due course of time, 2 of them left the study.

Unfortunately, the reasons for attrition could not be conclusively determined due to their continuous absence from the school itself during the three month long duration. However, reasons of attrition, after consultation with teacher, seemed to be family and financial problems for the male participant, and teen-age pregnancy for the girl participant.

4.5 Demographics

The 18 students (after attrition of 2 from 20) exhibited the following characteristics:

- Six (6) were male and twelve (12) were females.
- All eighteen (18) in the study were in ELL level 2.
- The students were in the age range of 14 to 17.
- All eighteen (18) students were of Hispanic ethnicity
- Many of them lived with family members such as uncles, aunts, and cousins; some did not live with their mothers or fathers, as shown in Figure 9.

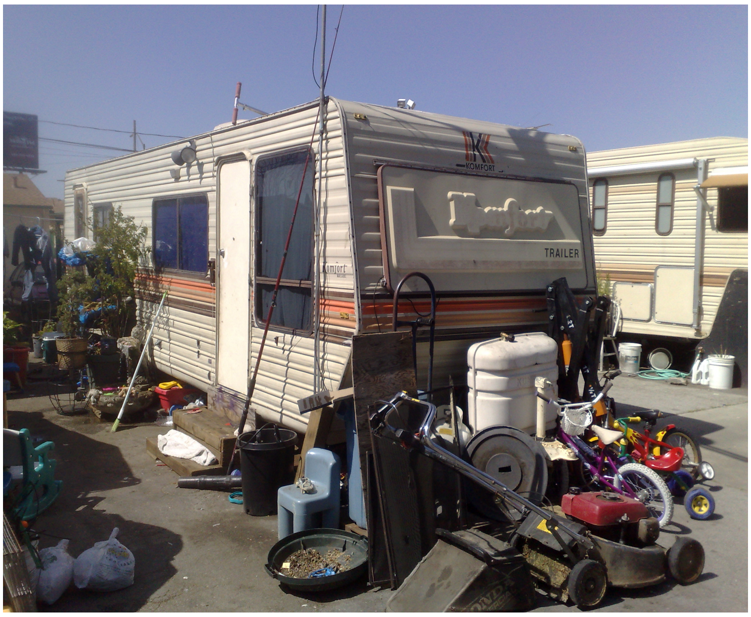


Figure 9 The house of one of the participants

- The fathers, uncles, and brothers held jobs working in a market, as a florist, washing cars, as a gardener, or other lower-end jobs. Few had younger/older brothers or sisters still in school.
- The mothers, aunts, and sisters had jobs that involved cleaning homes, babysitting, or no job at all.

Amongst the 18 children, many had ambitions of becoming a lawyer/attorney, doctors; attend university, a secretary, police officer, and a teacher.

16 of the 18 students either have a cell phone or access to a cell phone (from a family member) or only use it for texting or talking on the phone; none play the games on the phone.

When asked about what kind of games they played, students listed board games such as checkers to several Playstation games such as soccer (FIFA), Boxing, racing games, or some computer games, as shown in Figure 10.

Figure 10 Games most commonly played by the participants (clockwise), Digital: Boxing, Car-racing, Guitar Hero, and Mario. Board Games: Checkers

There were a small number of students who didn't play games at all, too. All except one student owns a computer or has access to a computer. Some do not have Internet access and sometimes play games such as solitaire, or do not use it at all (or it's broken).

Media exposure for all 18 students includes mainly the television and movies. All watch Spanish content and some watch English content with Spanish subtitles. For those who did watch English content without subtitles, they admitted that they did not understand the lyrics. For music, all the students listened to at least Spanish music. For those who listened to English music, genres included hip-hop, pop, country, and a diverse range; a few students only listened to English music because of the music and did not understand the lyrics.

When it comes to learning English, all the students pointed out vocabulary acquisition and pronunciation/speaking as their key issues; other issues were reading and writing.

All the students except one recognize the importance of learning of English, so they can attain better job prospects and communicate better. However, the teachers also mentioned that there is some resistance to learning English because these students are surrounded by a community of other Spanish-speaking peers and lowers their incentive to learn. These students also mentioned peer pressure because they did not want to sound silly when they mispronounce English words. Evidently, there are issues with intrinsic motivation.

While lack of intrinsic motivation is a discouraging factor, the extrinsic motivation is also lacking. While, the children want to succeed and aim high for their life, there are not many good examples available in their close-knit community of success, as suggested by them. The lack of motivation by the family compounds the problem.

Furthermore, for illegal immigrants, avenues for higher education and professional growth are virtually non-existent. This reduces the motivation of some of the students to try harder because they know that they will eventually get low skilled and low-waged jobs like their parents.

5. Design

This sections describes how I designed the study and the associated apparatus and content for a successful implementation. I begin by explaining the current curriculum taught at the school to the user group and how I derived a syllabus for the study. Next I explain the methodology behind the game designs and end with a description of the implementation and system design.

5.1 Curriculum Design

A student in the ELL Level 2 spends roughly 3 hours in the ELL classroom daily. This includes instruction and teaching, drills, practice sessions, silent readings, and tutor-time. I developed the curriculum worth teaching 7 percent of the entire vocabulary, for the entire academic year, in about 10 minutes session once a week after discussing with the ELL teacher for the class. This represents quite a negligible self-learning time. This is explained below

1 Week in Traditional Approach

spends hours daily

In 5 days, this is 15 Hours = **900 Minutes**

1 Week in SPRING

spends hours (10 Minutes X 2 times/Week)

In 5 days, this is **20 Minutes**

So, in 2.22% extra time, student learns 3.33% extra words

The students at ELL Level 2 at the chosen school attend classroom teaching by an experienced teacher, aided by audio-visual media to improve the attention and understanding. They follow the curriculum designed according to the textbook “Milestones California Edition” as shown in Figure 11.

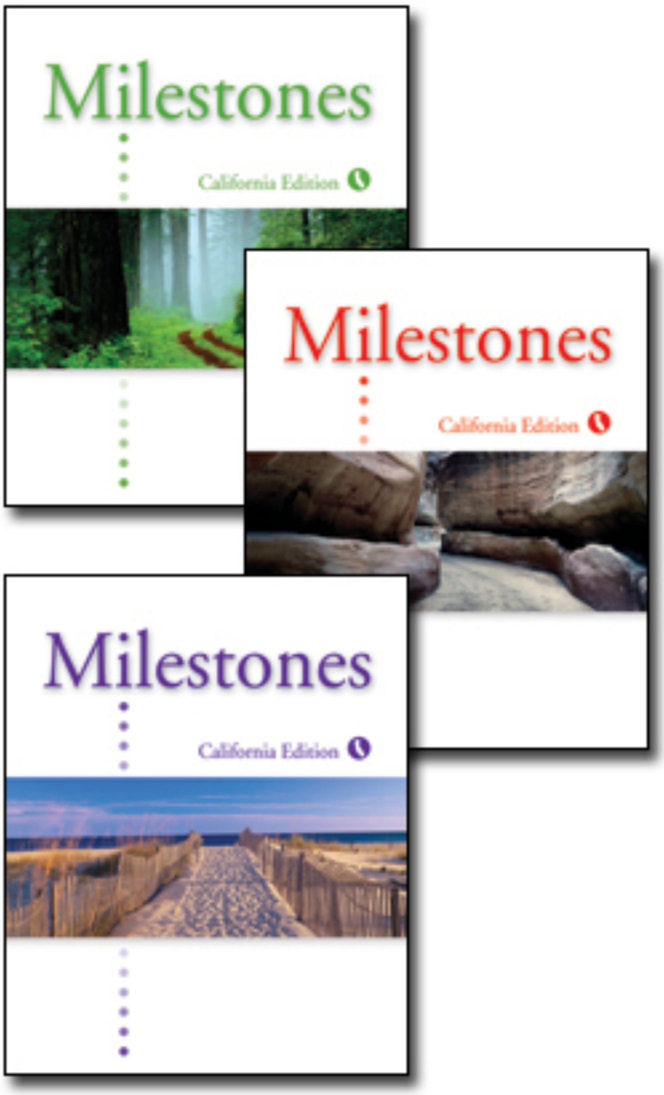


Figure 11 Milestones Books, student Editions A, B, and C

The book is divided into six units, each describing a different facet of life like “Family Connections”, “Dreams”, and “Survival” etc. This curriculum is heavily based on reading, vocabulary, and grammar lessons in content, and exercises. However, it offers limited opportunities to speak English formally, as shown in the Figure 12 .

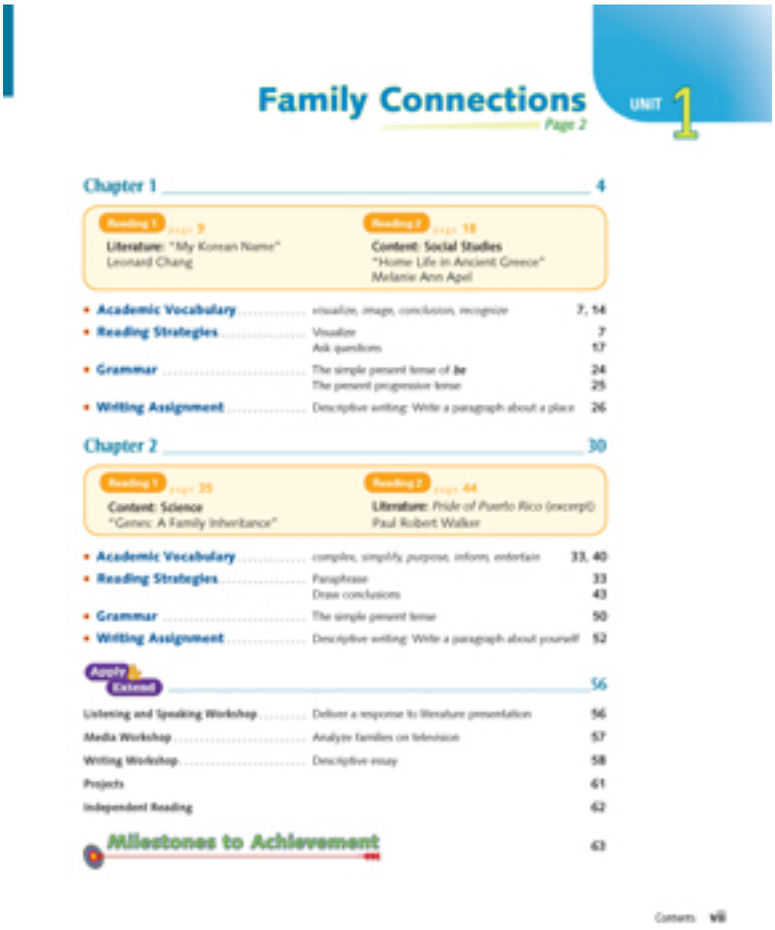


Figure 12 Chapter Family Connections showing the skills covered in the book but missing pronunciation

Although, the official language of the class is English and students are expected to converse in English with each other, most of them instead choose to converse in their native language, Spanish. They tend to do so because it is easier and more comfortable for them to speak in Spanish than English. On the other hand, each ELL level requires a certain minimum level of knowledge of English vocabulary. These words are discussed in the class but spoken & pronunciation correction drills of these words do not happen at the class or an individual level. The only opportunity that these children have at listening these words are when used by the teacher in the class during the discussions.

The “Milestones” book includes a list of about 300 words from the 6 units that are expected to be known by the students at the end of the academic year. I divided these 6 units into 3 parts:

Group A: Units 1 and 2, which had been taught by the teacher in the class before I began the study;

Group B: Units 3 and 4, which were being taught during the study;

Group C: Units 5 and 6, which had not been taught during the duration of the study.

I randomly chose 10 words out of each Group (A, B, and C), giving 30 words, a 10% sample set out of the pool of 300 words and created a syllabus of the study based on them. The aim to divide the words into the groups was to investigate if the games caused significant deviation between the learning gains of pre-existing knowledge (Group A), or unknown knowledge (Group C), or aided what is being taught (Group B). Some of the words in the sample set included Menacing, Attic and Soggy etc. These are tabulated in Table 6 below.

Table 6 List of words used for SPRING

Unit 1 & 2	Unit 3 & 4	Unit 5 & 6
Glowed	Wagon	Anchor
Mothball	Violence	Wreckage
Swirling	Carpenter	Soggy
Attic	Chimney	Haul
Circular	Eagle	Oar
Weapon	Barren	Menacing
Leap	Cupboard	Gnawed
Overflow	Bear	Locomotive
Crouch	Porcupine	Equipment
Peek	Curling	Aluminum

The study was designed to test the pronunciation ability of this sample set of words by the users, teach the users how to pronounce those words using a game, and then finally testing to detect the effects, if any.

5.2 System Design

The entire game logic for both the games was written in flash Action script. The game and the speech recognizer was eventually deployed an Ubuntu Linux installation.

Details of the individual pieces are as follow:

Speech recognizer

For the purposes of the speech recognition, the CMU Sphinx-III speech recognition engine was used. However, instead of using it in decoding mode it was used in forced-alignment mode.

In force-alignment, rather than being given a set of possible words to search for, the search engine in the recognizer is given an exact transcription of what is being spoken in the speech data.

The system then aligns the transcribed data with the speech data, identifying which time segments in the speech data correspond to particular words in the transcription data.

This mode worked well with the game concepts because at any point in the games, the word for which the participant was expected to give an utterance was always known. Another reason for using the force-alignment mode was to be able to obtain scores at the level of individual phonetic units. This information was then used to point out which part of a particular word was uttered incorrectly.

Speaker adaptation

An issue with speech recognizers is that their outputs (acoustic scores) vary with the texture and accent of voices.

Since, the games had to give feedback after comparing to standard American accented pronunciations, the recognizer was trained on large corpuses of data (the total size of the files used for training was more than 15GB in raw format) from American accented speakers. It was known that Spanish accents would be scored strictly on such a setup, but training the recognizer on Spanish accents would have defeated the purpose of the experiment, which was to bring the pronunciations of the participants closer to actual American pronunciations by making some of their phonetic sounds more understandable.

However, the change in texture from a male to female voice was accounted for. Audio utterances from 2 American males and 2 American females were recorded and MLLR (Maximum Likelihood Linear Regression) transforms were used to adapt the recognizer to male and female voices as and when required.

Data exchange between game and speech recognizer

Since the recognizer code was in C and the game logic was written in Action script, socket connections were used to make the game exchange data with the recognizer.

Some Perl and Python scripts were also used to implement some low level I/O that was required to store and maintain data logs of the user actions, audio files of pronunciations by the users, and the generated scores for these pronunciations.

Feedback routine

The recognizer could generate acoustic scores, but they had to be compared against standard American accented pronunciations, before giving feedback to the participants on how they did on a particular word utterance.

Therefore, a library was coded in Action script, that read from a large pool of acoustic scores generated from the training data and created a normal distribution out of them. The library also plotted the acoustic scores for the word utterance from the participant on this normal distribution, and returned back a Likert scale (1-3) rating for each phonetic unit in the word under consideration. This rating could then be used to give feedback to the participant.

Graduated interval recall

The game logic for Voz.Guitar was implemented in a way that the syllabus queue was chosen according to a well-established algorithm called Graduated Interval Recall. (Pimsleur, 1967). The algorithm helps in determining the order of the questions, given a syllabus.

It is modeled in a way that performance on a particular question determines the number of times it will be posed in the near future. It is commonly used to cause long-term retention of syllabus items.

The algorithm recognizes the lowest scores for each word-trial by the users and compares with the pending words in the queue. After one pass of the vocabulary is made, the second pushes the lowest scoring words to the front of the queue and continues to iterate in this fashion by showing the lowest scoring words much more frequently and omitting the successful words from the queue slowly over the multiple passes.

The game concept of Voz.Guitar had an aspect of repetition, as opposed to Zorro, which allowed the player to explore an exciting but static and pre-defined game world. Therefore this algorithm was used just for Voz.Guitar and not for Zorro.

However, the lack of repetition in Zorro was countered by making the participants play the game again. Moreover, it was ensured that the time for which the participants are getting instructed (also playing the game) stays constant across the two games.

System Architecture

An overview of various system components described above is shown in Figure 13.

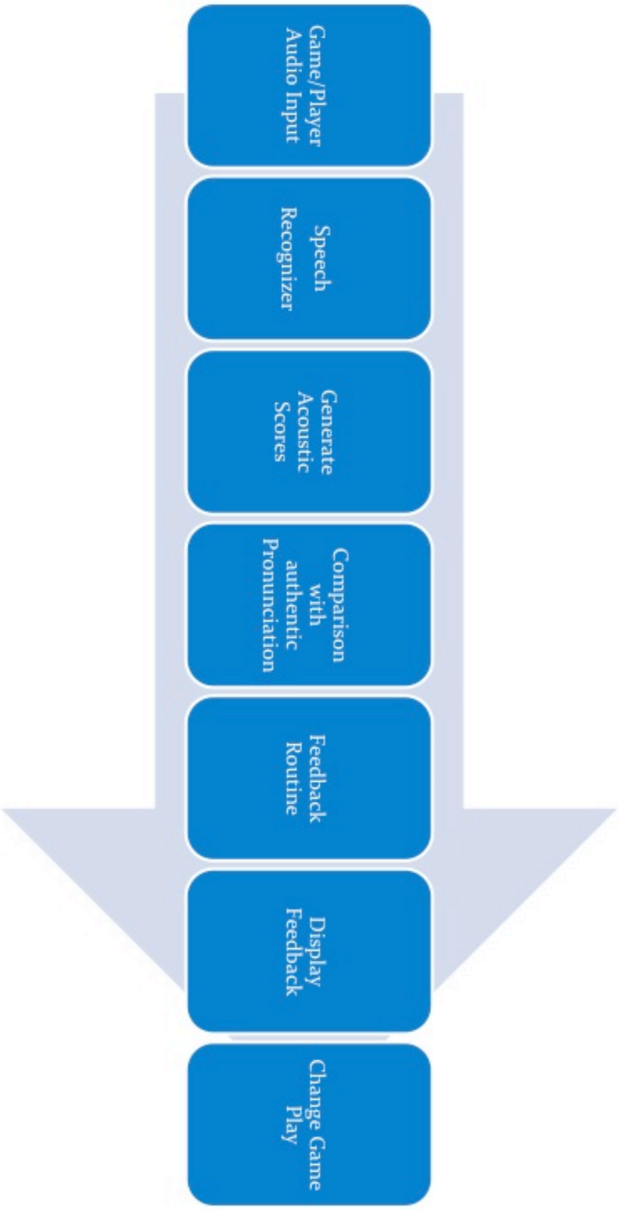


Figure 13 System overview of the multiple components

5.3 Game Design

The aim of the study was to design, and create games, enriched with pedagogy that might motivate the players to play them, despite the challenges posed by the learning material in the game.

So, I based the design of the games on the following resources:

1) Demographic interviews of the children clearly indicated a penchant for certain types of games.

2) Popular and best selling commercial software available in the market.

This gave me an advantage of creating games that were likely to peak interest of the children while they incorporated the best practices of game design and elements from existing games.

As was evident from the Demographic interviews, mobile phones were not a popular choice of the system for the participants to play games on. So, I had to move away from the initial proposal about creating mobile games, and instead focusing on creating games that could be played on rather larger displays like a computer screen or a television.

Using this knowledge, I decided to create two games: Zorro (based on Mario), and Voz.Guitar (based on Guitar Hero) for chiefly the following reasons:

1) **Activity based vs. Arcade based:** The demographic interviews pointed out the predilection for two different genres: card games and action games. However, in either genre, the children preferred fast paced, non-time restricted gaming sessions.

2) **Novel vs. Comfortable:** I based the design on two popular and proven games Mario and Guitar Hero. The demographic interviews indicated the previous playing experience of most of the participants with Mario, while only one knew about Guitar Hero. So, I decided to give them a mix of a comforting known game and a novel, and hopefully exciting, game.

3) **Adaptive vs. Non Adaptive**: I chose Mario based game because it is non-adaptive and gives a consistent experience of play, with onus on the player to act fast. On the other hand, Guitar Hero based game was adaptive and had an element of surprise.

Both games followed the principle of teaching, drill, immediate feedback, scores, and repetition. Both games feature the word, associated playable American accented female voice, and spelled-out-pronunciation to aid the users. The spelled-out-pronunciations were obtained from the online dictionaries [online dictionary] and then modified accordingly by a trained linguist with five years of experience. For example, the words used in Zorro can be seen here in the Table 7 along with their spelled out pronunciations.

Table 7 The words used in Zorro, with their associated spelled-out-pronunciations

Word	Spelled-out Pronunciation by Dictionary.com	Spelled-out pronunciation corrected by the linguist
Circular	sur-kyuh-ler	s r-kyih-l r
Swirling	Swurl-ih-ng	sw r-lih ng
Mothball	mawth-bawl	maw th-baw l
Overflow	oh-ver-floh	oh-v r-floh
Wreckage	rek-ij	reh k-ih j
Aluminum	uh-loo-muh-nuh m	uh-luu-mih-nuh m
Menacing	men-is-ng	meh-nih-sih ng
Equipment	ih-kwip-muh nt	ih-kwih p-mih nt
Locomotive	loh-kuh-moh-tiv	loh-kuh-moh-tih v
Glowed	gloh d	gloh d

As is evident, the spelled-out-pronunciations of the words are not the same between the online versions and the versions created by the linguists. The reason for this is that the online version is based on an accent-less pronunciation. However, this pronunciation would not have matched with the voice of the Californian female used to enunciate these words in the game. Hence, to ensure the consistence between the spoken sounds and the expected pronunciations, a version produced by a trained linguist was necessary.

Zorro

Zorro is a character based arcade game, which involves moving Zorro, the main character, of the game from left to the right of the scene using arrow keys until he reaches the castle. As shown in the Figure 14 on the way, he encounters five closed chests, dangerous animals, tricky terrain, and obstacles, which must be overcome.

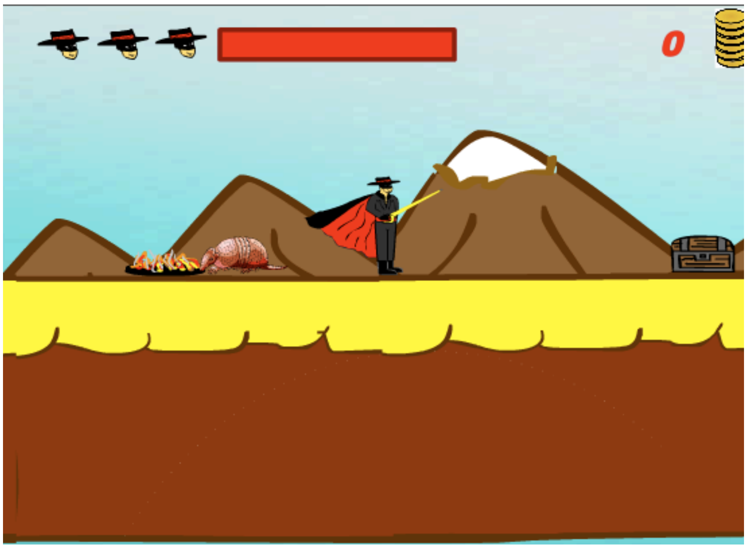


Figure 14 Zorro game showing Zorro, controlled by the keyboard by the player

The obstacles can only be overcome by opening up the chests. Each chest contains a word, associated pronunciation, and the associated audio pronunciation coupled animated spelled-out-pronunciation as shown in Figure 15.

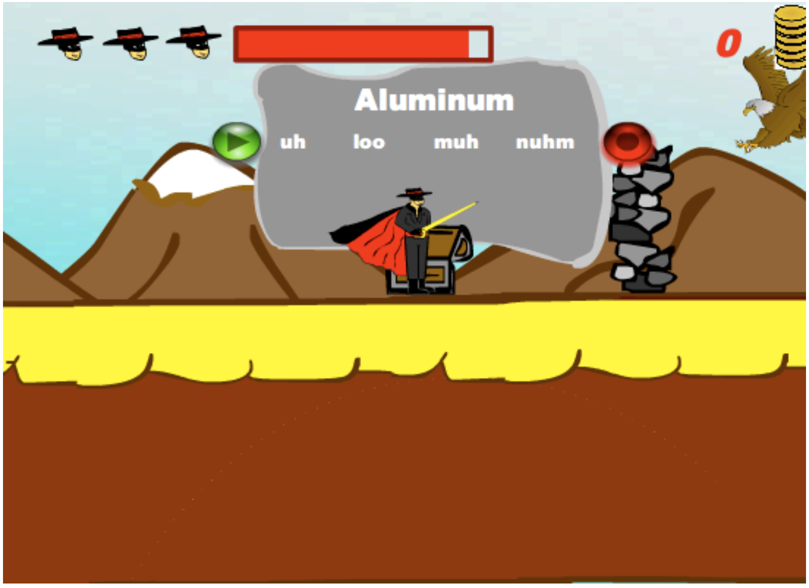


Figure 15 The chest showing the word Aluminum, the spelled-out-pronunciation, the play button, and the record button

The word is pronounced three times every time it is played. Next, the user gets an opportunity to record their pronunciation of the word by the click of a button.

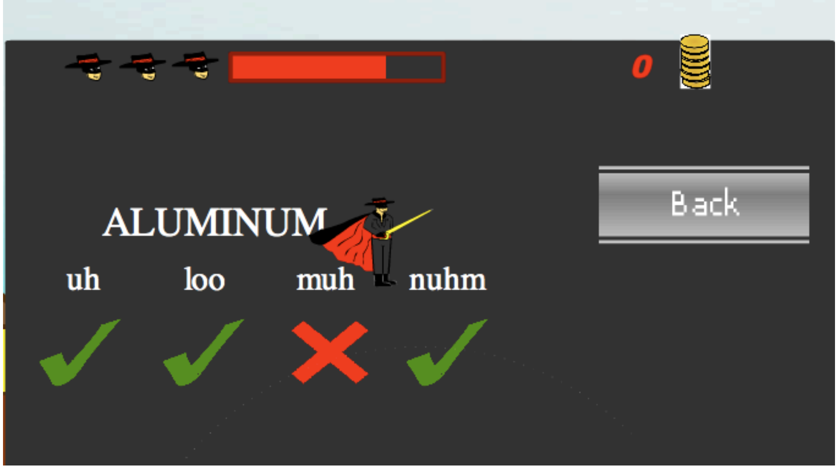


Figure 16 The feedback screen displaying Ticks as the right pronunciations and Crosses as the wrong Pronunciations

A feedback screen as shown in Figure 16 showing the correct and wrong parts of the pronunciation, and the associated score follows this. She also hears her own pronunciation and the intended pronunciation. This is important because she knows the exact part that she committed mistake at (for example at “muh” here). She also hears what she said, and she also hears what she should have said – giving an opportunity to hear again the whole word while focusing on the incorrect utterance.

After crossing the five obstacles by practicing the five words and avoiding the deadly animals, the user gets to win the game. In case, she finished short of 10 minutes, she is obliged to play the game again.

The different game elements of Zorro have been briefly summarized below:

Zorro can be moved left, right, and up using the corresponding arrow keys on the keyboard. It has 3 lives shown by 3 icons at top left corner.

Health Bar shows the Health of the Zorro.

Animals like and Birds are in constant motion while Fire Pits stay put on the ground. Zorro must avoid touching them to save the Health. Different Animals follow different paths to create challenging situations.

Chests need to be opened using Space-bar to see and ▶ listen to the word Aluminum uh loo muh nuhm, followed by ● self-pronunciation and get a Feedback screen with ✓ and ✗ for correct and incorrect pronunciations, respectively. Correct pronunciations earn Coins and are shown at the top-right corner of the screen as the scorekeeper of the session.

Obstacles cannot be crossed until the chest before them can be crossed. There are five obstacles in the game: Rock Wall, Bridge over Spears, Water Wave, Fire, and the Castle. Touching them may decrease the Health.

Landscape changes over time to create challenges and keep the game fun to play.

Since, the aim of the game, is not to hurry the participants, it is not timed – instead, the in-built game elements as described above coupled with the health ensure that the participants do not idle away, lest they might get killed by a bird or an animal. The students continue to be motivated by their scores and are encouraged to go back and try a wrongly pronounced word again to better their scores.

Voz.Guitar

Voz.Guitar is an activity-based game that displays the word, associated spelled-out pronunciation, and plays the associated pronunciation.

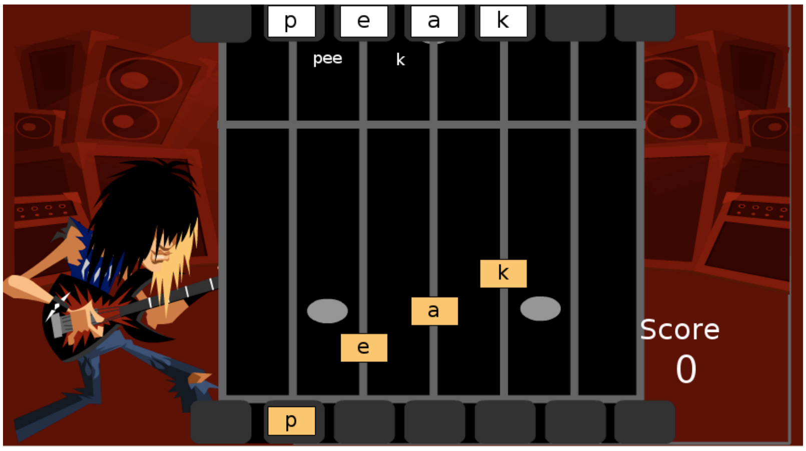


Figure 17 Voz.Guitar showing word peak falling down

As shown in the Figure 17, next it allows the users to hit the falling alphabets of the words at the right time. Next, the user is obliged to pronounce the word after the counter shown in Figure 18 is over.

The words taught by Voz.Guitar included the following:

ANCHOR, GNAWED, HAUL, OAR, SOGGY, LEEP, PEEK, CROUCH, ATTIC, WEAPON.

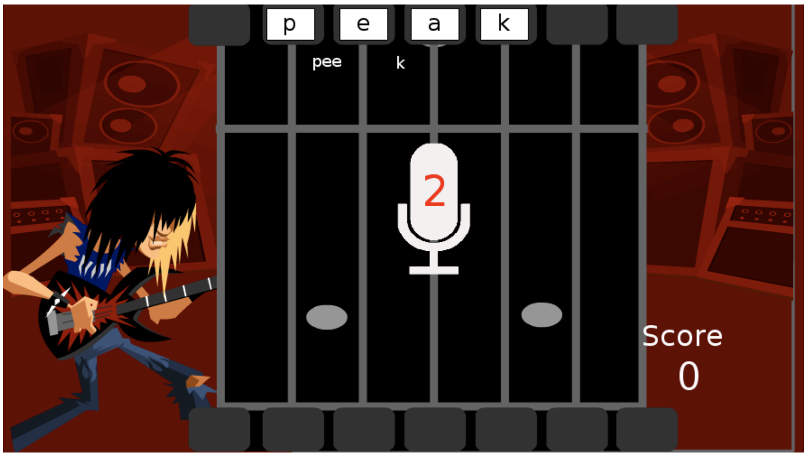


Figure 18 Voz.Guitar Mic Counter prompting the user to speak the word

The feedback screen, shown in Figure 19, displays the hits and misses in the spelled-out-pronunciation and corresponding errors, shown by the smiley. The user hears her pronunciation followed by the intended pronunciation.

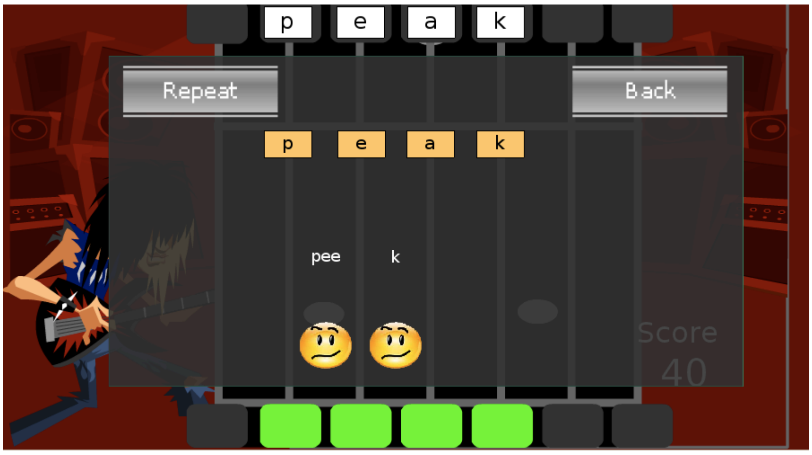


Figure 19 Feedback screen in Voz Guitar with Smiley

The game is adaptive and hence, tends to automatically repeat the words, which have not received a satisfactory pronunciation response from the players. Each positive utterance increases the score of the users.

The session continues until the time limit of the session reaches.

These games were created after three iterations of rapid-DIA cycles and usability testing. While the essence of the games stayed the same, and had to be kept constant to prevent any changes from affecting the results of the study, some issues had to be resolved and paid attention to. These are described in the next section.

5.4 Evolution of Games

During the deployment of the games, notes and video recording were made to understand the challenges faced while designing games that included pedagogical concepts also. The videotapes were transcribed to observe the usability issues faced while playing these games and to modify them in subsequent versions. These issues are discussed here and may serve as pointers to any future pedagogical game designs:

Reduce Game Elements when Pedagogy kicks in

In Zorro, when the chest is open, disable fire pits, animals, birds etc. It was originally thought that this would make it more challenging, but led to loss of interest after continuous deaths due to decrease in health of Zorro by touching these elements. Also, it might be prudent to ensure that the user stands that the game is frozen while the learning is happening by providing visual cues and restricting game based controls during learning.

So, it is important to mute features that distract the attention away or cause unnecessary anxiety while the learning is happening. The game can still be made challenging by tweaking other patterns, for example like creating the path of these “harmful” elements dynamic – rendering a challenging situation within the game, once the learning is over.

Motivation using in-built game pedagogy elements

In both the games, while a good pronunciation was being visually shown to the users – there was a lack of appreciation. Adding auditory cues like “Good Job!” or a clapping at the end of correct pronunciations or “You can do better!” at the end of incorrect pronunciations might help improve the experience and probably, the motivation as well.

Also, the visual elements can be made better. For example, one of the male participants suggested brining the castle down at the end of Zorro or blasting the words in Voz.Guitar at the end of good pronunciations is more satisfying and visually enriching to suggest an accomplishment.

Feedback at the end of session

The games were changed to provide a possibility for the users to speak one last time all the words learnt in the game to be able to finish it. While, this captures the learning effects at the end of the game – but does not provide a possibility to know the effects at the end of a session after playing multiple games.

So, a script was written to deduce the changes between the attempts in the first play and the last play to document the changes. At the end of the session and after the post-session interviews, the users were shown the changes. It was important to do this after the post-session interviews to ensure consistence in the responses in the interviews.

5.5 Study Sessions

As previously mentioned, the study was designed across three groups of words, for two sample sets of population: Control Group, and Experimental Group.

The sessions lasted for around two hours per day, three times per week, and four weeks in a row. There were two types of sessions: Pre/Post Test sessions, and Learning Sessions.

A 2-hour Learning session was typically structured as follows: preparing the game database with the pre-determined group of words (A, or C), arrival at the school premises, arrangement and setup at a quiet location in one of the pre-arranged labs, greeting with the teacher, a list exchange of the students needed for that day, escorting a student to the lab, explanation of how the game is played, the goals, and a demonstration, the gaming/learning session for 10 minutes, a post game session qualitative interview, escorted return of the student, and bringing back the next student.

A Pre/Post Test session involved the same as above except the student faced the test instead of the gaming session.

6. Quantitative Observations

6.1 Metrics

Before explaining the quantitative findings from the experiment, the metrics that used to gauge the change in pronunciation need to be defined. The following two metrics were used:

Acoustic score gain percentages (ASGP)

These were numerical scores generated by the CMU Sphinx-III speech recognizer.

A batch decoding of all the audio samples (pre-test and post-test) that from the participants and generated acoustic scores to quantify the quality of pronunciation was performed. It should be noted that the acoustic scores take all aspects of spoken language into account (like intonation, fluency, clarity etc). Moreover, the acoustic scores were generated for each phonetic unit in a word, and hence judge the actual quality of each phonetic unit. These individual phonetic unit level scores can be added together to generate word level scores.

As stated earlier, the speech recognizer was trained on American accented speaker and the scores were generated in comparison to that. So any increase in acoustic scores meant a positive change in quality of the utterance and it also meant greater proximity to the actual American pronunciation of the word.

The differences in the total acoustic scores across all the words, from the pre-test to the post-test were measured. Then this difference was expressed as a percentage of the total pre-test score across all words for a particular participant. Since acoustic scores are log probabilities, they tend to be large negative numbers, which are hard to understand, and that was the reason percentages were used to represent the gain.

Word gain (WG)

The recognizer only decoded the audio utterances that were close to the actual word. If the utterances were sounds that the recognizer was not trained on, it would reject the same.

Since the recognizer was trained on the sounds of the words in the curriculum, it readily recognized any utterance that was close to the actual sound of any of the words in the curriculum. Therefore, the word gain was nothing but the difference in the number of words that the recognizer could decode during the pre-test and the post-test.

In simple words, this metric is a high-level representation of the number of words a participant learned to pronounce (with acceptable pronunciation) during the course of the experiment.

It should be noted that I had initially divided the 20 words in the curriculum (that was taught), into two parts. As explained earlier the first part came from pool of words they had already encountered in class and the second part came from pool of words that were completely unfamiliar to them.

When the post-test and pre-test data was analyzed, it was found that the correlation between the category (familiar or unfamiliar) of the word and average gain on the same over the duration of the experiment was negligible.

Quantitatively speaking, the correlations between the average scores (both ASGP and WG) across all participants and the category (familiar/unfamiliar) was <=0.27 for all the 20 words.

Moreover, this was true for both, control and the treatment group. Hence it was decided to group the results together, and analyze the gains across all the 20 words.

6.2 Post-Test Gains

In each experiment, a standard statistical *t-test* was used to compare the gains of the treatment and the control group. This test yields a *p*-value indicating how significant the difference is between the means of the two groups. A two-tailed t-test on the pre-test scores of the treatment and the control group yielded a p-value of 0.25, which shows that there was not a statistically significant difference between the means of the two groups before the start of the experiment.

Acoustic Score Gain Percentages (ASGP)

After the post-test, the mean acoustic score gain percentage for the control group was -0.68 (σ=2.77, n=9) and that for the treatment group was 1.41 (σ=1.72, n=9). The ASGP are small numbers because they are percentages of total pre-test scores across 20 words (more than 110 phonetic units).

However, a two-tailed t-test between the ASGP for the control and the treatment group yielded a statistically significant p-value of 0.08. Table 8 lists the ASGP for participants in the control and treatment group.

Table 8 ASGP Scores for the Control and the Treatment Group

Control Group		Treatment Group	
CG1	1.24%	TG1	0.26%
CG2	-1.72%	TG2	0.67%
CG3	1.15%	TG3	1.02%
CG4	0.71%	TG4	-0.13%
CG5	1.02%	TG5	3.14%
CG6	-7.40%	TG6	1.32%
CG7	-1.68%	TG7	-0.12%
CG8	0.71%	TG8	1.43%
CG9	-0.12%	TG9	5.14%

It should be noted that a negative percentage denotes that the participant’s total acoustic score for the post-test was lower than her acoustic score for the pre-test, and therefore the increase was actually negative. Overall, however the effect was positive as shown in Figure 20.

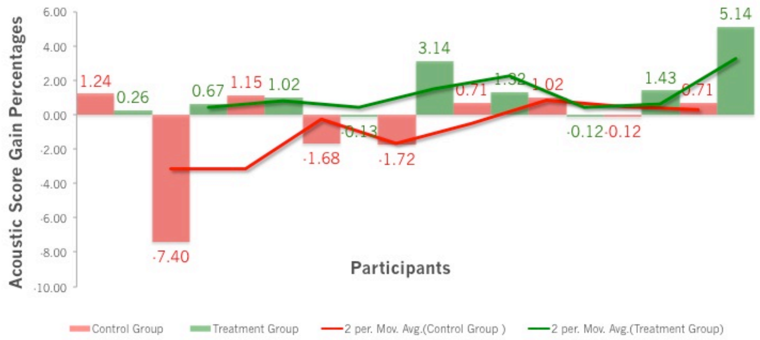


Figure 20 Contrast between the change caused by SPRING in ASGP scores of the Control and Treatment Group

Moreover, there are only 9 participants in both the groups (18 in total), as opposed to the 20 enrolled in the study. This is because of attrition due to children dropping out of school. I had the pre-test data for both the participants, but they had dropped out of school by the time the study ended, and hence were unavailable for the post-test.

Word Gains (WG)

After the post-test, the word gain scores had a mean of 0 (σ=0.71, n=9) for the control group and a mean of 1.11 (σ=1.54, n=9) for the treatment group. This means that the participants in the control group on an average did not learn to pronounce any new words, from the 20 syllabus words, whereas the treatment group students on an average learned to pronounce (to acceptable levels) more than one new word across the duration of the experiment. This gain was in addition to the improvement in the quality of the pronunciations that is represented by the ASGP. Tables 9 and 10 list out the words attempted in pre-test, post-test and the resulting WG for the control and the treatment groups.

Table 9 Word Gain (WG) for the Control Group

Participant ID	Number of words attempted in pre-test	Number of words attempted in post-test	Word Gain
CG1	13	14	1
CG2	15	15	0
CG3	16	15	-1
CG4	12	12	0
CG5	16	16	0
CG6	17	16	-1
CG7	19	19	0
CG8	12	12	0
CG9	17	18	1

There wasn’t a significant difference in the number of words the control and the treatment group could pronounce to some extent at the start of the experiment. The t-test on the number of words attempted at the start of the experiment yielded a value of 0.42.

Table 10 Word Gain (WG) for Treatment Group

Participant ID	Number of words attempted in pre-test	Number of words attempted in post-test	Word Gain
TG1	20	18	-2
TG2	19	20	1
TG3	16	17	1
TG4	15	16	1
TG5	12	15	3
TG6	13	14	1
TG7	17	20	3
TG8	19	19	0
TG9	15	17	2

However, the t-test on the number of words attempted at the end of the experiment (by the control and the treatment group) yielded a value of 0.06, which is statistically significant. It also points to a possible confidence boost during the study in terms of pronouncing less familiar and more complex words, as shown in the Figure 21.

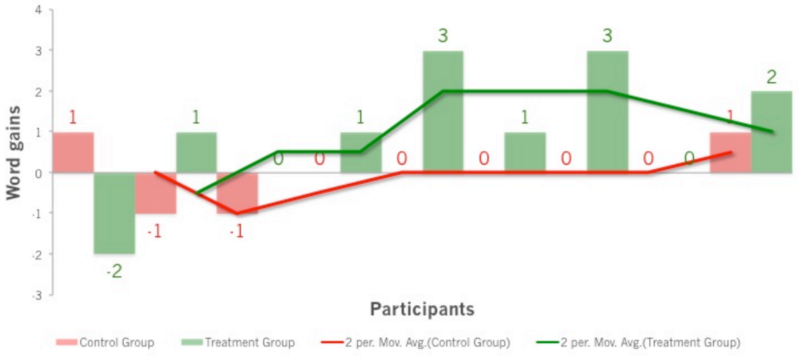


Figure 21 Changes in the Word Gain (WG) and contrast between the Treatment and the Control Group

Moreover, a two-tailed t-test on the WG values for the control and the treatment group yielded a p-value of 0.07. This shows that there was a statistically significant difference between the WG of the control and the treatment group.

To summarize, on an average, all the participants improved in terms of the quality of the pronunciations of the words in the syllabus. An additional consequence was that on an average, all the participants learned to pronounce some new words.

In a way SPRING affected both, intelligibility and active oral vocabulary of the users.

6.3 Gender related findings Post-Test Gains

As stated earlier, the control and treatment group had the same distribution in terms of gender. Both the groups contained 3 boys and 6 girls. Since the distribution was uniform across groups I also did some analysis to quantitatively measure the influence of gender on game play and learning.

The correlation between gender and ASGP for the control group (0.65) suggests that boys performed worse than the girls overall, over the period of the experiment. However, the correlation between gender and ASGP for the treatment group (0.32) suggests that gender did not influence the improvement in pronunciation quality that was exhibited by the participants, after playing the games.

This is in contrast to the findings from similar ESL (English as a Second Language) acquisition studies in the more underprivileged parts of the world. (Matthew Kam, 2009)

6.4 Effects of pre-test on post-test gains

The correlation between pre-test scores and ASGP for the treatment group was 0.11 and the correlation between pre-test scores and WG was 0.25. This shows that the participants in the study showed similar learning gains across both metrics irrespective of their performance on the pre-test.

Therefore, there was no notion of bimodality as suggested by similar ESL acquisition studies in developing parts of the world. (Matthew Kam, 2009) This might be happening due to various different factors like better ESL levels, prior exposure to technology, and access to education. It is imperative to explore these factors in greater detail in the next few iterations of the research.

6.5 Learning gains during game play

Though the post-tests conducted were delayed post-tests, as they were conducted a few days after the actual syllabus was taught, data logs of how the participants performed during a session were also collected.

Across a total of 10 (one of them dropped out of the school before the post-test) participants and a total of 40 game sessions the treatment group exhibited an average ASGP of 12%.

Calculating the differences in acoustic scores of the first and last instance of a particular word in a single game session and averaging it across all participants in the treatment group resulted in these percentages.

Though not much value can be attached to this measure, it definitely shows that there is a large window for future work, which would include efforts towards converting these short-term gains into long-term gains.

Lehr-und Forschungsgebiet Informatik 9
„Computerunterstütztes Lernen" an der
RWTH Aachen University

7. Qualitative Observations

In addition to the quantitative sources of data, the videos also served as an important part of the analysis. Approximately 600 minutes (10 hours of video) were recorded. After transcription and qualitative coding of the data the following major qualitative findings were deduced:

7.1 Player Profiles

Through the duration of the study, several key distinctions in the pool of subjects were observed. The first major separation appeared in gender difference.

The females appeared to be indifferent to the game play and were more focused on the speech/voice features of the game. Some used two hands to move Zorro on the arrow keys (which showed less exposure to gaming/computer); most used one fingers to type in the key-interaction in Voz.Guitar. Females also needed more assistance with the games compared to males, whether it were additional verbal cues or helping them with the obstacles in the game.

When a translator was used for one female, the two put together were more engaged with playing both Zorro and Voz.Guitar; the two laughed, gestured and were more focused on game play in addition to the speech/voice features.

The males were more focused on game play than the speech/voice features; for example, when they opened the chest with Zorro and the feedback screen appeared, the males were still playing with the Zorro character (trying to move it around). When the males did interact with the speech/voice features, they said the words with more confidence than the females and had less stuttering and hesitation.

For both the male and females, they exhibited a certain learning curve when playing, and that was true for both games. They both did not understand the feedback screen when there were happy and sad faces, and they did not pay attention, did not care, or did not even read the phonetic mimes/breakdowns of the word.

Also, almost none of the players used the "repeat" feature in the Zorro game, as opposed to Voz.Guitar that forced them by having them go through each word again.

In addition, although both genders found the games entertaining as a whole, they did occasionally display gestures of frustration including rolling their eyes and hand waving to brush off mistakes.

It was felt that these gestures were partly attributed to general game playing and demonstrate the student's attention and involvement in the game, which is a positive factor. It was also felt that these gestures were at times due to usability problems in the games, whether it be confusion on directions or displays.

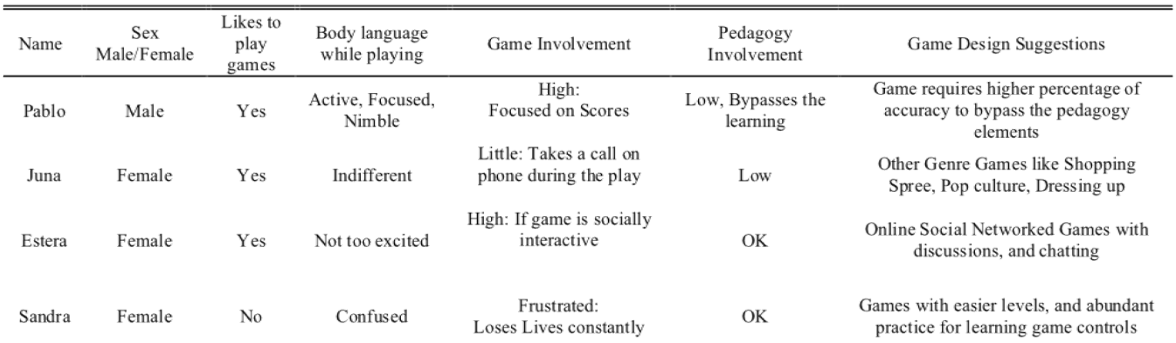
Accordingly, modifications were made to improve the games. However, all of these modifications were minor fixes, usually pertaining to the aesthetic organization of the game.

I further broke down the subject pool and found four specific player profile classifications in the subject population. They are represented by the following four names: **Adam, Juna, Courtney, and Sandra**. These profiles have been summarized below in Table 11.

Adam represents a teenage male who regularly plays computer and video games; hence he is quick and adept with the keyboard, character actions and movements, and overcoming obstacles. When he plays the Zorro game, he leans forward, fully focused and his fingers ready to move Zorro away from attacking animals. His focus is primarily to maximize his score. He is extremely involved in the game aspect of the process and less so in learning pronunciations.

Thus for future research, to better incorporate these profiles, modifications should be done to the game to stress the learning and education portion. For example, the game will require the player to achieve a certain percentage of accuracy before allowing him to proceed instead of the current model where simply saying the word is sufficient in passing that chest. The game screen could also be faded to emphasize and reduce distraction from word pronunciation portion.

Table 11 The Player profiles deduced from the video analysis and should be kept in mind while designing pedagogical games



Name	Sex Male/Female	Likes to play games	Body language while playing	Game Involvement	Pedagogy Involvement	Game Design Suggestions
Pablo	Male	Yes	Active, Focused, Nimble	High: Focused on Scores	Low, Bypasses the learning	Game requires higher percentage of accuracy to bypass the pedagogy elements
Juna	Female	Yes	Indifferent	Little: Takes a call on phone during the play	Low	Other Genre Games like Shopping Spree, Pop culture, Dressing up
Estera	Female	Yes	Not too excited	High: If game is socially interactive	OK	Online Social Networked Games with discussions, and chatting
Sandra	Female	No	Confused	Frustrated: Loses Lives constantly	OK	Games with easier levels, and abundant practice for learning game controls

Juna is a teenage female who has experience playing and watching others play computer and video games. She is familiar with the goals, obstacles, and other formal elements of the game, but is indifferent to her performance and score. While playing, she would take out her cell phone to text a friend, and then resume the game play. I believe that her indifference could be attributed to the type of game she played.

More specifically, both Voz.Guitar and Mario turned more male-oriented. Although, female voice recordings for all audio playbacks in the game were used, but that did not help a lot in this direction.

Thus, in order to spur her interest in playing games to learn pronunciation, female interest games such as ones involving shopping sprees or pop culture should be introduced to this player instead.

Courtney, like Juna, is also familiar with game playing, however, she prefers multi-player games such as chess and Mario Kart. She seeks social interaction. When she plays, her best friend is looking over her shoulder, and they both giggle at the animals or when Zorro falls into the water.

Accordingly, for Courtney, a game that involves multiple people where players can compare and discuss their scores, experiences, and findings might be more beneficial.

Sandra, the last player profile, is on the other end of the gaming spectrum. She has never or rarely played computer and video games, so she struggles with the arrow keys and movement, timing, and obstacles. When she plays, she's confused and does not understand what to do with Zorro or which direction to move.

She does not understand many of the basic components of the game such as the goal of avoiding monster characters, walls, and fire. She is often frustrated and loses lives (fails) a number of times before finally completing the game successfully.

To enhance her gaming and learning experience, one could provide demonstrations, practice rounds, and more instruction before having her start playing the game that incorporates the pronunciation portion. One would want to ensure that she first becomes fluent with the keys and arrow motions.

7.2 Learning to use SPRING

As stated earlier, two users said they didn't have cell phones or access to cell phones; the other 16 had exposure to cell phones and computers. They were already experienced with text messaging on cell phones to friends and playing computer games at home.

The users were given a 5-minute walkthrough to introduce them to the game; one user required the use of a translator. I continued by explaining and demonstrating the games one-on-one to each user since certain keys had specific functions (i.e. opening a treasure chest with the s

pace bar); none of them asked any question during the session.

For Voz.Guitar, users were shown that they needed to quickly use the keyboard when the letters fall down. For Zorro, users became adept at the keys and functions by the second or third chest, since they had the task of opening each chest and perform the pronunciation tasks; initially, some would forget how to open the chest and move past it.

For one user, additional verbal instructions and guidance were needed without prompt because facilitators noticed she did not know which direction to move the character in the game and made Zorro jump off the cliff (in the wrong direction) several times.

7.3 Speech

The Sphinx speech recognition capability appeared to work successfully with the student's accents: it was able to accept a majority of words that were pronounced. Based on the feedback mechanism implemented in SPRING, the students learned to recognize which syllables/mimes they were not pronouncing correctly. SPRING also generated narratives that spoke to the students by comparing their pronunciation of a word to the proper pronunciation of a word.

7.4 Curriculum Content

Overall, both Voz.Guitar and Zorro engaged both males and females. Both games required the students to interact with the games to build points, such as keyboard interaction and speed for Voz.Guitar and avoiding fires and attacking animals in Zorro.

However, as stated earlier, it was noticed that males were more interested in game play than the speech/pronunciation aspect of the game; at a treasure chest check-point, the males would make the Zorro character jump up and down to avoid an attacking animal, whereas the females would pay attention to the pronunciation; some females did not have a good notion of Health Points (HP) for Zorro.

However, the two games seemed to be geared more towards males; during one session, a female took out her cell phone to send/read a text message to a friend. This was not intentional, as the choice of games was informed by the interviews in the initial phases of the study.

Moreover, since the quantitative results show that the learning gains were gender independent, it is not a factor that hampered the results. However, an important lesson that was learned was that to create an engaging experience, the game world artifacts should be modified to suite the culture, context and gender. Omission of any of these factors might result in lack of engagement.

7.5 Game Play Sessions

The students played the game for an average of 15 minutes. Voz.Guitar was a game that did not have an end since it would recycle words for the students to pronounce while Zorro had a definite end point after five (5) chests. For Voz.Guitar, it was noticed that some students were tired after the 3rd or 4th repetition of the same word; Zorro did not have this same effect since no words were re-used 3 or 4 times, as 15 minutes of game play was equal to 2 iterations through the complete game.

7.6 Pronunciation Measures

The pronunciation measures were tested and identified using a speech recognizer. Since all the processing was happening off the field and on a dedicated machine, accurate scores were received. However, to bring more credibility and to add more human aspect to the research, I would like to seek help from trained linguists. Moreover, I would want to get Likert scale readings from general American population, like a housewife or a salesman at a supermarket. The overarching goal is to better the pronunciations to a level that is socially acceptable. Using the recognizer for evaluation is the first step, but using human inputs from various different sources would be beneficial.

7.7 Other Findings

In the post-game play session interview, 7 out of 9 participants reported that they felt they were learning pronunciations during the game, the rest said they did not know if they learned.

I also asked who they would want to help them with pronunciations. 6 out of 9 participants said they would want help from both their teacher and SPRING. The rest of the 3 participants said they would want to learn from the game only. Since this is self-reported data, I don't want to attach a lot of value to it, but it definitely points out that SPRING was a pleasant change for a majority of the students.

When asked which game did they like more, 6 out of 9 said they liked Zorro better than Voz.Guitar, the rest of the 3 said the opposite. This was intuitive because Voz.Guitar had a lot of repetition and Zorro was exciting. It would be good to mix these two factors in the next phases of the study. It would have been hard to mix game play and pedagogical concepts right from the start, but now I can use the current phase of study and the design decisions I took to inform the next phase of the study and design.

Lehr-und Forschungsgebiet Informatik 9
„Computerunterstütztes Lernen" an der
RWTH Aachen University

8. Challenges Faced

8.1 Motivation

The community I worked with was a very complex one, because they were a Spanish speaking community. They could go about their lives without learning English, because all their major interactions with people were in Spanish. Therefore, there was little or no motivation for them to acquire English as a Second Language. Through these games, I was trying to break this barrier to entry.

The aim was to develop games that are inherently more engaging and have pedagogical concepts merged into game concepts. Throughout the duration of the study I constantly tried to keep up the interest levels of the students I was working with. This was generally done through interface changes. I made sure that I modify any interface element that causes a loss of perception, or is frustrating to the participants. This required iterative design and rapid prototyping.

8.2 Technical Challenges with Speech

Use of speech had a lot of attached technical challenges to it. As discussed earlier, male and female voices were hard to adapt to, but it was accomplished by using MLLR transforms for speaker adaptation.

Speech recognition systems are generally very sensitive to background noise and environment changes; therefore one had to be careful about keeping the environment constant and stable across various sessions.

I also used a high quality noise reduction microphone to capture audio during game play and during tests, to minimize effects of background noise.

I also modified the recognizer code to ignore noisy readings, so that it is more usable in real world settings. There were instances in the games when I had to generate a narrative that included utterances inputted by the participant.

It is easy to do similar operations with text, but I had to create routines that would play audio files one after the other in a way that it sounds natural.

8.3 Administrative Challenges

I also faced some administrative challenges, which I would wish to share with researchers who are working with similar communities in the developed world. Before starting the research the IRB asked us for an approval for a study from a school.

However, when I approached some potential schools, they asked for an IRB approval. None of the agencies was at fault in this case, but it created a deadlock. However, having or developing contacts in a school administration generally helps in such cases. I also used the final applications as demos to the school authorities, to make sure the research appears to be credible to them. It also helps to develop good relations with local stakeholders like teachers.

The study was only possible because one of the ELL/ESL (English as a Second Language) teachers was excited to see the applications and tried hard to fit the experiment schedule into the class schedule, in a non-disruptive fashion. I also reached a consensus on the syllabus, with the teacher. This ensured that there were no confounding variables influencing the learning gains of the participants.

9. Conclusion and Future Work

This work presents a rare effort and rare approach towards English language learning. It is a rare effort because the focus of the research community, especially the Information and Communication Technology for Development (ICTD) community, has long been the developing parts of the world as opposed to the underprivileged or under-resourced children in the developed world. Though I hope the focus would still stay the same, I believe that this work would be seen as a proof of concept research towards use of technology for development in parts of the developed world. This work investigates the effectiveness and viability of use of speech technologies and games (which have not been used together in the past) in giving pronunciation feedback to Hispanic children. There is hope that the positive results shown in this piece of research would inspire other such efforts in the field and underprivileged communities in the developed world would also start benefitting from the field of ICTD.

As explained previously while the intervention led to statistically significant changes in the level of understandability of the pronunciation of the participants, one should be cautious because these changes might have occurred due to multiple other uncontrolled factors like: extra attention by the researchers, or socially normative behavior to perform better in front of the people perceived to be authority figures etc. Further research on a long-term basis might shed more light in this respect.

As stated earlier, the community I worked with did not necessarily have an intrinsic motivation to learn English. Therefore similar future ventures should involve efforts towards motivational games. Also the use of speech offers a lot of advantages. Emotional analysis of speech can be used to detect the motivation levels or emotional states of the children. This information can then be used to start emotional conversations with the students. Games with conversational agents seem to be a good fit in such cases.

The qualitative results point to some common player profiles that were observed during the game sessions. One would want to cater to all the profiles through the future games. There is a possibility of developing adaptive games, which try to gauge the profile of the player based on her interactions with the game and match that accordingly.

Moreover, I found that a “one size fits all” approach doesn’t work in term of gender. However, the model of any piece of ICTD research should be replicable. Therefore, there is a need for games that have multiple story lines, characters, goals, reward structures and endings.

In such cases interactive fiction seems like a good fit, where story, characters and plots could change based on the personality of the player. The kind of decision he/she takes in a game session would then determine the overall direction of the game. This would result in games that are still “one size fits all”, but are considerate of gender, context and culture.

The qualitative findings suggest that there was some demand for multiplayer or collaborative games in the future. Future research should try to explore implications of speech and games in the domain of shared learning. In such games, the players can collaborate and help each other with pronunciations.

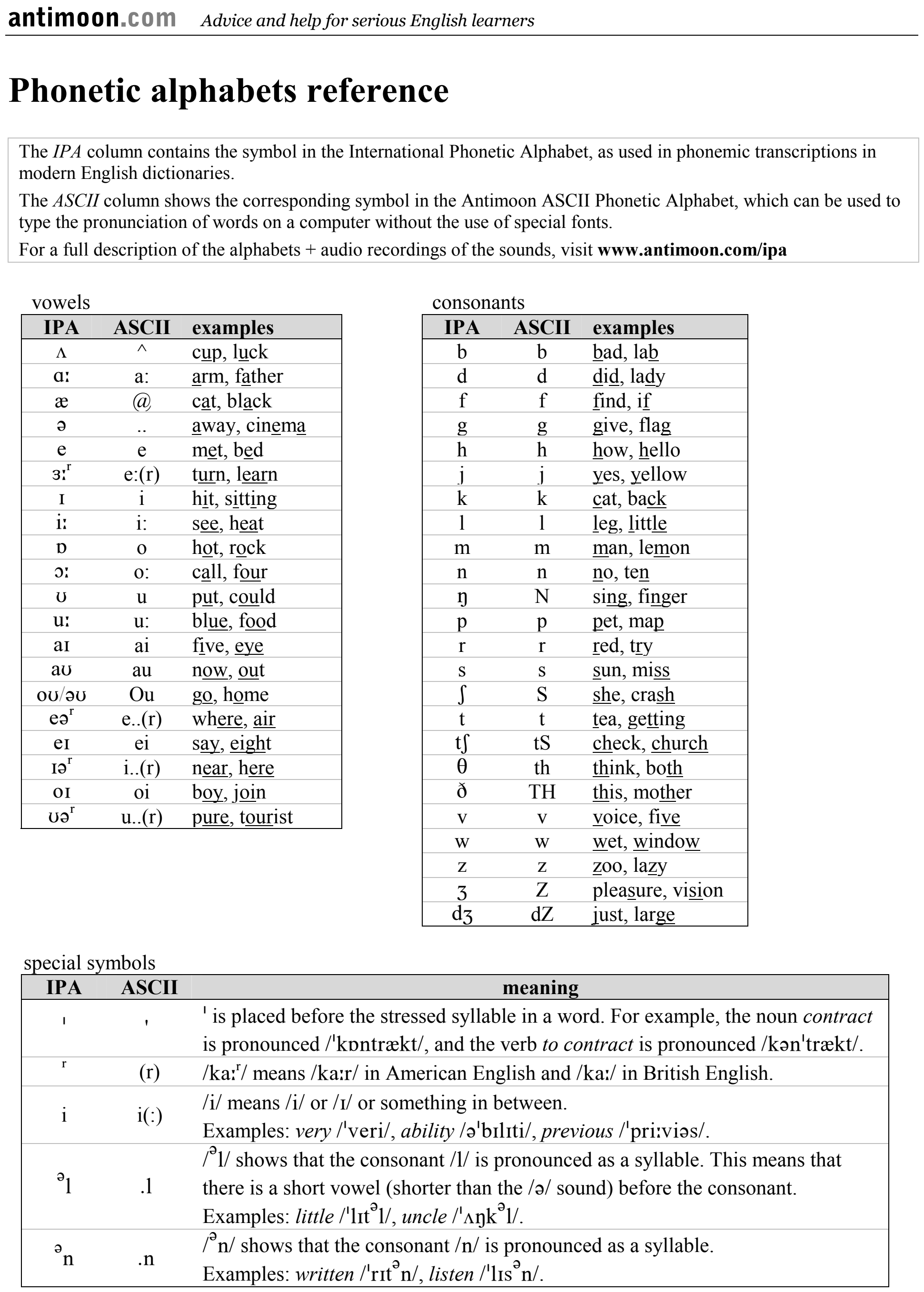
Moreover, I would also want to work with younger children in the future. Human pronunciations are more susceptible to change for younger children. This work has already shown statistically significant gains for high school students, and would want to test out these concepts with pre-school to middle school students too.

One would also want to conduct long-term experiments, with more hours of English instructions. One would also want to use sound pedagogical concepts to ensure retention. I was using the graduated interval recall principle for this phase, but one would want to explore methodologies or practices that are specific to speech and pronunciation.

The most important future work would be to look into the domain of mobile devices. With the increase in the processing power of the phones, it is possible to run Speech recognizers on cell phones It is already possible to port the CMU Sphinx-III speech recognition engine to mobile devices (Nokia N810), and the performance is comparable to the computer version (average time taken in decoding one word on Sphinx III is 0.92 seconds, and average time taken in decoding a word on the ported mobile version is 2.2 seconds). As stated earlier, one could use the recognizer in force-alignment mode and therefore it is even faster.

I believe that with some or all of these changes incorporated into the next phase of research, one will be able to cause a greater change.

Appendix A



antimoon.com *Advice and help for serious English learners*

Phonetic alphabets reference

The *IPA* column contains the symbol in the International Phonetic Alphabet, as used in phonemic transcriptions in modern English dictionaries.

The *ASCII* column shows the corresponding symbol in the Antimoon ASCII Phonetic Alphabet, which can be used to type the pronunciation of words on a computer without the use of special fonts.

For a full description of the alphabets + audio recordings of the sounds, visit **www.antimoon.com/ipa**

vowels

IPA	ASCII	examples
ʌ	^	cup, luck
ɑː	a:	arm, father
æ	@	cat, black
ə	..	away, cinema
e	e	met, bed
ɜː^{r}	e:(r)	turn, learn
ɪ	i	hit, sitting
iː	i:	see, heat
ɒ	o	hot, rock
ɔː	o:	call, four
ʊ	u	put, could
uː	u:	blue, food
aɪ	ai	five, eye
aʊ	au	now, out
oʊ/əʊ	Ou	go, home
eə^{r}	e..(r)	where, air
eɪ	ei	say, eight
ɪə^{r}	i..(r)	near, here
ɔɪ	oi	boy, join
ʊə^{r}	u..(r)	pure, tourist

consonants

IPA	ASCII	examples
b	b	bad, lab
d	d	did, lady
f	f	find, if
g	g	give, flag
h	h	how, hello
j	j	yes, yellow
k	k	cat, back
l	l	leg, little
m	m	man, lemon
n	n	no, ten
ŋ	N	sing, finger
p	p	pet, map
r	r	red, try
s	s	sun, miss
ʃ	S	she, crash
t	t	tea, getting
tʃ	tS	check, church
θ	th	think, both
ð	TH	this, mother
v	v	voice, five
w	w	wet, window
z	z	zoo, lazy
ʒ	Z	pleasure, vision
dʒ	dZ	just, large

special symbols

IPA	ASCII	meaning
ˈ	'	ˈ is placed before the stressed syllable in a word. For example, the noun *contract* is pronounced /ˈkɒntrækt/, and the verb *to contract* is pronounced /kənˈtrækt/.
^{r}	(r)	/kɑː^{r}/ means /kɑːr/ in American English and /kɑː/ in British English.
i	i(:)	/i/ means /i/ or /ɪ/ or something in between. Examples: *very* /ˈveri/, *ability* /əˈbɪlɪti/, *previous* /ˈpriːviəs/.
^{ə}l	.l	/^{ə}l/ shows that the consonant /l/ is pronounced as a syllable. This means that there is a short vowel (shorter than the /ə/ sound) before the consonant. Examples: *little* /ˈlɪt^{ə}l/, *uncle* /ˈʌŋk^{ə}l/.
^{ə}n	.n	/^{ə}n/ shows that the consonant /n/ is pronounced as a syllable. Examples: *written* /ˈrɪt^{ə}n/, *listen* /ˈlɪs^{ə}n/.

Appendix B

Goals for the Usability Test

Overall Goal:

To ensure ease of use and meaningful learning, it is important that the application is:

> Simple enough for young children to use.

> Sufficiently engaging to capture children's limited attention spans.

> Provides an element of practice for learning pronunciation.

student	needs	feature	behavior
	to learn pronunciation of the word	Visual: Word is written with characters spaced by corresponding phonemes	associate the word, phonemes, and the sound together
		Audio: Word is spoken by an American accented native.	
	to speak the word at exact time rightfully	Made visible by the grey border around the word letters	
	Encouragements	Points given for speaking at the right moment	
	Feedback	Audio feedback for correct and incorrect pronunciation by repeating the correct and the incorrect pronunciation Visual feedback by the color coded word characters about overall health of word pronunciation and separate color coded phonemes to show the correctness.	

Scenarios/User Tasks

1. Introductory animation focus:

___ Focused

___ Unfocused

2. Introductory animation reaction:

___ Positive

___ Neutral

___ Negative

3. Go button interaction:

___ No facilitator prompting

___ Objective prompting

___ Interaction walkthrough

___ Facilitator completed

Round One

4. Beginning game play

___ No facilitator prompting

___ Objective prompting

___ Interaction walkthrough

___ Facilitator completed

5. When did the child speak the word at the exact time?

___ First try ___ Second try

___ Third try ___ Fourth try

___ Fifth try ___ Never

6. Overall game focus

___ Focused

___ Unfocused

7. Overall game reaction

___ Positive

___ Neutral

___ Negative

8. Game Mastery – the child understood the following:

___ Game concept (speak the word correctly)

___ Temporal dimension of speaking the word at the right time

___ Did the child understand the feedback.

9.# of Errors

10. Smiles/Positive reaction?

Yes No

When?

____

____

____

11. Frown/Negative Reaction?

Yes No

When?

____

____

____

Total Activity Time __ minutes

Post Session Survey

Please fill out the following questions about the game:

	Yes, very much	Yes	Not really	Not at all
I understood the instructions.				
The game layout was fun (guitar)				
The words coming down the threads could be read clearly				
The little letters accompanying could be read clearly				
helped me 23				
The game was easy to play				
I want to play this game again				
I would like to play this game on mobile phone				
The visual feedback of the word was understandable				
The visual feedback of the phoneme was understandable				
The visual feedback was sufficient to identify where the mistake was				
The visual feedback was sufficient to identify what the mistake was				
The visual feedback was sufficient to identify how the mistake could be corrected				
The auditory feedback was sufficient to identify where the mistake was				
The auditory feedback was sufficient to identify what the mistake was				
The auditory feedback was sufficient to identify how the mistake could be corrected				
The feedback helped me correct my mistakes				

1. Did you get confused on how to play the games? Yes No

If so, how did you clear your confusion?

____

____

____

2. Did you ever get stuck on the games? Yes No

If yes, how did you get yourself out of the situation?

____

____

____

Did you continue or quit the game?

____

____

____

3. How was the speed and flow of the game?

____

____

____

4. Would you rather learn pronunciation through the game, Ms. Gilbert and Sue, or a computer language application, or a combination?

____

____

____

5. Was learning pronunciation from this game difficult

____

____

____

If yes, what made it difficult?

____

____

____

6. How can the feedback be improved?

____

____

____

7. What did you ……………..bb

8. What did you like dislike?

____

____

____

Thank you answering the questions.

Lehr-und Forschungsgebiet Informatik 9
„Computerunterstütztes Lernen" an der
RWTH Aachen University

Bibliography

Anna Hjalmarsson, W. (2007). Dealing with DEAL: A dialogue system for conversation training. *8th SIGdial Workshop on Discourse and Dialogue.*

Antimoon. (2010, April 20). *Printable IPA/ASCII Chart*. Retrieved December 20, 2010, from Antimoon: Printable IPA/ASCII Chart

Anuj Kumar, A. T. (2010). An Exploratory Study of Unsupervised Mobile Learning in Rural India. *CHI.*

Bloomfield, L. (1984). *Language.* Chicago: University of Chicago.

Carnegie Mellon University. (2010, November 30). *Sphinx*. Retrieved December 20, 2010, from Sourceforge: http://cmusphinx.sourceforge.net/

Chen Yu, D. H. (2004). A Multimodal Learning Interface for Grounding Spoken Language in Sensory Perceptions. *Applied Perceptions.*

D. Powers, R. L. (2008). PETA – a Pedagogical Embodied Teaching Agent. *PETRA.* USA: PETRA.

Dictionary.com. (2010, December 20). *Dictionary.com*. Retrieved December 20, 2010, from Dictionary.com: http://dictionary.reference.com/

Fallahkhair, S. (2004). Dual device Interface for Ubiquitous Language Learning: Mobile Phone and Interactive Television (iTV). *WMTE.*

Fritzsche, K. H. (1997). *Visual Feedback for a Student Learning Language Pronunciation.* Monterey: Naval Postgraduate School, Monterey, California.

G.Aist, J. a. (2001). Evaluating tutors that listen: An overview of Project LISTEN. In *Smart Machines in Education* (pp. 169-234). CAmbridge: AAAI/MIT Press.

International Bureau of Eduation. (2001). *Teaching additional languages.* UNESCO: IBE.

International Phoenetic Association. (1888, December 14). *IPA*. Retrieved December 14, 2010, from IPA: http://www.langsci.ucl.ac.uk/ipa/

J. E. Horowitz, L. J. (2006). *Evaluation of the PBS Ready to Learn cell phone study: Learning letters with Elmo.* USA: WestEd.

J.Mostow, G. H. (2003). Evaluation of an automated Reading Tutor that listens: Comparison to human tutoring and classroom instruction. *Journal of Educational Computing Research* (29(1)), 61-117.

J.Mostow, G. R. (2008). 4-Month evaluation of a learner-controlled Reading Tutor that listens. In V. M. (Eds.), *The Path of Speech Technologies in Computer Assisted Language Learning: From Research Toward Practice* (pp. 201-219).

Johnson, W. L. (2007). Serious Use of a Serious Game for Language Learning. *Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work.*

K.Reeder, M. (2005). TheRole of L1 in Young Multilingual Readers' Success With a Computer-Based Reading Tutor. *Fifth International Symposium on Bilingualism.* Barcelona.

M.Kam, A. S. (2009). Improving Literacy in Rural India: Cellphone Games in an After-School Program. *ICTD.*

Massaro. (2003). A Computer-Animated Tutor for Spoken and Written Language Learning,. *ICMI* (pp. 103-113). USA: ICMI.

Matthew Kam, A. M. (2009). Designing Digital Games for Rural Children: A Study of Traditional Village Games in India. *CHI.*

Migration policy Institute. (2010). *ELL Press Release.* Washington: MPI.

Migration Policy Institute. (2010, April 22). *Migration Facts, Stats and Maps*. Retrieved from Migration Policy Institute: http://www.migrationinformation.org/DataHub/state.cfm?ID=US

National Center for Education statistics. (2010). *Trends in High School Dropout and Completion Rates in the United States: 1972–2008.* Washington: National Center for Education statistics.

Pew hispanic Centre. (2010, April 22). *Hispanic report*. Retrieved from Pew hispanic Centre: http://pewhispanic.org/topics/?TopicID=16

Pimsleur, P. (1967). A Memory Schedule. *The Modern Language Journal* (51(2)), 73-75.

R. J. W. Sluis, I. W.-M. (2004). Read-It Five-to-seven-year-old children lean to read in a tabletop environment. *IDC.* ACM.

R. Poulsen, P. H. (2007). Tutoring Bilingual Students with an Automated Reading Tutor That Listens. *Journal of Educational Computing Research* , 191-221.

United States Census Bureau. (2000). *School Enrollment: 2000 Census Brief .* Washington: United States Census Bureau.

US Census Bureau. (2007). *American Community Survey.* Washington DC: US Census Bureau.

W. Lewis Johnson, H. V. (2005). Serious Games for Language Learning: How Much Game, How much AI? . *Conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology.*