EFFECTIVENESS OF ONLINE

LANGUAGE LEARNING SOFTWARE (DUOLINGO)

ON ITALIAN PRONUNCIATION FEATURES:

A CASE STUDY

By

MICOL MARTINELLI

Bachelor of Arts in Psychology and English

University of Hawaii, Manoa

Honolulu, Hawaii

2009

EFFECTIVENESS OF ONLINE

LANGUAGE LEARNING SOFTWARE (DUOLINGO)

ON ITALIAN PRONUNCIATION FEATURES:

A CASE STUDY

Thesis Approved:

Dr. Stephanie Link

Thesis Adviser

Dr. Nancy Caplow

Dr. Carol Moder

.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my thesis committee for their continuous support, patience, and immense knowledge (and perhaps a few sleepless nights). I would also like to extend a thank you to all my professors for their invaluable help and guidance, and for the wonderful experience that was learning in their classes.

A special acknowledgment goes to my parents for their love, support, and kindness: everything that I have ever accomplished is thanks to their wisdom and their fearlessness. Also, thanks to my husband for his tenacity, his humor and intelligence, and for the late night thesis-writing sessions he never once complained about.

Last but not least, I would like to thank my participants and language learners all over the world; the process of learning a language, of being curious about another culture, and rejoicing in our diversities and similarities, is one of the few things that makes the world a better place.

Name: MICOL MARTINELLI

Date of Degree: JULY, 2016

Title of Study: EFFECTIVENESS OF ONLINE LANGUAGE LEARNING

SOFTWARE (DUOLINGO) ON ITALIAN PRONUNCIATION FEATURES: A CASE

STUDY

Major Field: ENGLISH

Abstract: This study aimed to evaluate the effectiveness of online language learning software for the acquisition of pronunciation macro skills (comprehensibility and accentedness) and micro skills (gemination and voice onset time); specifically, this study focused on *Duolingo*, a data-driven online language learning software created in 2011. Designed as a case study, it followed five participants through four weeks using *Duolingo* to learn Italian. Data were collected through both quantitative measures, such as *Duolingo* reports, acoustic analysis in Praat, and native speaker ratings, and qualitative measures, such as interviews, observations, and questionnaires. Findings from the native speaker ratings on comprehensibility and accentedness revealed a statistically significant improvement in one of the participants. Further, findings from the analysis of practice data and the acoustic analysis showed that those participants who practiced more, through the repetition of lessons, were more likely to improve their pronunciation, while those who tended not to repeat lessons, showed very little signs of improvement. However, the results also indicate that language and linguistics training, L2 and consequent languages typology, and recency of language study also impact levels of improvement.The findings from the perceptions of the participants on the effectiveness of *Duolingo* for the acquisition of pronunciation skills further corroborate the finding that online language learning software like *Duolingo* do implement certain pedagogically sound *practices*, which is certainly a step forward in the development of sound, valid, programs. However, the results of the study also reveal the need to calibrate such practices through modifications to design. The implications of these results relate to both theory and practice, and highlight the facts that with appropriate research, technology can become - and has possibly already begun to be - an engaging and efficient tool for language learning and that the future of online language learning looks bright.

TABLE OF CONTENTS

Chapter   Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The last few decades have seen major advances in technology, with new devices and software being created every year. A prominent sector of this new technology is that of online language-learning software or OLL software, language programs and courses developed entirely, or in major part, for the web. Today, there are hundreds of dedicated brands on the market and while some software focus on marketing language learning as fast and fun, such as Rosetta Stone ("The Fastest Way To Learn A Language"), Berlitz ("Today Berlitz, tomorrow the world"), Rocket Languages ("Speak and understand a new language faster"), or Babbel ("Fresh, fast, fun & easy"), others focus more on its accessibility. One example is *Duolingo*, a free language-learning platform created by Luis von Ahn and Severin Hacker in 2011, whose slogan reads, "Free language education for the world." These qualities, being fast, widely accessible, and at times cheap, are the main reasons for the success of these software applications, as they appeal to those that need to learn language on a schedule, on the go, and for an affordable price.

Regardless of marketing strategies and of the important research carried out on the beneficial aspects of computer-assisted language learning (CALL), such as its ability to provide conscious knowledge of the second language (L2), interaction with a larger community of

learners, and highly interactive materials (Blake, 2011a; Hubbard, 2006; Kern, Ware &

Warschauer, 2008; Tozcu & Coady, 2004), evaluators of OLL software are still skeptical, often

with good reason, about the effectiveness of said programs. Many of the issues investigated have

to do with costs (Bush, 2008; Dorwick, 2002), assessment (Tarone, 2015), the comparison

between language software and traditional classroom instruction (Lord, 2015), and most

prominently CALL's relationship to second language acquisition (SLA) theory (Buendgens-

Kosten, 2013; Bush, 2008; Chapelle, 2009; Chun, 2012; Peterson, 2013; Yamazaki, 2014). The

situation is further exacerbated by the fact that very little empirical research has been conducted,

even though *Duolingo* did take into consideration a few of these concerns in its latest report

(Vesselinov, 2012). Of particular interest to this study are the areas of pronunciation, often

undervalued in the evaluation of OLL software, and user experience, critical to studies of

interactions between humans and technological tools (Garrett, 2010; Karapanos, 2013;

Kuniavsky, 2003) but often unacknowledged by more SLA-oriented research.

Overall, there seems to be a pressing need for a comprehensive evaluation of OLL

software that takes into consideration the many facets of technology, while upholding the

importance of theorized language acquisition. It is with the intent to further the literature in this

direction that this exploratory study was designed. Indeed, this study aims to evaluate

effectiveness of online language learning programs, with special focus on user experience and a

concentration on pronunciation, through the use of *Duolingo*, chosen for its original stance in the

online language learning software community and its growing number of users.

Participants in this study were followed through a period of four weeks in which they

used *Duolingo* to learn Italian, the target language, and, in particular, selected features of

pronunciation. Based on the literature on Italian phonology and on common errors produced by

second-language learners of Italian, the focus here is on gemination and voice onset time (VOT),

two features considered common mistakes made by Italian as a second language learners and

salient characteristics of native-like production. Specifically, the presence of contrastive

gemination in Italian and its absence in English and the unaspirated quality of word-initial

voiceless stops in Italian and their aspiration in English can make it challenging for learners of

Italian as a second language to produce the feature appropriately.

The study also looked at native speaker ratings for comprehensibility and accentedness,

which were used to determine improvements in overall intelligibility, and practice reports from

the program, used to investigate the amount of repetition for each lesson of each participants.

The implications of this study reflect both in theory and practice. As the necessity for a

deeper connection between SLA theory and CALL applications grows stronger, it is important to

evaluate new applications in light of and through theorized language acquisition. This study

suggests that usage-based accounts of acquisition are a successful model for the evaluation and

the design of these platforms. Further, this study highlights the importance of a focus on

pronunciation as well as one on user experience: the former because of its prominent position in

both the language curriculum and in the mind of the learners and the latter because of the

behavioral data which can emerge from its investigation.

The next chapter, Chapter II, will explore the literature that informed and inspired this

study, while Chapter III will explain the methodology adopted to conduct it. Chapters IV and V

will present an overview of the results for each research question, and Chapter VI will provide a

discussion and conclusion, inclusive of some recommendations for future research.

CHAPTER 2

REVIEW OF THE LITERATURE

As defined by Levy in 1997, computer-assisted language learning, or CALL, is ""the search for and study of applications of the computer in language teaching and learning" (p. 1). Many advances in technology have occurred since then, as technology now undeniably permeates our everyday lives. It is no surprise, then, that CALL research has flourished over the years, even though many of its areas still need further exploration.

Because of those same innovations, CALL is also characterized by ongoing changes, which create the opportunity to review findings and carry out new research, while often challenge any beliefs about teaching and learning (Beatty, 2003). One of the most recent of these opportunities lies in language learning software, a booming new market which counts hundreds of dedicated brands. In particular, as our society is more and more dominated by connectivity and the Internet, it is relevant that we look for such opportunity in language software that is available online (OLL - Online Language Learning). As exciting as this new technology is, however, it is important to remember that CALL is interested in the evaluation of these technologies, which is what the current study aims to do through the use of *Duolingo.*

This chapter will give an overview of the theoretical background of this study, followed by a discussion on current issues with language learning platforms, with the aim of highlighting a weak connection between theorized acquisition and the evaluation of said applications, a scarce focus on pronunciation, and a lack of investigation into user experience. The last few sections will make the case for the importance of investigating pronunciation and user experience.

2.1   Usage-based emergentism and skill acquisition theory

Emergentist approaches to language acquisition posit that language learning is a gradual process of association between elements of language that occur together, and this probabilistic knowledge overtime leads to fluent language performance. In other words, language acquisition emerges from simple developmental processes which are exposed to complex environments (Ellis, 1999). Emergentist approaches reject the assumption that language learning is predispositional to or inbuilt in humans. Rather, proponents of an emergentist view of language acquisition believe that the complexity of language is learnt primarily through a statistical analysis of the input.

In the majority of emergentist approaches input (or usage) plays an important role and is considered the main source of learning (O'Grady, Lee & Kwak, 2000). When a learner encounters a form in the input and is able to successfully map a function onto it, a *mapping* is created and, if we view language acquisition as rational contingency learning, then acquisition is simply the gathering of information about the "relative frequencies of form-function mappings" that occur in the input (Ellis, 2006, p. 1). The reason that information about relative frequencies is conducive to acquisition is because human learning is sensitive to frequency. In other words, a stimulus (or input), or rather the mapping between a form and its meaning, is processed faster and more accurately the more times it is encountered (Ellis, 2006). The power function that relates accuracy and prior occurrence frequency is known as the power law of learning and is one of the main tenets of skill acquisition theory.

Drawing from cognitive psychology, skill acquisition theory emphasizes that the process through which we learn the majority of skills available to us is incredibly similar in terms of the advancement from an initial representation of knowledge to a "fluent, spontaneous, largely effortless, and highly skilled behavior" (VanPatten & Williams, 2014, p. 94). This similarity can be explained through an analysis of certain basic principles common to the acquisition of all skills, chief among them the cycle of practice. This process provides opportunities for declarative knowledge (knowledge of facts) to be developed into procedural knowledge (knowledge of how to do something) as the use of the language becomes more and more automatic, and is divided into three stages of development: declarative, procedural, and automatic (Taatgen, Huss, Dickison & Anderson, 2008).

The declarative stage is a period during which the learner acquires knowledge *about* a skill, sometimes without ever using it, through observation, analysis, or through verbal form from an expert. The declarative stage is then followed by the procedural stage, a period in which the knowledge becomes practice, and in which "reaction time" (time needed to complete the task), "error rate" (the amount of mistakes committed during the task) and "robustness" (the amount of attention needed to perform the task) decrease (Taatgen et al., 2008, p. 548) to the point where the task becomes more automatic. The idea that that both reaction time and error rate decrease as a consequence of practice is part of an important concept in skills acquisition theory called *power law*. This concept highlights the fact that there exist some fundamental learning mechanisms regardless of the type of skill being learned - this 'law' pertains to all skill acquisition processes. At the end of the procedural stage, the learner enters the automatic stage, or the stage where the behavior becomes fluent and skilled, although much overlap between stages can exist.

Clearly no degree of practice is going to lead to 'perfect' knowledge, but if and when practice is defined as "opportunities for meaningful language use […] and for thoughtful, effortful practice of difficult linguistic features, then the role of practice is clearly beneficial and even essential" (Lightbown, 2000, p. 443). In order to understand wether we can call certain

activities and tasks "opportunities for meaningful language use" or "effortful practice," we can turn to theorized acquisition, although this analysis seems to be missing in the vast majority of studies and evaluations, alongside a lack of focus on pronunciation and user experience.

To this end, through the lens of usage-based emergentism and, in particular, skill acquisition theory, this research aims to investigate the applications of CALL in the recent and highly relevant sector of OLL software, as the majority of these software emphasize practice as one of their main features. Further, it focuses on determined pronunciation features and explores aspects of user experience.

2.2   Overview of online language-learning software

OLL software are language programs and courses developed entirely, or in major part, for the web. They are available either as subscriptions, packages, or apps, and contain a variety of online tools, such as games, video tutorials, forums, and tests. There are hundreds of online language software available as of this year (2016), but the six most popular and consistently voted best among their category are *Berlitz, Rosetta Stone, Rocket Languages, Fluenz, Babbel,* and *Duolingo,* in chronological order of their foundation (shown in Figure 1).
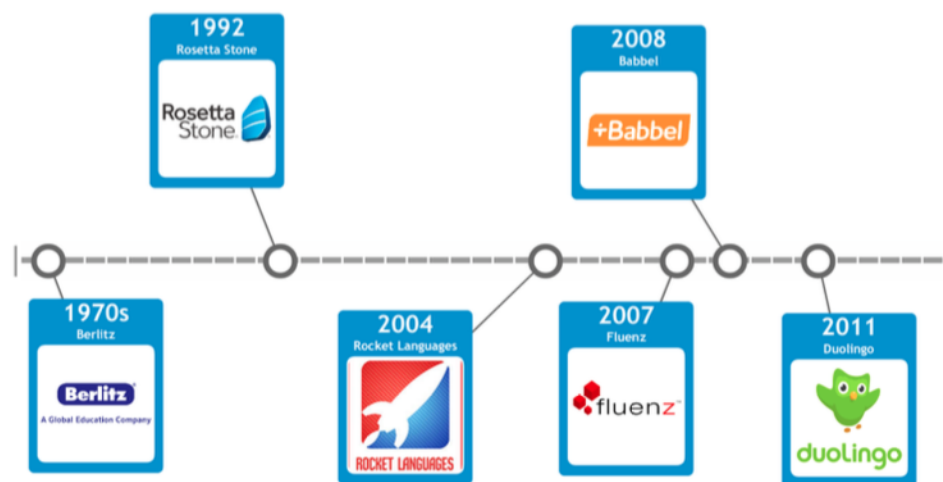
Figure 1. Chronology of the six most popular language software companies.

Berlitz, Rosetta Stone, and Rocket Languages all use the same theoretical approach in their programs: teaching through the use of the target language only. However, they vary in the ways in which this approach translates to their specific activities. Berlitz and Rocket Languages aim to allow students to work out grammatical rules from the input language provided without necessarily being able to explain the rules overtly, while Rosetta Stone uses images, text, and sound in spaced repetition to teach words *and* grammar.

Fluenz and Babbel, instead, use what Fluenz calls "leverage" rather than immersion, where the programs *leverages* the learner's previous knowledge in order to create connections to the new language. This leverage is achieved through tutor led explanations in Fluenz and themed lessons in Babbel, which aim to teach relevant content that would lead to immediate communication.

Overall, these language programs use common online language learning software structures and approaches (i.e. target-language only approach, tutor-led videos, repetition etc.) to help learners progress through their language learning journey, although they may vary in terms of how much focus is placed on which skill. Further, whether the software is accessed through a monthly subscription, a one-time download, or a physical software package, all five programs require some form of payment, ranging from 30 to 400 dollars.

The last of the six most popular programs is *Duolingo*, which was selected for this study because quite different from the other software. The program makes use of innovative features like gamification. Gamification is the concept of applying game mechanics and game design techniques to engage and motivate people to achieve goals. It uses underlying principles which empower the learner in more ways that non-gamified experiences can. Among the many principles are the *Well-ordered Problems* principle (Gee, 2007), which states that if learners face problems early on that are too free-form or too complex, they often form creative hypotheses about how to solve them, but hypotheses that don't work well for later problems. Therefore, the problems learners face early on are crucial and should be well designed to lead them to

hypotheses that work well, not just on these problems, but as aspects of the solutions of later, harder problems, as well. Another of these principles is the *Information 'On Demand' and 'Just in Time'* principle (Gee, 2007), which highlights the fact that human beings are quite poor at using verbal information (i.e. words) when given lots of it out of context and before they can see how it applies in actual situations. They use verbal information best when it is given 'just in time' (when they can put it to use) and 'on demand' (when they feel they need it) (Gee, 2007).

Overall, mechanics, design, and gamification principles create an environment conducive to high engagement and empowerment levels, which are implemented in *Duolingo* through a simple and interactive design, a points/rewards system, and on demand information, as shown in Figure 2.
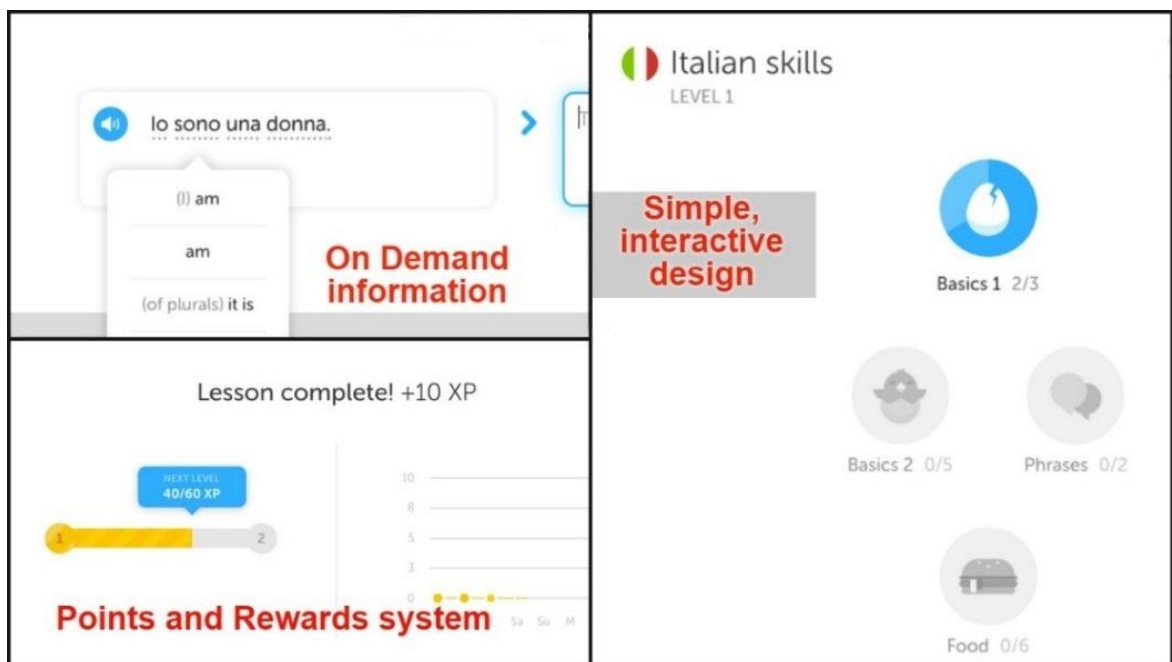
Figure 2. Gamification principles applied in *Duolingo*.

Further the software has 54 different language courses across 23 languages, with 28 additional courses in development, is one of the only free online language-learning software available, and the only free one that counts so many users and so many awards, and has one of the

largest user community in this sector, with over 100 million registered users worldwide (Protalinski, 2015).

2.3   Current issues with language-learning platforms

2.3.1 Weak connection to SLA theory

As mentioned earlier, one of the tenets of CALL is to study its applications, and one of the important elements of this process is an investigation of current issues. Although Chapelle (2009) highlights the benefits of using SLA approaches for the development and evaluation of CALL materials, research exists and has revealed that the relationship between language-learning software and SLA theory is quite weak. The idea that there is a need for a more robust connection between SLA theory and CALL applications is shared by many researchers (Buendgens-Kosten, 2013; Bush, 2008; Chapelle, 2009; Chun, 2012; Peterson, 2013; Yamazaki, 2014) and, as stated by Colpaert (2006), the core of the issue lies in the gap that exists between technology and language pedagogy.

One of the reasons for this seeming disconnect between SLA theory and CALL applications is that, as Garrett (1991) states, "technology that can be taken for granted is already light years ahead of the profession's ability to integrate a principled use of it into the classroom and the curriculum" (p. 74), or in other words advances in technology outpace advances in language learning practices. Another reason is that, as Tarone (2015) highlights, online learning research has followed the negative trend of focusing on achievement rather than on proficiency. In other words, research has aimed at creating and evaluating online language learning by measuring the knowledge of facts of the learners rather than their ability to use the knowledge and "do."

A successful merging of the gap between theorized acquisition and technological applications would results in more effective programs and is therefore an important aim of CALL research on online language learning. Therefore, the question posited in 1991 by Garrett, asking

what kind of technology-based learning activities integrated how, at what level of language learning, for what kind of language learner are likely to be effective for what specific learning purposes, is still relevant today and still requires a deeper exploration.

2.3.2 Lack of focus on pronunciation and user experience

According to Hubbard and Bradin (2004), CALL is effective when there is a balance of conscious knowledge of the L2, active interaction with the materials, and frequent and meaningful communicative practice. Much of the literature on CALL applications and online-language learning software (Blake, 2011a; Kern et al., 2008; Lee & Lee, 2011; O'Brien & Hegelheimer, 2007; Tozcu & Coady, 2004; Yip & Kwan, 2006; Zhang, Song & Burston 2011) indeed highlights the presence and the benefits of such features. However, while making strong cases for a variety of skills, this body of research lacks data on the effectiveness of OLL software for the acquisition of pronunciation skills and seems to pay little attention to the experience of individual learners, concentrating rather on global improvement rates.

Most researchers have focused on vocabulary knowledge and found that a variety of CALL applications, such as mobile phones (Zhang et al., 2011), Tutorial CALL (Tozcu & Coady, 2004), or online language software and games (Yip & Kwan, 2006) are either as effective as other traditional classroom instruction tools or outperform them. However, most of these results are based on either achievement or proficiency measures, which while often satisfactorily measuring possible progress in the acquisition of vocabulary knowledge, also lack deeper investigation into the psycho-social reasons behind said progress. Indeed, in these studies there is no data on the experience of each individual towards one or the other activity, engagement levels, and perceptions. Data of such types could have revealed ulterior benefits from CALL applications, such as higher motivation levels or self-initiated study, or hidden issues, such as difficulty of the tasks or short-term retention rates. Further, vocabulary knowledge was not linked to successful pronunciation of said vocabulary – knowledge was tested mostly in writing.

The importance of this "hidden" data on user experience is highlighted in a study by Lee and colleagues (Lee et al., 2011) on speaking and listening, where participants in the experimental group had meaningful interactions with intelligent programs in an immersive environment. In comparison to the control group, which received more traditional classroom instruction, no significant difference was found in terms of listening. However, the experimental group's speaking skills improved significantly and, further, the activities promoted and improved students' "satisfaction, interest, confidence, and motivation" (Lee et al., 2011, p. 25). Also, blogging (Kern et al., 2008) and podcasting (O'Brien & Hegelheimer, 2007) were found to be activities that not only provide rich input but also foster autonomy and self-confidence. Nonetheless, data on user experience in said studies are in aggregated form, and shed little light on individual variation. Further, as most measures of vocabulary knowledge rely on writing or matching, an investigation into progress in pronunciation is scarce or missing.

Vocabulary knowledge, and speaking and listening, along with writing and reading comprehension, were also investigated in *Duolingo* specifically in a study by Vesselinov (2012). Native English speakers studying Spanish were tested at the beginning and the end of an eight-week period, though a test of general listening, reading, and writing skills (WebCAPE). The results indicated that with 34 hours of study on the program, a learner would acquire as much language as in one semester of college. However, the study lacks pronunciation data and focused on aggregated data from 600 participants. Therefore, it is neither possible to conclude that *Duolingo* is beneficial for pronunciation, nor what aspects of *Duolingo* are beneficial for what kind of learner.

Overall, these studies suggest that CALL applications can be effective in some aspects of language learning, although they focus mostly on vocabulary, listening and speaking, reading comprehension and lack investigation into the experiences of individual learners, focusing instead on aggregated data. Further, still very few comparative studies exist with reference to OLL and L2 learning (Blake, 2011a), and those too focus mostly on grammar and vocabulary (Peters,

Weinberg, and Sarma, 2009; Warschauer & Grimes, 2007) or reading and listening comprehension (Chun 2006; Cobb, 2007).

2.4   Importance of pronunciation in the investigation of CALL platforms

Although pronunciation teaching has gone through constant fluctuation regarding its position in different language teaching methods and approaches, it has kept a steady place since the 1990s. In the early 2000s, pronunciation featured in dedicated issues of *TESOL Quarterly* and *Prospect* (Ketabi & Saeb, 2015). Even some prominent debates among scholars surfaced in that period, such as the ones regarding intelligibility versus nativeness, segmentals versus suprasegmentals, and the idea of Lingua Franca Core (Jenkins, 2000), or a set of pronunciation features critical to intelligible communication between speakers of different first languages that use English as the communicative medium of choice (Seidlhofer**,** 2011).

Nowadays, however, it seems that researchers have reached the consensus that pronunciation teaching is indeed important and that the central concern, rather than anything else, should be what features to teach learners in a way to enable them to communicate most effectively (Couper, 2008; Levis, 2005; Moghaddam, Nasiri, Zarea, & Sepehrinia, 2012; Zeilinski, 2006).

The goal of more effective communication can be achieved by focusing on intelligibility, or "comfortable intelligibility" (Abercrombie, 1991, p. 93) by a native speaker, which is "the extent to which a speaker's utterance is actually understood" (Munro, Derwing, & Morton, 2006, p. 112). This variable has been used extensively in contemporary research on pronunciation and is considered a critical criterion to measure improvement. Further, intelligibility is the interplay of comprehensibility, or the "listeners' estimation of difficulty in understanding an utterance," and accentedness, or "the degree to which the pronunciation of an utterance sounds different from an expected production pattern" (Munro, Derwing, & Morton, 2006, p. 112).

While comprehensibility has been considered significant for a while now, even accentedness has found its way to a more prominent position, with much research (Boula de

Maerüll & Vieru-Dimulescu, 2006; Brennan & Brennan, 1981; Shiri & Boaz, 2010) demonstrating the importance of the perceptions of accentedness by native speakers, and suggesting that degree of accentedness also plays an important role in communication. Further, although accentedness is mainly related to pronunciation and comprehensibility to pronunciation and non-pronunciation factors (Saito, Isaacs & Trofimovich, 2015), these two scales are equally important. As mentioned in the description of the pre and post-test, comprehensibility is rated overall in sentences and passages, and not in single words, and as Munro and Derwing (1995a) point out this kind of comprehensibility judgment is related to pronunciation.

In this context, and through the use of these criteria, a genuine interest in the use of technology for the teaching of pronunciation has emerged with positive results. In fact, research on the relationship between Automated Speech Recognition (ASR) systems and CALL (Burgos, Cucchiarini, van Hout, & Strik, 2014; Golonka, Bowles, Frank, Richardson, & Freynik, 2014) and CALL and pronunciation (Carranza, Cucchiarini, Burgos, & Strik, 2011; Gambari, Kutigi, & Fagbemi, 2014; Liakin, Cardoso, & Liakina, 2015; Young & Wang, 2013) suggests that CALL is generally effective for pronunciation improvement across languages as oral concepts seem to be taught and learned better thanks to a more propitious learning environment (Liakin et al., 2015). Indeed, CALL provides a private, stress-free, autonomous environment for learners to practice pronunciation (Hismanoglu, 2006), which is often not possible in traditional classroom settings. This environment proves to be extremely important for learners, as many studies (Baran-Łucarz, 2014; Liu, 2006; Wilson, 2006; Woodrow 2006) have demonstrated that it is a "concern over pronunciation mistakes that is particularly likely to cause embarrassment and apprehension", which in turn lower the learners' willingness to communicate (Baran-Łucarz, 2014, p. 445).

Thus, the present study will focus on specific pronunciation features and will use comprehensibility and accentedness as criteria against which to determine the effectiveness of the program for the acquisition of these Italian pronunciation skills. In addition, the study will also

acoustically analyze changes in the production of the features and approximation to native-speaker-like production. The latter was used as a criterion for understanding acquisition mostly because it provided data on whether native-like production is necessary.

The pronunciation features investigate in this study were gemination and voice onset time (VOT), selected from literature on Italian phonology and second language acquisition.

2.4.1 Gemination

A geminate refers to a consonant to a long or "doubled" consonant that contrasts phonemically with its shorter singleton counterpart. Stevens (2011) points out that Italian is different from other romance languages having maintained a consonant length contrast, suggesting that the main phonetic correlate for gemination is an increase in consonant duration, which non-native speakers often struggle with in learning Italian. Indeed, while English has double consonants in the spelling of certain words, as in 'winner', this is not contrastive gemination as there is no variation in the duration of the sound and no minimal pair.

Variation in the duration of geminates has been a main object of research for various languages (Agostiniani 1992; Bastien De Clercq, Simon, & Crocco 2014, Celata & Costamagna 2011, Ericsdotter 1998, Harris 2010, Kabak, Reckziegel, & Braun, 2011, Pickett, Blumstein, & Burton 1999 for Italian; Arvaniti 1999 for Cypriot Greek; Local & Simpson 1999 for Malayalam; Louali & Maddieson 1999 for Berber; Hansen 2004 for Persian; Mah & Archibald 2003 for Japanese) and generally the duration contrast between geminates and non-geminates has been reported to be robust

Further, Payne (2005) reported that Type 1 geminates (those that contrast with non-geminates and are determined lexically) are 1.56 times longer than their singleton counterparts, providing further evidence that there is a robust distinction between singleton and geminate, or in other words a salient contrast in pronunciation.

By analyzing the ratio of single versus geminate stop produced by second language learners and comparing it to the ratio of native speakers, most of these studies have concluded

that learners do produce a length contrast, indicating that they build the required phonological representation. However, they do not implement this contrast in a native-like fashion, suggesting that they have yet to master "the correct phonetic implementation strategies" (Mah & Archibald, 2003, p. 211). While the native speakers in these studies produced all geminates with durations which were more than twice as long as the singletons, the learners' length differences were much smaller.

Most of the mentioned research on gemination has focused on plosives and nasals. However, a few researchers have also looked at /l/, including Payne (2005) who concluded that only /t/ and /l/ are consistently strong in their durational difference. Thus, this study focused on /n/ as the token for nasals, /t/ as the token for plosives, and /l/ as the token for liquids, but did not include /r/ as this feature presents the further challenge of being trilled which could potentially alter the data.

Further, as most studies suggest (Agostiniani 1992, Bastien De Clercq et al. 2014, Celata & Costamagna 2011, Ericsdotter 1998, Harris 2010, Kabak, et al., 2011, Pickett et al. 1999), it is not the actual closure duration measurements that are to be compared, but the ratio of single versus geminate stop produced by each speaker. For nasals, Kabak et al. (2011) suggest using information on the waveform and wide-band spectrogram, such as drops in amplitude and reductions in energy. As an example, Figure 3 shows the boundaries for the geminate /n/ in the word 'gonna' [gɔnːa], *skirt*.
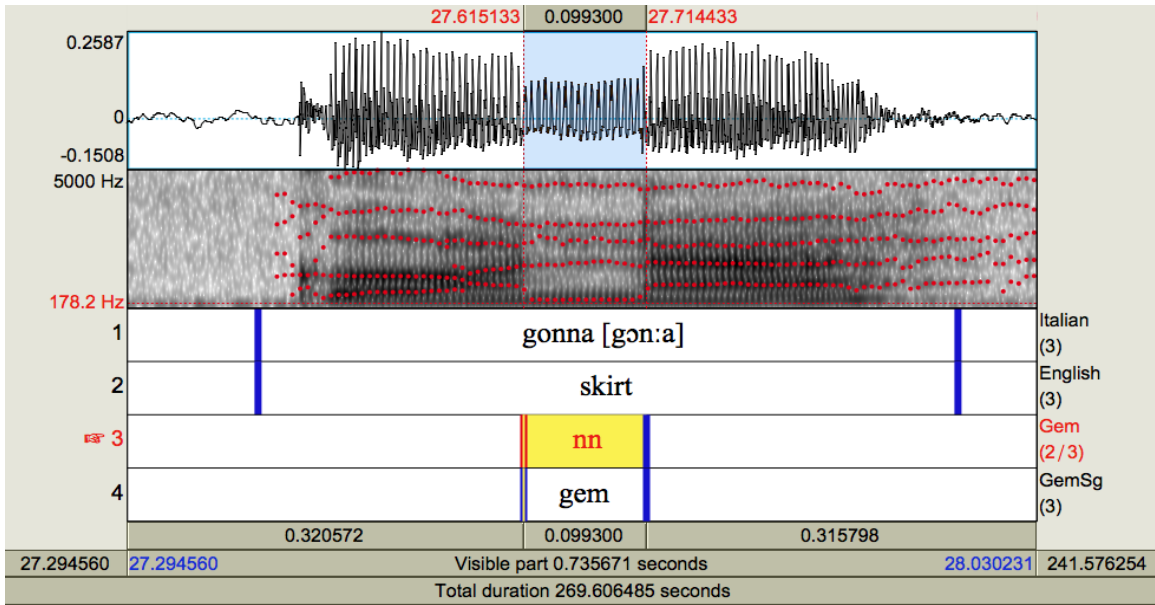
Figure 3. Boundaries of geminate /n/ in the word 'gonna' [gon:a], *skirt.*

For plosives, many of the studies measured closure duration, said to be one of the most reliable cues for geminate-singleton distinctions, based on changes in amplitude and in the waveform. This was the methodology selected here - information in waveform was used to mark boundaries which were then adjusted to the nearest zero-crossing. Figure 4 is an example of this type of measurement with the word 'gatto' [gat:o], *cat*.
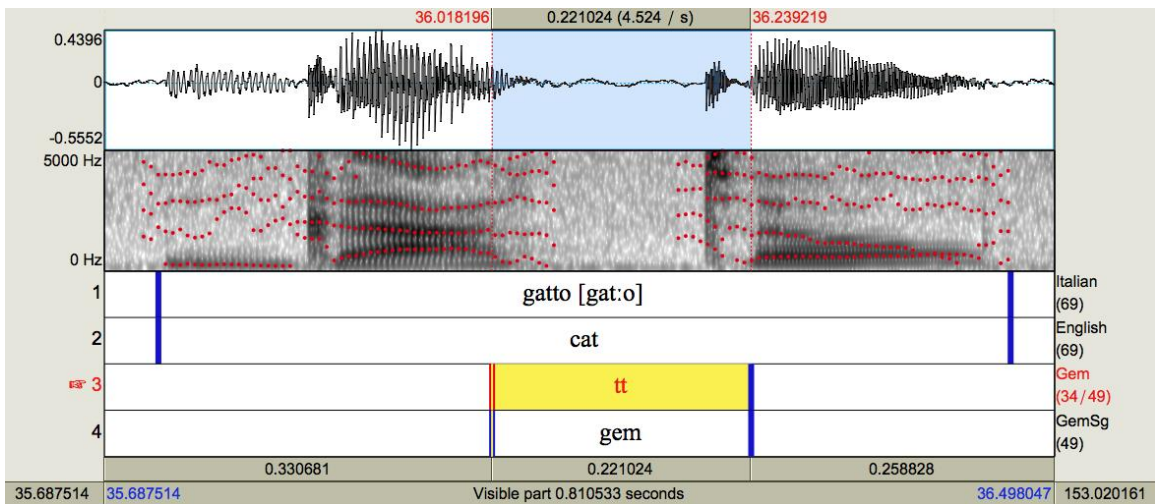
Figure 4. Boundaries of geminate /t/ in the word 'gatto' [gat:o], *cat.*

17

Further, many of the aforementioned studies compared the ratios of the learners to those produced by selected native speakers. As one of the criteria to understand the effectiveness of the program is improvement, and that one of the ways to measure this improvement is the approximation of native-like production, this study also compared results between learners and a selected native speaker.

2.4.2 Voice Onset Time

Voice Onset Time (VOT) is a feature of the production of stop consonants, or plosives. It is defined as the period of time between the release of the stop, or burst, and the onset of vocal fold vibration. When the onset of vocal fold vibration coincides with the plosive release it is said that there is Zero VOT. When there is a delay in the onset of vocal fold vibration after the plosive release, then the VOT is Positive. Lastly, when the onset of vocal fold vibration precedes the plosive release the VOT is said to be Negative. In English, word-initial voiceless stops are produced with Positive VOT, meaning that they are aspirated. Although it is subject to variations, the literature (Klatt, 1975; Lisker & Abramson, 1964; Zlatin, 1974) suggests that the average ranges are between 47 and 65 ms for /p/, 67 to 75 ms for /t/, and 70 to 85 ms for /k/. In Italian, word-initial voiceless stops are produced on average with a voice onset time of less than 30 or 40 ms (Nagy & Kochetov, 2013), or in other words they are unaspirated. Indeed, the average VOTs of the native speaker in this study were 14 ms for /p/, 15 ms for /t/, and 30 ms for /k/, and the average VOTs of the *Duolingo* voice were 32 ms for /p/, 30 ms for /t/, and 43 ms for /k/.

The fact that in English, /p, t, k/ are aspirated makes it difficult for native English speakers learning Italian to produce these same stops as unaspirated in Italian. These phones tend to be late acquired by those native English speakers learning languages that have unaspirated stops, as stated by Kissling (2013) in her research on explicit pronunciation instruction. Her study focused on learners of Spanish who tend to produce /p, t, k/ with overly long voice onset times. One of Kissling's (2013) secondary conclusions was that aspiration, or lack thereof, does affect

18

comprehensibility and this finding, alongside the limited research on VOT in Italian, prompted the interest in voice onset time for Italian as a second language learners in this study.

Further, the previous literature that does exist for VOT in Italian (Celata & Costamagna, 2011; Harris, 2010; Reeder, 1998) reveals that only VOT varies significantly between monolingual Italian and English, as compared to closure duration and burst amplitude. Therefore, this study measured VOT only, and concentrated on word-initial voiceless stops, as has other literature (Harris 2010; Reeder, 1998). Also in line with previous studies, and because of the complications that would arise in analysis, word-initial voiceless stops, which are part of a cluster with /r/, were excluded from the sample.

Thomas (2011) suggests it is best to follow Klatt's (1975) method of measuring from the point where vocal pulses begin in F2 and higher formants. Thus, the measurements were taken on the waveform from the beginning of the aperiodic energy signaling the release burst, with attention paid to point where vocal pulses begin in F2 and higher formants, to the first peak of the sudden appearance of periodic vibration which corresponds to the onset of the following vowel, as suggested by Harris (2010). Further, as recommended by Cho & Ladefoged (1999), in the case of multiple release bursts, the release will be established at the final release burst. As an example, Figure 5 shows the boundaries for the voice onset time of the word-initial voiceless stop /t/ in the word 'torta' [torta], *cake*.
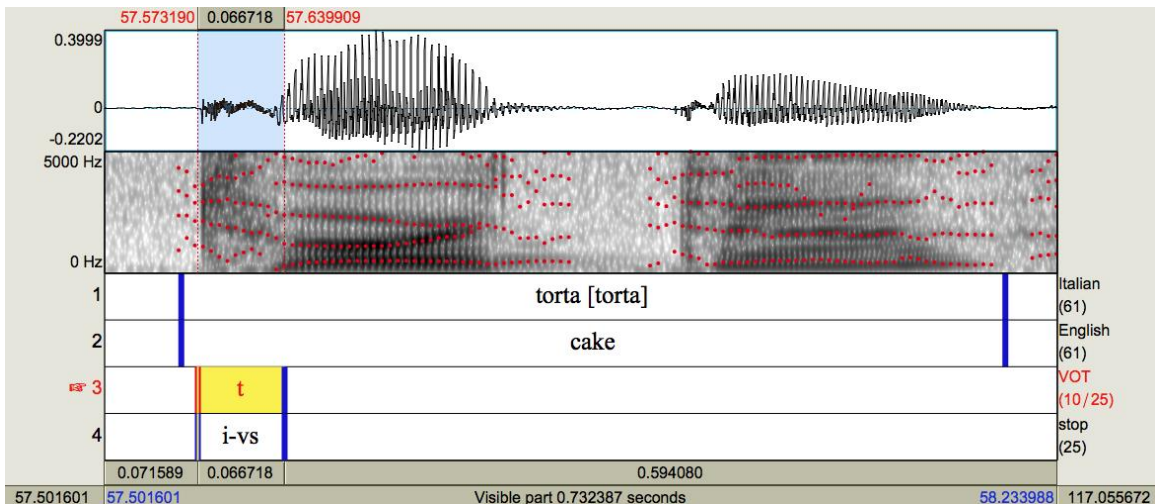
Figure 5. VOT boundaries of the word-initial voiceless stop /t/ in the word 'torta' [torta], *cake*.

It is important to note, also, that since there is some indication that the measure of voice onset time is also sensitive to the place of the stop closure, this could cause an apparent overlap in distributions if stops that have the same manner of articulation but different place are considered together (Lisker & Abramson, 1964). In their seminal work on voicing in initial stops, Lisker and Abramson (1964) suggest keeping the data for the three positions of stop closure separate to avoid such overlapping, as was done in the current study.

2.5   Importance of user experience in the investigation of CALL platforms

The other element lacking in most research is that of user experience. This term is a CALL variation of what is called "learner fit" (Chapelle, 2009) in traditional classroom instruction. Learner fit centers mostly around those factors of language acquisition that relate strictly to the learner, such as the level of language required for the learner to access the selected input and the nonlinguistic benefits that the learner may obtain from work on the input (Chapelle, 2009), but also including the aptitude of the student (DeKeyser, 2010).In CALL, learner fit has been renamed user experience, influenced by the terminology of computer science and technology, and encompasses similar factors. Also, just as the exploration of learner fit has been beneficial for traditional instruction, the monitoring of user experience has led to important discoveries about the effectiveness of certain designs, tasks, and gamified applications especially in terms of engagement.

Many researchers (Lai & Gu, 2011; Munday, 2016; Son, 2007) found higher levels of active engagement in students' self-initiated use of technology, which often translated in substantial improvement. In other words, these applications were found to keep students engaged if and when other activities should failed to do so. Other studies also found an increase in satisfaction, confidence, and motivation (Lee et al., 2011), but even overall time on task, with those learning online spending more time engaged with second language material than those in face-to-face

20

situations (Grgurovic & Hegelheimer, 2007). Also, in her recent study on *Duolingo*, Munday (2016) states that some of the reasons for these higher levels of engagement and interest are due to the accessibility of the input on a variety of platforms, including mobile devices, its gamification aspect, and the variety of tasks.

However, monitoring engagement levels through time on task and improvement, as done by the majority of studies, is often not deep enough. Indeed, user experience can also incorporate the *perceptions* of users, which are currently very much researched in the field of interactions between humans and technological tools (Kuniavsky, 2003; Garrett, 2010; Karapanos, 2013) for other purposes. The main tool for investigating user experience in technology-related areas is through interviews which, as Kunaivsky (2003) suggests, should ask about the broader aspects first and more detailed ones after. The initial focus should be on attitudes, expectations, and assumption of the general category of the product, while the deeper focus stage should concentrate on a specific product and "what it does, how it does it, whether they can use it, and what their immediate experience with it is" (Kuniavsky, 2003, p. 118).

This study will thus specifically address perceptions through the use of interviews starting with a broader focus on technology learning in general and a deeper focus on technology for language learning and OLL software, while also investigating previously acquired knowledge in terms of language background.

2.6 Research Questions

Overall, it is evident that more research on the effectiveness of CALL applications, OLL software in particular, is needed, especially in terms of pronunciation, proficiency, and user experience. Based on the choice of features extrapolated from the literature on Italian phonology and Italian as second language learners' common errors, the following research questions guide this study:

(1) How effective is *Duolingo* for the acquisition of the pronunciation macro skills of accentedness and comprehensibility for Italian as a second language learners?

(2) How effective is *Duolingo* for the acquisition of the pronunciation micro skills of geminate contrast and voice onset time (VOT) for voiceless stops in word initial position for Italian as a second language learners?

(3) What are the learners' perceptions (user experience) of the effectiveness of *Duolingo* for the acquisition of pronunciation skills for Italian as a second language learners?

CHAPTER 3

METHODOLOGY

This section describes the methodology used to conduct this study. The section will illustrate the type of design selected for the study, the role of the researcher, the participant selection process, and lastly the data collection and analysis procedures.

As mentioned, the purpose of this study is to investigate the effectiveness of online language-learning (OLL) software for the acquisition of the pronunciation macro and micro skills through an evaluation of *Duolingo*, both in terms of actual improvements in the participants' pronunciation of certain features (i.e. acoustic analysis, reports, ratings) and in terms of their perceptions (i.e. interviews, questionnaires, observations). Three questions guide this study: (1) How effective *is Duolingo* for the acquisition of the pronunciation macro skills of accentedness and comprehensibility?; (2) How effective *is Duolingo* for the acquisition of the pronunciation micro skills of geminate contrast in word-medial position and voice onset time (VOT) for voiceless stops in word initial position for Italian as a second language learners?; (3) What are the learners' perceptions (user experience) of the effectiveness of *Duolingo* for the acquisition of pronunciation skills?

3.1 Type of Design

Within a post-positivist epistemology, a qualitative case study design was selected to conduct this study in order to gather in-depth information of specific interactions with language software and further understand each participant's experience. As defined by Merriam (1988), a qualitative case study is an "intensive, holistic description and analysis of a single entity, phenomenon, or social unit" (p. 16). Yin (2003) further states that a case study must investigate phenomena within its "real-life context" (p. 13). This study seeks to explore the use of online language software in a natural setting (i.e. the home). As Duff (2008) points out, the data in qualitative case studies shows the changes that occur in the behavior and knowledge of a learner including the influence of other factors. Although relying on linguistic analysis and improvement scores, this study focuses on the overall experience of each individual participant, with the goal of understanding these changes in behavior and knowledge as they naturally occur. Further, this study will support the in-depth analysis of each participant's changes through the triangulation of different perspectives. For a complete overview of the data gathering and data analysis methods, refer to Table 4 in Section 3.6, Data Analysis.

The main disadvantage of the way this study was designed is the fact that, because it is exploratory and a case study, the results are difficult to generalize. However, Stake (2005) suggests a different view of this issue, stating that one can describe a study as an "instrumental case study if a particular case is examined mainly to provide insight into an issue or to redraw a generalization. The case is of secondary interest, it plays a supporting role, and it facilitates our understanding of something else" (p. 445). He clearly underlines the importance of scrutinizing and analyzing but suggests that there is a possibility for implications outside of the restricted amount of cases under study.

Even considering Stake's (2005) approach, a more suitable way of describing the hopes of this research is through the term *transferability.* The idea behind this term is that it "assigns the

responsibility to readers to determine whether there is congruence, fit, or connection between one study context [...] and their own context [...]" (Duff, 2008, p. 51).

3.2   Role of the Researcher

In qualitative studies, the researcher is considered an *instrument* of data collection (Creswell, 2014; Denzin & Lincoln, 2005). In order to fulfill this role of "human instrument" (Simon, 2011, p.1), it is important for those who read qualitative studies to know about this instrument and it therefore important for the researcher to describe relevant aspects of self, including any biases or assumptions, any expectations, and experiences to qualify his or her ability to conduct the research (Creswell, 2014; Greenbank, 2003).

My interest in this study has developed through my own experiences as a second language learner and through contact, both professional and personal, with other second language learners in different settings. Throughout my learning journey, I was fortunate enough to attend both high school and college, both at the undergraduate and graduate level, in the United States. As I am originally from Italy, this allowed me to improve my English daily, in settings that spanned from academic writing classes, to working for the advising office on campus, to interactions with friends. I was also living alone, which "forced" me to use the English I knew at the doctor's office, at the DMV, at the post office, and so on.

However, most of the people that I know who are learning English or Italian as a second language, do not have access to all this, and not just because they might be English as a foreign language (EFL) learners; even those who live in the country whose language they are learning often struggle to find good and rich environments in which to practice and acquire their language skills. Consequently, many rely on the Internet and programs such as *Duolingo,* which is free and accessible on many different types of electronic devices. However, it is important that those learners who do not have access to more extensive language learning systems be provided with quality tools. Therefore, I am inspired by own journey to investigate these new technologies.

Throughout the study, my role was etic. I interacted with the participants mostly during a pre-interview and a post-interview, and minimally during observations. Being a non-participant observer allowed the learners to focus on their improvement in the setting they preferred (i.e. at home, at a coffee shop, in the office), and for the amount of time they preferred, allowing for more authentic data.

3.3   Participant Selection Procedures

In qualitative research, participant selection is purposeful (Creswell, 2014); participants are selected "who can best inform the research questions and enhance understanding of the phenomenon under study" (Sargeant, 2012, p. 2). Thus, a crucial step in the design phase is to identify appropriate participants based on the research questions, theoretical perspectives, and evidence informing the study (Creswell, 2014; Sargeant, 2012). In this study, the most important condition was that the participants be true beginners of Italian. Hence, the participants were selected on the basis of their previous knowledge of Italian, and in order to have a variety of L2s and experience with technology for language learning. Information for the study was sent through the email system to the faculty and staff of the English Department and those willing to participate were screened for the aforementioned basic requirements. A total of five participants responded and were thus selected for the study.

3.3.1 Participants

Table 1 is an overview of the demographic data for the five participants in the study (all participants were given a pseudonym) and information on their native language and L2s, L3s, etc, followed by a detailed description of each of the participants. The participants range in age from 23 to 51 years old, are all females, and with at least a Master's degree. Four out of five participants are native-speakers of English, while the fourth is an English as a second language (ESL) speaker. Three participants are from Texas, one from Kansas, and one from Europe (Ukraine).

Table 1

Participant demographic data

Participant	Age	Sex	P.O.B.	Education	L1	L2	L3
Barbara	51	F	Texas	Masters, History	English	French (6-7 years)	
Sabrina	26	F	Kansas	Masters, Literature	English	Spanish (15 years)	
Luciana	25	F	Texas	Masters, TESL	English	Portuguese (10 years)	Arabic (1-2 years)
Emma	31	F	Ukraine	PhD, TESL	Ukrainian/ Russian	English (10 years)	German (5 years)
Carolina	23	F	Texas	MFA, Poetry	English	Spanish (5 years)	

3.3.1.1 Emma

Emma is a 31-year-old Ukrainian student from Kiev. Besides knowing Russian and Ukrainian, she is also quite fluent in German, learned in college, and has lived in Germany for one year. She is also fluent in English, which she studied during school with help from her mother who is an English teacher. Further, she was an exchange student in South Carolina during high school and is currently enrolled in an American university pursuing her PhD in Teaching English as a Second Language (TESL). Although she has no experience with Italian, she traveled to Italy for one week, but mostly spoke English with her travel companions. She is not currently studying a language. Her motivations for the study were her curiosity for languages and interest in technology, although she stated that she is skeptical of technology for language learning.

3.3.1.2 Barbara

Barbara is warm and talkative 51-year-old woman from Bowie, Texas. She has some language experience with French in high school and college. Although she was an Ancient and Medieval history major, and very interested in Europe, she has never been to Italy and has never

been exposed to Italian, with the exception of some common words or media portrayal. Prior to the study, she was not studying a language because of lack of time and regreted not having occasion to practice her French. Her interest in the study stemmed from her interest in Europe, and in the Italian language. She felt like most studies and opportunities always focus on the same languages, such as Spanish, and that this study was a good chance to practice a different language.

3.3.1.3 Luciana

Luciana is a 25-year-old Teaching English as a Second Language (TESL) student from Woodlands, Texas. She has ten years of experience with Portuguese, especially as a language requirement during her homeschooling. For Portuguese, she has also had an exchange student neighbor with which she had opportunity to talk, and she also visited Brazil. Luciana has also taken two semesters of French in college, but she says she wasn't able to practice that much speaking. Since 2011, she has also studied some Arabic, with two semesters in college, and has ongoing opportunities to speak with and listen to some Arabic speaking friends. However, she was not currently studying it, or any other language. Although she has no experience with Italian or travel to Italy, her motivations for the study were based in her interest for the language and on her short previous experience with *Duolingo*, which she found fun and helpful.

3.3.1.4 Sabrina

Sabrina is a friendly and outgoing 26-year-old Literature major. Although she has some experience with Italian pronunciation due to her background in singing, she has never spoken or used it outside of that context. She does have a lot of experience with Spanish, however, in elementary, middle, and high school. She also took two semesters in college although her last classroom experience with Spanish dates back to 2010. She also knows a little Danish and German from a study abroad experience of six months in Denmark. Her motivations for joining

the study included the fact that she likes Italian a lot and wanted to get familiar with language learning software; however, she admitted that she is a bit anxious around new technology.

3.3.1.5 *Carolina*

Carolina is a shy but sweet 23-year-old from Houston, Texas. She has extensive experience with Spanish, including four years in high school and two semesters in college, but has not been able to practice it for a few years now. However, she still feels confident about her fluency in Spanish and states that her best skill is pronunciation. She also has some experience with Hebrew, although no formal academic training in it, and has only practiced reading it. She was not currently learning a language prior to the study. Although she has never been to Italy, she believes Italian is "really cool" and was interested in learning more about it.

3.4   Raters

Naïve or novice raters have been used in many studies on accentedness and comprehensibility (Munro & Derwing, 1995a; 1995b). Indeed, the experience with foreign-accented speech of a novice rater is very similar to that of an average listener (Isaacs & Thomson, 2013). Further, it is easier to control for the experience (meaning very little) with foreign-accented speech of novice raters as opposed to non-novice raters. Therefore, using novice raters is a valuable and accepted method for the investigation of improvement in the intelligibility of non-native speech. In this study, two novice raters were used to rate read speech samples on both accentedness and comprehensibility.

Angelo is a 70-year-old retired judge, native Italian speaker from Rome, while Angela is a 55-year-old postal worker from a coastal town in Florence. Both have completed graduate level college, and are specialized professionals in their respective jobs. While Angelo has less experience with foreign accents than Angela, he has more experience with foreigners in general. Both raters have some basic knowledge of English, mostly in reading and writing.

3.5   Data Collection

This section contains an overview of the materials used in this study followed by a description of the procedures for data collection. For a chart of how each data collection procedure ties into the research questions and to various data analysis procedures, refer to Table 4 (Data Triangulation) under section 3.7, Data Analysis Procedures.

3.5.1 Materials

The materials for this study included: (i) the *Duolingo* software, (ii) a test of Italian pronunciation containing three sections (words, sentences, passage), (iii) pre and post-interview questions, (iv) questionnaires, which inquired about the perceptions of the participants in terms of language learning, technology, and *Duolingo*, (v) observations recordings, (vi) reports from the *Duolingo* software, which tracked each participant's progress through the program, and (vii) rating scales, which were given to the raters in order to rate the read speech samples from the participants in terms of comprehensibility and accentedness. The following sections will give a detailed description of each of these materials.

3.5.1.1 Duolingo Software

Duolingo uses a data-driven approach to lesson planning and a gamified skills tree model as the basis for its language-learning program. As previously discussed gamification is a combination of mechanisms, design, and principles which aid the user reach his or her goals. In *Duolingo* the first evidence of gamification can be found in the graphics and set-up, as shown in Figures 6 (the "tree" model) and 7 (gamification through levels). The simple design, the word "level," the circled tiles in a tree-like formation, all help the learner visually perceive the program as a game. This perception helps to activate competitive instincts and a higher level of motivation in the learners, while the simplistic visuals help them navigate through the program easily.
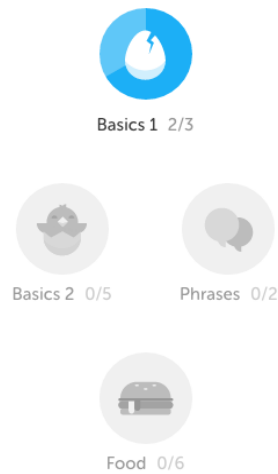
Figure 6. Gamified graphic and set-up (example 1).

Basics 1

Figure 7. Gamified graphic and set-up (example 2).

Gamification also transpires in the ability of the users to gain experience points, or XP, for each completed lesson (Figure 8). Points and rewards systems, while quite basic, are an important component of gamified experiences. Firstly, they help the motivation of the user which is now not only aiming to be successful at the task but also aiming at being successful enough to gain a certain amount of points or certain rewards. This process often pushes users to go the

distance because of the satisfaction that comes from being rewarded. This system is similar to that used in traditional instruction in the form of grades; however, while grades are a direct expression of the quality of the performance, points and awards can also be achieved by time spent on task, number of consecutive days the program is accessed, number of features used, etc. In other words, a points and rewards systems *rewards* other aspects of effort besides the quality of the final result, often aiding in keeping motivation and satisfaction high.
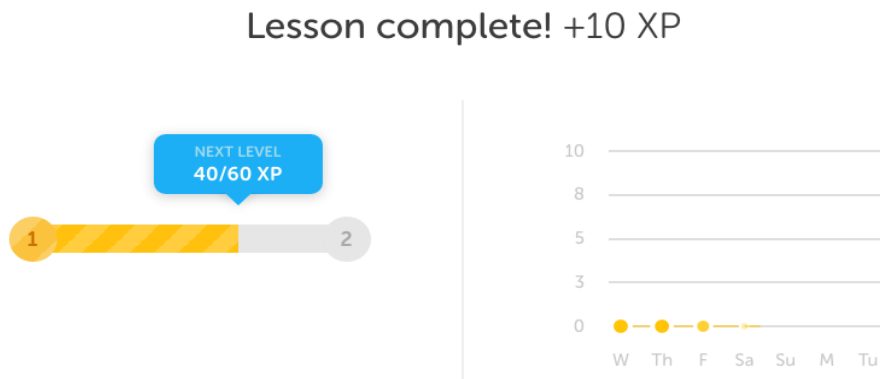
Figure 8. Screenshot of XP points.

In *Duolingo* skills are considered "learned" when users complete all the lessons associated with the skill, although through the use of a "strength bar", the programs estimates the strength of a memory (Figure 9).
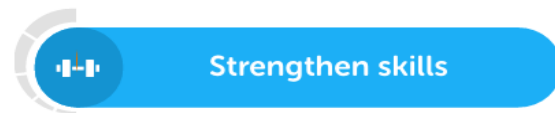
Figure 9. Strength bar.

After a certain duration of time, this bar fades signaling to the user that it is time to refresh certain lessons or skills. Inbuilt functions help the system keep track of which questions the user is struggling with and what mistakes they are making and uses that data in a variety of ways: it presents those questions or skills to the user more frequently; it suggests certain lessons in the strength bar activities, and it feeds them into its program in order for these lessons or skills to appear again in other lessons, when possible.

The idea of a strength bar is also modeled on game mechanisms, where the user can or must keep track of his or her *vitals*. By making the user conscious of what skills might be fading, the user is empowered to refresh certain lessons. In a traditional classroom setting, older or less practiced knowledge might either be ignored, or becomes a point of weakness in further tests or evaluations.

The program also contains a variety of activities, with more types of activities at the higher levels. The basic activities range from matching picture to words, translating, typing what is said by the computer, and selecting an option from a dropdown menu. A few are depicted below, in Figures 10 (Matching Picture), Figure 11 (Translation), Figure 12 (Dictation), and

Figure 13 (Select the Missing Word). Again, simple design, a progress bar at the top, and the collecting of points, all fit into the gamification framework and help in keeping the user engaged and motivated. Further, implicit pronunciation is embedded through these activities, and perhaps more overtly in the *Click the Microphone and Say* activity (shown in Figure 14), as the program runs the audio for all words and phrases anytime they are presented and on hovering, an option detailed in the next paragraph.
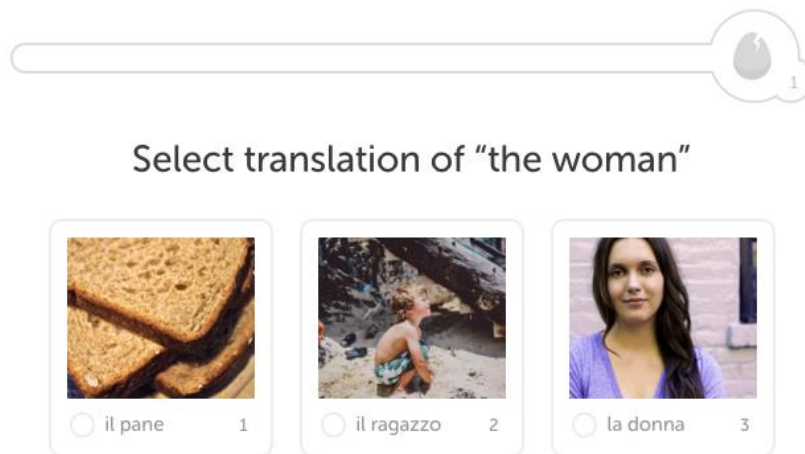
Figure 10. Matching picture to text activity.
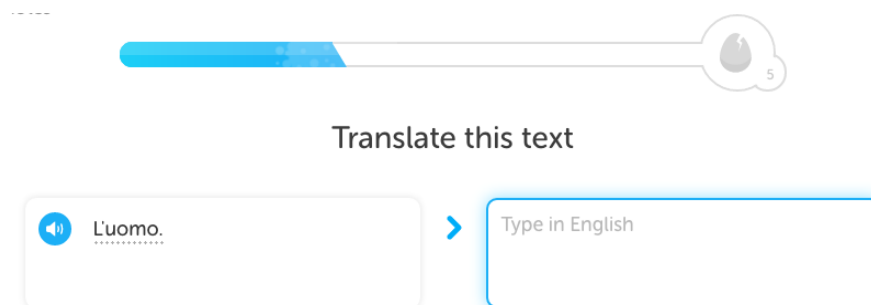
Figure 11. Text translation activity.

Type what you hear

Type in Italian

↑ À Á È É Ì

Í Ò Ó Ù Ú
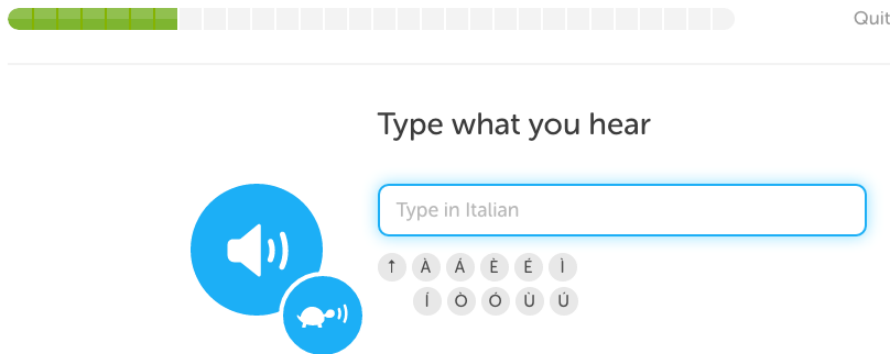
Figure 12. Type what you hear activity.

Tips & notes   Quit

Select the missing word

Io sono un (Select word ∨) .
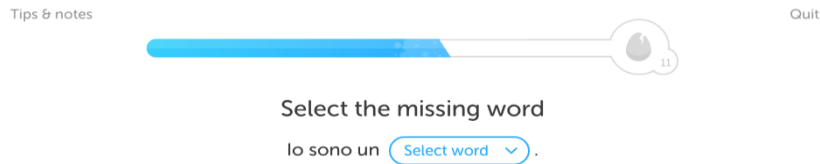
Figure 13. Select the missing word activity.

Click the microphone and say:

Buongiorno
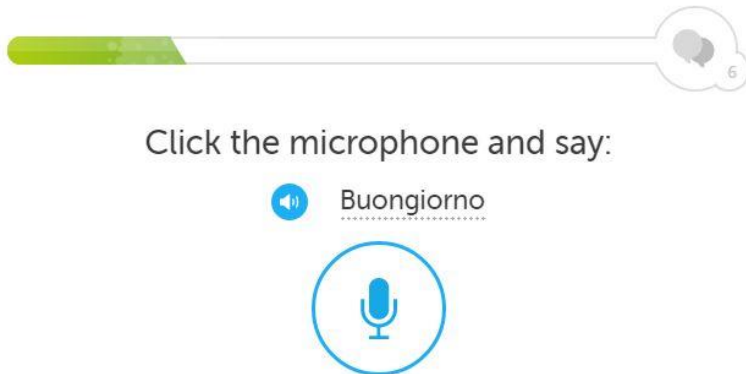
Figure 14. Click the microphone activity.

The software also has an array of features, including hover, conjugate, tips and notes, discuss sentence, and scorecard. These features fall under the previously mentioned gamification

principle of 'on demand' or 'just in time' information. In other words, certain information is not

presented as an introduction to a task or as a separate lesson, but rather is made available right

when the user needs it. As it is difficult for human beings to process much verbal and written

information in large quantities, dividing the information so that it is presented when most needed

is one possible way to make such information more useful, more salient, and more

comprehensible.

Indeed, the *hover* feature allows the user to hover with the mouse over the words on the

screen in order to listen to their pronunciation and read the translation right when the word

appears. This allows for continuous, immediate access to both the pronunciation and the

translation of the words in an "on demand" fashion, giving the user many opportunities for

refreshing this knowledge. Included in the *hover* feature, the *conjugate* feature allows the user to

access the entire conjugation table for the selected verb. The *tips and notes* feature is a collection

of notes from the session that is being studied, while *discuss sentence* allows users to participate

in discussions about the specific sentence being studied with other users worldwide. These

functions contain the information related to the task at hand, rather than a collection of all the

information needed by the user to complete all the levels.

The following screenshots illustrate each feature: Figure 15 (Hover feature), Figure 16

(Conjugate feature), Figure 17 (Tips & Notes feature), and Figure 18 (Discuss Sentence feature).

Translate this

Figure 15. Hover feature.

Figure 16. Conjugate feature.
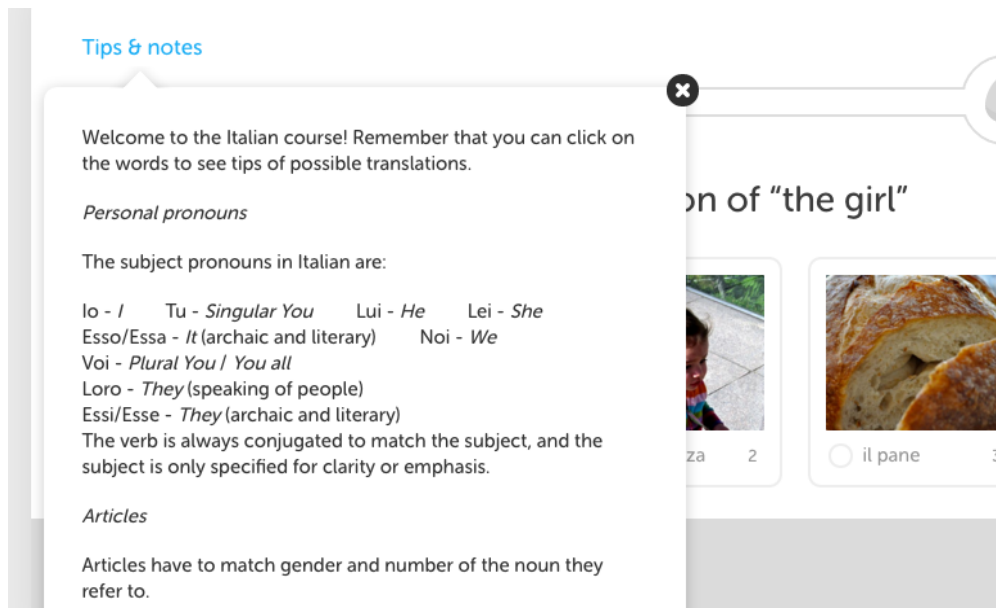
Figure 17. Tips & Notes feature.
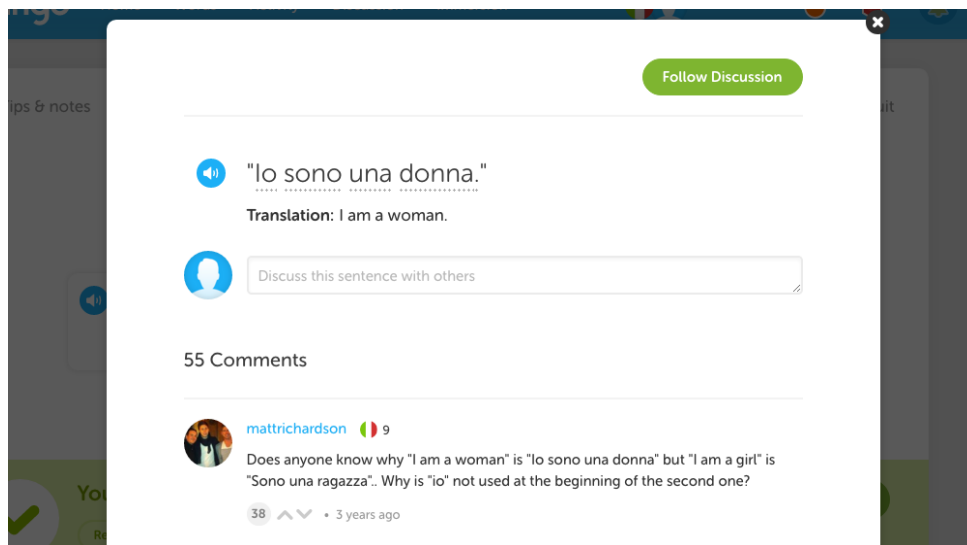
Figure 18. Discuss Sentence feature.

Lastly, the scorecard function is a tool that tracks the answers to each question and shows the user which were answered correctly and what the answer was, and which were answered incorrectly. Figure 19 (Scorecard feature) shows an example of the *scorecard* feature. By allowing the users to keep track of their own progress, including their mistakes, empowers them while also helping them to keep a record for future reference.
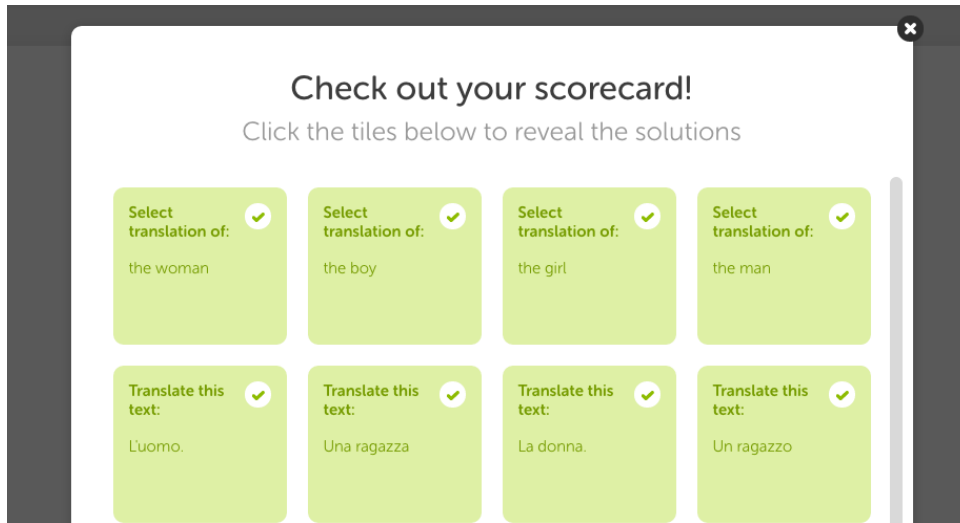
Figure 19. Scorecard feature.

Duolingo is divided into units, levels, and lessons. Units are the overarching categories of knowledge, which tend to approximate with level of proficiency. Each of these units is comprised of various levels, which deal with various topics – be they about structures, grammar, or content. Examples of grammar levels are *Plurals, Possession,* and so on, while examples of structure levels are *Phrases* or *Formal You.* Lastly, examples of content levels are *Animals, Household,* or *Politics*. In turn, each of these levels is composed of a series of lessons, which build one on the other to help the user understand all the concepts needed to complete the entire level.

The target level of completion was set to the fourth level of the first unit, where the first unit is tailored to beginners and the four levels are basic concepts (*Basics 1* and *Basics 2*), *Phrases,* and *Food*. Table 2 shows the number of lessons for each of these four levels.

Table 2

Duolingo modules selected for this study

Target: 4th Level (within the first unit)	Lessons per Level
Level 1 – Basics 1	3 lessons
Level 2 - Basics 2	5 lessons
Level 3 - Phrases	2 lessons
Level 4 - Food	6 lessons

3.5.1.2 Pre- and post-test

Participants were administered identical pre- and post-tests (Appendix A), which tested their Italian pronunciation skills. The test contained three sections: one testing pronunciation in words, one in phrases, and one in a longer passage. The decision to have these three sections is based on the fact that pronunciation varies when pronouncing just a single word as compared to that same word in a phrase. Further, comprehensibility and accentedness ratings are easier to give when listening to a phrase or a longer passage, rather than a single word, out of context.

In the test, the participants were exposed to both words and phrases from *Duolingo* and words and phrases that did not appear in the software. The words that did not appear in the software were taken from a corpus of written Italian called CORIS, a collection of authentic and commonly occurring texts chosen for their representativeness of modern Italian (http://corpora.dslo.unibo.it/coris_eng.html). The words selected from the corpus were chosen on the basis of their linguistic similarity to the words appearing in the software and their frequency in the corpus. For example, one of the geminates that appears in *Duolingo* is "donna" (woman). The word chosen as a similar variant to it is "gonna" (skirt), which participants will not have encountered in their *Duolingo* lessons but that has a high frequency rate in CORIS and shares the same type of gemination. In the longer discourse passage, a selection of words and phrases, all

displaying the features under investigation, was made into a short story about a restaurant experience, based on the fact that the first four levels of the program focus mainly on basic expressions and food.

The first section of the test was composed of 22 words with words representative of each category of interest: three geminates appearing in the *Duolingo* program, three geminates from CORIS, three singletons from *Duolingo* used for comparison with the geminates, three words with word initial position voiceless stop (/p, t, k/) from *Duolingo* and three from CORIS, and six distractors both from *Duolingo* and from CORIS. Distractors are words which do not contain the target features and are used to *distract* the participants and hide the features being targeted).

The second section of the test is composed of a list of 18 short phrases. Six sentences appeared in the *Duolingo* modules, while seven did not. The five distractor sentences were mixed (both from *Duolingo* and not). Each sentence contained one of the features (gemination or voiceless stop in word-initial position), excluding distractor sentences. Further, the sentences that were not taken from the *Duolingo* modules were created with the idea that they should follow a similar syntax and have the same level of complexity as the ones taken from the software, although the subject or vocabulary may differ. For example, "Io mangio un biscotto" (I eat a cookie) appears during the basics module in *Duolingo*, while "Lui batte il libro" (He types the book) does not. However, the two sentences share many similarities in terms of sentence construction. Both sentences start with a pronoun, then have a verb in the present tense, and then an object, composed of a singular, masculine article (one indefinite and one definite), and a singular, masculine noun. Further, both sentences contain gemination of /t/.The last section of the test presents a short passage. The passage contains some of the structures and vocabulary presented in the *Duolingo* modules, but also unfamiliar ones that however still contain the same focus features.

Both the words and the sentences on the test were randomized with data randomization software and divided into blocks of 4 or 5 before being presented to the participants. The

41

randomization of the words and the distractors serves the purpose of limiting the possibilities that the participants pick up on what is being tested and of any learner effect.

3.5.1.3 Pre- and post-interview

The pre-interview (Appendix B) asked about demographics, such as gender, age, level of education, and place of birth, and investigated the participant's language background, with questions regarding her experience learning Italian or visiting Italy, and experience with other languages. The aim of this section of the interview was to describe the participants in as much detail as possible in terms of their experience with language in order to better analyze the presence or absence, and quality, of acquisition. Lastly, the rest of the interview dealt with three main perceptions, or a priori themes: on language learning, especially in terms of the participant's perceptions on the best ways in which language is learned, on technology for language learning, with focus on examples of successful or unsuccessful experiences, and on language software and *Duolingo* specifically. The reasoning behind the a priori themes of language learning, technology, and online language-learning software, was that in order to be able to understand the participants' perceptions on the effectiveness of *Duolingo* for pronunciation, it was important to first understand each participant's perceptions of language learning in general and of how technology could fit into that framework, and only then to investigate whether or not *Duolingo* specifically aligned with those perceptions and frameworks.

The post-interview (Appendix D) asked participants to reflect on their answers in the pre-interview and to comment on any change in perceptions, if any, with a special focus on pronunciation with explicit questions on the topic. Questions about the perceptions on *Duolingo* were in fact mostly concentrated on pronunciation (e.g. "Which activity do you feel has improved your pronunciation skills the most?), as were the more general language information ones (e.g. What aspect of pronunciation did you struggle with the most (e.g. a specific sound, stress, rhythm, etc. ?). At this point of the study, the participants were informed about the focus on pronunciation and it was important to collect their impressions on this specific aspect.

3.5.1.4 Questionnaire

A questionnaire (Appendix C) was administered during the second week of the study with the aim of explicitly asking the participants to report on their perceived improvement and their opinions on what activities where most beneficial for said improvement, if any. The questionnaire asked about the use of *Duolingo* during the week and the participants' perceptions on their personal improvement in Italian overall. In order to understand which tasks seemed to be more effective, the questionnaire also contained questions on which activities participants' considered more or less effective. Question 2 asked about improvement and used a Likert scale from 1 to 9, with one being a marked improvement and 9 being no improvement at all. Questions 3 and 4 asked about the usefulness of the different types of activities included in the *Duolingo* modules, and Questions 5 and 6 were open answer questions and allowed for the participants to freely describe with which content they had had more or less difficulties.

3.5.1.5 Observation

As suggested by Schmuck (1997), observations can provide a way to collect data on nonverbal expression of feelings, interactions, or time spent on various activities. Further, participant observation allows researchers to "check definitions of terms that participants use in interviews, [...] and observe situations informants have described in interviews, thereby making them aware of distortions or inaccuracies in description provided by those informants" (Kawulich, 2005, section 4). Kawulich (2005) also suggests that observations can increase the validity of the study, especially with the use of additional data collection sources such as interviews, document analysis, questionnaires, or other more quantitative methods.

In order to collect such information (i.e. time spent of activities, situations described in the interviews) and to observe whether anything different than what expressed by the participants in the interviews occurred, the participants were asked to come in and use the program on the Lab computers. The participant's interaction with *Duolingo* was recorded using QuickTime's screen

recording function and replayed to the participant to gather further information on the session and elicit and record any comments.

3.5.1.6 Duolingo Reports

Duolingo reports were retrieved from the virtual classroom on *Duolingo*. Each report contained information about the number of times each learner accessed each lesson with a date stamp, the lessons completed and those not yet completed, and the amount of XP obtained for each lesson. The XP corresponds to the amount of points obtained for getting an answer correct in each activity. For example, if an activity asks for the user to select the appropriate article for a noun and the user chooses the correct one, he or she will receive a certain amount of points, or XP. On the other hand, if the user chooses the incorrect article, he or she does not gain any, and is presented with the same question again later. This process continues through the level until either the user has passed the level or has answered incorrectly too many times and has to restart the level.

It is therefore possible to look at each individual lesson and see how many times each learner accessed which lessons, or, in other words, how much practice they had on what topics/lessons. This was useful in analyzing which sections were more or less challenging for each participant, and to investigate any connection between practice and improvement. However, the reports did not provide information for time on task, and thus an analysis of this variable was not possible.

3.5.1.7 Ratings

Each rater was sent a total of 154 samples from both the participants and the native speaker. The audio files were randomized using Qualtrics' block randomization feature both intrablock and inter-block. In other words, the samples were randomized and divided into blocks of five samples, from different participants, each (and one block of four). Then the blocks were also randomized so that each rater would receive different blocks and in a different order.

Raters were also given two scales on which to rate samples (Appendix E). These scales were adapted from a study (Levis, Link, Sonsaat, & Barriouso, 2016) on the impact of native speaker status on learner performance in pronunciation, and originally discussed in Munro and Derwing (1994). The first scale was about accentedness, defined as the degree of the markedness of foreign accent, and ranged from 1 to 9, with 1 being no foreign accent and 9 being very strong foreign accent. The second scale was about comprehensibility, defined as the ease of understanding the single utterances, and ranged from 1 to 9, with 1 being extremely easy to understand and 9 being impossible to understand. The raters were told to complete the ratings in one sitting, through Qualtrics, and with the use of noise-cancelling headphones. They were also given a detailed description of the scales and two training examples and instructed to listen to the each audio file only once.

3.5.2 Procedures

Figure 20 gives an overview of the data collection procedures for each participant. The study was divided into four weeks: week one for the pre-interview and pre-test, week two for the questionnaire, week three for the observation, and week four for the post-test and post-interview. The items in the lighter color were analyzed through qualitative coding methods, while the darker colored items were analyzed through Praat and descriptive statistics.
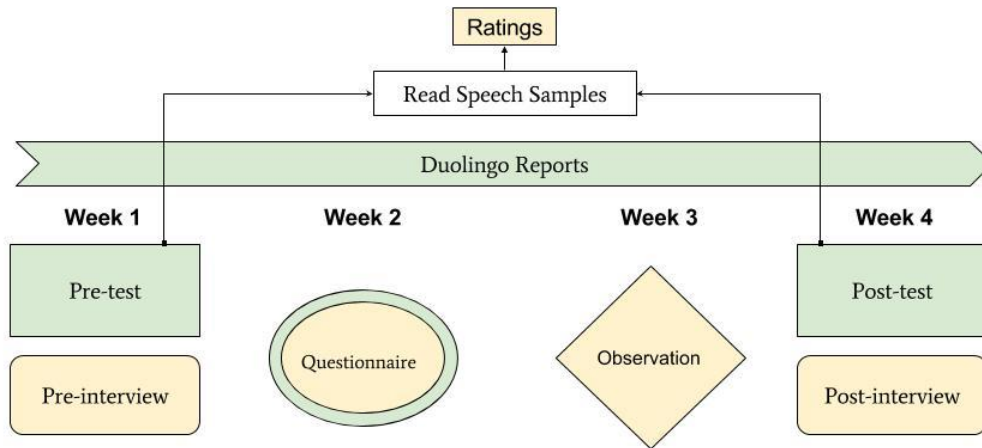
Figure 20. Overview of data collection procedures for each participant.

After recruitment, the pre-interview was administered and took an average of 30 minutes. The participants were then given the pre-test, which was recorded in the Linguistic Laboratory and took an average of 15 minutes, excluding volume and setting checks[1]. Although the focus was on specific pronunciation features, the subjects were not made aware of this, or even that the focus was on overall pronunciation. This allowed the participants to be involved in the language learning process across the various skills, and not only concentrate on pronunciation. It also prevented the adding of any pressure to the already natural uneasiness that comes from speaking in an unfamiliar foreign language. However, the participants were given this information at the conclusion of the study and allowed to comment specifically on the journey with pronunciation.

Further, both the words and the sentences on the test were randomized with data randomization software and divided into blocks of 4 or 5 tokens before being presented to the participants. The randomization of the words and the distractors serves the purpose of limiting learner effect, also known as practice effects or carry-over effects. Learner effect results in

---

[1] Participants were recorded with a Marantz Professional PMD 660 recorder and a Behringer ECM8000 Ultra-Linear Measurement Condenser Microphone.

increased performance in certain parts of the test as a result of the participant understanding the format or the purpose of the test, developing strategies, and in general becoming more familiar with the test as he or she progresses through it.

The subjects were then asked to access *Duolingo* on their personal computers. They were instructed on the level they were required to reach by the completion of the study, but were not given a specific timetable. In other words, the participants were allowed to use the software for as long and as often as they desired. The only requirement was that they reach the target level of completion. The reason why this study did not include a detailed timetable is because the aim was to replicate as much as possible an authentic self-study environment in which the learner decides when and how much learning will take place during the day. Also, monitoring these patterns helped in the analysis of each participant's behavior with the software and helped interpret perceptions towards the software.

According to the latest study on *Duolingo*'s effectiveness for Spanish, conducted by Vesselinov and Grego (2012), a "person with no knowledge of Spanish would need between 26 and 49 hours (or 34 hours on average) to cover the material for the first college semester of Spanish" (p. 1). Based on these calculations, it was estimated that to complete the four levels of this study, each participant would have had to spend about two hours per week on *Duolingo*. This was considered an acceptable and appropriate amount of work for a four-week study, in which the participants are not already enrolled in a class or other program but make use of the software on their own time.

At the beginning of the second week, a questionnaire was administered through Google Forms and was calculated to take an average 10 minutes. During the third week of the study, the participants were not given a questionnaire. Instead, they were asked to come in for an observation. This allowed for the collection of spontaneous and recalled reactions to the software, which might have not surfaced in the structured questionnaires. The observation was designed as a stimulated recall. Participants were instructed to use the software from where they had left off

for about 15 minutes and this interaction was screen recorded. During the following 15 minutes, the participants were asked to watch the recording and reflect on what they had done. Both elicited comments and possible spontaneous comments were recorded for analysis.

At the conclusion of the study, or at the end of the fourth week, participants were asked to come in for the post-test whose procedures were similar to the pre-test. The participants were also administered the post-interview in a similar fashion to the pre-interview.

Both the samples from the pre-test and the samples from the post-test were then sent through Qualtrics to the novice raters, after being normalized at 0.0 db to ensure they were all at the same volume and pace and with limited external noise. Each rater was instructed on the procedures to follow before receiving the samples and was given two training examples. The raters were given two scales (Appendix E): one for comprehensibility and one for accentedness for each sound file (N =154).

3.6   Data Analysis

The data was analyzed using various methods appropriate to the source the data came from. Table 3 illustrates the relationship between the research questions, the data collection methods, and the data analysis methods. While the various data sources attend to one of the three research questions more so than to the others, each source will carry information that helps in answering all the questions. Further, it is the interrelated analysis of all the data sources that will be most successful in providing answers to the questions.

Table 3

Data Triangulation

Research Question	Data Collection Method	Data Analysis Method
1. How effective *is Duolingo* for the acquisition of the pronunciation macro skills of accentedness and comprehensibility?	• *Duolingo* Reports • Comprehensibility / Accentedness Ratings	• Descriptive Statistics • Paired *t*-test
2. How effective *is Duolingo* for the acquisition of the pronunciation micro skills of geminate contrast and voice onset time (VOT) for voiceless stops in word initial position for Italian as a second language learners?	• *Duolingo* Reports • Read Speech Samples	• Acoustic Analysis with Praat
3. What are the learners' perceptions (user experience) of the effectiveness of *Duolingo* for the acquisition of pronunciation skills?	• Interviews • Observations • Questionnaires	• Open Coding • Axial Coding • Selective Coding

3.6.1 Research Questions 1 and 2

The first two research questions asked about the effectiveness of *Duolingo* for the

acquisition of the pronunciation macro skills of accentedness and comprehensibility and the

pronunciation micro skills of geminate in word-medial position contrast and voice onset time

(VOT) for voiceless stops in word-initial position for Italian as a second language learners.

3.6.1.1 Duolingo Reports

Each report was analyzed with the coding of three categories of data: the number of times

each learner accessed each lesson with a date stamp, number of lessons completed and number of

lessons not yet completed, amount of XP obtained for each lesson. The latter two categories were

excluded from the final analysis since all participants reattempted any lesson where incorrect

answers were given, and therefore also received the maximum amount of XP possible for each

lesson. Therefore, the analysis concentrated on computing how much practice, or number of times

a specific lesson was completed, each participant had. These data were then compared to other

data sources, and specifically to the acoustic analysis measurements and the ratings, in order to understand if there existed any connection between the amount of practice and improvement in macro (comprehensibility and accentedness) and/or micro pronunciation (gemination and VOT) skills.

3.6.1.2 Ratings

The rating scores were collected from Qualtrics, and exported into an excel worksheet. The data was then imported into ReCal 0.1 Alpha for two Coders, a software that calculates interrater reliability among two raters. Further, in order to better understand where the raters were most in agreement, single interrater reliability measures were taken for each of the four subgroups of ratings: (1) pre-test comprehensibility ratings, (2) post-test comprehensibility ratings, (3) pre-test accentedness ratings, (4) post-test accentedness ratings. The results of this analysis are summarized in Table 4 and expressed in percentage agreement.

Table 4

Interrater reliability

		Percent Agreement
Comprehensibility Ratings	Pre-test	68.9%
	Post-test	73.3%
Accentedness Ratings	Pre-test	60.0%
	Post-test	82.2%
Overall		71.1%

Although some of the agreements are not too strong, the average agreement rate is significant (71.1%). Also important to note are the higher agreements rate in the post-tests (73.3% for comprehensibility and 82.2% for accentedness) as compared to the pre-tests (68.9% for comprehensibility and 60.0% for accentedness), a fact which can be explained by the high amount of fluctuations in the pronunciation of true beginners in pre-tests. Therefore it is possible

to base conclusions, albeit tentative, based on the ratings of the two novice raters used in this study. The native speakers ratings were also analyzed though paired two-tailed *t*-tests in order to understand whether any improvements in ratings were statistically significant. An equal variances assumption was adopted when applying the *t*-test because the difference in the standard deviations of the sampled distributions was rather small. Further, effect size was calculated using an eta squared statistic calculator and interpreted using the guidelines proposed by Cohen (1988), or: .01=small effect, .06=moderate effect, and .14=large effect.

3.6.1.3 *Acoustic Analysis with Praat*

The samples, a total of 330 tokens, then underwent acoustic analysis in Praat, a scientific computer software package for the analysis of speech in phonetics. In order to gather the results from the data, a Praat script was created and run[2]. The script measured the duration of each singleton and geminate marked on the relative TextGrid and reported the information annotated on the tiers: (1) token number, (2) word in English and Italian, (3) feature name (in this case gemination), (4) singleton or geminate, and (5) the duration of the C in milliseconds (ms). Similarly, another Praat script was created in order to measure VOT. It recorded information annotated on the following tiers: (1) token number, (2) word in English and Italian, (3) feature name (in this case VOT), (4) duration of VOT in milliseconds (ms). Lastly, both the voice onset time averages and the geminate to singleton ratios of the learners were compared to those of the selected native speaker, as one of the criteria to evaluate acquisition was the approximation to native-like production.

3.6.2 Research Question 3

The third research question asked about the learners' perceptions (user experience) of the effectiveness of *Duolingo* for the acquisition of pronunciation skills.

---

[2] The Praat script was created by Dr. Nancy Caplow (Oklahoma State University).

3.6.2.1 Interviews, Observations, Questionnaires

The following data sources (interviews, observations, and questionnaires) were analyzed through interpretive analysis, which is comprised of three stages: deconstruction, interpretation, and reconstruction (Cohen, Manion, & Morrison, 2011; Sargeant, 2012). *Deconstruction* is the process of breaking down data into the various parts that compose it so as to better see what these parts are and what is or is not included, while the i*nterpretation* process that follows, refers to making sense of and understanding the coded data (Sargeant, 2012). This can be done in numerous ways, such as coding for themes. As Sargeant (2012) suggest, it is also important to compare findings with those of other studies and to look for theories which might explain relationships among themes. Lastly, *reconstruction* refers to recreating the codes and themes so as to show the relationships and insights derived in the interpretation phase and "explains them more broadly in light of existing knowledge and theoretical perspectives" (Sargeant, 2012, p. 2).

The interpretation of the data was done here through open, axial, and selective coding (Creswell, 2014; Strauss & Corbin, 1990). The main aim of this analysis was to provide an in-depth account of each participant's experiences and perceptions of language learning, and of technology for language learning and *Duolingo*, specifically. While the first step of the analysis was completed through open coding, the aforementioned elements were set as a priori themes (i.e. perceptions on language learning and perceptions of technology for language learning and *Duolingo*). Once the a priori themes were coded, the following step was to find sub-themes that emerged from the participants' discussions of the major themes, through the use of axial coding. These sub-themes were then cross-referenced in all three data sources in order to gain more insight into each sub-theme from different perspectives and through different mediums, and axial coding was used to extrapolate significant quotes to be used as descriptors of the themes and categories. Then, all the resulting data were confronted with information and findings from the literature and a set of key findings was shaped. Lastly, these findings were compared to the findings from the tests, *Duolingo* reports, and ratings with the aim of comparing *Duolingo* for

pronunciation both in terms of quantitative effectiveness (acoustic analysis, ratings, reports) and perceptual effectiveness (user experience).

3.6.2.2 *Member Checks*

Member checking is a common procedure used in qualitative research used to check the interpretations and reconstructions that the researcher has developed by having field participants review the conclusions and statements for both their accuracy and their completeness (Gall, Gall, & Borg, 2003). Comments concerning the major sub-themes that emerged from the interviews which were either in some part incomplete or for which an interpretation needed more information were selected and compiled for member checking. Participants were asked to comment on what they received so as to validate or confute my interpretations and add to the depth of the overall analysis.

3.7   Chapter Summary

This section discussed the methodology for this study. Since the main aim of this research was to provide an in-depth, holistic analysis of each participant's perceptions of online language software through the use of *Duolingo,* a qualitative case study methodology was selected. The main aim of the study is to measure and describe any improvement, or lack thereof, in acquiring macro and micro pronunciation skills through both quantitative methods and qualitative methods, so as to compare the actual effectiveness of *Duolingo* for pronunciation to its perceived effectiveness by the participants. Through the collection of data obtained from tests, interviews, observations, reports, and ratings, triangulation was ensured. After analysis, a selection of findings was sent to participants as a part of member checking. The following chapter will overview the results of the acoustic analysis, the ratings, and the *Duolingo* reports, and provide a discussion of the main findings in relationship to the first two research questions: (1) How effective *is Duolingo* for the acquisition of the pronunciation macro skills of accentedness and comprehensibility; (2) How effective *is Duolingo* for the acquisition of the pronunciation

micro skills of geminate contrast in word-medial position and voice onset time (VOT) for

voiceless stops in word-initial position for Italian as a second language learners?

CHAPTER 4

RESULTS: EFFECTIVENESS OF DUOLINGO

This chapter outlines the results of the *Duolingo* reports, the ratings, and the acoustic analysis for each participant. Each section begins with an overview of the participants' language backgrounds, in order to understand whether some influence or transfer is possible or present. Then an analysis of the *Duolingo* reports, which were used to measure the amount of practice for each participant, aims to establish a connection between repetition and improvement.

The sections then continue with an overview of the results of the ratings and the paired *t*-tests conducted on the ratings, which are used to answer the first research question, which asked about the effectiveness of *Duolingo* for the acquisition of the pronunciation macro skills of accentedness and comprehensibility for Italian as a second language learners. Further, to attempt to account for the results obtained from the ratings, each section will contain a description of the results from the acoustic analysis in Praat, which are used to answer the second research question that asked about the effectiveness of *Duolingo* for the acquisition of the pronunciation micro skills of geminate contrast and voice onset time (VOT) for voiceless stops in word initial position for Italian as a second language learners.

As mentioned before, this study measures effectiveness in terms of three main factors: (1) intelligibility; (2) acoustic evidence of gemination contrast and shorter voice onset time as compared to English; (3) the approximation to native-like production of the gemination contrast and voice onset time.

The data for each analysis is displayed for all participants through graphs: each measurement is displayed as a vertical line with three points (i.e. the lowest measurement, the mean, and the highest measurement). Each figure also contains the measurements of the native speaker (NS). Unless indicated on the x-axis data label line, the plots are based on four tokens of each target sound, with the exception of measurements for singletons in sentences and all features in the passages, which are based on two tokens.

4.1   Native Speaker

As mentioned, the following sections contain the acoustic analysis for each participant. As all the measurements are shown in comparison to those of the native speaker, it is important to overview the characteristic of the production of the geminate contrast and of the voice onset time of the voiceless stops of the native speaker.

For the contrast between singleton /t/ and geminate /tt/, the native speaker's singletons averaged 105 ms while for the geminate 215 ms – that is, the native speaker produced geminates which were almost exactly double the singletons. Very similarly, in the case of the n:nn contrast, the native speaker's singletons were on average 67 ms long, while his geminates were 152 ms long – more than double. The sharpest contrast was between singleton /n/ and geminate /nn/. The native speaker's /n/ was on average 66 ms long, while his /nn/ was 176 ms, or, in other words, the native speaker produced geminates which were almost triple the duration of the singletons.

As for the second feature under investigation, the native speaker's voice onset time for both /p/ and /t/ was on average 18 ms. Slightly longer, his average voice onset time duration for /k/ was 33 ms. Overall, though, he produced all voiceless stops in word-initial position tested in study as unaspirated, or with a voice onset time of less than 35 ms.

56

4.2   Emma

Emma's native languages are Russian and Ukrainian. She is also fluent in German, which she learned in college and through a one year stay in Germany, and English, which she studied during school with help from her mother who is an English teacher and through her residency in the United States as a graduate student. Further, she is a PhD student in Teaching English as a Second Language (TESL) which means she is trained in and knowledgeable about language acquisition processes.

The literature on these languages indicate that in Russian there is some contrastive gemination, although double consonants are not always realized as phonetically long. The phonological environments in which double consonants are more commonly lengthened in production are in consonant sequences that span a morpheme boundary within a word or in a phrase (concatenated geminates) and on the boundary of prefixes, although at time also in intervocalic position (Dmitrieva, 2012). Further, "there is a high frequency of geminated [n] in Russian" (Dmitrieva, 2012, p. 60). Ukrainian also has consonant gemination usually in intervocalic position. In the case of West Germanic, consonant gemination is usually before j (Ham, 1997). Further, Russian and Ukrainian produce voiceless stops as unaspirated (Nagy & Kochetov, 2013), while the vast majority of German dialects show long-lag VOT in p, t, k, meaning they are aspirated (Jessen & Ringen, 2002).

Emma accessed *Duolingo* 15 times, accessing levels from three to eight time, and completing 43 lessons. Table 5 shows on overview of her report, including repeated lessons marked with an asterisk.

Table 5

Emma's Duolingo Report: Overall days of practice, number of lessons completed, and number of times each lesson was repeated

	Days	N. lessons[a]	Repeated Lessons[b]															
			Basics			Basics 2					Phrases		Food					
Levels			*1*	*2*	*3*	*1*	*2*	*3*	*4*	*5*	*1*	*2*	*1*	*2*	*3*	*4*	*5*	*6*
Emma	15	43	*	**	**	**	**	*	*	*	*	**	**	**	**	**	**	**

[a] *Note*: number of lessons completed including repeated lessons
[b] *Note:* the asterisk indicates a repeated lesson; the number of asterisks indicates how many times

4.2.1 Ratings

For comprehensibility, Emma showed a difference in ratings of 1.2, going from a score of 3.5 in the pre-test to a score of 2.3 in the post-test. However, the difference between the pre-test (M=3.50, SD=1.91) and the post-test (M=2.72, SD=1.70) was not statistically significant; $t(17) = 1.59$, $p > 0.05$. As for accentedness, Emma showed a difference in ratings of 1.3. This difference between the pre-test (M=5.88, SD=2.08) and the post-test (M=4.55, SD=1.65), though, was also not statistically significant; $t(17) = 1.82$, $p > 0.05$.

Overall, Emma did not show improvements in terms of the native speaker ratings, although she is the participant who practiced the most. The following section reviews the acoustic analysis on Emma's read speech samples, which reveals that there was improvement in voice onset time, although it did not translate to improvement in intelligibility.

4.2.2 Acoustic Analysis

4.2.2.1 Gemination in words

Figure 21 shows the data for Emma. The data show that Emma was already making some contrast in the pre-test and that the contrast did not change much in the post-test. Further, the geminate data in the post-test indicates that the geminates are still too short (M=138 ms) as

compared to the native speaker (M=213 ms) and have considerable overlap with the singletons, which have a high value of 137 ms.
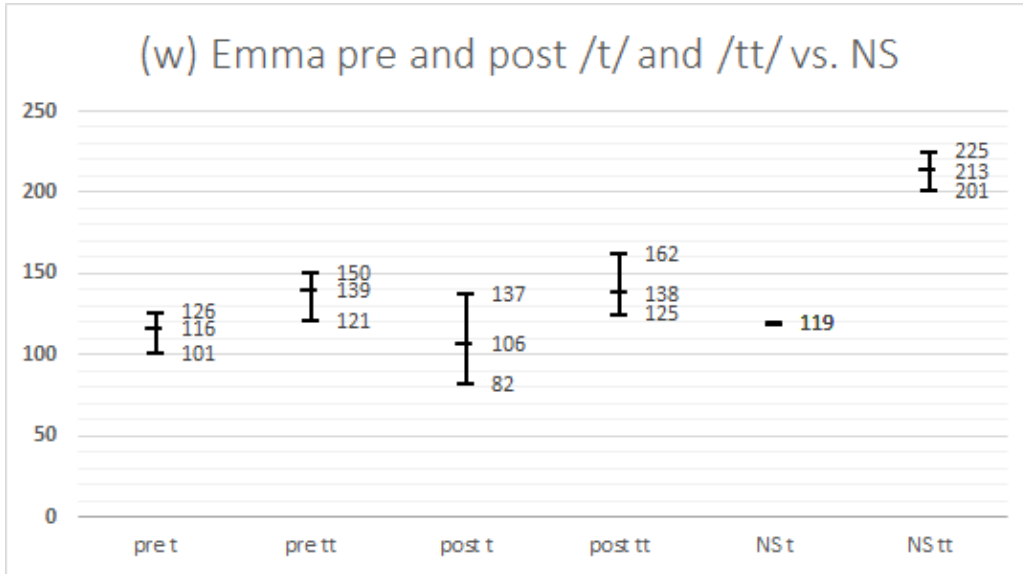
Figure 21. Emma's pre and post /t/ and /tt/ in words (w) vs. native speaker (NS) data (msec).

Emma's data for /n/, shown in Figure 22, indicate that there was improvement in the production of a contrast, as in the pre-test Emma did not seem to make any distinction between the singletons and the geminates. In the post-test, although the highest value for the singleton overlaps with the lowest value of the geminate, there is quite a difference in means: 64 ms for the singleton and 95 ms for the geminate. However, the contrast is still not as pronounced as that of the native speaker whose geminate (M=143) is double the singleton (M=75) in duration.
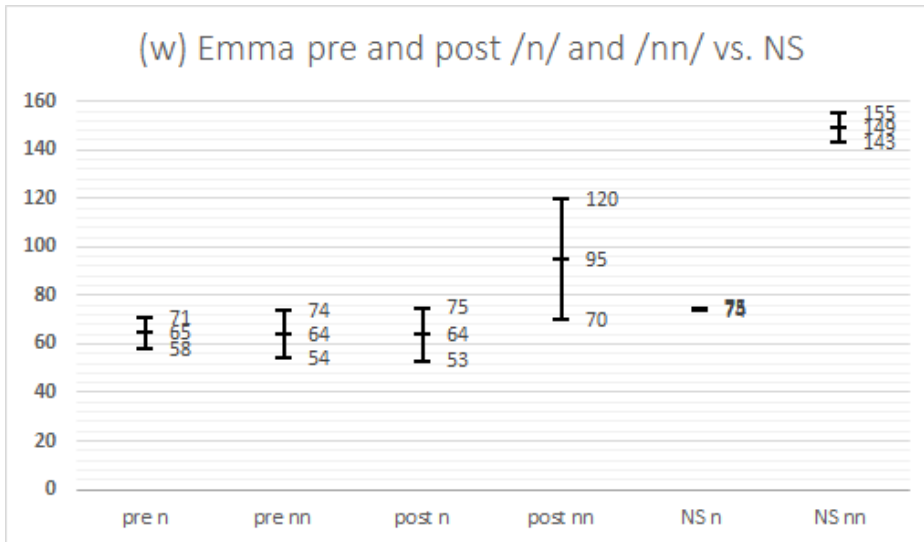
Figure 22. Emma's pre and post /n/ and /nn/ in words (w) vs. native speaker (NS) data (msec).

Lastly, Emma does seems to produce some contrast between the singleton /l/ and the geminate /ll/ (Figure 23) in both tests, though her geminates still did not approach the duration produced by the native speaker. Indeed, the geminates did not change much from the pre- (M=94 ms) to the post-test (M=91 ms) and are almost half the duration of those of the native speaker (M=170 ms).
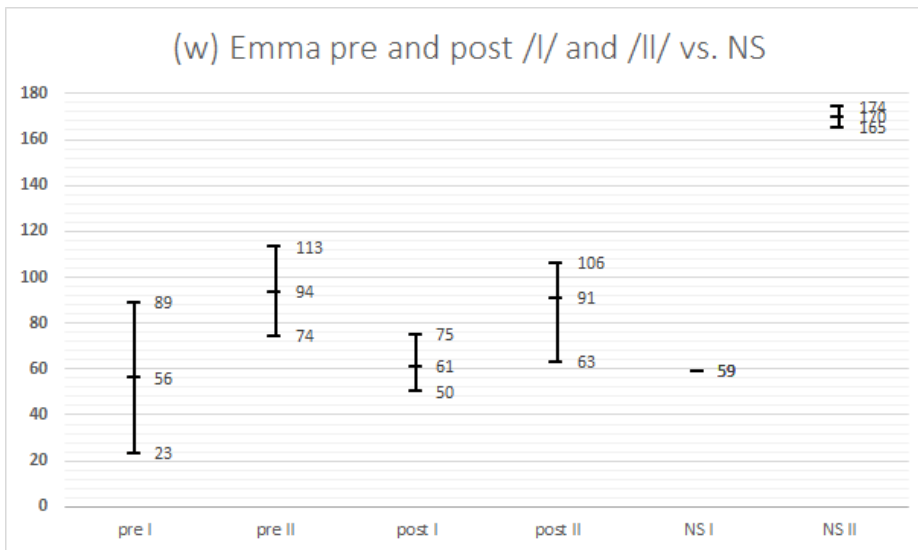
Figure 23. Emma's pre and post /l/ and /ll/ in words (w) vs. native speaker (NS) data (msec).

Emma's ratios for the contrast in words is shown in Table 6. The data from the three ratios match what was discovered in the acoustic analysis, which is little to no improvement in the tt to t and the ll to l ratio, and some improvement in the nn to n ratio. Further, all the ratios are still smaller than those of the native speaker.

Table 6

Emma's geminate contrast ratios vs. native speaker ratios in words

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	**Post**		**Pre**	**Post**		**Pre**	**Post**	
Emma			1.79			1.99			2.87
	1.20	1.29		0.97	1.46		1.67	1.49	

4.2.2.2 Gemination in sentences

Emma shows a contrast between /t/ and /tt/ in sentences (Figure 24) in the pre-test but not in the post-test, with much overlap between the geminates and the singletons. As for the words, no improvement seems to have been made.
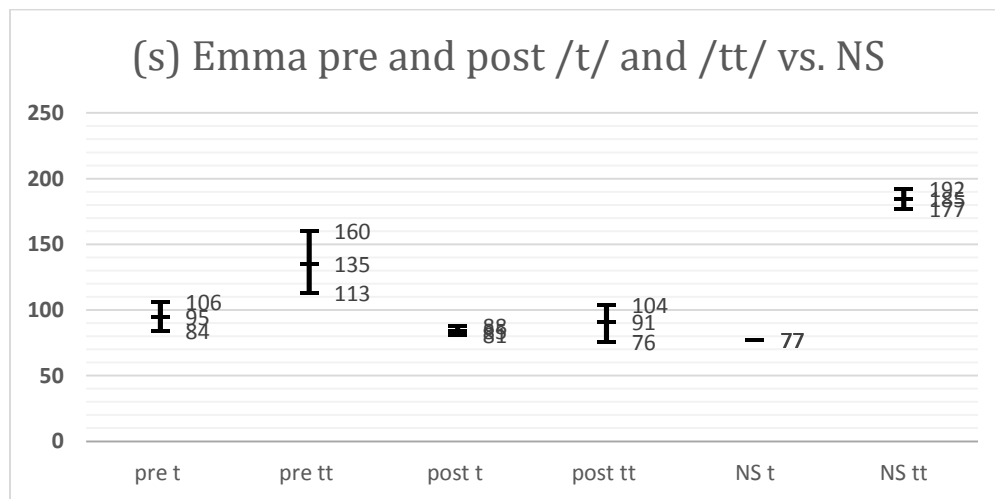
Figure 24. Emma's pre and post /t/ and /tt/ in sentences (s) vs. native speaker (NS) data (msec).

The data on the nn to n contrast, shown in Figure 25, instead show a significant improvement. Not only did Emma produce a stronger contrast in the post-test, she also tightened the ranges and reached those of the native speaker.
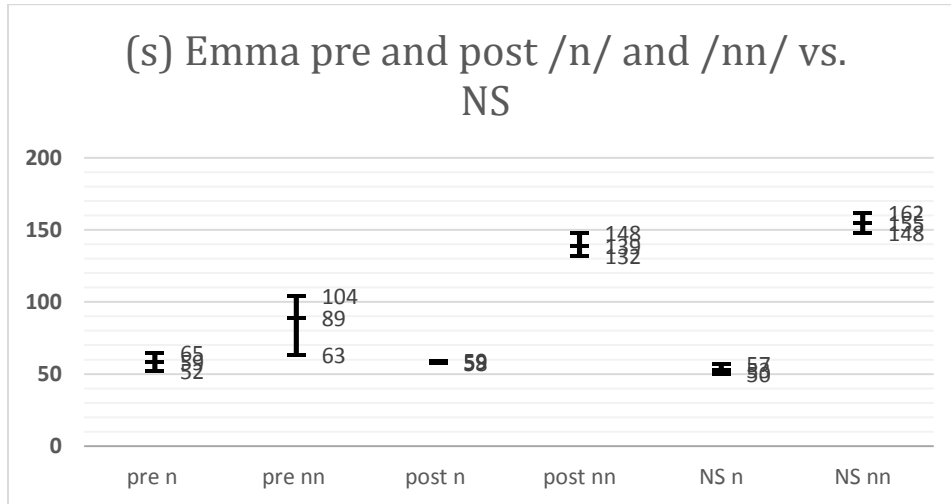
Figure 25. Emma's pre and post /n/ and /nn/ in sentences (s) vs. native speaker (NS) data (msec).

Lastly, the data for the ll to l contrast in Figure 26 show the presence of a contrast in the post-test. However, since the data on the pre-test geminate are missing, due to the fact that it was produced as a [j], it is not possible to determine whether there was improvement or not.
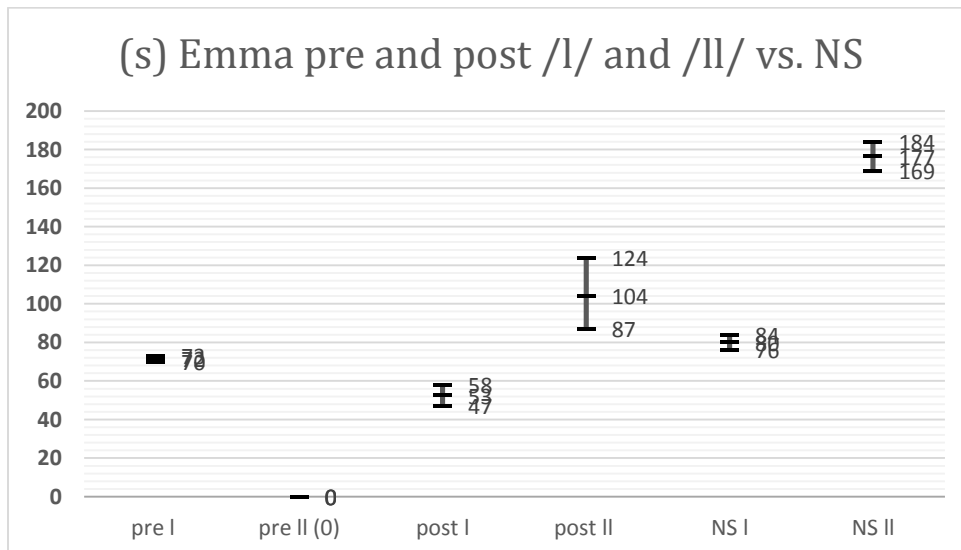
Figure 26. Emma's pre and post /l/ and /ll/ in sentences (s) vs. native speaker (NS) data (msec).

The ratios shown in Table 7 corroborate the data highlighted in the graphs. The only contrast in which there was improvement is that between singleton /n/ and geminate /nn/. Indeed, the ratio in the pre-test of 1.50, becomes a ratio of 2.35 in the post-test, indicating that geminates are more than double the singletons in duration. However, the other two contrast show no signs of improvement, although in part for lack of data as well.

Table 7

Emma's geminate contrast ratios vs. native speaker ratios in sentences

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	**Post**		**Pre**	**Post**		**Pre**	**Post**	
Emma	1.42	1.07	2.40	1.50	2.35	2.92	n/a	1.96	2.21

4.2.2.3 Gemination in passage

The data shown in Figure 27, relative to the contrast between /t/ and /tt/ in the passage, indicate that there was no improvement in the production of this feature. Indeed, in the post-test, the singleton and geminates seems to have approximately the same duration with a mean of 95 ms for singletons and a mean of 90 ms for geminates.
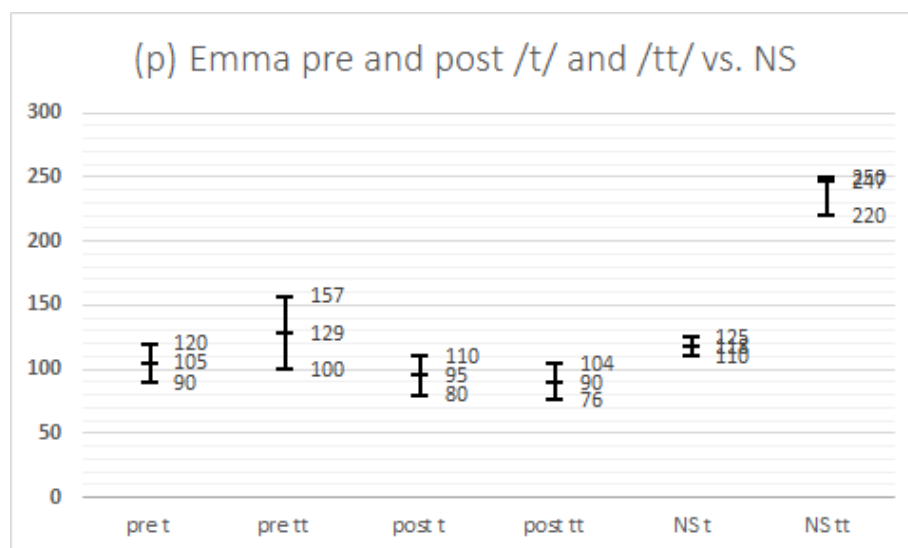
Figure 27. Emma's pre and post /t/ and /tt/ in passage (p) vs. native speaker (NS) data (msec).

As for the n to nn contrast, show in Figure 28, Emma already produced an appropriate contrast in the pre-test. However, she produced a more pronounced contrast in the post-test by lengthening the duration of her geminates.
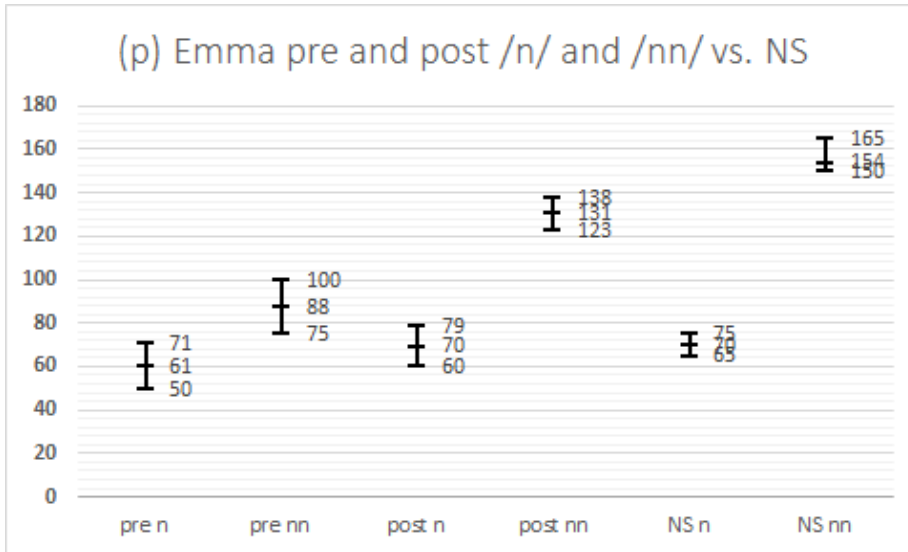
Figure 28. Emma's pre and post /n/ and /nn/ in passage (p) vs. native speaker (NS) data (msec).

Lastly, in the contrast between /l/ and /ll/, shown in Figure 29, Emma produced a contrast in both pre- and post-test, although the duration of her geminates shortened in the post-test. The contrast is still not as sharp as that of the native speaker, however.
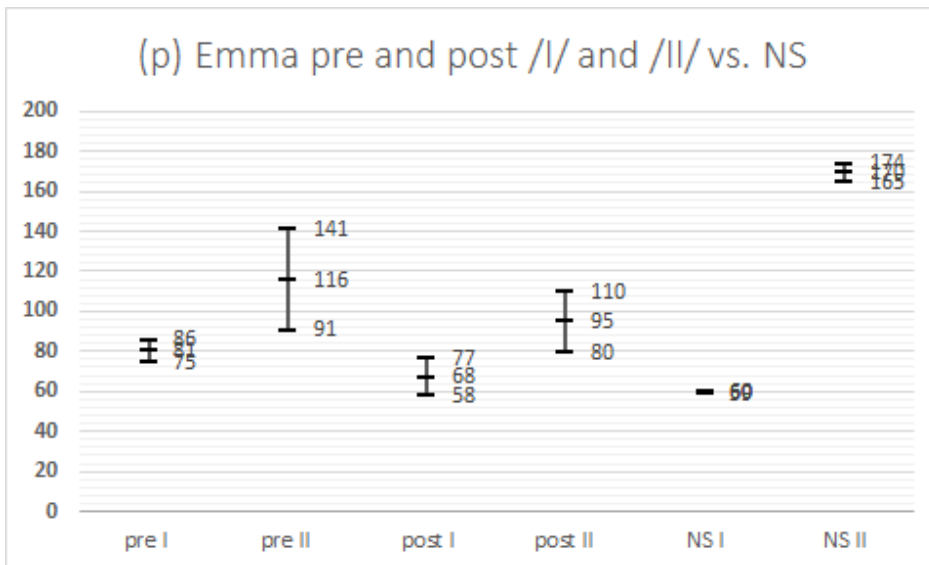
Figure 29. Emma's pre and post /l/ and /ll/ in passage (p) vs. native speaker (NS) data (msec).

64

The ratios in Table 8 indicate similar results. While there is no evidence of improvement in the tt:t ratios, the other two ratios indicate that either the contrast is more pronounced or the range of values have tightened indicating some improved control.

Table 8

Emma's geminate contrast ratios vs. native speaker ratios in passage

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	Post		Pre	Post		Pre	Post	
Emma	1.22	0.94	2.10	1.44	1.87	2.05	1.43	1.39	2.85

4.2.2.4 Voice Onset Time in words

Figure 30 shows the data for Emma's /p, t, k/ voice onset times in words. Overall, Emma decreased the duration of aspiration. For /p/, it seems as if Emma tightened the range in the post-test, showing more control over the feature, and approached the native speaker production, with a post-test mean of 26 ms (as compared to the native speaker's mean of 14 ms). For /t/, the data indicates that she shortened her average by 25 ms, going from a pre-test mean of 69 ms to a post-test mean of 44 ms. However, her post-test range is still quite large, even if her lowest value is quite close to the measurements for the native speaker, indicating that there is no significant improvement. Lastly, Emma's data for /k/ shows that she improved the most out of the five participants, going from a pre-test mean of 75 ms to a post-test mean of 55 ms, and keeping tight ranges. Although her measurements are still longer than those of the native speaker, the evidence seems to point to improvement in both /p/ and /k/.
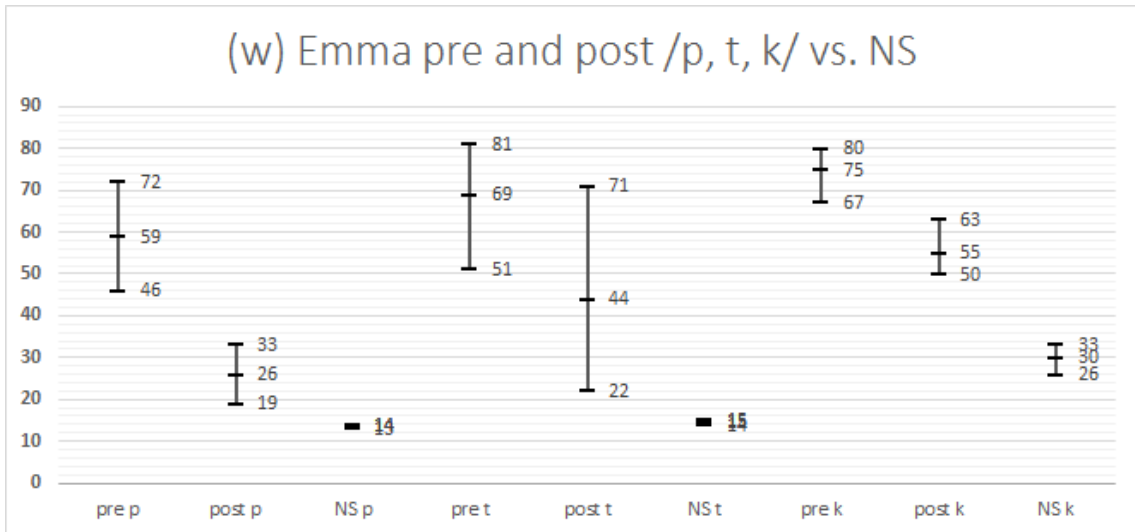
Figure 30. Emma's pre and post /p, t, k/ in words (w) vs. native speaker (NS) data (msec).

4.2.2.5 Voice Onset Time in sentences

Figure 31 shows Emma's voice onset time measurements for /p, t, k/ in sentences. In the pre-test /t/ one outlier of (360 ms) is not displayed in the graph so as to be able to focus on the rest of the data. The improvement in /p/ in words does not seem to have been carried out in the sentences. However, both /t/ and /k/ improved considerably more than in words. Not only did the ranges tighten, voice onset time duration reduced by up to 85 ms. Emma, thus, seems to produce less aspirated stops here, too.
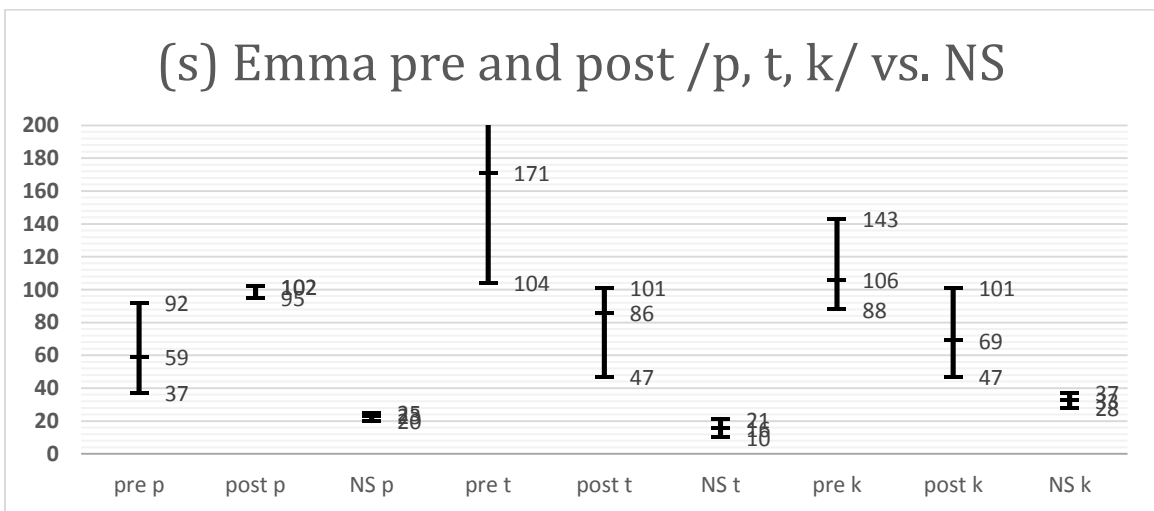
Figure 31. Emma's pre and post /p, t, k/ in sentences (s) vs. native speaker (NS) data (msec).

66

4.2.2.6 Voice Onset Time in passage

Figure 32 shows the data for voice onset time of /p, t, k/ for Emma in the passage. The improvement of the voice onset time of /p/ seen in the words does not seem to have happened in the passage either. However, both the VOT of /t/ and of /k/ considerably improved in the passage, similarly to the sentences.
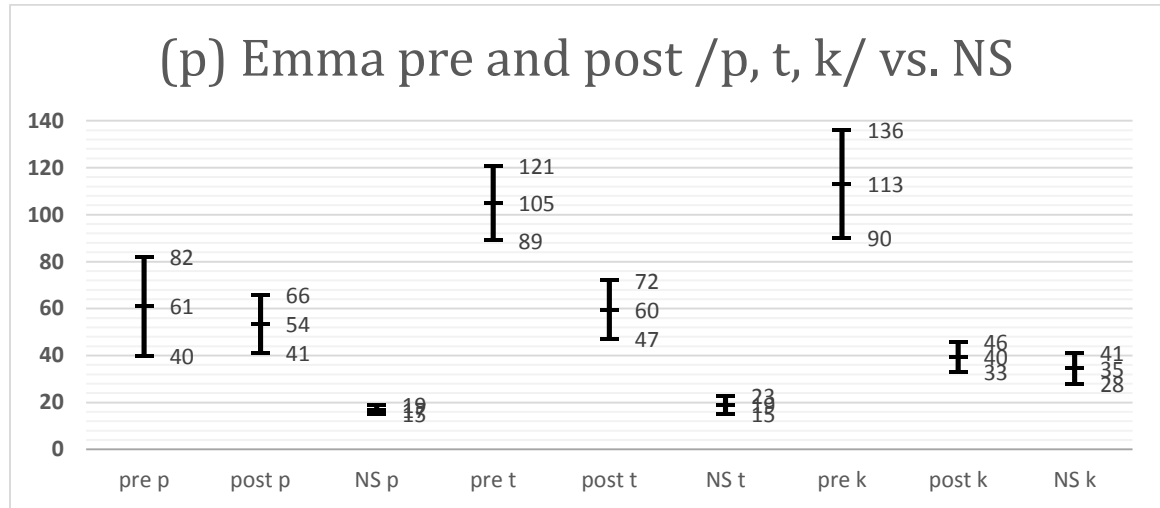
Figure 32. Emma's pre and post /p, t, k/ in passage (p) vs. native speaker (NS) data (msec).

4.2.3 Summary (Emma)

Overall, Emma's ratings indicate that there was no statistically significant improvement in either comprehensibility or accentedness. However, the acoustic analysis did show much improvement in the production of the voiceless stops. This improvement could be due to the fact that in both Russian and Ukrainian voiceless stops are always produced as unaspirated, like in Italian, and partly to her practice. No improvement, though, was seen in the production of a contrast between singletons and geminates with the exception of the nn to n contrast. This lack of improvement cannot be fully explained by her language background; even if gemination is not as frequent in Russian and Ukrainian as it is in Italian, it does exist and often occurs intervocalically, as in the words of the test in this study.

Barbara's language background is in French, which she studied both in high school and college. In the majority of French variations, double consonants do appear in the orthographic form of many French words, but they tend to be relatively rare in pronunciation. An example of this is the verb *to learn,* 'apprendre', [aprɑ̃ndrə] in which the geminate /p/ is not lengthened. Further, a few cases of phonetic gemination do not correspond to double consonant letters in the orthography. An example of this occurs when the vowel /e/ is deleted between to identical consonants such as in the word for *similarly*, 'mêmement', [mɛmmɑ̃] (Tranel, 1987). On the other hand, similarly to Italian, French voiceless stops are unaspirated (Caramazza & Yeni-Komshian, 1974; Tranel, 1987).

The *Duolingo* report for Barbara showed that she accessed *Duolingo* ten times, accessing each level from two to four times, and completing 17 lessons. Table 9 shows on overview of her report, including repeated lessons marked with an asterisk. It is important to notice that the "Test" function is not counted as a repetition of the lesson. The function does not repeat the entire lesson, but rather extracts from the lesson and offers no interactive functions or help functions (such as *hover, conjugate, explain, discuss sentence,* and *tips and notes*).

Table 9

Barbara's Duolingo Report: Overall days of practice, number of lessons completed, and number of times each lesson was repeated

	Days	**N. lessons**[a]	**Repeated Lessons**[b]															
			Basics			Basics 2					Phrases		Food					
Levels			*1*	*2*	*3*	*1*	*2*	*3*	*4*	*5*	*1*	*2*	*1*	*2*	*3*	*4*	*5*	*6*
Barbara	10	17				*												
[a]*Note*: number of lessons completed including repeated lessons [b]*Note:* the asterisk indicates a repeated lesson; the number of asterisks indicates how many times																		

4.3.1 Ratings

In terms of comprehensibility, or the "listeners' estimation of difficulty in understanding an utterance" (Munro, Derwing, & Morton, 2006, p. 112), Barbara showed a difference in ratings of 1, going from a score of 5.3 in the pre-test to a score of 4.3 in the post-test. However, the difference between the pre-test (M=5.27, SD=2.58) and the post-test (M=4.33, SD=1.84) was not statistically significant; $t(17) = .97$, $p > 0.05$. In terms of accentedness, or "the degree to which the pronunciation of an utterance sounds different from an expected production pattern" (Munro, Derwing, & Morton, 2006, p. 112), the difference in Barbara's pre-test score (M=7.11, SD=1.87) and post-test score (M=7.16, SD=1.94) was also not statistically significant, $t(17) = -.07$, $p > 0.05$.

Overall, Barbara did not show improvement in the macro pronunciation features and almost never repeated lessons. To corroborate this lack of improvement, the following acoustic analysis also reveals that Barbara did not produce gemination contrasts and improved in voice onset time only in the case of the voiceless stop /p/ in sentences.

4.3.2 Acoustic Analysis

4.3.2.1 Gemination in words

Figure 33 shows the data from Barbara's contrast between singleton /t/ and geminate /tt/ in words. It appears as if Barbara has an opposite contrast to what expected, with long singletons and much shorter geminates. Indeed, as compared to the native speaker, who has a mean singleton measurement of 119 ms and mean geminate measurement of 213 ms, Barbara has much longer averaging 96 ms much shorter geminates averaging 181 ms.
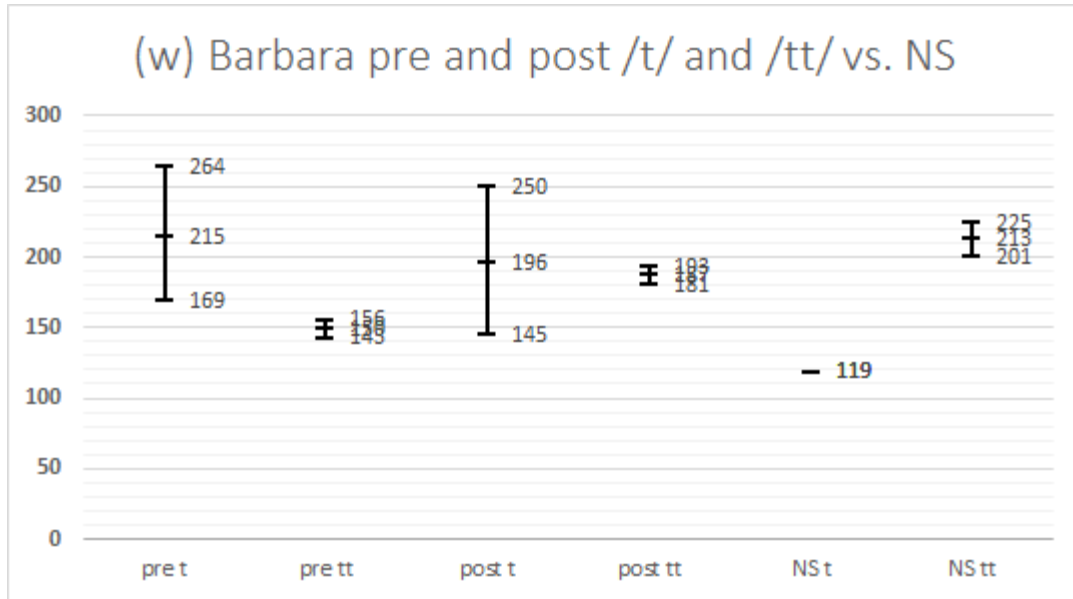
Figure 33. Barbara's pre and post /t/ and /tt/ in words (w) vs. native speaker (NS) data (msec)

In the case of the singleton /n/ and geminate /nn/ contrast, Barbara's data, shown in Figure 34, is incomplete. Both tokens of the geminate /nn/ in the pre-test could not be used as they were pronounced as [ñ]. However, it is still clear that the singletons in the post-test are longer that the geminates and that therefore Barbara is not producing the appropriate contrast, which is illustrated by the mean and range for the native speaker.
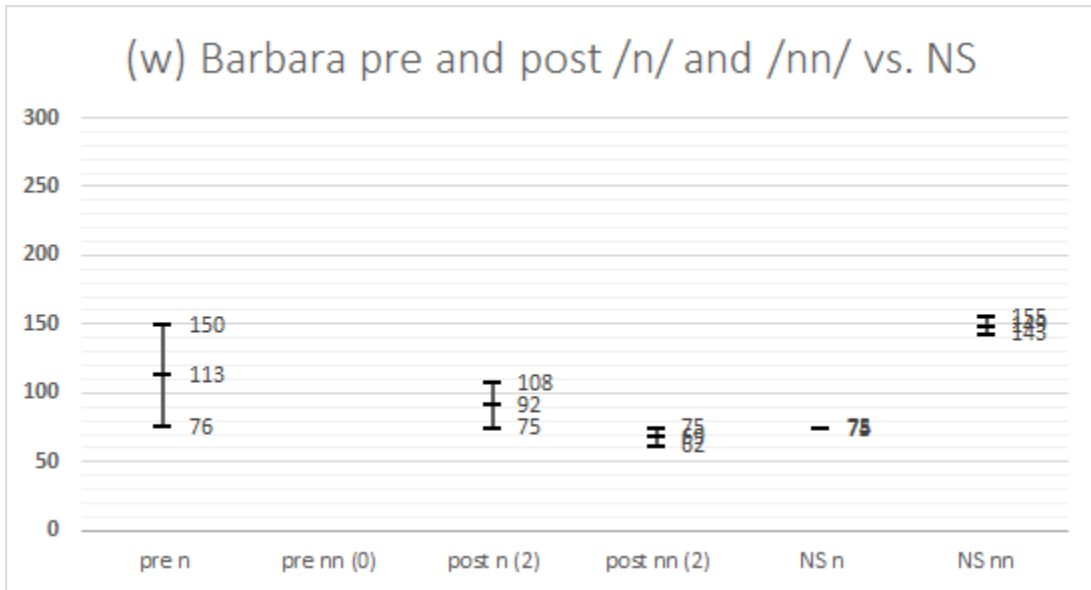
Figure 34. Barbara's pre and post /n/ and /nn/ in words (w) vs. native speaker (NS) data (msec).

The data on Barbara's singleton /l/ and geminate /ll/ is shown in Figure 35. It seems clear that Barbara did not produce a contrast in either test. Also, the duration of Barbara's geminates is much shorter than that of the native speaker.
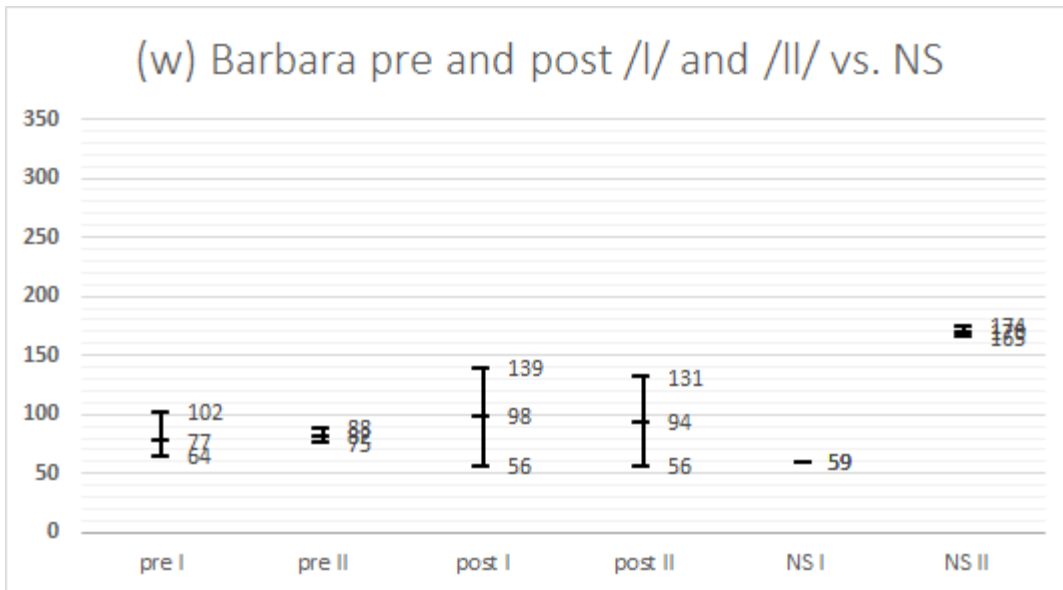
Figure 35. Barbara's pre and post /l/ and /ll/ in words (w) vs. native speaker (NS) data (msec).

Since much of the literature also looks at ratios when analyzing the contrast between singletons and geminates, Table 10 is an overview of ratios in words for Barbara. Also included are the ratios of the native speaker. The ratios corroborate the data shown in the plots above. As the ratios gravitate around 1.0, and are sometimes lower, there is no evidence of an appropriate contrast. Further, the contrast that does exist is often opposite of what expected, with singletons longer than geminates.

Table 10

Barbara's geminate contrast ratios vs. native speaker ratios in words

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	Post		Pre	Post		Pre	Post	
Barbara	0.70	0.96	1.79	n/a	0.75	2.00	1.06	0.87	2.87

4.3.2.2 Gemination in sentences

In sentences, Barbara's contrast between singleton /t/ and geminate /tt/, shown in Figure 36, does not seem to be defined. However, in the post-test, Barbara has narrowed down the range of the singletons very tightly, which indicates precision, and the mean and highest value for the geminates are longer than for the singletons, suggesting that she is attempting some contrast.
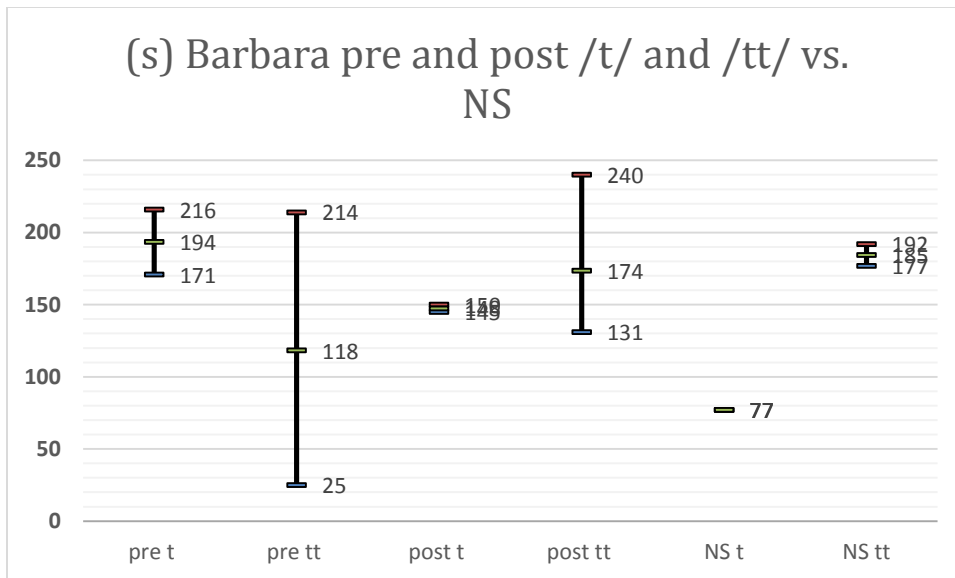
Figure 36. Barbara's pre and post /t/ and /tt/ in sentences (s) vs. native speaker (NS) data (msec).

In Barbara's data for singleton /n/ and geminate /nn/, shown in Figure 37, the geminates overlap the singleton, although they do tend to be longer at their highest value. In the presence of such overlap, in which the full range of the singleton is within the range of the geminate, it is clear that no clear contrast is being produced.
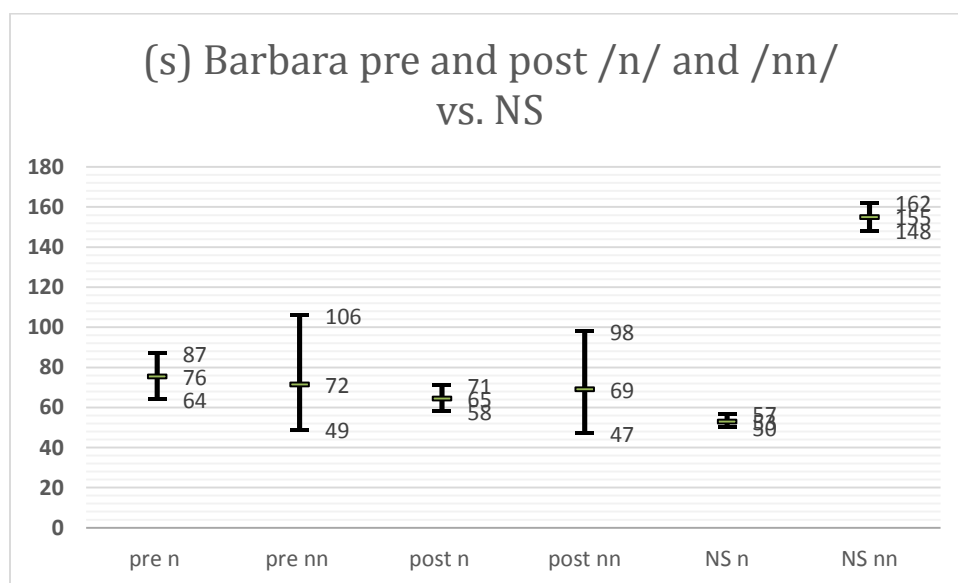
Figure 37. Barbara's pre and post /n/ and /nn/ in sentences (s) vs. native speaker (NS) data (msec).

73

In Figure 38, Barbara's data for the singleton /l/ and geminate /ll/, it is clear that although the singleton ranges are quite tight, the geminates once again overlap them. Further, the means for the geminates are consistently lower that the lowest values of the singletons: for example, in the post-test, the mean of the geminates is 65 ms, and the lowest value of the singleton is 72 ms. Overall, no contrast is shown.
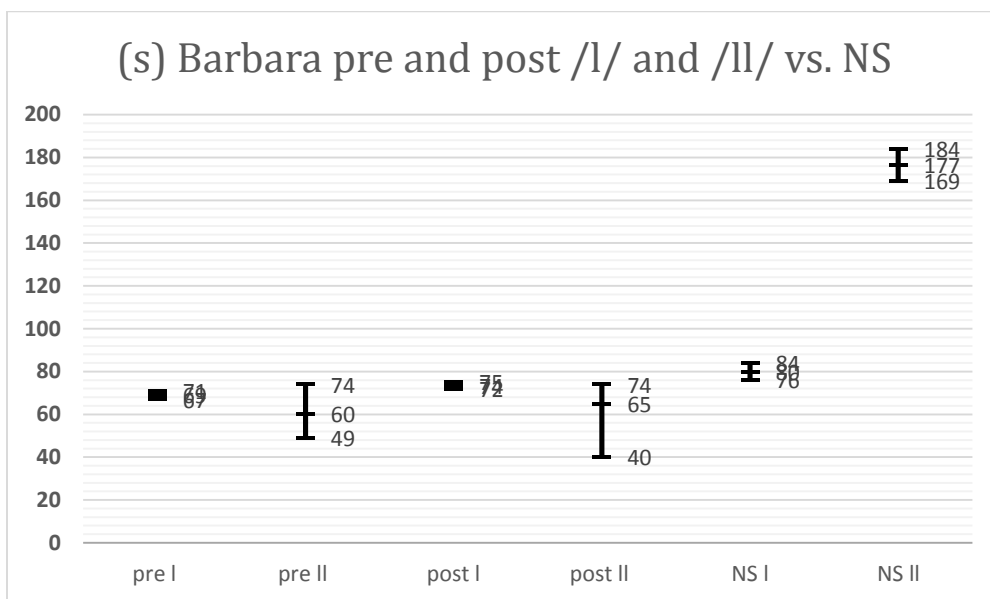
Figure 38. Barbara's pre and post /l/ and /ll/ in sentences (s) vs. native speaker (NS) data (msec).

The ratios for sentences are shown in Table 11. They corroborate the data contained in the graphs, as the ratios shows that the singletons and geminates tend to be equal (ratio of 1). The only exception is the /tt/ to /t/ ratio of 1.17, in which, as shown in the graph, the geminates high values are much longer in duration that the singletons. However, none of the ratios are significantly close to the native speaker ones.

Table 11

Barbara's geminate contrast ratios vs. native speaker ratios in sentences

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	**Post**		**Pre**	**Post**		**Pre**	**Post**	
Barbara			2.40			2.92			2.21
	0.60	1.17		0.94	1.06		0.86	0.87	

4.3.2.3 Gemination in passage

The data for the tt to t contrast in the passage for Barbara are shown in Figure 39. As for the words and sentences, the data indicate that she did not produce any contrast.
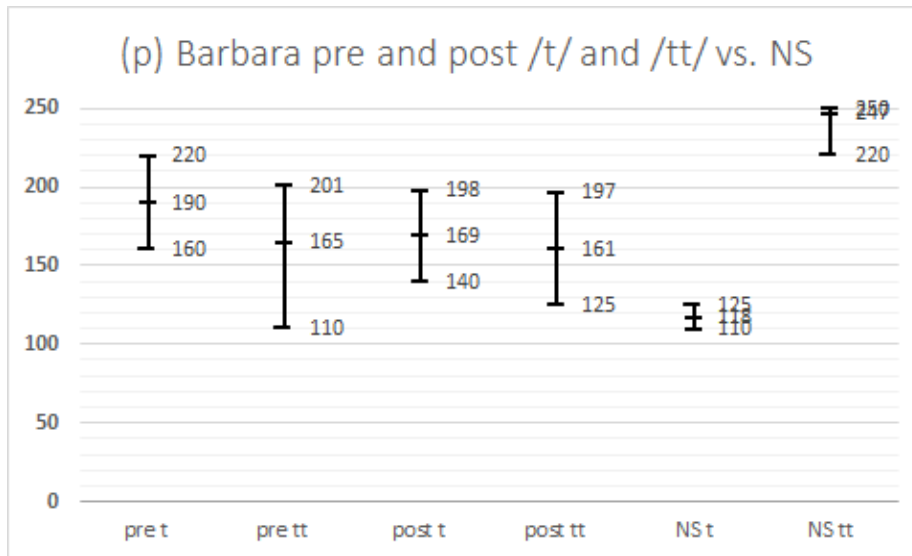
Figure 39. Barbara's pre and post /t/ and /tt/ in passages (p) vs. native speaker (NS) data (msec).

As for the nn to n contrast, the data shown in Figure 40 indicate that although the high values in the geminates indicate the possibility of a contrast in some cases, overall there is no meaningful contrast with significant overlap.

75
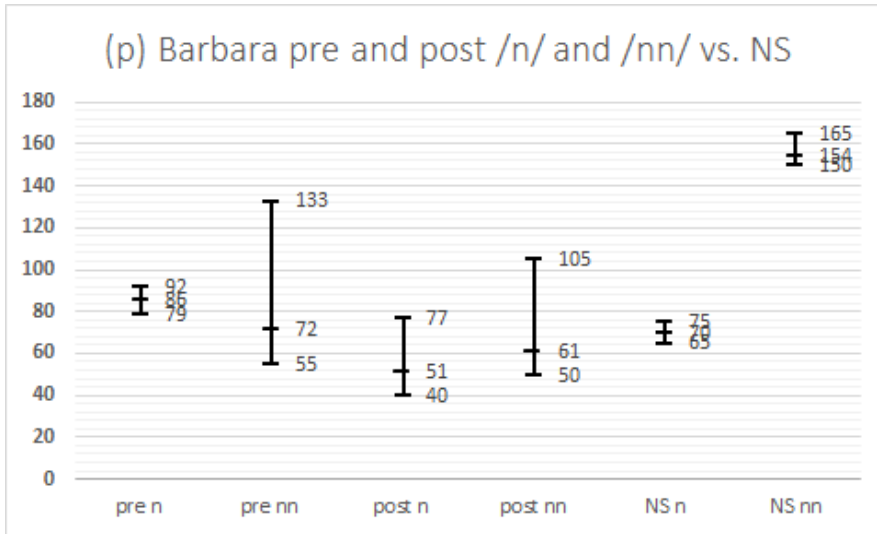
Figure 40. Barbara's pre and post /n/ and /nn/ in passages (p) vs. native speaker (NS) data (msec).

Lastly, there seems to be no contrast in the production of /l/ and /ll/, either, as shown in Figure 41. Although the singletons are close to the native speaker ranges, the geminates are not long enough in either test for there to be a contrast.
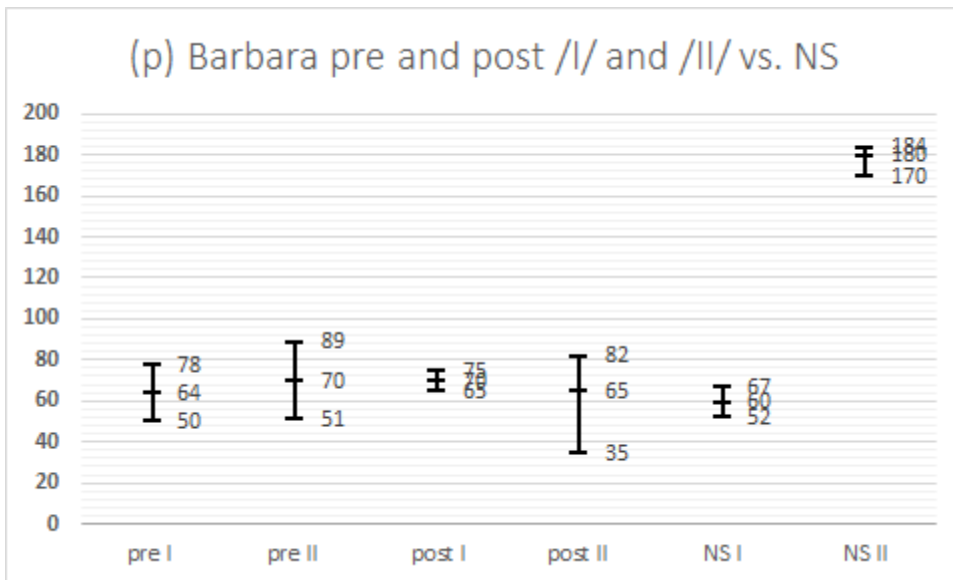
Figure 41. Barbara's pre and post /l/ and /ll/ in passages (p) vs. native speaker (NS) data (msec).

The ratios for sentences are shown in Table 12. They seem to corroborate the data contained in the graphs, as the ratios show that the singletons and geminates tend to be equal (ratio of 1). Further, none of the ratios are significantly close to the native speaker ones.

Table 12

Barbara's geminate contrast ratios vs. native speaker ratios in passage

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	Post		Pre	Post		Pre	Post	
Barbara			2.10			2.05			2.85
	0.86	0.95		0.83	1.19		1.09	0.92	

4.3.2.4 Voice Onset Time in words

Figure 42 shows the measurements of voice onset time in /p, t, k/ for Barbara. Although it is clear that the range tightened in the post-test, the duration of the voice onset time of /p/ is more than three times longer than that of the native speaker. For /t/, although Barbara seems to have shortened her voice onset time for /p/ by 14 ms on average, her post-test range is still considerably longer than that of the native speaker. Lastly, for /k/, it is clear that no improvement was made, and although the post-test lowest value approaches the native speaker pronunciation, the overall range is quite high with a mean of 57 ms. Overall, although producing the stops as less aspirated than what is typically found in English, it is still nowhere close to the native speaker ranges.
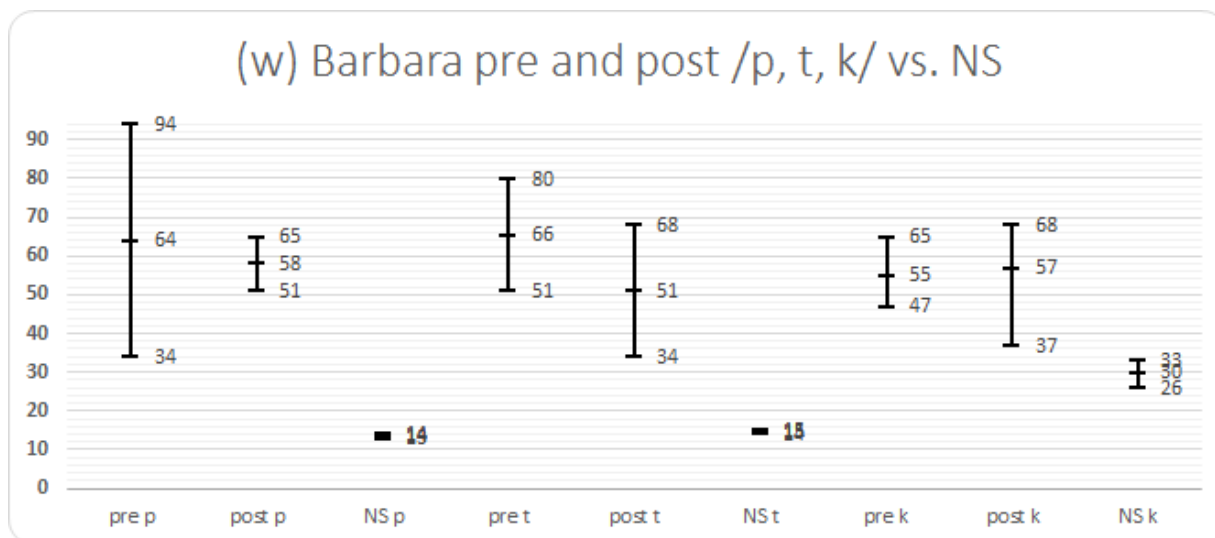
Figure 42. Barbara's pre and post /p, t, k/ in words (w) vs. native speaker (NS) data (msec).

4.3.2.5 Voice Onset Time in sentences

The data for Barbara's voice onset time in sentences in shown in Figure 43. Barbara tightened the range of the post-test /p/, similarly to what happened in words. In this case, though, she was also able to reduce the duration of the voice onset time. However, the mean (36 ms) post-test time is still twice as long as that of the native speaker (14 ms). For /t/, it is clear that Barbara actually widened the range for her voice onset time of /t/ and lengthened the duration rather than shortening it, although in both cases she is producing the features as unaspirated. or /k/, the data indicates no improvement from the pre- to the post-test; although some of the lowest values indicated that she produced some of the tokens as unaspirated, the ranges are not tight enough to indicate control or improvement.
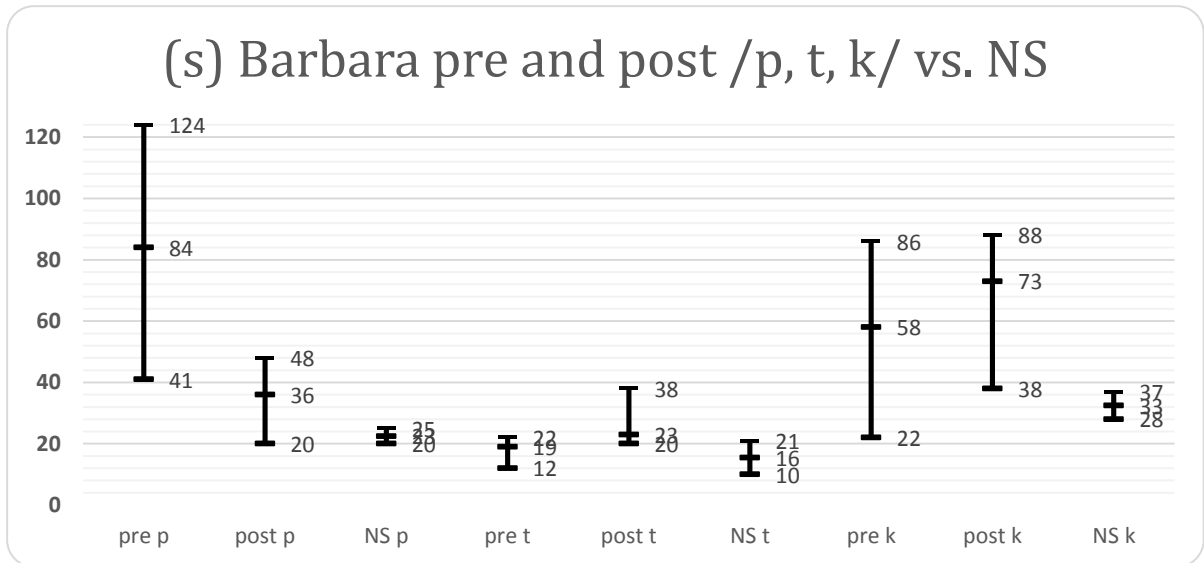
Figure 43. Barbara's pre and post /p, t, k/ in sentences (s) vs. native speaker (NS) data (msec).

4.3.2.6 Voice Onset Time in passage

The passage data, shown in Figure 44, are similar to the data from the words and indicate no improvement in the production of voice onset time. Although some of the values, particularly for /t/ and /k/, indicated a plain production, Barbara does not seem to tighten the ranges in the post-test and at times still produced the stops as aspirated (up to 76 ms).
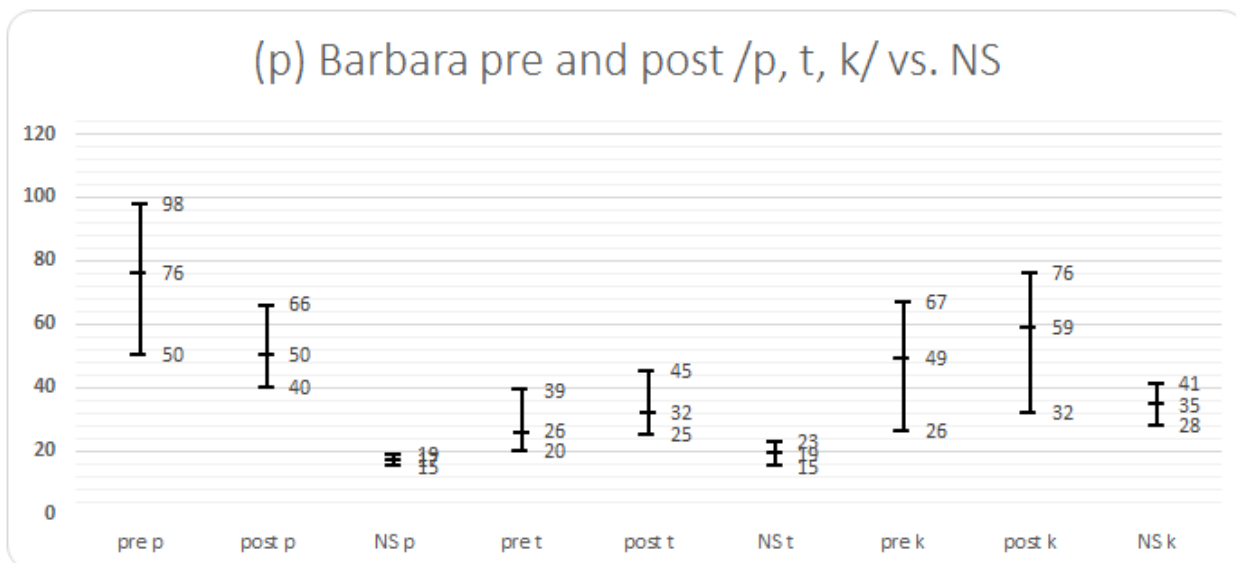
Figure 44. Barbara's pre and post /p, t, k/ in passage (p) vs. native speaker (NS) data (msec).

4.3.3 Summary (Barbara)

In the native speaker ratings, Barbara did not improve in either comprehensibility or accentedness. This is supported also by the lack of improvement in the micro features: Barbara did not grasp the contrastive gemination and did not reduce aspiration in word-initial stops, although in some cases she did produce the stops as plain. The inability to improve in the production of geminates could be due to Barbara's background in French, in which orthographic gemination rarely translates to pronunciation. As for voice onset time, the absence of improvement cannot be traced back to Barbara's background, as in French voiceless stops are unaspirated, and could be explained in part by the little practice.

4.4 Luciana

Luciana's language background is in Portuguese, which she studies for ten years, and Arabic, which she started studying recently in college. The literature on these languages indicates that Arabic has contrastive gemination (Davis & Ragheb, 2014), while Portuguese does not. Further, Portuguese voiceless stops are produced "with gradience from unaspirated (< 35 ms) to

aspirated (60 ms)" (Alves, 2011, p. 5). What seems to affect the presence or absence of aspiration the most in Portuguese is the place of articulation, with bilabials having the shortest VOT and velars having the longest (Alves, 2011). In Arabic, voiceless stops are aspirated before stressed vowels (Kopczynski & Meliani, 1993).

Luciana entered *Duolingo* seven times, accessing the level from one to four times, and completing a total of 24 lessons. Table 13 is an overview of her *Duolingo* report, inclusive of which lessons were repeated and how many times.

Table 13

Luciana's Duolingo Report: Overall days of practice, number of lessons completed, and number of times each lesson was repeated

	Days	N. lessons[a]	Repeated Lessons[b]														
			Basics			Basics 2					Phrases		Food					
Levels			*1*	*2*	*3*	*1*	*2*	*3*	*4*	*5*	*1*	*2*	*1*	*2*	*3*	*4*	*5*	*6*
Luciana	7	24	**	***	*	*	*	*			*	*						
[a] *Note*: number of lessons completed including repeated lessons [b]*Note:* the asterisk indicates a repeated lesson; the number of asterisks indicates how many times																		

4.4.1 Ratings

Luciana's ratings for comprehensibility went from 3.8 in the pre-test to 1.6 in the post-test, resulting in a difference of 2.2. The *t*-test results indicate that there was a statistically highly significant difference in the scores from pre-test (M= 3.77, SD= 2.18) and the scores in the post-test (M=1.61, SD=.97); t(17) = 5.23, p < 0.001 * (two-tailed), indicating an improvement. The eta squared statistic (.54) indicates a large effect size. Similarly, the ratings for accentedness show a statistically significant difference from the pre-test scores (M=6.61, SD=1.88) to the post-test scores (M=4.37, SD=2.32); t(17) = 3.63, p < 0.01**. Here too, the statistical significance indicates actual improvement and the eta squared statistic (.47) indicates a large effect size

Overall, Luciana improved significantly in both comprehensibility and accentedness and is one of the two participants who practiced the most. However, when looking at her acoustic analysis, describe in the following section, there is only evidence of improvement in gemination and not much at all in voice onset time. This indicates that other factors have influenced the native speaker ratings, and that, as seen in Emma's case, improvement in certain micro features does not necessarily translate to improvement in intelligibility.

4.4.2 Acoustic Analysis

4.4.2.1 Gemination in words

Figure 45 shows Luciana's data for the singleton /t/ and geminate /tt/. It indicates that Luciana tightened the ranges for both the singleton /t/ and the geminate /tt/ and increased the durational contrast, reaching an almost native like production.
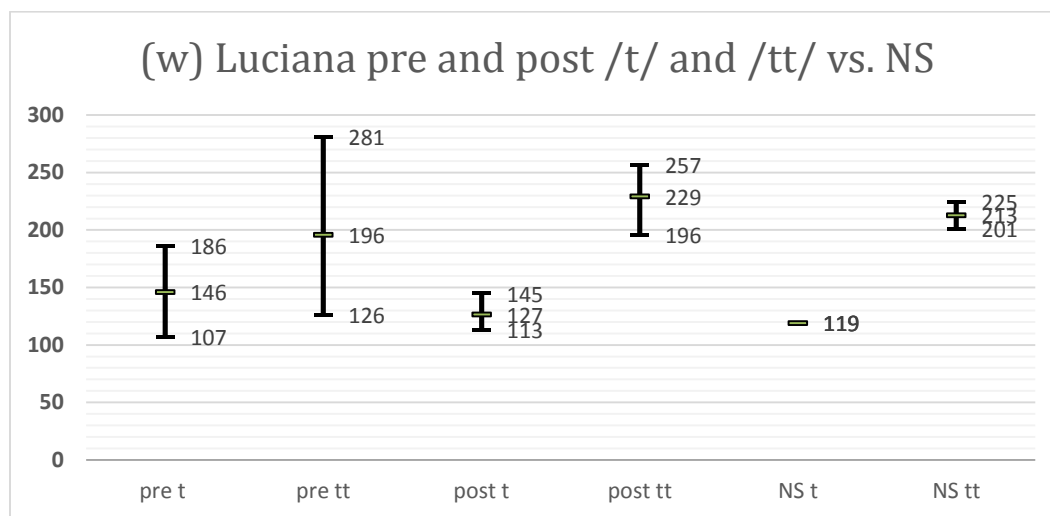
Figure 45. Luciana's pre and post /t/ and /tt/ in words vs. native speaker (NS) data (msec).

Luciana's data for /n/ in Figure 46 shows that she produced a significant contrast between singleton /n/ and geminate /nn/ in both tests, although her measurements for the geminate are much longer in duration to those of the native speaker.
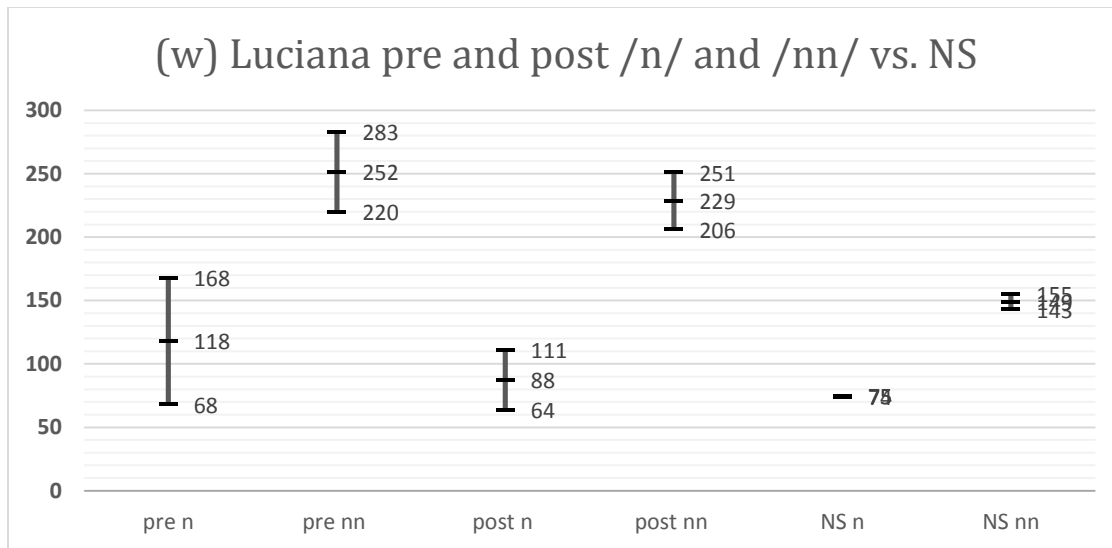
Figure 46. Luciana's pre and post /n/ and /nn/ vs. native speaker (NS) data (msec).

Likewise, Luciana's data for /l/ and /ll/, shown in Figure 47, shows that she produced a significant contrast between the singleton and the geminate. Further, she improved that contrast in the post-test by tightening the range and adjusting the duration of the geminates in particular, going from a mean of 267 ms to one of 195 ms. However, the duration of both the singleton and the geminate are still quite a bit longer than those of the native speaker.
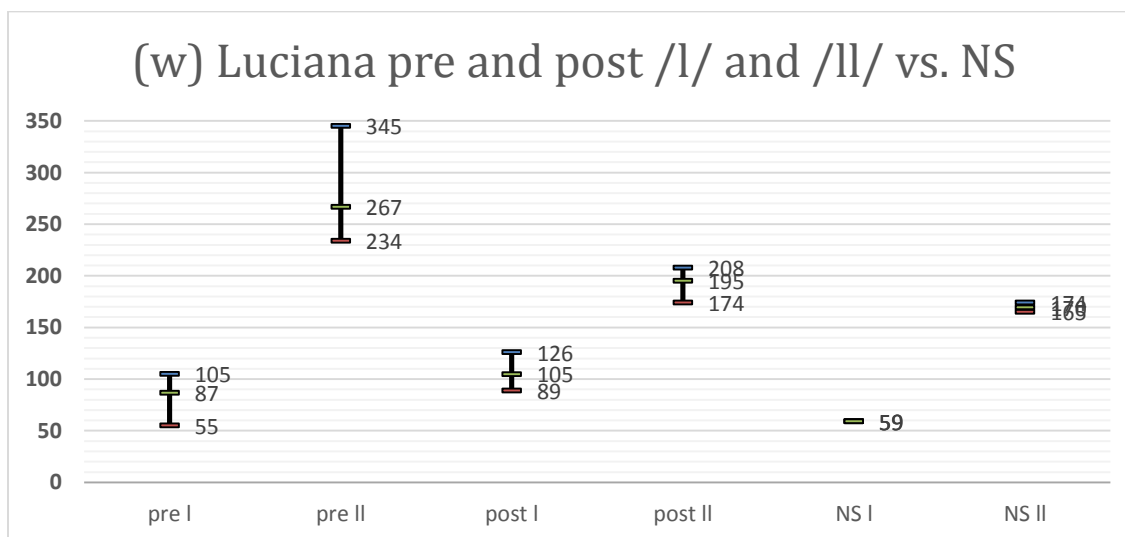
Figure 47. Luciana's pre and post /l/ and /ll/ in words vs. native speaker (NS) data (msec).

The ratios for the above three contrasts are shown in Table 14. The ratios for tt to t has improved from the pre-test to the post-test, almost matching that of the native speaker. However, the ratios for the other two contrast can be deceiving: although they seem to be less similar to the native speaker, this does not mean the production is not improved. Indeed, the changes in the ratios are caused by the tightening of the ranges, which on one hand might make the ratio shift further from the native speaker, but on the other hand highlights a better control over the feature.

Table 14

Luciana's geminate contrast ratios vs. native speaker ratios in words

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	Post		Pre	Post		Pre	Post	
Luciana	1.34	1.81	1.79	2.14	2.67	2.00	3.07	1.86	2.87

4.4.2.2 Gemination in sentences

In Figure 48, Luciana's data for the tt to t contrast shows no improvement. Although the range for the geminate /tt/ tightened, it also shortened in overall duration, making the contrast between singletons and geminates less meaningful.
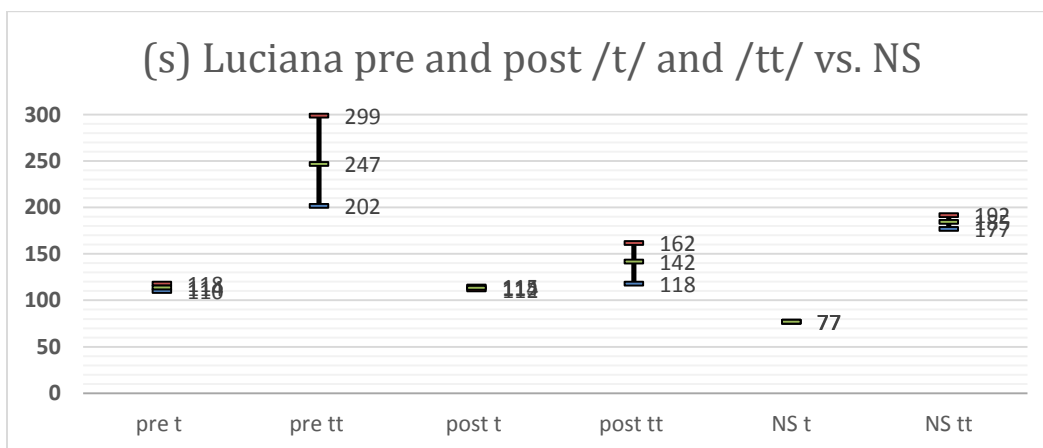
Figure 48. Luciana's pre and post /t/ and /tt/ in sentences (s) vs. native speaker (NS) data (msec).

As for the contrast of /n/ and /nn/, shown in Figure 49, Luciana seems to have adjusted the length of her geminates, although the overlap in the post-test indicates that at times a contrast might not be present.
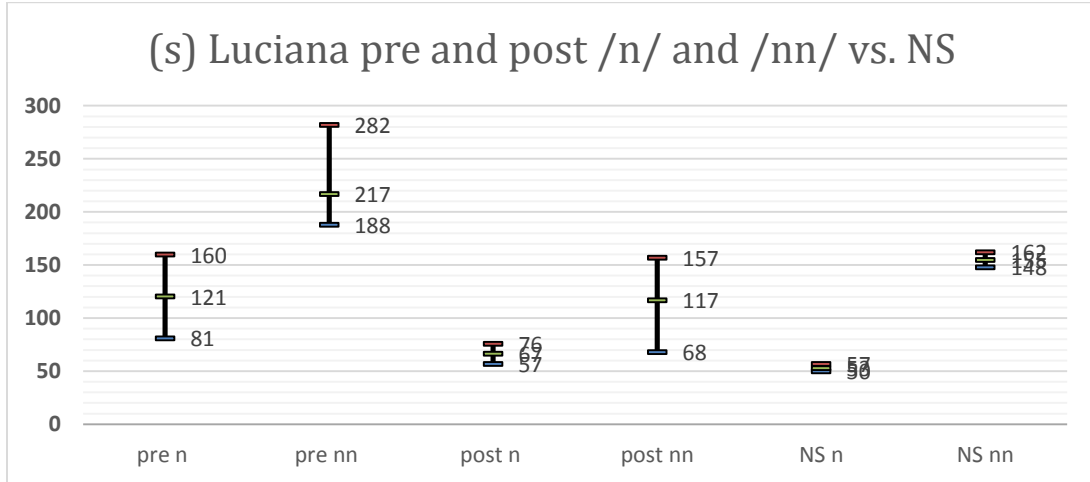
Figure 49. Luciana's pre and post /n/ and /nn/ in sentences (s) vs. native speaker (NS) data (msec).

As for the previous two contrasts, the data on the one between /l/ and /ll/, shown in Figure 50, indicates that Luciana shortened her geminates. While this has consistently made her geminates shorter than those of the native speaker, it does also indicate that she has tightened her ranges, or, in other words, has better control over the production.
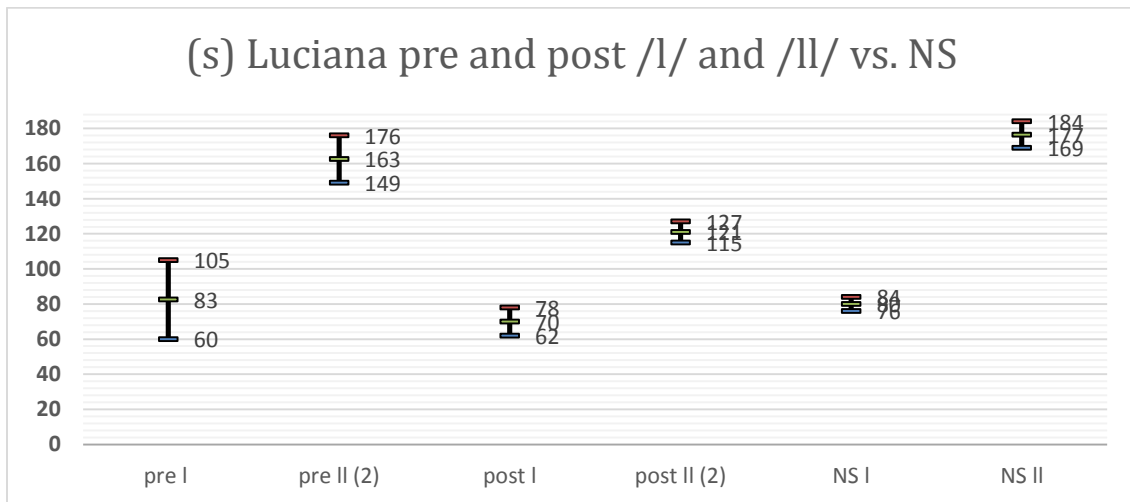
Figure 50. Luciana's pre and post /l/ and /ll/ in sentences (s) vs. native speaker (NS) data (msec).

The ratios for the above three contrasts are shown in Table 15. As in the case of words, the ratios can be deceiving. The values do tend to be less similar to the native speaker from the pre- to the post-test, but they hide the fact that Luciana has gained better control over the features.

Table 15

Luciana's geminate contrast ratios vs. native speaker ratios in sentences

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	Post		Pre	Post		Pre	Post	
Luciana	2.16	1.24	2.40	1.79	1.74	2.92	1.96	1.72	2.21

4.4.2.3 Gemination in passage

Figure 51 shows the data from Luciana's passage in terms of the contrast between singleton /t/ and geminate /tt/. No improvement is shown as she was already making a clear distinction in the pre-test and approximating the durations of the native speaker.
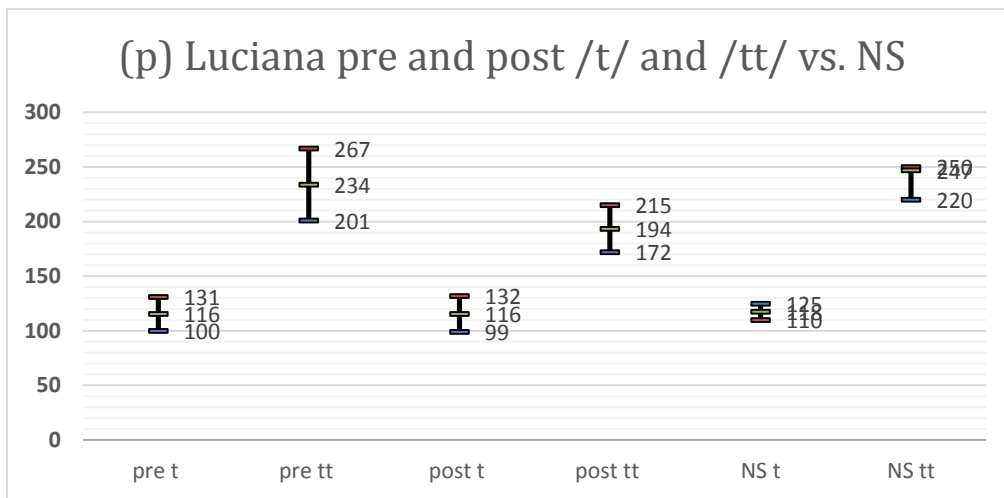
Figure 51. Luciana's pre and post /t/ and /tt/ in passage (p) vs. native speaker (NS) data (msec).

For the contrast between /n/ and /nn/ (shown in Figure 52) and between /l/ and /ll/ (shown in Figure 53), instead, a clear improvement can be seen. Although the contrast seem to be already

present, the ranges adjust in the post-test and even closely approach those of the native speaker.

This improvement echoes that in the words and sentences.

Figure 52. Luciana's pre and post /n/ and /nn/ in passage (p) vs. native speaker (NS) data (msec).

Figure 53. Luciana's pre and post /l/ and /ll/ in passage (p) vs. native speaker (NS) data (msec).

The ratios, shown in Table 16, support the data shown in the graphs, with a clear improvement in ratio of nn to n, which went from 1.72 to 2.28, and in the ratio of ll to l, which went from 1.43 to 2.72.

Table 16

Luciana's geminate contrast ratios vs. native speaker ratios in passage

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	**Post**		**Pre**	**Post**		**Pre**	**Post**	
Luciana	2.01	1.67	2.10	1.72	2.28	2.05	1.43	2.72	2.85

4.4.2.4 Voice Onset Time in words

Luciana's data for /p, t, k/, shown in Figure 54, indicates that she caught on to the idea that initial voiceless stops are not aspirated only in certain cases, such as for /k/. However, her post-test measurements are still more than double the duration of those of the native speaker.
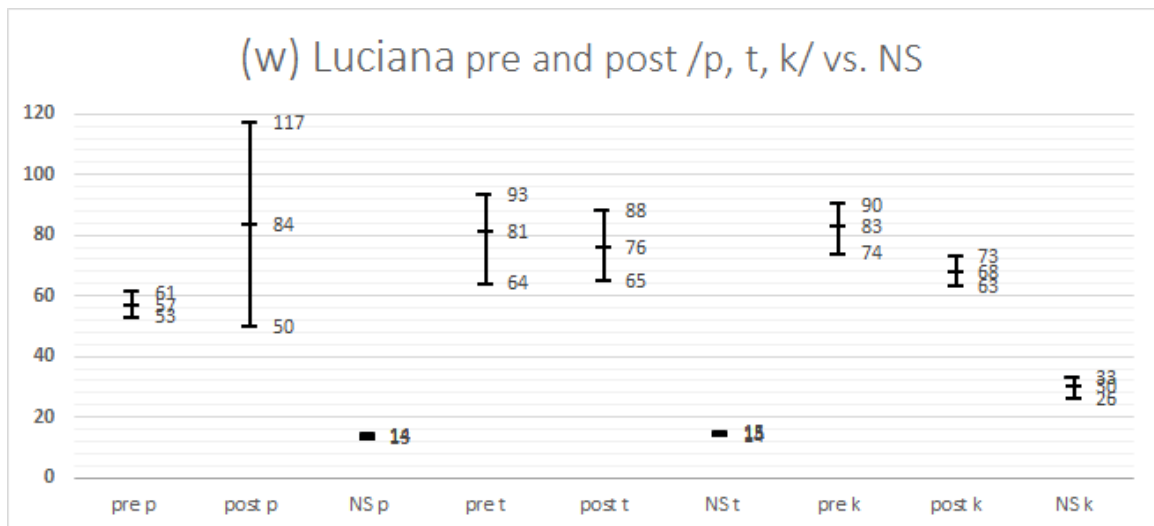
Figure 54. Luciana's pre and post /p, t, k/ in words (w) vs. native speaker (NS) data (msec).

4.4.2.5 Voice Onset Time in sentences

The data from Luciana's sentences, in Figure 55, shows that she produced both /p/ and /k/ as unaspirated (compared to English), and that improved in the control of both the voice onset times of /p/ and /t/, although not /k/.

88
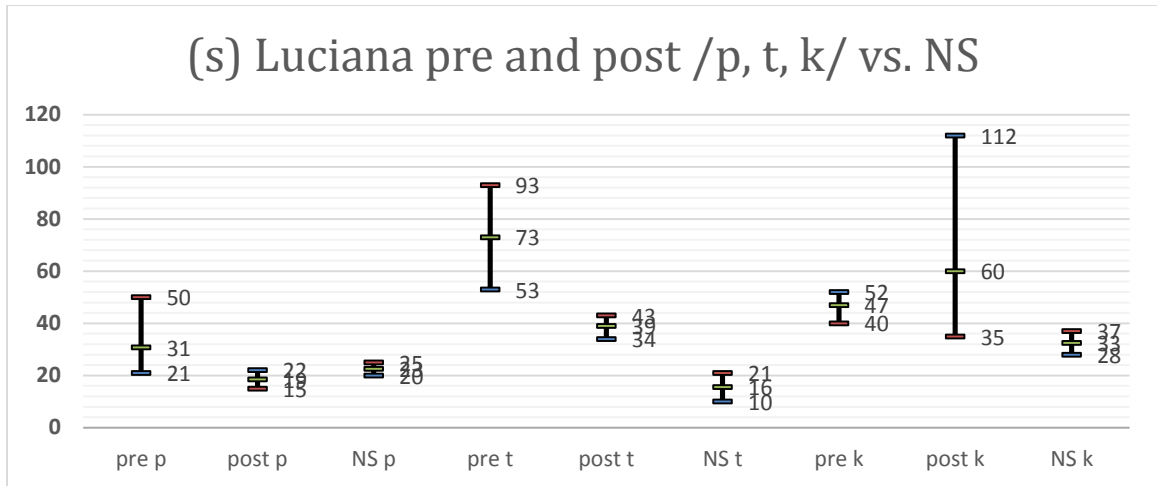
Figure 55. Luciana's pre and post /p, t, k/ in sentences (s) vs. native speaker (NS) data (msec).

4.4.2.6 Voice Onset Time in passage

Figure 56 shows the data for the voice onset time measurements of /p, t, k/ for Luciana in the passage. Here, too, she showed some improvement in /p/, but not in /t/ or /k/.
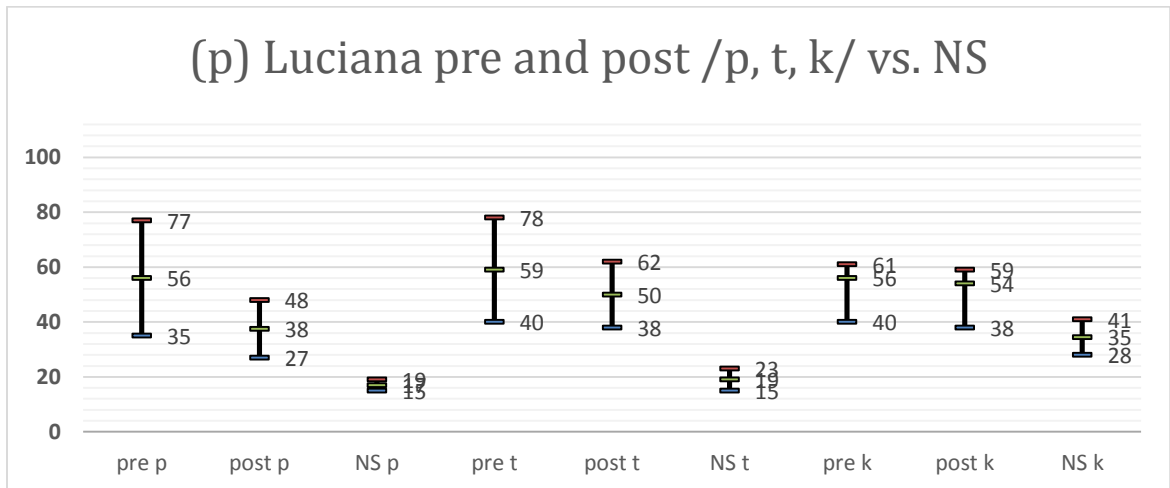
Figure 56. Luciana's pre and post /p, t, k/ in passage (p) vs. native speaker (NS) data (msec).

4.4.3 Summary (Luciana)

Overall, Luciana seems to have consistently improved in geminate contrasts across contexts – words, sentences, and passages. Luciana's pre-existing contrast and their improvements could be due to her experience with Arabic, which has consonant contrasts, and to

89

practice. Only the contrast between /t/ and /tt/ in sentences and passages seems to have not followed this pattern of improvement. This could be due to the fact that in English the stop consonant /t/ becomes a flap between two vowels, where the first vowel is stressed and the second is not. This is the environment where the geminate /tt/ occurred in the words of the test, such as in the word for cat, *gatto* ['gat:o]. In terms of voice onset time, Luciana showed evidence of improvement only in sentences, and only for /p/ and for /t/. Although it does not account for the reason why Luciana's improvement in VOT is relegated to the sentences, the reasons behind her lack of improvement in /k/ could also be due to the influence of her knowledge of Arabic. Arabic, in fact, although sharing the characteristic of having a two way distinction for voice onset time with Italian, tends to have longer ranges for /k/, from 40 to 60ms (Mitleb, 2009; Yeni-Komshian et al., 1977), as compared to other languages that share the same feature.

4.5   Sabrina

Sabrina's language background is mostly in Spanish, which she studied for many years in elementary, middle, and high school and through two semesters in college. Standard Spanish is often described as a language that lacks geminate consonants, with the only exception of the tap/trill rhotic contrast, which is found only in word-internal intervocalic position, as in caro 'expensive' vs. carro 'cart' (Scarpace, 2014), although many other variations of Spanish maintain consonant gemination, such as western Cuban Spanish (Rivas, 2000). However, similarly to Italian, in Spanish voiceless stops are always unaspirated (Lisker & Abramson, 1973).

Her practice data, shown in Table 17, indicates that she used *Duolingo* six times, accessing the levels from one to two times, and completing a total of 18 lessons. An overview of the data collected in the report is shown Table 17 along with an indication of which lessons were repeated and how many times. As always, the "Test" function was not counted as a repetition of the lesson, as it does not repeat the entire lesson, but parts of it, and it does not contain any of the interactive functions of the program.

Table 17

Sabrina's Duolingo Report: Overall days of practice, number of lessons completed, and number of times each lesson was repeated

	Days	**N. lessons[a]**	**Repeated Lessons[b]**															
			Basics			Basics 2					Phrases		Food					
Levels			*1*	*2*	*3*	*1*	*2*	*3*	*4*	*5*	*1*	*2*	*1*	*2*	*3*	*4*	*5*	*6*
Sabrina	3	18	*															
[a] *Note*: number of lessons completed including repeated lessons [b]*Note*: the asterisk indicates a repeated lesson; the number of asterisks indicates how many times																		

4.5.1 Ratings

For comprehensibility, Sabrina showed a difference in rating of 1.4. However, the difference from the pre-test (M=4.22, SD=2.83) to the post-test (M=2.83, SD=2.43) is not statistically significant; $t(17) = 1.42$, $p > 0.05$. In terms of accentedness, too, Sabrina did not show improvement as the difference between the pre-test (M=6.16, SD=1.85) and the post-test (M=5.77, SD=2.23) was not statistically significant: $t(17)=.61$, $p > 0.05$.

Overall, Sabrina did not improve on either scale, and was one of the three participants who practiced less. Her acoustic analysis, described in detailed in the following section, corroborates these findings as it indicates that there was very minimal improvement in either the geminate contrast or voice onset time (only in certain features and in certain contexts).

4.5.2 Acoustic Analysis

4.5.2.1 Gemination in words

Figures 57, 58, and 59 show the data on the /t/ to /tt/, /n/ to /nn/, and /l/ to /ll/ contrasts respectively. It is clear that no apparent improvement was made in the production of these contrast – often with singletons being equal in duration to the geminates.
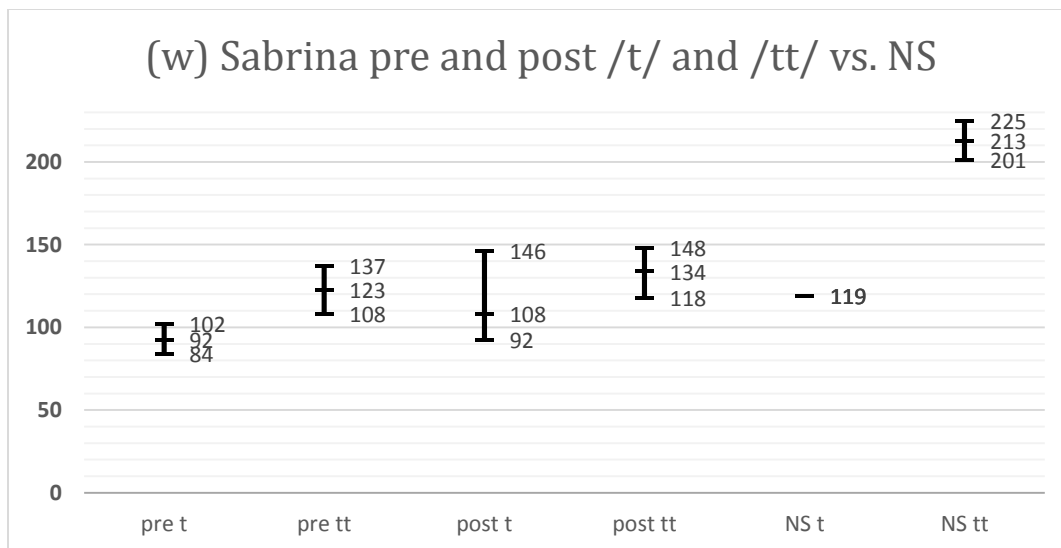
Figure 57. Sabrina's pre and post /t/ and /tt/ in words (w) vs. native speaker (NS) data (msec).
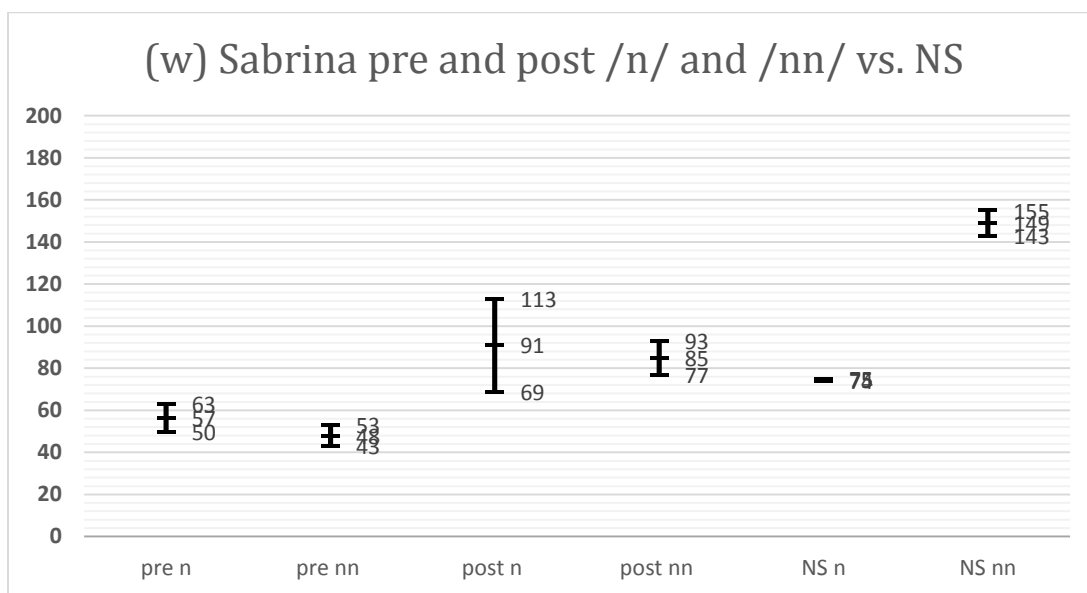
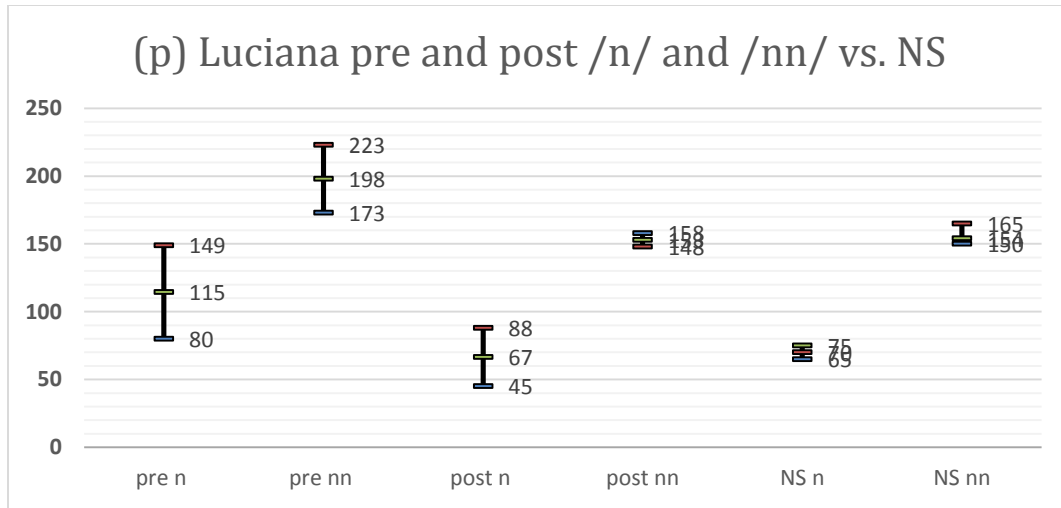*Figure 58.* Sabrina's pre and post /n/ and /nn/ in words (w) vs. native speaker (NS) data (msec).
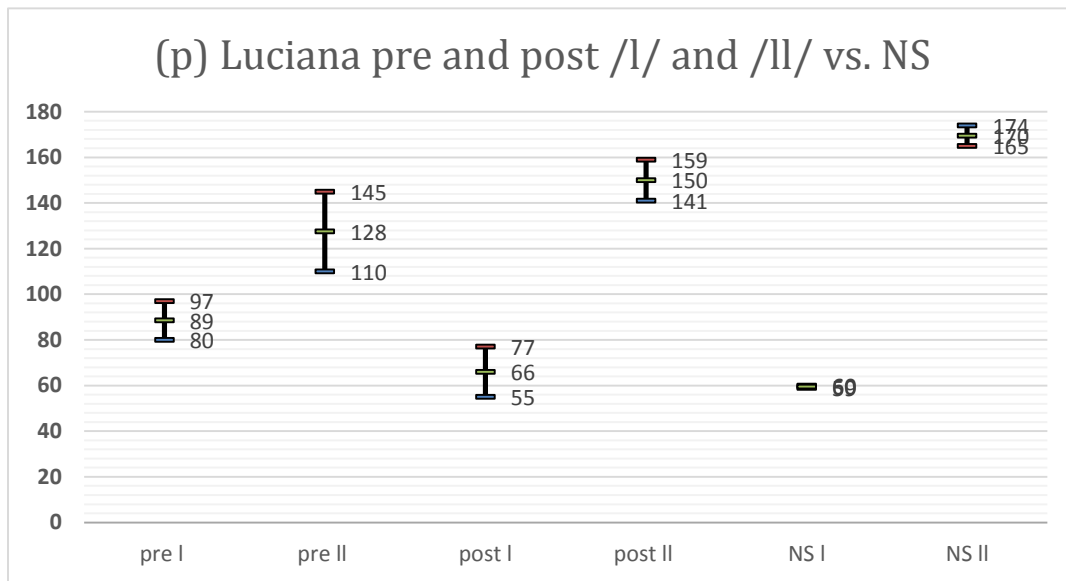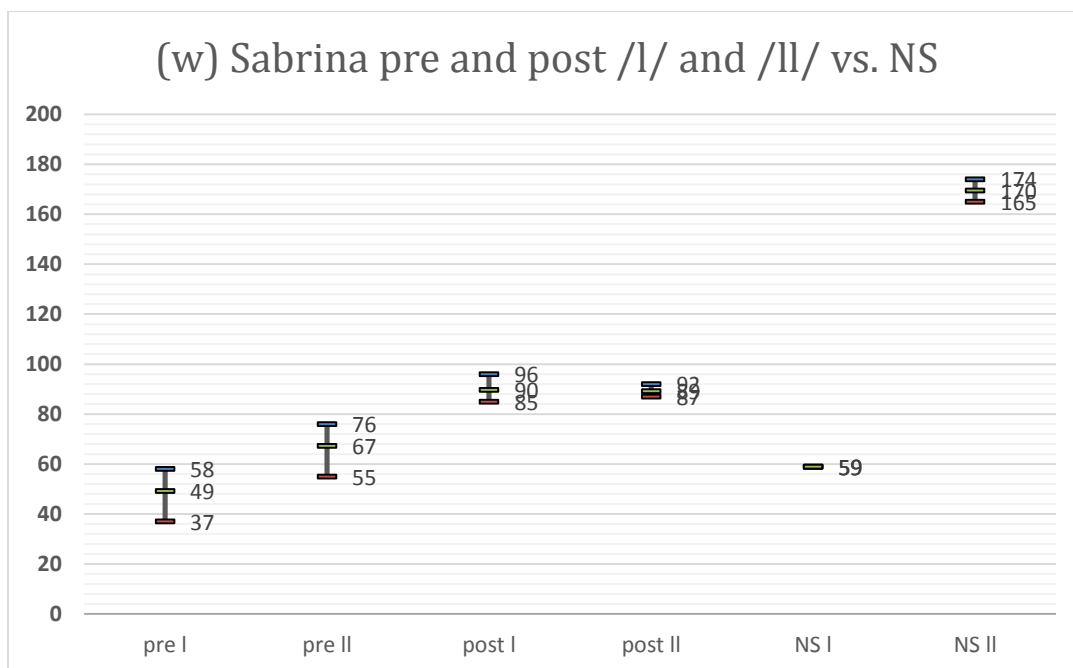
Figure 59. Sabrina's pre and post /l/ and /ll/ in words (w) vs. native speaker (NS) data (msec).

The ratios, shown in Table 18, further support the results shown in the graphs, with value often being lower in the post-test than in the pre-test indicating less contrast between the singletons and the geminates.

Table 18

Sabrina's geminate contrast ratios vs. native speaker ratios in words

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	**Post**		**Pre**	**Post**		**Pre**	**Post**	
Sabrina	1.33	1.24	1.79	0.84	0.93	2.00	1.36	0.98	2.87

4.5.2.2 Gemination in sentences

Figures 60, 61, and 62 show the data for the three geminate contrast for Sabrina in sentences. While the /t/ to /tt/ and the /l/ to /ll/ contrasts show no improvement, the /n/ to /nn/

93

shows the presence of a contrast in the post-test. Further, the measurements in the pre-test were

taken over two tokens only as Sabrina pronounced two of the /nn/ tokens as [ñ].
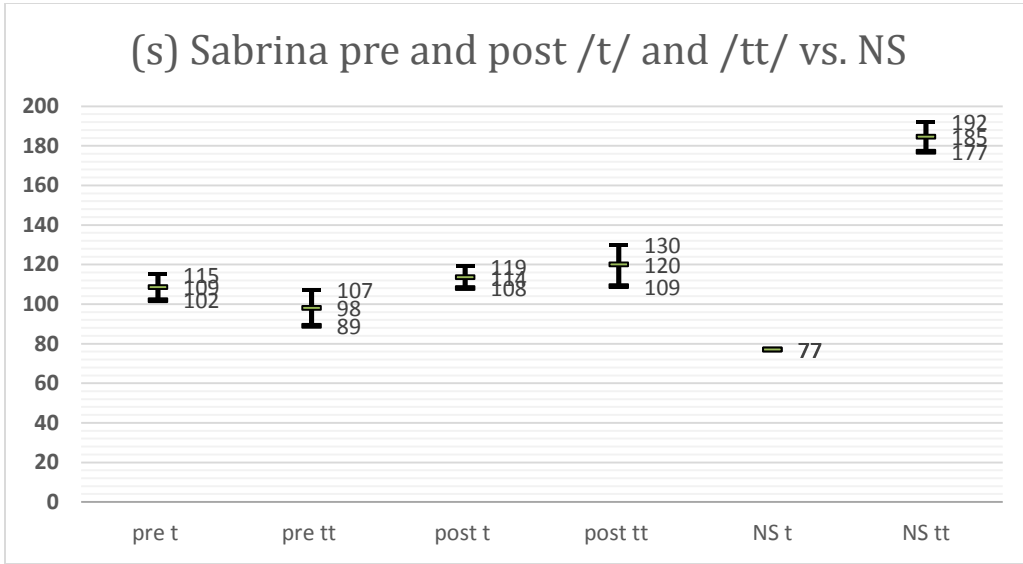
Figure 60. Sabrina's pre and post /t/ and /tt/ in sentences (s) vs. native speaker (NS) data (msec).
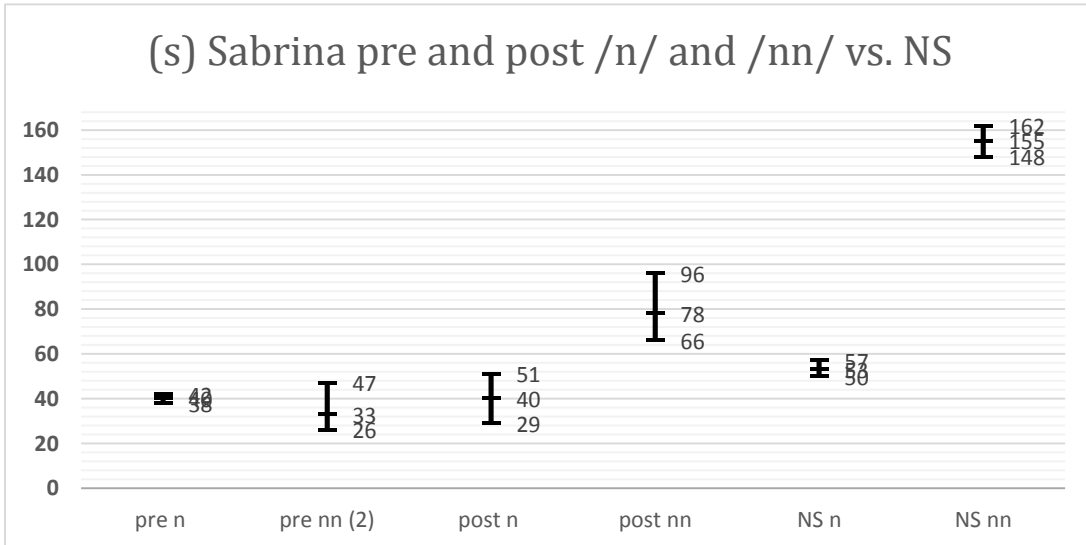
Figure 61. Sabrina's pre and post /n/ and /nn/ in sentences (s) vs. native speaker (NS) data
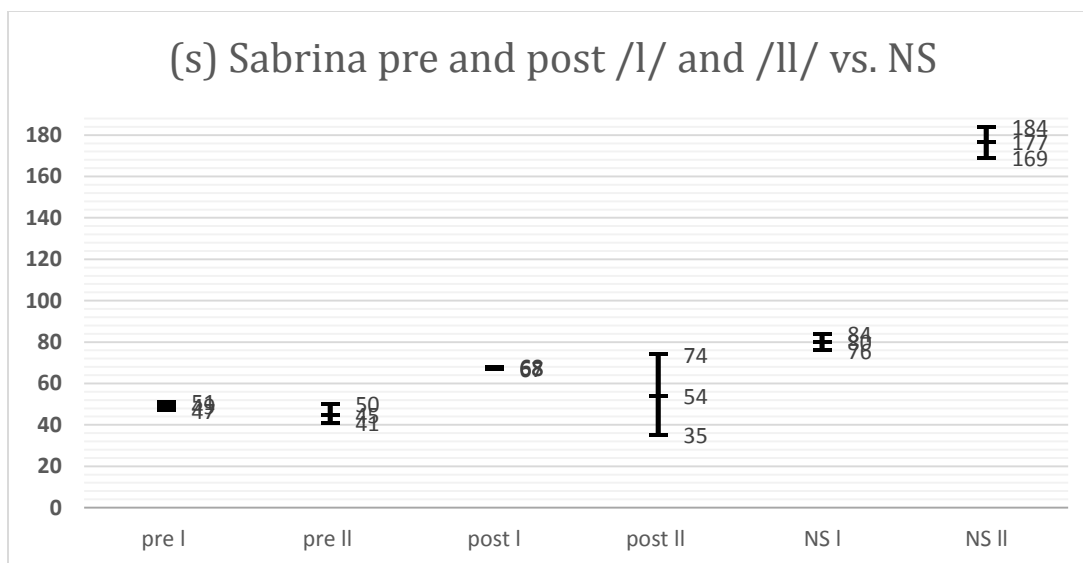
(msec).

Figure 62. Sabrina's pre and post /l/ and /ll/ in sentences (s) vs. native speaker (NS) data (msec).

The ratios, shown in Table 19, seem to show the same thing – that Sabrina improved greatly in the /n/ to /nn/ contrast, but not in the other contrasts.

Table 19

Sabrina's geminate contrast ratios vs. native speaker ratios in sentences

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	**Post**		**Pre**	**Post**		**Pre**	**Post**	
Sabrina	0.92	1.05	2.40	0.82	1.95	2.92	0.91	0.79	2.21

4.5.2.3 Gemination in passage

Figures 63, 64, and 65 show the singleton to geminate contrasts for Sabrina in the passage. Similarly to the other context, with the exception on /n/ to /nn/ in sentences, here too Sabrina showed no improvement. While the singletons tend to be native-like in duration, the geminates are consistently too short to make a contrast.
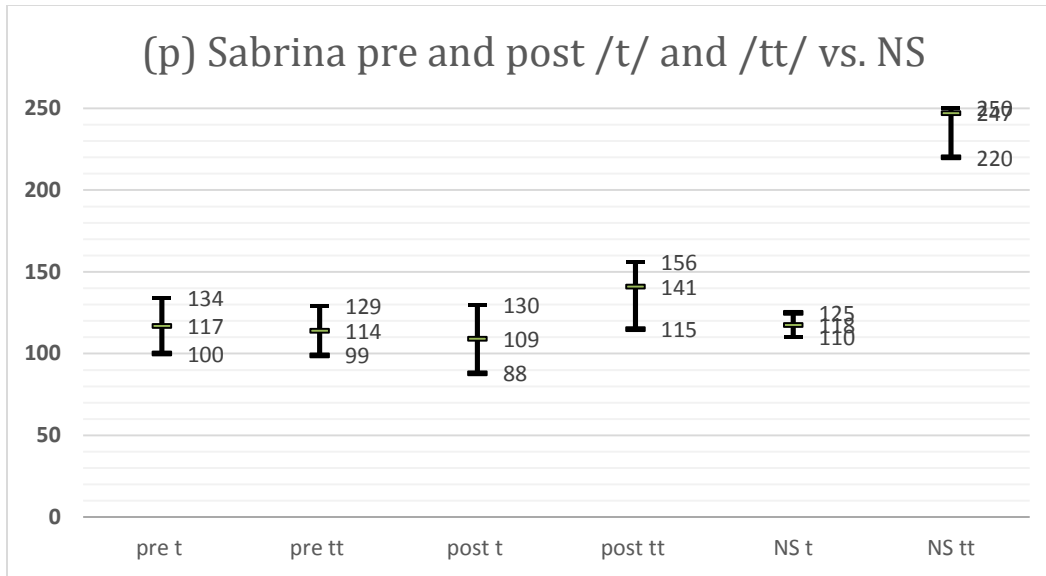
Figure 63. Sabrina's pre and post /t/ and /tt/ in passage (p) vs. native speaker (NS) data (msec).
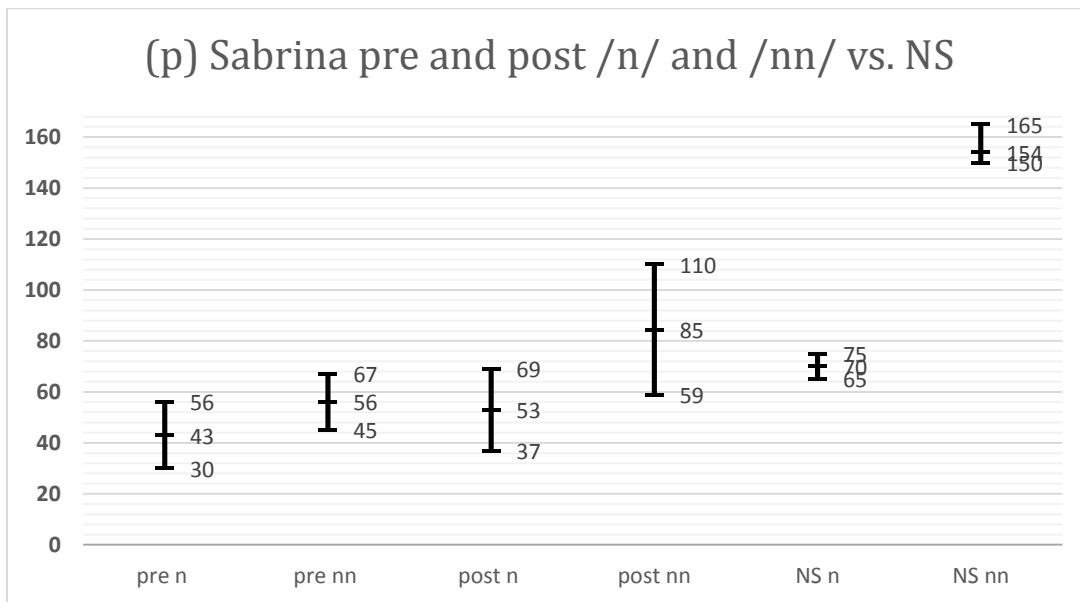
Figure 64. Sabrina's pre and post /n/ and /nn/ in passage (p) vs. native speaker (NS) data (msec).
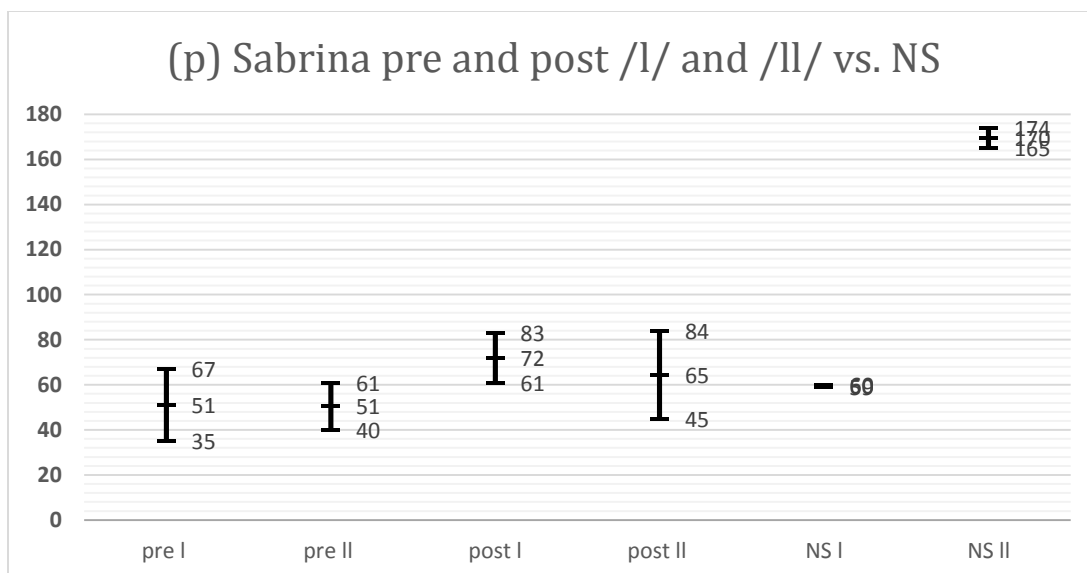
Figure 65. Sabrina's pre and post /l/ and /ll/ in passage (p) vs. native speaker (NS) data (msec).

As in previous cases, the ratios show the same results as the graphs - only a slight improvement in the nn to n ratio. The data is shown in Table 20.

Table 20

Sabrina's geminate contrast ratios vs. native speaker ratios in passage

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	**Post**		**Pre**	**Post**		**Pre**	**Post**	
Sabrina	0.97	1.29	2.10	1.30	1.60	2.05	1	0.90	2.85

4.5.2.4 Voice Onset Time in words

Figure 66 shows the data for the voice onset time measurements of /p, t, k/ for Sabrina in words. Although the means lowered in all three stops, and especially in /p/ (58 ms to 36 ms), the ranges are very wide, showing that Sabrina still had little control over the feature and did not figure out that the stops are plain.
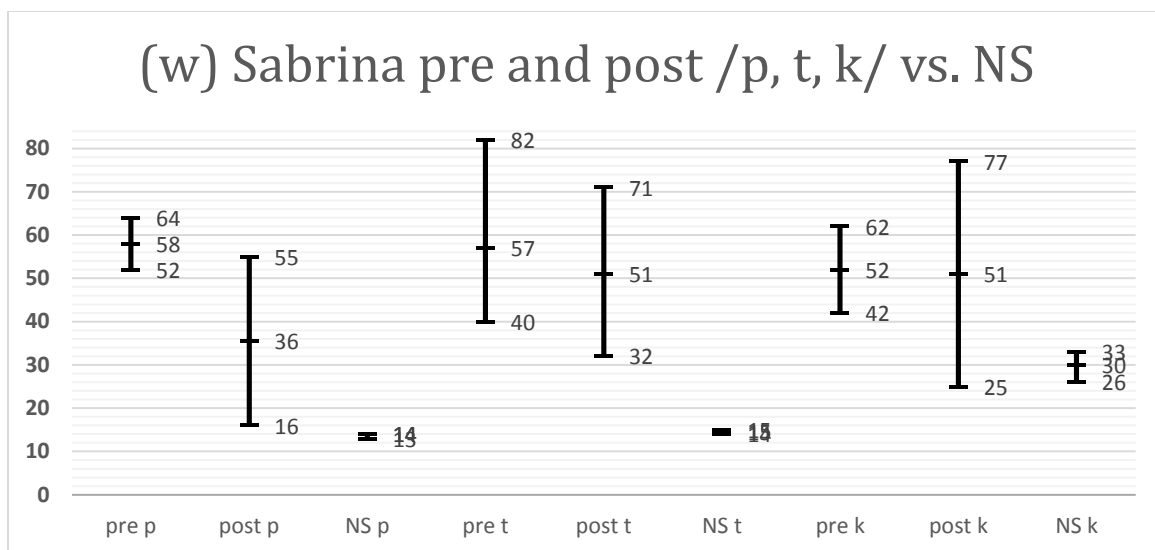
Figure 66. Sabrina's pre and post /p, t, k/ in words (w) vs. native speaker (NS) data (msec).

4.5.2.5 Voice Onset Time in sentences

The data for the stops in sentences is shown in Figure 67. For the voice onset time of /p/ and /t/ there is very slight improvement, as the means show a reduction in duration. However, for /k/ although the lowest value of 30ms is within the range of the native speaker, or of an unaspirated stop, the post-test range is very wide and has a high value of 86 ms, almost three times the duration of the native speaker.
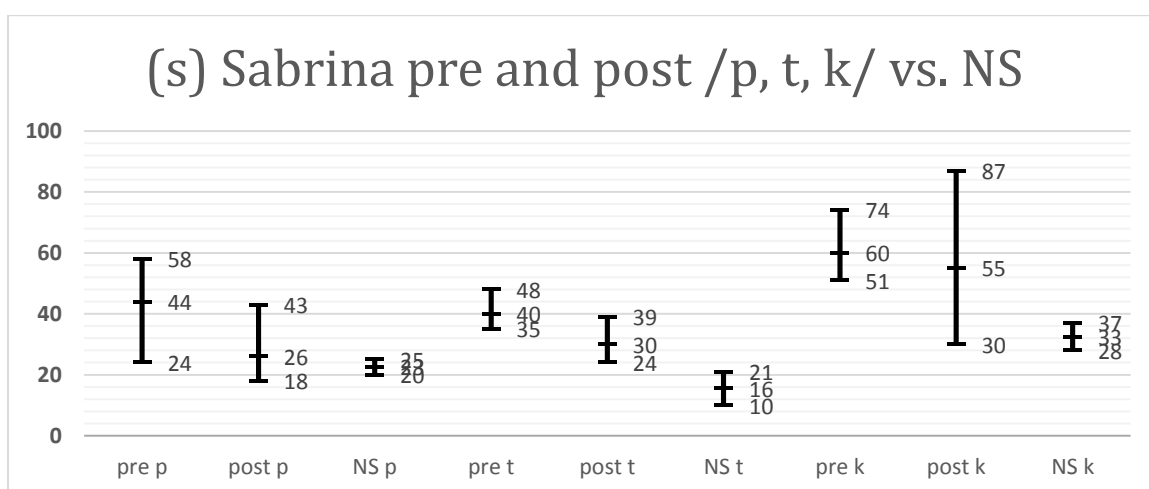
Figure 67. Sabrina's pre and post /p, t, k/ in sentences (s) vs. native speaker (NS) data (msec).

4.5.2.6 Voice Onset Time in passage

The VOT measurements of /p, t, k/ in the passage are shown in Figure 68. As in the case of the sentences, although some of the means are lower, there seems to be no clear improvement in the voice onset time of /p/ and /k/. However, there is evidence of some improvement in the voice onset time of /t/, as the full range in the post-test is shorter in duration than that in the pre-test.
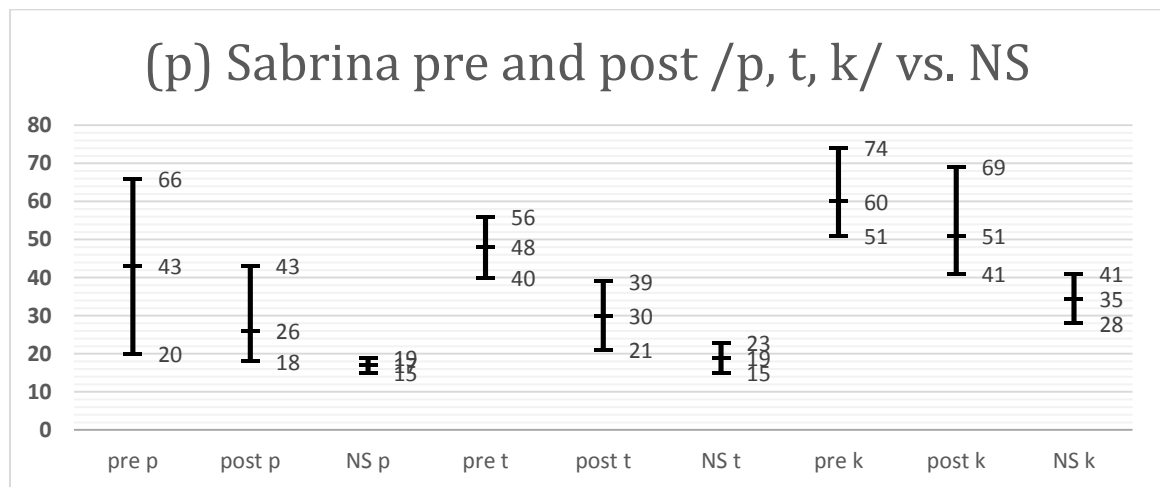
Figure 68. Sabrina's pre and post /p, t, k/ in passage (p) vs. native speaker (NS) data (msec).

4.5.3 Summary (Sabrina)

Overall, the data shows that Sabrina did not improve in neither the macro feature nor the two micro features. The difference in ratings from pre- to post-test are not statistically significant, and there is very little evidence of improvement in the acoustic analysis of her pronunciation features, as detailed below. Only the /n/ to /nn/ contrast in sentences and the voice onset time of /t/ in the passage seems to have improved. The lack of improvement could be due partly to the fact that Sabrina tended not to repeat lessons. Further, lack of improvement in gemination could be explained by the fact that neither of the languages she has experience with (English and Spanish) have contrastive gemination, and that in her native language (English) voiceless stops are aspirated in word-initial position.

99

4.6 Carolina

Carolina has a language background in Spanish, including four years in high school and two semesters in college. As mentioned in Sabrina's language background, Spanish lacks geminate consonants (except for the tap/trill rhotic contrast), but has unaspirated voiceless stops like Italian.

The *Duolingo* report for Carolina showed that she accessed *Duolingo* eight times, and each level once, completing a total of 16 lessons. Data from the report is shown in Table 21, and shows that no lesson was repeated. Similarly to both Sabrina and Barbara, the "test" function was not counted as a repetition of a lesson

Table 21

Carolina's Duolingo Report: Overall days of practice, number of lessons completed, and number of times each lesson was repeated

	Days	N. lessons[a]	Repeated Lessons[b]														
			Basics			Basics 2					Phrases		Food					
Levels			*1*	*2*	*3*	*1*	*2*	*3*	*4*	*5*	*1*	*2*	*1*	*2*	*3*	*4*	*5*	*6*
Carolina	8	16																

[a] *Note*: number of lessons completed including repeated lessons
[b] *Note:* the asterisk indicates a repeated lesson; the number of asterisks indicates how many times

4.6.1 Ratings

The *t*-test run on the comprehensibility scores for Carolina show that there was no significant difference between the pre-test (M=4.90, SD=2.56) and the post-test (M= 3.50, SD= 2.54); t(17)=1.64, p > 0.05. Similarly, for accentedness, the difference between the pre-test (M=7.1, SD=2.06) and the post-test (M=6.7, SD=1.76)) is not statistically significant; t(17)=.574, p > 0.05 These results indicate a lack of improvement.

Overall, Carolina did not show improvement in the native speaker ratings on either scale. She also tended not to repeat lesson, similarly to Barbara and Gianna. Her acoustic analysis,

outlined in the following section, corroborates these findings: no evidence of improvement was found in either the geminate contrast (except for the tt to t contrast in words) or the voice onset time of the stops.

    4.6.2 Acoustic Analysis

    4.6.2.1 Gemination in words

    Figure 69 shows Carolina's data. Although the geminate data from the post-test (M=194) was an improvement from the geminate data from the pre-test (M=165), the singletons are still quite long as compared to the native speaker. Overall, it appears that Carolina does appear to make some sort of contrast in the post-test.

(w) Carolina pre and post /t/ and /tt/ vs. NS
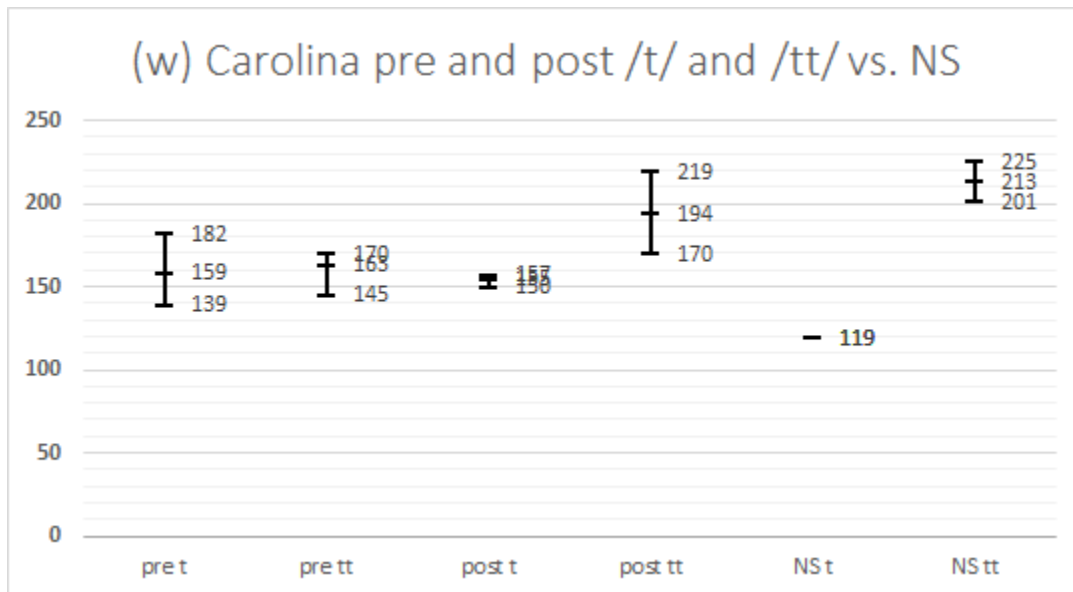
Figure 69. Carolina's pre and post /t/ and /tt/ in words (w) vs. native speaker (NS) data (msec).

    Figure 70 shows Carolina's data. It appears that she produced a small contrast and that the highest value of the geminate in the post-test (144 ms) reached the range of the native speaker. However, her singleton /n/ is still quite long and at its highest value (106 ms) overlaps the geminate range.
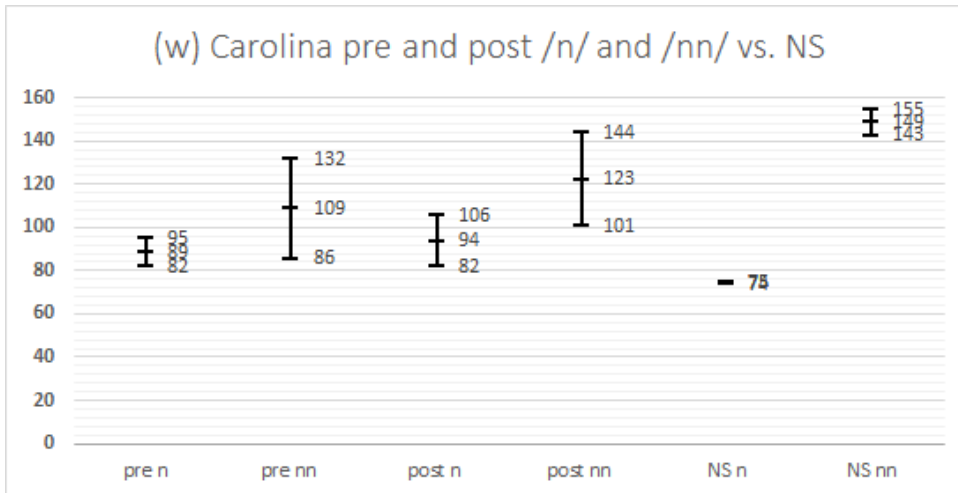
Figure 70. Carolina's pre and post /n/ and /nn/ in words (w) vs. native speaker (NS) data (msec).

Figure 71 shows Carolina's data for /l/ and /ll/. It seems that in her case too there is no evidence of the production of a contrast. Similarly to Barbara, the durations of Carolina's geminate /ll/ are much shorter than those of the native speaker.
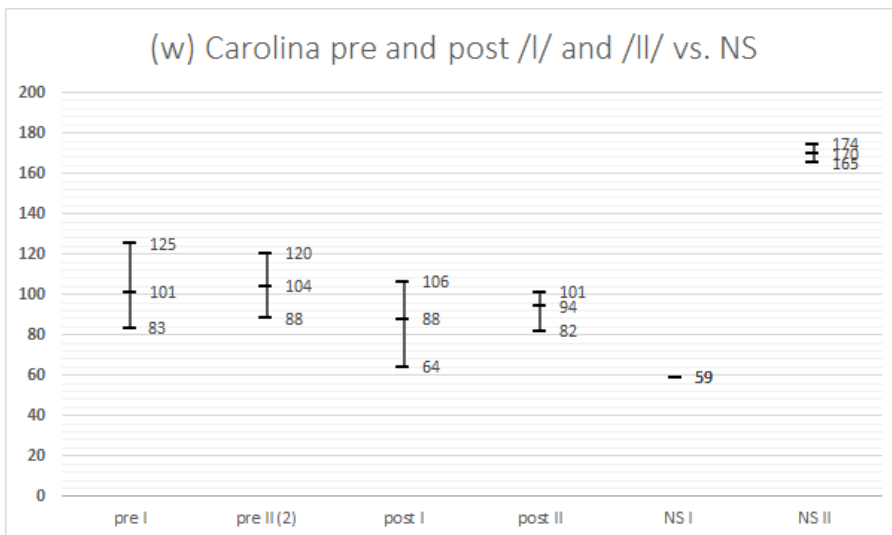
Figure 71. Carolina's pre and post /l/ and /ll/ in words (w) vs. native speaker (NS) data (msec).

The ratios, shown in Table 22, align with the data shown in the graphs and indicate a small improvement in the tt:t ratio, but no improvement in the other two contrasts. Also, the ratios

show that Carolina's measurements are far from those of the native speaker, as anticipated in the graphs.

Table 22

Carolina's geminate contrast ratios vs. native speaker ratios in words

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	Post		Pre	Post		Pre	Post	
Carolina	1.03	1.38	1.79	1.22	1.30	2.00	1.02	1.06	2.87

4.6.2.2 Gemination in sentences

Figure 72 shows the data measurements from the pre- and post-test for the contrast between singleton /t/ and geminate /tt/. Although the range of the geminate /tt/ slightly increase, there is no evidence of a contrast in either pre- or post-test.
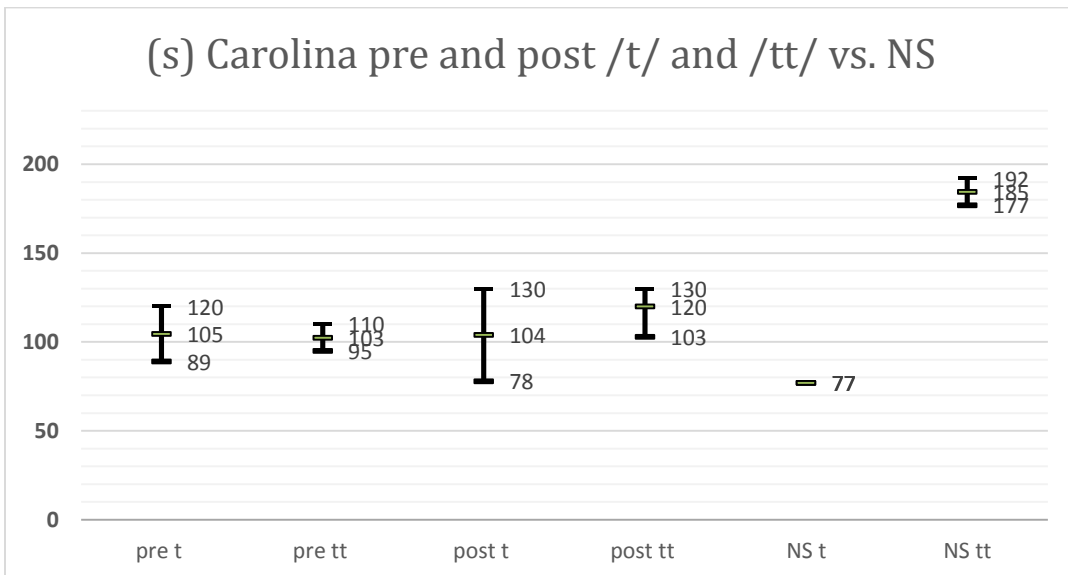
Figure 72. Carolina's pre and post /t/ and /tt/ in sentences (s) vs. native speaker (NS) data (msec).

In the case of the contrast between singleton /n/ and geminate /nn/, there seems to be a slight improvement in the post-test, although a clear contrast is still absent (see Figure 73). Most

of the values from the singleton and the geminate overlap and both are much shorter than the

singleton of the native speaker.
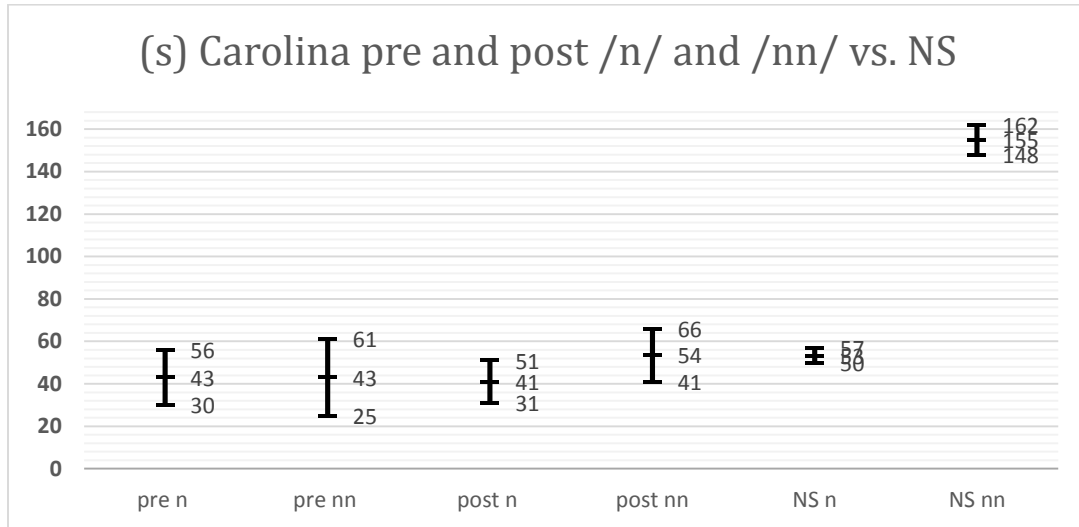
Figure 73. Carolina's pre and post /n/ and /nn/ in sentences (s) vs. native speaker (NS) data

(msec).

Data from the contrast between singleton /l/ and geminate /ll/, in Figure 74, also shows

no evidence of a contrast in either test, even though the singleton range lowered in the post-test.
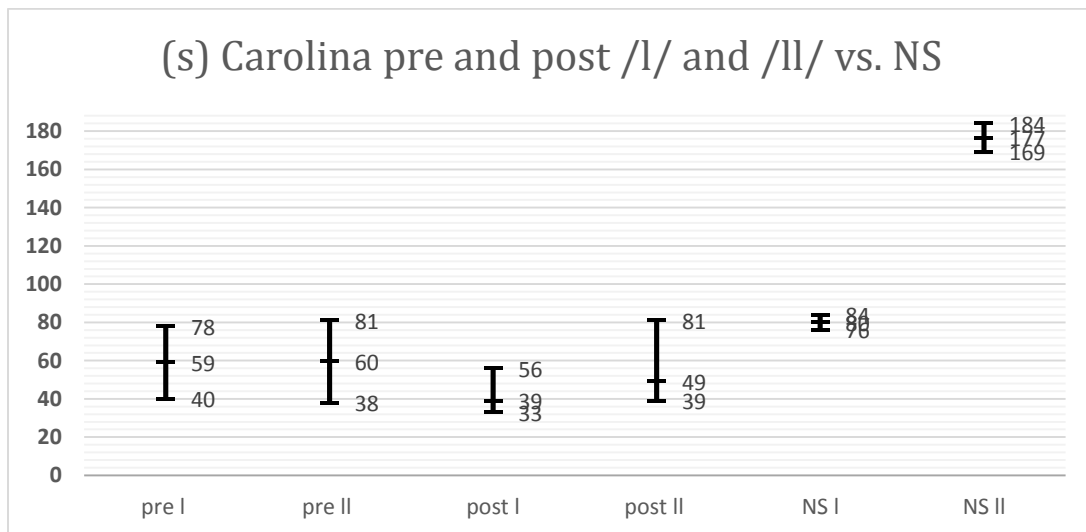
Figure 74. Carolina's pre and post /t/ and /tt/ in sentences (s) vs. native speaker (NS) data (msec).

The ratios reflect what was displayed in the graphs and are shown in Table 23. As the ratios seem to gravitate around 1, they indicate the singletons and geminates tend, on average, to be similar in length. The only slight improvement seems to be in the nn:n ratio, although it is not a clear contrast.

Table 23

Carolina's geminate contrast ratios vs. native speaker ratios in sentences

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	Post		Pre	Post		Pre	Post	
Carolina	0.98	1.15	2.38	1.00	1.31	3.10	1.01	1.25	2.31

4.6.2.3 Gemination in passage

Figures 75, 76, and 77 show the data for the three contrasts in the passage. All three graphs, and in all three contrasts, the measurements remain stable from singletons to geminates. The similar means indicate a lack of contrast in both pre- and post-test, although the singletons seem to be in similar ranges, although wider, to those of the native speaker.
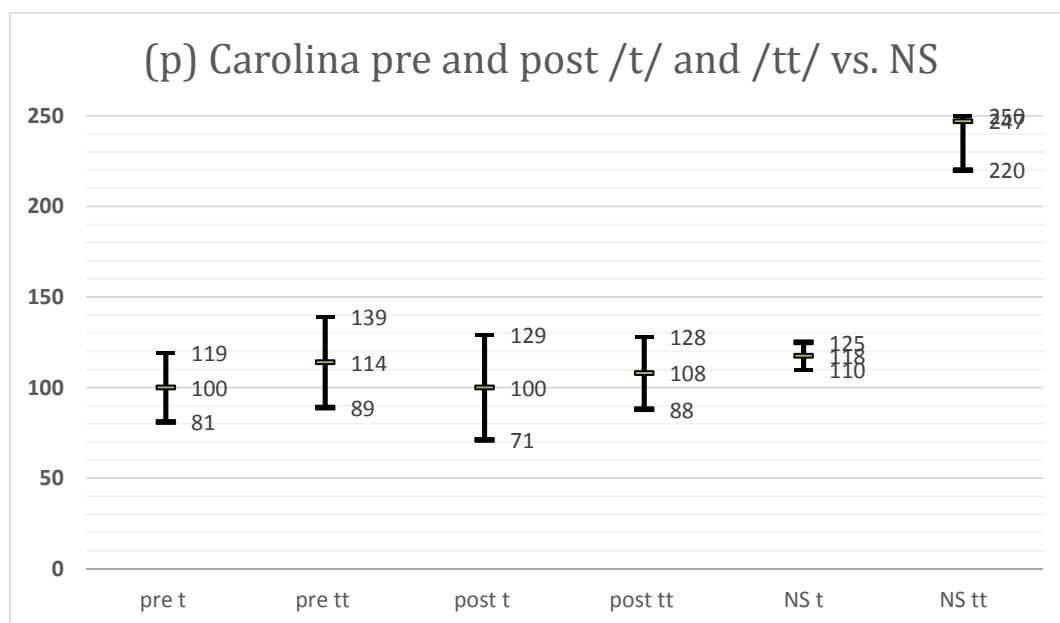
Figure 75. Carolina's pre and post /t/ and /tt/ in passage (p) vs. native speaker (NS) data (msec).
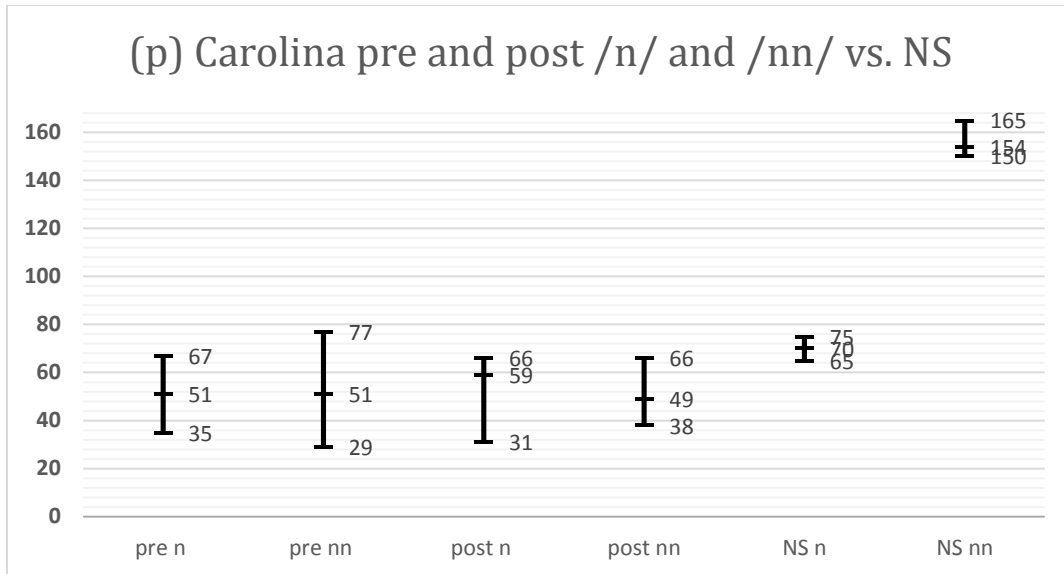
Figure 76. Carolina's pre and post /n/ and /nn/ in passage (p) vs. native speaker (NS) data (msec).
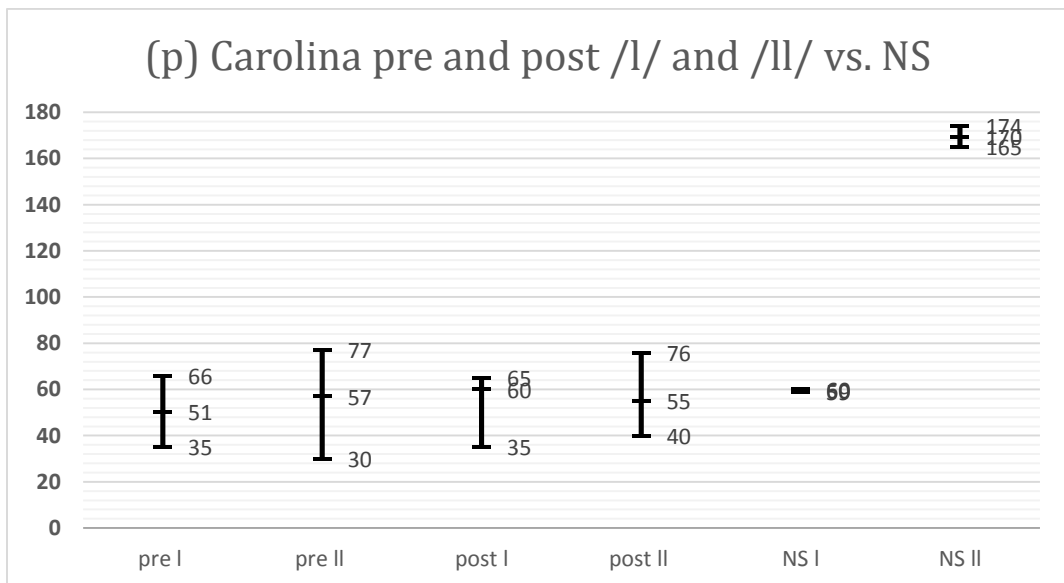
Figure 77. Carolina's pre and post /l/ and /ll/ in passage (p) vs. native speaker (NS) data (msec).

Regardless of the fluctuations in them, however, the ratios, shown in Table 24, also show that there is no evidence of contrast in the passage.

Table 24

Carolina's geminate contrast ratios vs. native speaker ratios in passage

Participant	tt:t ratio		NS Ratio	nn:n ratio		NS Ratio	ll:l ratio		NS Ratio
	Pre	Post		Pre	Post		Pre	Post	
Carolina	1.14	1.08	2.10	1.00	0.83	2.05	1.13	0.92	2.85

4.6.2.4 Voice Onset Time in words

Figure 78 shows the data for the voice onset times of /p, t, k/ for Carolina in words. It is clear from the data that there was no improvement, as the ranges – although shifting some – do not change much from the pre- to the post-test. This means that Carolina did not catch on to the idea that these stops are unaspirated. Further, the means are considerably higher than those of the native speaker.
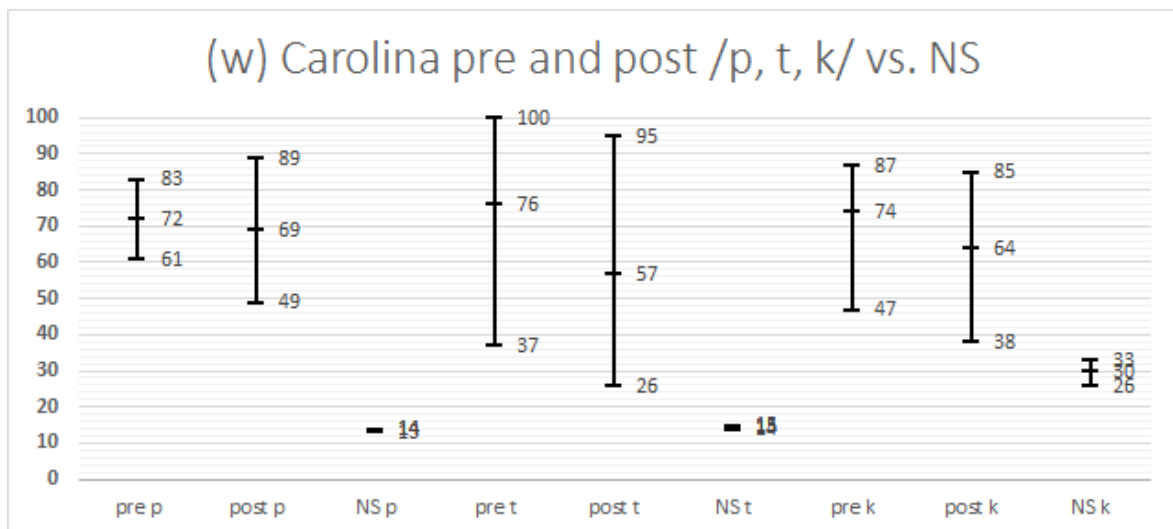
Figure 78. Carolina's pre and post /p, t, k/ in words (w) vs. native speaker (NS) data (msec).

4.6.2.5 Voice Onset Time in sentences

The data for the sentences, in Figure 79, shows a similar picture. No evidence of improvement is seen for the voice onset time of /p/, /t/, or /k/ with means remaining much higher than those of the native speaker, although some fluctuation can be seen.
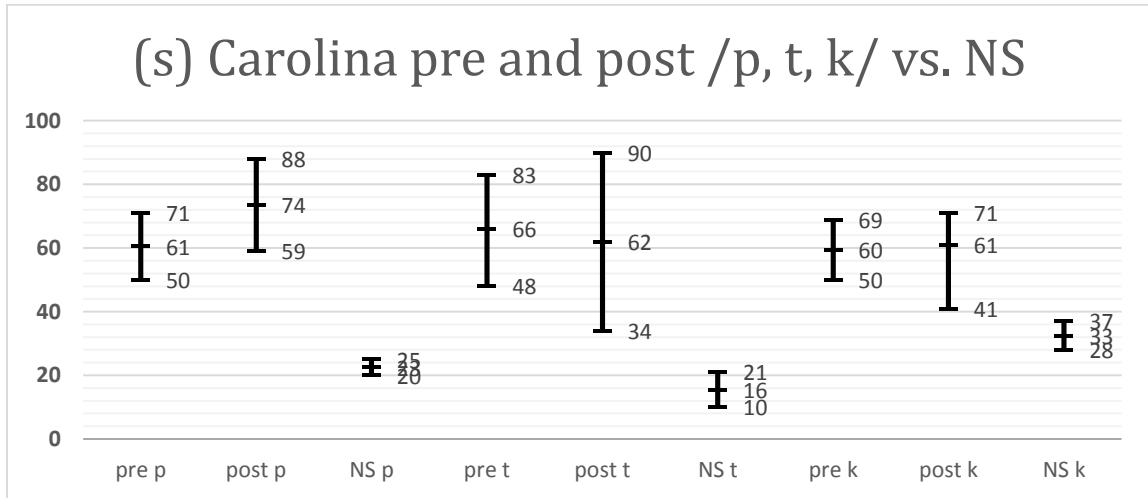
Figure 79. Carolina's pre and post /p, t, k/ in sentences (s) vs. native speaker (NS) data (msec).

4.6.2.6 Voice Onset Time in passage

Similarly to the words and sentences, the data from the passage (shown in Figure 80) indicates that there was no improvement in the voice onset times of /p/, /t/, /k/, which are still produced as aspirated.
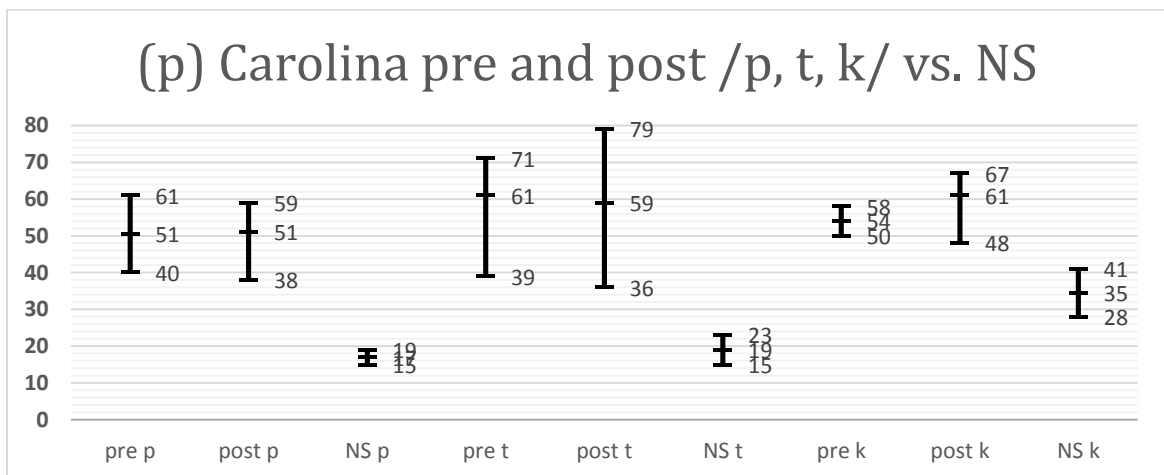
Figure 80. Carolina's pre and post /p, t, k/ in passage (p) vs. native speaker (NS) data (msec).

4.6.3 Summary (Carolina)

Carolina's ratings and acoustic analysis match. No improvement was measured with the native speaker ratings, and this is also seen in the acoustic analysis of her read speech samples. With the exception of the contrast between singleton /t/ and geminate /tt/ in words, there seems to have been no improvement in the production of the contrasts or the voice onset times of /p/, /t/, and /k/. As for Sabrina, lack of improvement in gemination could be due to her background with Spanish, and English. However, also like Sabrina, lack of improvement in the voice onset time of the voiceless stops could be due to English but not Spanish, as the latter produces the same stops unaspirated.

4.7   Chapter Summary

The ratings from the native speakers show that Luciana achieved statistically significant improvement on both comprehensibility and accentedness, while the rest of the participants did not. Further, the *Duolingo* reports indicate that Luciana and Emma are the two participants who practiced the most.

The results from the acoustic analysis of the geminate contrasts (shown in Table 25) indicate that both Luciana and Emma already produced at least some of the contrasts in their pre-tests. However, they were also the only two participants that were able to either refine the contrasts (i.e. making it more pronounced and/or tightening the ranges) or keep them consistent in almost all contrasts, with a few occasions of actual improvement (i.e. no contrast in the pre-test but present in the post-test). Barbara neither produced the contrasts in the pre-test nor improved in their production, and Sabrina and Carolina only improved in two context-specific contrasts each, with no refinement or improvement in the rest.

Table 25

Overview of acoustic analysis of gemination results for all participants

	Gemination								
	Words			Sentences			Passage		
	t-tt	n-nn	l-ll	t-tt	n-nn	l-ll	t-tt	n-nn	l-ll
Emma	†N	**Y**	†N	N	**Y**	**	N	†Y	†Y
Barbara	N	N	N	N	N	N	N	N	N
Luciana	**Y**	†N	†Y	N	†Y	†Y	†Y	†Y	†Y
Sabrina	N	N	N	N	**Y**	N	N	**Y**	N
Carolina	**Y**	**Y**	N	N	N	N	N	N	N
† the speaker was already producing the appropriate contrast									
†Y indicates that the degree of contrast may be more pronounced in the post-test or the range of values might have tightened, indicating some improved control									
†N indicates that the speaker had the contrast in the pre-test, still has it in the post-test, and does not show change. It cannot be claimed that gemination is something the speaker learned through *Duolingo* as she was already aware of it.									
** no data available for the pre-test									

Further, Luciana and Emma are the two participants with the strongest language backgrounds, and also the only two participants with knowledge of language acquisition processes through academic training.

The difference between Luciana and Emma in terms of their ability to improve and refine their contrasts probably lies in their language studies. Although both participants have L1s (English, Russian, Ukrainian) that tend to have gemination that is not produced phonetically, Luciana's latest experience with language is with Arabic, a language that is rich of contrastive gemination, while Emma's most recent language studies involve German which has much fewer instances of gemination and only in specific environments (such as before /j/).

The rest of the participants not only had limited knowledge of language acquisition and no formal training, they also had experience with either Spanish or French - the former being a language with no gemination whatsoever, and the latter having limited gemination and inconsistent across oral production and orthography.

In terms of voice onset time, Table 26 shows the results from the acoustic analysis of voice onset times. In this case, while Luciana did improved in a few occasions, Emma seems to

have consistently improved across phones and across contexts, with Barbara and Sabrina

improving in only one instance each and Carolina showing no improvement at all.

Table 26

Overview of acoustic analysis of VOT results for all participants

| | Voice Onset Time (VOT) | | | | | | | | |
| | Words | | | Sentences | | | Passage | | |
	P	T	K	P	T	K	P	T	K
Emma	**Y**	N	**Y**	N	**Y**	**Y**	N	**Y**	**Y**
Barbara	N	N	N	Y	‡	N	N	N	N
Luciana	N	N	Y	Y	Y	N	N	N	N
Sabrina	N	N	N	‡	‡	N	N	Y	N
Carolina	N	N	N	N	N	N	N	N	N
‡ Inconclusive data. The speaker's VOT in the pre-test was already relatively short. In some cases it was similarly short in the post-test, and in some cases even lengthened. However, it is not possible to draw any conclusions: we cannot determine whether the speaker was actually attempting to produce unaspirated voiceless onsets in Italian.									

The difference between Emma and Luciana could again be due to their different

experiences with language. Although Emma most recently studied German, which has aspirated

word-initial voiceless stops, her native languages both have plain ones, which could have helped

her hear the plainness of the stops in the program. Luciana, on the other hand, has English as her

L1, which has aspirated word-initial voiceless stops, and Portuguese as her L2 which has a range

of aspiration. Her experience with the languages could have made it harder for Luciana to pick up

on the lack of aspiration in Italian.

The rest of the participants, however, come from languages that have plain word-initial

stops just as Italian does. It is possible that the fact that they have not studied these languages in a

while (some even years) could have impacted their ability to realize this similarity. Further, their

lack of training in language acquisition could have made them less likely to pick up on this

feature.

Overall, these findings indicate that language acquisition training and language

acquisition processes awareness, typological similarity of L1s and consequent languages and

recency of study of said languages, along with practice in the program, were all possibly important variables in the success of the acquisition of gemination and voice onset time for these Italian as a second language learners.

However, the reason why Luciana showed statistically significant improvement in the native speaker ratings and Emma did not, although having similar training and practice patterns and similar consistency either in the refinement or in the improvement of one or the other features, is not apparent from the results of this study. Further, the typological similarities and differences of their L1s and L2s, and in some cases L3s, with Italian also do not provide sufficient reasons for this difference. A possibility, thus, could lie in the features that were not investigated in this study, such as other segmental features (i.e. vowel quality, rhotic r) or suprasegmental ones (i.e. stress, intonation).

The first research question asked about the effectiveness of *Duolingo* for the acquisition of the pronunciation macro skills of accentedness and comprehensibility for Italian as a second language learners. The data from this study seems to indicate that the first four levels of the *Duolingo* program do not directly affect improvement in accentedness and comprehensibility. However, since improvement, or lack thereof, in the features investigated here is usually considered to have an important bearing on perceptions of accentedness and comprehensibility by native speakers, it is possible that others features played a role.

The second research question asked about the effectiveness of *Duolingo* for the acquisition of the pronunciation micro skills of geminate contrast and voice onset time (VOT) for voiceless stops in word initial position for Italian as a second language learners. The data shows that many variables are at play, including training in language acquisition processes, L1 and consequent languages, recency of study, and practice. Further, the data seems to suggest that if these variables are aligned, either improvement or refinement of already acquired features can be possible through the use of the program. However, as demonstrated in the first research question,

improvement in these feature may not translate to improvement in overall intelligibility, which is

the overarching goal of pronunciation training.

CHAPTER 5

RESULTS: PERCEPTIONS ON THE EFFECTIVENESS OF DUOLINGO

As mentioned in Chapter II, user experience, also called "learner fit" (Chapelle, 2009; Hubbard, 2006), is an important factor in the evaluation of CALL applications (Garrett, 2010; Karapanos, 2013; Kuniavsky, 2003). One of the elements of user experience is perceptions (Derwing & Murray, 2005), which, if explored, grant a more detailed account of the effectiveness of CALL platforms. The third research question, thus, asked about the perceptions of Italian as a second language learners on the effectiveness of *Duolingo* for the acquisition of pronunciation skills.

The following chapter is an overview of the results for the interviews, questionnaires, and observations, along with a discussion of these results in relation to one another and the literature. As the interest of this study was in exploring perceptions overall, the results divided by themes, rather than by participant. The three major themes are, in the pre-interviews, that of language learning and technology for language learning, and in the post-interviews those of *Duolingo* for language learning and of *Duolingo* for pronunciation.

The reasoning behind the a priori theme of language learning and technology for language learning in the pre-interview was that in order to be able to understand the participants' perceptions on the effectiveness of *Duolingo* for pronunciation, it was important to first understand each participant's perceptions of language learning in general and of how technology

could fit into that framework, and only then to investigate whether or not *Duolingo* specifically aligned with those perceptions and frameworks. In the post-interview, the participants were asked to comment on the same theme but with their experience with the tool in mind in order to assess if anything had changed in their perceptions. Lastly, they were asked to specifically address pronunciation and their perceptions of the effectiveness of the tool for this specific skill.

5.1   Language Learning and Technology for Language Learning

There were three sub-themes that emerged from the discussions about general language learning in the pre-interviews. The most prominent was the idea that foundational knowledge must precede more communicative or varied instruction. Four out of the five participants mentioned that they believe that successful language learning starts with a foundation of vocabulary and grammar at the beginning (Luciana, Sabrina), but also of pronunciation (Emma, Sabrina), and Carolina also believes it is important to learn these basic concepts in academic settings. A second sub-theme was the need for immersion, conversation, or the contact with native speakers, mentioned by three of the five participants and often referred to as a very crucial step in learning a language. The third main sub-theme is that of motivation, and specifically, as Emma states, that "[one] must already possess strong intrinsic motivation - it's a must - nothing can happen without it."

A few other topics were also discussed in the interviews, such as the balance between skills, mentioned by Emma when saying that "also very crucial [...] is extensive reading and writing, but listening to TV shows, radio, movies can also be very helpful; overall, all four skills must be practiced together," the ability to be a "good noticer" (Luciana), and the value of feedback mentioned by Barbara.

Two of the participants had no experience with technology for language learning, while the other three agreed that technology cannot be a primary source of instruction when learning a language and that either the combination of traditional classroom instruction and technology, or various other sources and technology would be more appropriate and more effective. This balance

115

between instruction/sources and technology seemed to be the dominating theme in the discussion of technology of language learning. It is important to note that the participants refer only to certain types of technology, the ones they have experience with, which include mostly tutorial videos and subtitled videos, online discussion boards, and online tests.

Barbara pointed to the classroom instruction component by saying that "the mix of teacher and software is the best sort of use of technology in the classroom or in general for language learning" and that she is "[...] all for it, but the human component is crucial." Likewise, Emma stated that technology can be helpful, but that "you really have to be in the country to learn a language and tech can't be the primary source of learning." Emma also highlighted the need for a foundation, a theme that had already emerged in the discussion for language learning, saying that one needs "to have a foundation before using tech, for example be able to already read something and have some understanding of the grammar."

Out of the five participants, only Luciana referred specifically to OLL software, and coincidentally to *Duolingo,* when discussing her perceptions of technology for language learning, However, similarly to the other participants, Luciana stated that "[...] technology is just a starting point, one must add things to it the more one keeps going."

5.2   *Duolingo* **for Language Learning**

The rest of the participants had never used any language software, and three of them had never heard of *Duolingo* before. The perceptions on *Duolingo* were therefore almost entirely explored in the post-interviews, in the form of a reflection on the elements of language learning or technology for language learning they had mentioned in the pre-interviews and a discussion how they related them to their experience with *Duolingo*. Thus, the three sub-themes of language learning (i.e. foundational knowledge, conversation, and motivation), and the main sub-theme of technology for language learning (i.e. technology as a primary source of instruction) are presented again in the following section in light of user experience.

5.2.1 Foundational knowledge

Emma and Carolina mentioned that they felt the need for introductory materials. Emma explained that although she felt as if she had gained a lot of confidence in most areas, some verbs were hard to understand or remember, and added "I would've liked a list." Although a list is present in *conjugate* function, a sub-function of the *hover* feature which allows the users to view a table of the various conjugations of the verb, during the observation it was clear that Emma had not explored the function and was not aware of it. During the observation, she used the *hover* function but never clicked on the *conjugate* function, indicating that she might not have figure out it was there or what is was. Carolina also had a similar reflection, affirming that she felt the lack of a "comprehensive list of the vocabulary or rules," adding that she "couldn't really go back and consult anything" and "had to redo a lesson if anything."

These statements corroborate what was expressed in the questionnaires. Four participants cited that the most difficult aspects of Italian were grammar rules and the translation from English to Italian, which align with the numerous mentions of lack of introductory or foundational materials in which the basic grammar and syntax rules are explained explicitly. Further, in the questionnaires, most participants also mentioned that the most helpful activity was translating words and phrases. For example, Carolina acknowledged that the most useful activity for her, although also the hardest, was the translation from Italian to English, and stated, "I struggled with those, even in Spanish to English, but that was the most helpful thing for me. I can do the Italian to English, Spanish to English, but backwards is really hard for me."

This need for a foundation seems to also be corroborated by the observations, in which participants pointed to their struggles with two main activities: the translation from English to Italian and an activity in which a sentence is presented with a dropdown menu with a list of possible choices for verb forms in the place of the verb (see Figure X). During the observations, all five participants spent more time on these exercises than any of the other activities (i.e. type

what you hear, matching pictures, etc.), although they often commented that because of the challenge they considered these very helpful activities.

Only Barbara and Sabrina seemed to not feel the lack of introductory materials, at least after having gained more confidence with the program. Barbara explained that rather than having a certain amount of background knowledge, which you must learn before starting any activity, she enjoyed the immediate access to information - "all information is very organized and immediate -" that could be consulted at any time while "actually doing things." Similarly, Sabrina said that "at the beginning maybe you might feel the lack of a foundation, but specifically in the fourth level, I felt like I didn't have to look up anything," explaining that she had "just simply internalized things" and adding that she felt much more confident and "didn't feel the lack of formal instruction or foundational stuff at all."

Overall, it seems as if the participants perceived *Duolingo* as being ineffective in terms of the declarative stage, or the period in which one acquires knowledge *about* something. However, this perception seems to be quite different when discussing *Duolingo*'s effectiveness for pronunciation, as detailed in section 5.3 (this chapter).

5.2.2 Conversation

Barbara, Carolina, and Emma explained that they felt they were missing the opportunity to talk and listen, while Sabrina stated that she anticipates that learners at higher levels might miss conversing, although she acknowledges the fact that the practice she did receive in *Duolingo* might be sufficient at her level:

> As far as conversation goes, I think that if I were to have a conversation with what I learnt in present tense I probably could converse fine, maybe some basic needs. I think that especially with the speaking into the microphone, and maybe my own desire to pronounce properly, it was helpful. Maybe at higher levels, I would begin to miss conversation, though.

118

However, all three participants (Barbara, Carolina, and Emma) agreed that the consistency of the program made up in part for this lack. As a matter of fact, Barbara explained that the lack of more human-like variation, "one time you say it like this, the other time you say it like that [so it makes] you feel safer and boosts your confidence when you do get it right" at least when you are first approaching a language. Carolina also added that "a good thing though, which was better than having a teacher, was that you could have the program repeat things as many times as you wanted," explaining that achieving this "would be weird with a teacher or a native speaker."

Overall, the perceptions of the participants on conversation were that although the consistency of the program is effective for beginners and grants a certain amount of flexibility which might not be available when speaking with a teacher or a native speaker interlocutor, the program lacks communicative practice, which might be a critical problem at higher levels especially.

5.2.3 Motivation

Emma and Sabrina confirmed their initial analysis of motivation as an important element for language learning, stating that there needs to be strong and specific motivation to keep using the program. Emma explained that she believes the experience was successful overall, but that "that motivation is extremely important, even if gamification helps a lot." Likewise, Sabrina also emphasized the importance of motivation stating that she felt that one needs "a certain type of motivation and a strong one, like the hover option might give an excuse not to learn, but if you use it well it makes you very independent." She also adds that this motivation should not come from the purpose of achieving a grade, as the program "wouldn't work for homework because students would cheat maybe, but if you have other motivations, like not a grade, and you are smart about it, it works very well." Thus, these perceptions seems to indicate that, although the *Duolingo* can be effective for certain types of learners, especially through its gamified

experience, the program could also be ineffective for learners that have low motivation or are motivated by obtaining a grade.

5.2.4 Technology as a primary source of instruction and role of the teacher

In the post-interviews, both Barbara and Sabrina stated that they were surprised that they did not miss the presence of a teacher. Barbara, in particular, reported that she thought the instructor was crucial, but that she had changed her mind. "It can be done without," she stated, "you can also look stuff up, take breaks, or take longer without interrupting the flow of a class. Makes you feel safer and boosts your confidence." Similarly, Sabrina explained that "not having a teacher was disorienting at first, especially since you jump right in, but despite that, as I kept going I did feel myself getting better and anticipating things." As she gained experience with the program, she proposes, "things were probably happening unconsciously, more naturally than I thought would, and I thought that maybe I don't really need the pure instruction." Carolina, too, said that if anything she lacked the introductory materials rather than a teacher.

Overall, the participants were split in terms of the necessity of foundational knowledge, but all of them agreed that technology cannot be a primary source of instruction. This seems to be due to the fact that the program lacks more communicative practice, which participants perceived as being crucial especially later in the learning process. However, most of them agreed that *Duolingo* is effective for beginner levels, especially thanks to the consistency and the repetition contained in the software, which at times surpassed what is available in a traditional classroom setting. Further, while a strong motivation is perceived by some as being an important element of a successful experience with *Duolingo*, the same participants also perceived the gamified structure as being very effective.

5.3   *Duolingo* for pronunciation

During the post-interviews, the conversation also included pronunciation specifically. In the questionnaire, the participants all indicated that the two easiest aspects of Italian to learn were

pronunciation and vocabulary and this was corroborated at the end of the study. Also, even in the very first week, all participants indicated that they felt they had improved by 2 or 3 points on a scale from 1 to 5, where 1 indicates no improvement at all and 5 indicates improved a lot. This was also confirmed in the post-interviews: when asked what they thought they had improved on most, all participants answered that they had mostly improved in pronunciation.

The merit for this perceived improvement was given especially to the *hover* function, a feature of *Duolingo* that allows the user to hover over any word and hear it, and the *type what you hear* activity. This was obvious in the observations too, as all five participants used the *hover* feature more than any of the other features combined. Figure 81 shows aggregated data on the amount of mouse clicks in a certain position on the screen. Each participant's interaction with the program was screen recorded through QuickTime. Once complied, the videos were edited so that they would play simultaneously during only the actual interaction and not before or after (i.e. login, logoff). Once the video were set to play together, a script was created to run through the editing software so that each time the cursor would "click" on a location on screen a red dot would appear. The actual lesson depicted is just an example to show where each function is located and mouse clicks which were followed by typing were excluded. The high amount of mouse click dots located on the sentence to the left is due to the high use of the *hover* feature, which is much higher than the number of clicks on the other features (i.e. *Tips and Notes, Discuss Sentence*).
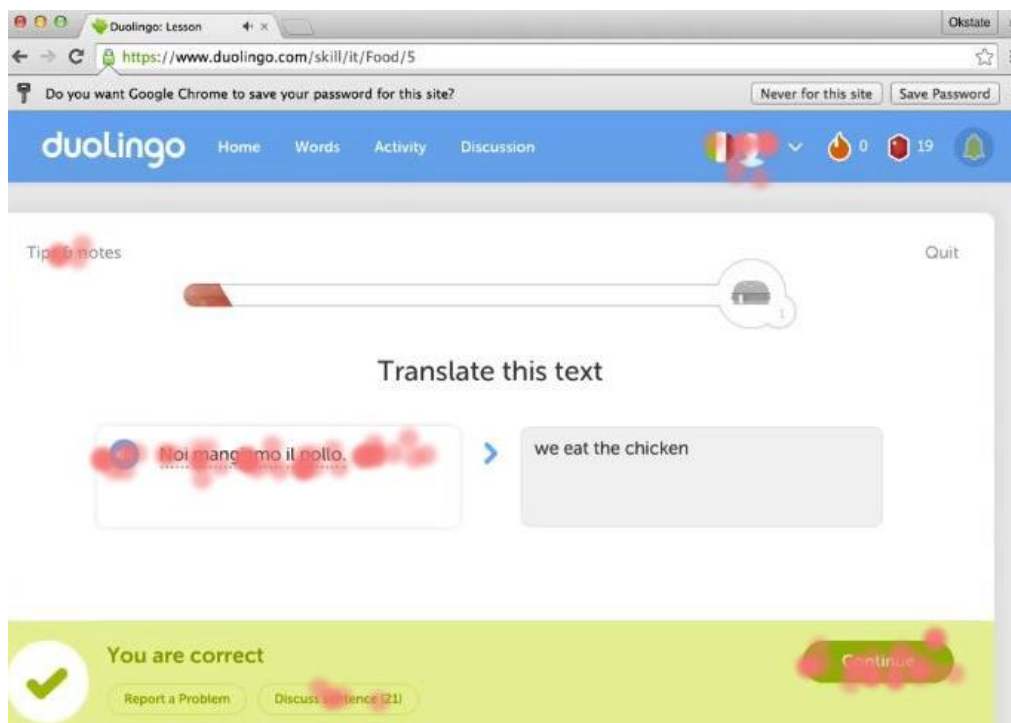
Figure 81. Aggregated data on the amount of clicks on screen during observations.

The *hover* feature and *type what you hear* activity are tied to the repetitive and the deductive nature of the program. Barbara, for example, states that "[...] the hover option makes pronunciation always accessible, you have immediate access" pointing to the repetitive nature of the activities, and adds that "pronunciation was so much easier to learn using Duolingo than in language classes taken before. There was no need for a foundation, or prep, most of the info was all there," highlighting the deductive learning that occurs using the program. Likewise, Sabrina highlights the benefits of repetition stating that she "loved the fact that you could hear the voice say the word or phrase as many times as you wanted" and adding that she "really used it a lot because [she] wanted to be able to pronounce things properly." Carolina made a similar remark about the repetitive mechanism of the software saying that "in terms of pronunciation, the biggest help was from the hover option, the repetition was great." Luciana also pointed out the deductive aspect of the program, explaining that "one positive thing is this technology uses deductive,

implicit learning, and I learn better when I figure out the patterns myself, especially in pronunciation."

5.4   Chapter Summary

The third research question asked about the learners' perceptions on the effectiveness of *Duolingo* for pronunciation. From the findings in this study, *Duolingo* for language learning in general seems to be perceived as efficient for lower proficiency levels and those with specific and high motivation. Further, the program's consistency and the repetition, alongside its gamified structure, were judged as very useful qualities, which at times can surpass what is available in a traditional classroom setting. In contrast, it is perceived as ineffective for those who feel the need of foundational knowledge before practice, and for higher levels of proficiency which need more communicative practice.

In terms of *Duolingo* for pronunciation, the participants perceived that the lack of foundational knowledge was not an issue, but rather the rich procedural stage (or practice section), with the addition of implicit learning methods and highly interactive materials, was very effective for the acquisition of pronunciation skills. Overall, in terms of user experience, the learners felt that the repetition, the availability of on demand audio, and the dictation-style activity (i.e. *type what you hear*) were all successfully implemented tools, which they felt were conducive to their improvement in Italian pronunciation.

CHAPTER 6

DISCUSSION AND CONCLUSION

6.1   Discussion

As technology becomes more and more prominent in our lives and starts to be heavily applied to language learning, work in computer-assisted language learning (CALL) seeks to use theories on how SLA happens in order to improve the design and evaluate technology for language learning (Chapelle, 2009), often using evaluations to improve design. However, this is easier said than done, and as Colpaert (2006) proposes, a crucial problem lies in the gap between language pedagogy and technology. This gap seems to be caused especially by the fact that technology seems to be improving much faster than acquisition can be theorized, and often because the acquisition that is theorized is "normal", while designers need to also know in which ways this normal process can and should be modified to help students learn languages faster and better (Chapelle, 2009).

These difficulties are especially salient with brand new technology, such as online language-learning software, language programs and courses developed entirely, or in major part, for the web. The situation is further exacerbated by the fact that very little empirical research has been conducted, least of all by the companies who create these software which only conduct annual "reports," and then only in the last few years, which have been often criticized for their operationalization of assessment.

Tarone (2015) highlights that the trend of online learning research is to focus on achievement rather than on proficiency[3], a negative trend due to economical and organizational issues with testing proficiency online. On the other hand, more academic research often ignores the investigation of user experience, which now occupies a prominent role in the field of interactions between humans and technological tools (Garrett, 2010; Karapanos, 2013; Kuniavsky, 2003). Some research already suggests that user experience is important both in terms of engagement with the material and in terms of improvement (Lai & Gu, 2011; Son, 2007), but it often fails to investigate individual data and to use the findings to make suggestions for improvements in design.

The aim of this study, therefore, was to evaluate online language learning software through the lens of theorized acquisition, a focus on proficiency rather than achievement, and with the inclusion of an investigation into user experience; specifically, this study focused on the use of *Duolingo*, a new and fast growing project, which now counts more than 100 million users, for the acquisition of the pronunciation skills for Italian as a second language learners.

In order to understand the possible connections to SLA theory, in particular skill acquisition theory, data on the amount of practice for each user was gathered through *Duolingo* reports. This data was then integrated in the analysis of the first two research questions: one about the macro skills of accentedness and comprehensibility and the other about the micro skills of gemination and voice onset time, both aimed at focusing on proficiency rather than achievement.

Proficiency in pronunciation is related to intelligibility or the "extent to which a listener actually understands an utterance," which in turn is a combination of accentedness, or a "listener's perception of how different a speaker's accent is from that of the L1 community", and comprehensibility, or a "listener's perception of how difficult it is to understand an utterance" (Derwing & Murray, 2005, p. 385). Therefore, this study used native speakers in order to

---

[3] Proficiency refers to the ability to do, to function, and achievement refers to knowledge of specific information (Language Testing International, 2016).

understand improvement based on ratings of both accentedness and comprehensibility. Further, much research has shown that prosodic errors tend to have more influence on loss of intelligibility than phonetic ones (Munro & Derwing, 1995b). However, closely linked to intelligibility is the role of functional load. The "gravity of certain types of errors is believed to differ, depending on the functional load of the phonological contrast that the learner has incorrectly produced" so it is still valuable to look at segmental features to test predicted hierarchies (Derwing & Murray, 2005, p. 391-392). Further, though accentedness might have only a minor impact on comprehensibility and intelligibility (Derwing & Murray, 2009), learners are concerned nonetheless with reducing the accentedness of their speech (Derwing, 2003; Harlow & Muyskens, 1994; Timmis, 2002), perhaps because they are aware that native listeners sometimes judge accents negatively (Derwing & Murray, 2009; Eisenstein, 1983; Galloway, 1980; Shiri & Boaz, 2010). Therefore, this study focused on the segmental features of voice onset time and gemination.

Lastly, in order to address the element of user experience, or learner fit, this study investigated the perceptions of the learners, through interviews, observations, and questionnaires.

The following sections will discuss the findings of this study in terms of connections to SLA theory and improvements in proficiency, and in terms of user experience. A discussion on the limitations of this study and on suggestions for future research will be provided next, and lastly a conclusion.

6.1.1 Effectiveness of *Duolingo*: SLA theory and improvements in proficiency

The key findings of the first two research questions concentrate on the areas in which *Duolingo* seems most effective, those in which the program seems to be less effective, and the factors that influence these outcomes, such training, language background, and practice.

In terms of reaching the goal of intelligibility, the first four levels of the program seem ineffective for meaningful improvement towards increased comprehensibility and reduced accentedness. Although Luciana improved in both macro features, with statistically significant

126

differences between the pre- and the post-test, other variables (i.e. such as other segmental features or suprasegmental ones) are likely to have influenced this result beyond improvement in these two micro features. This is clear when comparing Luciana's acoustic analysis to Emma's.

Luciana and Emma improved or refined the production of each feature more than the rest of the participants. In terms of gemination, they showed evidence of more prominent contrasts between singletons and geminates, more control through tighter ranges of values, and consistency across consonants, and for voice onset time a reduction of aspiration and tighter ranges, even though both participants were far from the native speaker measurements. This, however, was expected; most studies on both gemination and voice onset time indicate that even at high proficiency levels learners do not produce the geminate contrast and unaspirated stops in a native-like fashion, even while showing the presence of the required phonological representation.

Further, their training and language backgrounds are comparable. Both participants have training in second language research, in language acquisition processes, and in SLA theory. Further, either their L1s or consequent languages contain at least one of the feature under investigation, so that both Emma and Luciana had equal chances of being exposed to at least on feature in their previous studies.

However, as mentioned, what is not comparable are the native speaker ratings: only Luciana obtained statistically significant improvements in the ratings of comprehensibility and accentedness. Thus, it seems that other variables and not the improvement in the micro features investigated here lead Luciana to improve in the ratings. We could possibly also exclude previous training, language background, and practice, as Emma and Luciana showed many similarities in these variables.

Regardless of the causes of the difference in the native speaker ratings, however, an important finding is that indeed only Luciana and Emma improved or consistently refined the production of the features. In other words, only those participants who had previous training and repeated each lessons the most times showed evidence of improvement. The other participants,

127

Barbara, Sabrina, and Carolina, who had no training and tended not to repeat lessons, showed very little evidence of improvement.

Hence, while knowledge and training seems to have played an important role, a positive relationship between practice and improvement seems to have emerged also. The power of repetition, especially in pronunciation, is supported by much of the literature (Blake, 2011b; Hubbard and Bradin, 2004; Hubbard, 2006; Kern et al., 2008; Tozcu & Coady, 2004) and can be explained through the automatization process in skill acquisition theory. Once the learner enters into the procedural stage, a period in which the acquired knowledge is put into practice, and as time goes by, the skill being acquired becomes more and more automatic until it transforms into a "fluent, spontaneous, largely effortless, and highly skilled behavior" (VanPatten & Williams, 2014, p. 94). Indeed, it seems that Emma and Luciana's procedural stage, composed of training but also of a large amount of repetition, influenced the features under investigation: from new and declarative, they became more fluent and spontaneous.

6.1.2 Perceptions on the effectiveness of *Duolingo*: User experience

Another key finding lies in the perceptions of the learners, and this study' investigation into user experience. At the end of the first week of the study, some of the participants felt that the declarative stage, the period in which the learner acquires knowledge *about* a skill through observation, orally or in a book, was insufficient. *Duolingo* seems to favor the procedural stages (i.e. the activities are at the center of the program) over the declarative stage (i.e. tips and notes are a secondary function), causing many of the participants to feel unprepared for practice. Indeed, a tenet of skill acquisition theory is that if the declarative stage is incomplete or the knowledge acquired is not appropriate, although the level of completeness and appropriateness varies from learner to learner, the procedural stage often becomes very complex for the learner.

However, an important finding of this study emerged from the interviews conducted at the *end* of the study. Almost all of the participants agreed that, at least in terms of pronunciation, the practice made up for the lack of foundational knowledge they felt at the beginning. Here, too,

skill acquisition theory can offer an explanation: if the practice stage is rich and engaging, the learner will feel more confident and gain skills faster (VanPatten & Williams, 2014). Further, practice can also decrease "reaction time" (time needed to complete the task), "error rate" (the amount of mistakes committed during the task) and "robustness" (the amount of attention needed to perform the task) making the learner feel more engaged and more successful (Taatgen et al., 2008, p. 548), even if there are shortcomings in the declarative stage.

Further, as stated by the participants, the most important type of practice seemed to have been the *hover* feature, a feature of *Duolingo* that allows the user to hover over any word and hear it as many times as needed. This element reflect the findings of many different studies in CALL. According to Hubbard and Bradin (2004), and with the support of other studies (Blake, 2011a; Hubbard, 2006; Kern et al., 2008; Tozcu & Coady, 2004), an extremely effective tool of CALL systems is the active interaction with the materials, in this case the ability to interact with the recorded speech through the *hover* feature. Also, the repetition of mappings of form to function, an important concept in usage-based emergentism, and a central tenet of frequency studies, has been proven effective for second language acquisition. Further, the *hover* feature also aligns with many core gamification principles, such as 'on demand' and 'just in time' information.

Therefore, although no declarative knowledge, or explicit instruction, was provided, the participants' perceptions on the effectiveness of *Duolingo* for pronunciation indicate that this knowledge is unnecessary. Indeed improvement was possible, as in Luciana's and Emma's case, without this information.

It is clear, however, from the results of the ratings and the acoustic analysis, that although favoring the procedural stage over the declarative stage can be effective, especially if there is interaction with the materials, effective implicit instruction, and a positive user experience, there needs to be a certain amount of practice for this structure to translate into significant improvement. This conclusion is also supported by the perceptions of the participants with regards to conversation practice. The fact that communicative activities are relatively non-existent

for beginners, a concern voiced by Lord (2015), left many participants feeling skeptical of the effectiveness of *Duolingo* at higher proficiency levels. Indeed, while practice seems clearly important for improvement, the *amount* of practice is just as important.

6.1.3 Implications

The findings of this study have three major implications related to theory, practice, and research. Firstly, the findings demonstrate that the issue of the gap between technology and language pedagogy mentioned by Colpaert (2006) can successfully be addressed when evaluating CALL systems. By implementing considerations of how the normal acquisition process is modified through technology, as suggested by Chapelle (2009), SLA theory and language pedagogy can be used as a "basis for decisions that go into the design and evaluation of technology for language learning" (Chapelle, 2009, p. 742). Some of these modifications are the new orders, sequences, and approaches to language learning that technology offers, and especially online language learning software - in the case of *Duolingo*, for example, the preference for a richer procedural stage and secondary explicit instruction. By the same token, although at times CALL applications can appear to put more emphasis on technological advances rather than pedagogically sound content, the findings indicate that online language learning software like *Duolingo* do implement certain pedagogically sound *practices*, which is certainly a step forward in the development of sound, valid programs.

Nevertheless, the findings also highlight the need to better calibrate the abovementioned modifications (i.e. orders, sequences, and approaches to language learning). Indeed, the structure of *Duolingo* seems to rely on the fact that the program contains a large amount of levels (66 levels total for Italian), therefore giving the learner numerous occasions for rich practice. Indeed, key concepts, grammar structures, collocations, and audio are repeated and connected often throughout all the levels, and the opportunity of repeating a lesson is always present. However, as demonstrated by the use of the participants in the study, and by the general trend reported by the company itself (Vesselinov, 2012), the population which *Duolingo* and similar OLL software

target often does not proceed across all the levels from finish to start, and within a determinate timeframe or with regularity. Instead, users often start levels but do not finish them, complete a few levels and then return to the program days, weeks, or months later, or never repeat levels more than once. Therefore, a revisitation of the design seems necessary to help even the more casual users, which make up a vast part of the community of online language learners, to achieve some improvement. Indeed, the participants in this study who tended not repeat lessons, and were not forced to do so by design, were not able to improve in the four week timeframe.

A third implication relates to research on the effectiveness of software for pronunciation. The importance of using ratings for the evaluation of improvement suggested in the literature (Derwing & Munro, 1995a) is corroborated here. It is only through the ratings of the native speakers that it is possible to claim that significant improvement has been made - in this case by Luciana - as the overarching goal of pronunciation training is to improve intelligibility, or the interplay of comprehensibility and accentedness. Further corroborated by this study, is the idea that "although strength of foreign accent is correlated with perceived comprehensibility and intelligibility, a strong foreign accent does not necessarily reduce the comprehensibility or intelligibility of L2 speech (Munro & Derwing, 1995b, p. 73). Therefore, while certain content and certain practices can be effective for improvements in accentedness, they must be also designed to support improvements towards intelligibility, even at the most basic levels.

However, the findings of this study also suggest that attention to acoustic analysis, in particular of segmental features, is still important. Although improvement in a given micro feature - gemination and voice onset time in this study - may not translate to improvement in comprehensibility, acoustic analysis can provide information of the features' functional load and provide information to test predicted hierarchies.

6.1.4 Limitations and Suggestions for Future Research

Some of the potential limitations of this study include the limited amount of pronunciation features investigated and of the levels of the program accessed by the participants.

Although an analysis of segmental features is clearly valuable, the suprasegmental features are really those that affect comprehensibility the most. While it is still possible to conclude that certain variables are positively related to improvement in micro features, the value of such improvement might be only marginal to that caused by control of more prosodic features instead, such as stress (Field, 2005) or intonation (Hillenbrand, 2003). Thus, a suggestion for the future is to also focus on prosodic features in research on the effectiveness of online language learning software for the acquisition of pronunciation skills in order to better understand the impact of OLL software practices on comprehensibility and provide insights into possible adjustments to design.

In terms of the number of lessons accessed, the benefits of a study which asks the learners to progress through more than four levels are numerous. Firstly, it would provide more data on the overall practice structure and determine with a higher confidence level if a positive relationships exists between said structure and improvement. Secondly, it would allow the researcher(s) to investigate more features, as more content will be supposedly be learnt, and test whether extended exposure to the program practice do in fact translate in improvement in intelligibility. Lastly, a study which includes more levels could provide data on whether the program successfully addresses individual variations. While true beginners might be relatively equal in their advancements, higher proficiency levels will start to show signs of variation, which might or might not be addressed by the program. Data on this aspect of the program could provide insights on how to adjust algorithms and design within the software, and to avoid the "'one size fits all' variety, designed to appeal to a mass market" mentioned by Derwing and Murray (2005, p. 391).

6.2   Conclusion

As technology is becoming more and more prominent in our lives, and has started to be heavily applied to language learning, the need for critical evaluations of the effectiveness of computer-assisted language-learning has emerged. While these evaluations are ongoing in the

private sector and in the CALL field, they are not always easy to conduct. This is due primarily to a disconnect between technology and language pedagogy, and because of the different pace at which technology and the theorization of acquisition proceed. This study investigated the effectiveness of one of these software, *Duolingo*, with the aim of deepening the relationship between SLA theory and OLL software. To do so, it collected information on practice, focused on proficiency rather than achievement, and included an investigation into user experience.

Thus, the first research question asked about the effectiveness of *Duolingo* for the acquisition of the pronunciation macro skills of accentedness and comprehensibility for Italian as a second language learners. This study seem to indicate that *Duolingo* may be ineffective in its initial levels for improvement towards accentedness and comprehensibility, although the program might have a positive impact on the upkeep of those features that most enhance intelligibility. However, more research needs to be conducted in this area.

The second research question asked about the effectiveness of *Duolingo* for the acquisition of the pronunciation micro skills of geminate contrast in word-medial and voice onset time (VOT) for voiceless stops in word-initial position for Italian as a second language learners. The findings indicate that the program does have a positive impact on the acquisition or refinement of these features, even though previous SLA training, amount of practice, and language background play a very important role in shaping the learner's improvement.

The third research question asked about the learners' perceptions (user experience) of the effectiveness of *Duolingo* for the acquisition of pronunciation skills. Although at first unsettled by the lack of a clear declarative stage, in which information *about* pronunciation is presented explicitly, most participants perceived *Duolingo* to be effective for the acquisition of pronunciation skills, especially thanks to its rich procedural stage and its gamified design.

Overall, although this study corroborates the current views on online language learning software - that they concentrate on accent reduction rather than comprehensibility - it also suggests that the underpinnings of these software (i.e. focus on practice) are and can be

pedagogically sound. With more empirical research providing concrete suggestion for the calibration of said practices, the future of OLL software could look bright. Further, this study adds to the continued effort to bridge the gap between technology and language pedagogy, and also between the private and academic sectors. While many educational programs developed within universities and institutes might have solid pedagogical foundations, they often lack the appeal of commercial software (i.e. gamification, instant access, interactive materials, etc). In turn, commercial software often lack insights into valuable theorized practices, which have been proven to enhance the learning experience and overall success rate of language learners. The benefits of coordinating these two sectors would far outweigh the potential complications, logistic and theoretical, of such an operation, and would therefore be an important step forward.

Overall, I have aimed in this study to establish that OLL software can become effective and pedagogically sound tools for language learning, as long as more research is conducted on their practices and content, and on the experience of those who use them. This study demonstrates that CALL should further its efforts in understanding and evaluating said programs through any tool at the field's disposal, linguistic analysis and qualitative methods included. Further, it contributes to the valuable research that aims to forward the exploration of ways to connect theorized acquisition to the possibilities of technology, with the hope of designing sounder tools for the future.

REFERENCES

Abercrombie, D. (1949). Teaching pronunciation. *English Language Teaching, 3,* 113-122.

Agostiniani, L. (1992). Su alcuni aspetti del 'rafforzamento sintattico' in Toscana e sulla loro importanza per la qualificazione del fenomeno in generale. *Quaderni del Dipartimento di Linguistica dell' Universita degli Studi*, 3, 1-28.

Alves, M. A. (2011). Production of English and Portuguese Voiceless Stops by Brazilian EFL Speaker. *Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Letras/Inglês e Literatura Correspondente.*

Alvord, S. (Ed). (2011). Proceedings from LARP5 2011: *The 5th Conference on Laboratory Approaches to Romance Phonology*. Somerville, MA: Cascadilla Proceedings Project.

Apple Classrooms of Tomorrow. (2008). *Apple Classrooms of Tomorrow - Today: Learning in the 21st Century.* Cupertino, CA: Apple Inc.

Arvaniti, A. (1999). Cypriot Greek. *Journal of the International Phonetic Association*, *29*(2), 173-178.

Baran-Łucarz, M. (2014). The Link between Pronunciation Anxiety and Willingness to Communicate in the Foreign-Language Classroom: The Polish EFL Context. *Canadian Modern Language Review, 70*(4), 445-473. doi: http://dx.doi.org/10.3138/cmlr.2666

Baart, J. (2010). A field manual of acoustic phonetics. Dallas, TX: SIL International.

Beatty, K. (2003). *Teaching and Researching Computer-assisted Language Learning.* London, England: Longman.

Blake, R. J. (2011a). Current Trends in Online Language Learning. *Annual Review of Applied Linguistics, 31*, 19-35. doi: 10.1017/S026719051100002X

Blake, R. J. (2011b). Second Language Instruction in different settings. *Annual Review of Applied Linguistics, 31*, 19-35.

Boula de Maerüll, P., Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica, International Journal of Phonetic Science, 63(*4), 247-267.

Brennan, E. M. & Brennan, J. S. (1981). Accent Scaling and Language Attitudes: Reactions to Mexican American English Speech. *Language and Speech, 24*(3), 207-221. Doi: 10.1177/002383098102400301

Buendgens-Kosten, J. (2013). Authenticity in CALL: three domains of 'realness'. *ReCALL, 25*(2). 272-285. doi: 10.1017/S0958344013000037

Burani, C., Paizi, D., Sulpizio, S. (2014). Stress assignment in reading Italian: Friendship outweighs dominance. *Memory & Cognition, 42,* 662-675.

Burgos, P., Cucchiarini, C., van Hout, R., & Strik, H. (2014). Phonology acquisition in Spanish learners of Dutch: error patterns in pronunciation. *Language Sciences*, 41, 129-142.

Bush, M. D. (2008). Computer-Assisted Language Learning: From Vision to Reality? *CALICO Journal, 25*(3), 443-470.

Caramazza, A.; Yeni-Komshian, G. (1974). Bilingual switching: The phonological level. *Canadian Journal of Psychology, 28*(3), 310-318. Doi: org/10.1037/h0081997

Carranza, M., Cucchiarini, C., Burgos, P., Strik, H. (Ed). (2011). Proceedings from EDULEARN14: *6th Annual International Conference on Education and New Learning Technologies*. Barcelona, Spain.

Celata, C., & Costamagna, L. (2011). Timing delle geminate nell'italiano L2 di apprendenti estoni delle geminate nell'italiano L2 di apprendenti estoni. *Quaderni del laboratorio di*

linguistica, 10(1), 1-27.

Chapelle, C. A. (2009). The Relationship Between Second Language Acquisition Theory and

    Computer-Assisted Language Learning. *The Modern Language Journal*, 93, Focus Issue,

    741–753.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18

    languages. *Journal of Phonetics, 27*, 207-229.

Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch:

    Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics,*

    33(2), 121-157. doi:10.1016/j.wocn.2005.01.001.

Chun, D. M. (2006). CALL technologies for L2 reading. *Calling on CALL: From theory and*

    research to new directions in foreign language teaching, *5*, 7-98.

Chun, D. M. (2012). Replication studies in CALL research. *CALICO Journal*, *29*(4), 591-600.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning &*

    Technology, *11*(3), 38-63.

Cohen, J. (*1988*). *Statistical Power Analysis for the Behavioral Sciences, 2nd ed.* Hillsdale, New

    Jersey: Erlbaum.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education.* London, UK:

    Routledge, 2011.

Collentine, J. & Collentine, K. (2015). Input and output grammar instruction in tutorial CALL

    with a complex grammatical structure. *CALICO Journal32*(2), 273-298.

Colpaert, J. (2006). Pedagogy-driven Design for Online Language Teaching and Learning.

    CALICO Journal, *23*(3), 477-497. Retrieved from http://www.jstor.org/stable/24156348

Couper, G. (2008). What makes pronunciation teaching work? Testing for the effect of two

    variables: socially constructed metalanguage and critical listening. *Language Awareness,*

    20(3), 159-182. doi: 10.1080/09658416.2011.570347

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods*

approaches. Thousand Oaks, CA: Sage Publications.

Davis, S., & Ragheb, M. (2014). Geminate representation in Arabic. In Farwaneh, S., & Ouali,

H. (Eds.), *Perspective on Arabic Linguistics XXIVXXV: Papers from the annual symposia on Arabic Linguistics, Texas, 2012 and Arizona, 2011.* Amsterdam, NL: John Benjamins

Publishing Company.

De Clercq, B., Simon, E., & Crocco, C. (2014). Rosa versus rossa: The acquisition of Italian

geminates by native speakers of Dutch. *Phrasis (Gent): Studies in Language and Literature, 2*(2), 3-29.

DeKeyser. R. (2010). Practice for Second Language Learning: Don't Throw out the Baby with

the Bathwater. *International Journal of English Studies, 10*(1), 155-165.

Denzin, N. K., and Lincoln, Y., S. (2005). *The Sage Handbook of Qualitative Research*.

Thousand Oaks: Sage Publications.

Derwing, T. M. (2003). What Do ESL Students Say About Their Accents? *Canadian Modern*

Language Review, 59(4), 547–567.

Derwing, T. M. and Murray, M. J. (2005). Second Language Accent and Pronunciation Teaching:

A Research-Based Approach. *TESOL Quarterly, 39*(3), 379-397.

Derwing, T. M. & Murray, M. J. (2009). Comprehensibility as a Factor in Listener Interaction

Preferences: Implications for the Workplace. *Canadian Modern Language Review, 66*(2),

181-202.

Dmitrieva, O. (2012). *Geminate typology and the perception of consonant duration* (Unpublished

Doctoral dissertation). Stanford University, California.

Duolingo. (2012, December). *Duolingo Effectiveness Study - Final Report.* New York and South

Carolina: Roumen Vesselinov & John Grego.

Eckerth, J. (2008). *Case Study Research in Applied Linguistics* - by Patricia Duff. International

Journal of Applied Linguistics, *18*, 302–306. doi: 10.1111/j.1473-4192.2008.00204.x

Eisenstein, M. (1983). Native reactions to non-native speech: A review of empirical research. *Studies in Second Language Acquisition*, *5*(02), 160-176.

Ellis, N. (1999). Cognitive approaches to SLA. *Annual Review of Applied Linguistics*, *19*, 22-42.

Ellis, N. (2006). Cognitive perspectives on SLA: The associative-cognitive CREED. *Aila Review*, *19*(1), 100-121.

Ericsdotter, C. (1998). Effects of stress and speaking rate on the Italian geminate contrast: durational measurements. *Conference paper at Fonetik 98*, Stockholm.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL quarterly*, *39*(3), 399-423.

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational Research: An Introduction* (7th ed.). Boston: Allyn & Bacon.

Galloway, V. B. (1980). Perception of the communicative efforts of American students of Spanish. *The Modern Language Journal, 64,* 428–433.

Gambari, A. I., Kutigi, U. A., Fagbemi, P. O. (2014). Effectiveness of Computer-Assisted Pronunciation Teaching and Verbal Ability on the Achievement of Senior Secondary School Students in Oral English. *Education & Learning Research Journal, 8,* 11-28.

Garrett, J. J. (2010). *Elements of User Experience: User-Centered Design for the Web and Beyond.* New York, NY: Pearson Education.

Garrett, N. (1991). Technology in the Service of Language Learning: Trends and Issues. *The Modern Language Journal, 75*(1), 74-101.

Garrett, N. (2009). Computer-Assisted Language Learning Trends and Issues Revisited: Integrating Innovation. *The Modern Language Journal*, 93, Focus Issue, 719–740.

Gee, J. P. (2007). Good video games and good learning.

Golonka, E. M., Bowles, A. R., Frank, V. M, Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their

effectiveness. *Computer Assisted Language Learning, 27*(1), 70-105. doi: 10.1080/09588221.2012.700315

Greenbank, P. (2003). The role of values in educational research: the case for reflexivity. *British Educational Research Journal, 29*(6), 791-801. doi: 10.1080/0141192032000137303

Grgurović, M., & Hegelheimer, V. (2007). Help options and multimedia listening: Students' use of subtitles and the transcript. *Language learning & technology*, *11*(1), 45-66.

Ham, W. H. (1997). A new approach to an old problem: Gemination and constraint reranking in West Germanic. *Journal of comparative Germanic linguistics*, *1*(3), 225-262.

Hansen, B. B. (2004). Production of Persian geminate stops: Effects of varying speaking rate. In *Proceedings of the 2003 Texas Linguistics Society Conference* (pp. 86-95). Somerville: Cascadilla Press.

Harlow, L. L., & Muyskens, J. A. (1994). Priorities for Intermediate-Level Language Instruction. *The Modern Language Journal*, *78*(2), 141-154.

Harris, K. (2010) Native English speakers' production of Italian /t/: The extent of phonetic learning in adult second language acquisition and the effect of native speaker input. *Leeds Working Papers, 15,* 40-73.

Hillenbrand, J. M. (2003). Some effects of intonation contour on sentence intelligibility. *The Journal of the Acoustical Society of America*, *114*(4), 2338-2338.

Hismanoglu, M. (2006). Current Perspectives on Pronunciation Learning and Teaching. *Journal of Language and Linguistic Studies, 2*(1), 101-110.

Hubbard, P. (2006). Evaluating CALL Software. In L. Ducate & N. Arnold (Eds.). Calling on CALL: From Theory and Research to New Directions in Foreign Language Teaching. San Marcos: CALICO. Pre-publication copy.

Hubbard, P., Bradin, C. (2004). Another look at tutorial CALL. *ReCALL, 16*(2), 448-461.

Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford: Oxford University Press.

Jessen, M., & Ringen, C. (2002). Laryngeal features in German. *Phonology, 19,* 189-218.

Kabak, B., Reckziegel, T., & Braun, B. (2011). Timing of second language singletons and geminates. In *ICPhS XVII* (pp. 994-997).

Karapanos, E. (2013). *Modeling Users' Experiences with Interactive Systems*. New York, NY: Springer Berlin Heidelberg.

Kawulich, B. B. (2005). Participant Observation as a Data Collection Method. *Qualitative Social Research, 6(*2). Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/466/996

Kern, R., Ware, P., Warschauer, M. (2008). Network-Based Language Teaching. In *Encyclopedia of Language and Education, 2nd Edition*. (Vol. 4, pp. 281-292). Springer Science+Business Media LLC.

Ketabi, S., & Saeb, F. (2015). Pronunciation Teaching: Past and Present. *International Journal Of Applied Linguistics And English Literature, 4*(5), 182-189. Retrieved from http://journals.aiac.org.au/index.php/IJALEL/article/view/1581/1521

Kissling, E. M. (2013). Teaching Pronunciation: Is Explicit Phonetics Instruction Beneficial for FL Learners? *The Modern Language Journal, 97*(3), 720–744.

Klatt, D. H. (1975). Voice Onset Time, Frication, and Aspiration in Word-Initial Consonant Clusters. *Journal of Speech, Language, and Hearing Research*, *18*, 686-706. doi: 10.1044/jshr.1804.686

Kopczynski, A. & Meliani, R. (1993). The Consonants of Arabic and English. *Papers and Studies in Constrastive Linguists, 27,* 193-203.

Kuniavsky, M. (2003). *Observing the User Experience: A Practitioner's Guide to User Research.* Burlington, MA: Morgan Kaufmann.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, OX, UK: Blackwell Publishers.

Lai, C. & Gu, M. (2011). Self-regulated out-of-class language learning with technology.

Computer Assisted Language Learning, 24(4), 317-335. doi:
10.1080/09588221.2011.568417

Language Testing International (2016). Understanding Proficiency with LTI. In *Understanding Proficiency.* Retrieved from http://www.languagetesting.com/understanding-proficiency

Lee, W. & Lee, E. (Eds). (2011). Proceedings from ICPhS XVII: The *17th International Congress of Phonetic Sciences.* Cambridge: Cambridge University Press.

Lee, S., Noh, H., Lee, J., Lee, K., Lee, G. G., Sagong, S., & Kim, M. (2011). On the effectiveness of Robot-Assisted Language Learning. *ReCALL, 23*(1). 25-58. doi:
10.1017/S0958344010000273

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*(3), 369-377.

Levis, J. M., Sonsaat, S., Link, S., & Barriuso, T. A. (2016). Native and nonnative teachers of L2 pronunciation: Effects on learner performance. *TESOL Quarterly*.

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

Liakin, D., Cardoso, W., Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal, 32*(1), 1-25.

Lightbown, P. M. (2000). Classroom SLA research and second language teaching. *Applied Linguistics, 21*(4), 431-462.

Lisker, L., Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word, 20*(3), 384-422.

Lisker, L., & Abramson, A. S. (1973). Physiological aspects of certain laryngeal features in stop production. *The Journal of the Acoustical Society of America*, *53*(1), 294-295.

Liu, M. (2006). Anxiety in Chinese EFL students at different proficiency levels. *System, 34*(3), 301-316.

Liu, P., Chen, C., Chang, Y. (2010). Effects of a computer-assisted concept mapping learning

strategy on EFL college students' English reading comprehension. *Computers & Education*, *54*(3), 436-445.

Local, J., & Simpson, A. P. (1999). Phonetic implementation of geminates in Malayalam nouns. *work*, *4*(92), 46, 595-598.

Lord, G. (2015). "I don't know how to use words in Spanish": *Rosetta Stone* and Learner Proficiency Outcomes. *The Modern Language Journal*, *99*(2), 401–405.

Louali, N., & Maddieson, I. (1999). Phonological contrast and phonetic realization: the case of Berber stops. *ICPhS99 San Francisco*, 603-606.

Mah, J., & Archibald, J. (2003). Acquisition of L2 length contrasts. In *Proc. 6th Generative Approaches to Second Language Acquisition Conference* (pp. 208-212).

Mello, H., Pettorino, M., & Raso, T. (Eds). (2012). Proceedings from GSCP 2012: *Speech and Corpora.* Firenze, Italy: Firenze University Press.

Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. San Francisco: Jossey-Bass.

Merriam, S. B., & Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass Publishers.

Mitleb, F. (2009). Voice onset time of Jordanian Arabic stops. *The 3rd International Conference on Arabic Language Processing (CITALA '09).* Rabat, Morocco, 133-135.

Moghaddam, M., Nasiri, M., Zarea, A., & Sepehrinia, S. (2012). Teaching Pronunciation: The Lost Ring of the Chain. *Journal of Language Teaching and Research, 3*, 215-219. Doi: http://dx.doi.org/10.4304/jltr.3.1.215-219

Munday, P. (2016). The case for using DUOLINGO as part of the language classroom experience. *RIED Revista Iberoamericana de Educación a Distancia, 19* (1), 83-101. Doi: http://dx.doi.org/10.5944/ried.19.1.14581

Munro, M. J. and Derwing, T. M. (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing 11*(3), 253-266.

Munro, M. J. and Derwing, T. M. (1995a). Processing Time, Accent, and Comprehensibility in

    the Perception of Native and Foreign-Accented Speech. *Language and Speech, 38*(3),

    289-306.

Munro, M. J. and Derwing, T. M. (1995b). Foreign Accent, Comprehensibility, and Intelligibility

    in the Speech of Second Language Learners. *Language Learning, 45*, 73–97. Doi:

    10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech.

    Studies in Second Language Acquisition, 28(1), 111-131.

    doi:10.1017/S0272263106060049

Nagy, N., & Kochetov, A. (2013). Voice onset time across the generations: a cross-linguistic

    study of contact-induced change. In Siemund, P. et al. (Eds.), *Multilingualism and

    Language Diversity in Urban Areas: Acquisition, identities, space, education* (19-38).

    Amsterdam, NL: John Benjamins Publishing Company.

O'Brien, A. & Hegelheimer, V. (2007). Integrating CALL into the classroom: the role of

    podcasting in an ESL listening strategies course. *ReCALL, 19*(2), 162-180. doi:

    10.1017/S0958344007000523

O'Grady, W., Lee, M., & Kwak, H. Y. (2009). Emergentism and second language acquisition.

    The new handbook of second language acquisition, 69-88.

Payne, E. (2005). Phonetic variation in Italian consonant gemination. *Journal of the International

    Phonetic Association, 35*(2), 153-181.

Perceptual Works. (2015). Details. In *WebCAPE Language Placement Exams.* Retrieved from

    http://www.perpetualworks.com/webcape/details

Peters, M., Weinberg, A., & Sarma, N. (2009). To like or not to like! Student perceptions of

    technological activities for learning French as a second language at five Canadian

    universities. *Canadian Modern Language Review*, *65*(5), 869-896.

Peterson, M. (2010). The Use of Computerized Games and Simulations in Computer-Assisted

Language Learning: A Meta-Analysis of Research. *Simulation Gaming, 41*(1), 72-93. doi: *10.1177/1046878109355684*

Peterson, M. (2013). Computer Games and Language Learning: Theoretical Rationales. In Bulfin, S., Johnson, N. F., & Begum, C. (Eds.), Critical Perspectives on Technology and Education (51-60). New York, NY: Palgrave Macmillan US.

Pickett, E. R., Blumstein, S. E. and Burton, M. W. (1999). Effects of speaking rate on the singleton/geminate consonant contrast in Italian. *Phonetica,* 56, 135-157.

Protalinski, E. (2015). T100M users strong, Duolingo raises $45M led by Google at a $470M valuation to grow language-learning platform. Retrieved from http://venturebeat.com/2015/06/10/100m-users-strong-duolingo-raises-45m-led-by-google-at-a-470m-valuation-to-grow-language-learning-platform/

Reeder, J. T. (1998). English Speakers' Acquisition of Voiceless Stops and Trills in L2 Spanish. *Texas Papers in Foreign Language Education*, *3*(3), 101-18.

Rivas, A. T. (2000). Enigmas of Cuban Spanish. *Newsletter of the National Association of Judiciary Interpreters and Translators,* Vol. IX, No. 3.

Rocca, P. D. A., & Marcelino, M. (1999). Some characteristics of VOT in plosives produced by speakers of English and Portuguese. *ICPhS99*. San Francisco, CA. 1425-1428.

Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: an instructed SLA approach to L2 phonology. *Language Awareness, 20*(1), 45-59.

Saito, K., Webb, S., Trofimovich, P., Isaacs, T. (2015). Lexical correlates of comprehensibility versus accentedness in second language speech. Bilingualism: Language and Cognition, 1-13. doi: 10.1017/ S1366728915000255

Sargeant, J. (2012). Qualitative Research Part II: Participants, Analysis, and Quality Assurance. *The Journal of Graduate Medical Education, 4*(1), 1–3. Doi: 10.4300/JGME-D-11-00307.1

Scarpace, D. (2014). Proceedings from the International Symposium on the Acquisition of

  Second Language Speech: Montreal, Quebec: COPAL.

Schmuck, Richard (1997). *Practical action research for change*. Arlington Heights, IL:

  IRI/Skylight Training and Publishing.

Seidlhofer, B. (2011). <u>*Understanding English as a Lingua Franca*</u>. Oxford: Oxford University

  Press.

Shiri, L., Boaz, K. (2010). Why don't we believe non-native speakers? The influence of accent on

  credibility. *Journal of Experimental Social Psychology 46*, 1093–1096.

Simon, M. K. (2011). Role-of-the-researcher.

Son, J. (2007). Learner experiences in web-based language learning. *Computer Assisted

  Language Learning, 20*(1), 21-36. doi: 10.1080/09588220601118495

Stake, R. E. (2005). Qualitative case studies.

Stevens, M. (2011). Consonant length in Italian: Gemination, degemination and preaspiration. In

  The 5th Conference on Laboratory Approaches to Romance Phonology (pp. 21-32).

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research* (Vol. 15). Newbury Park, CA:

  Sage.

Taatgen, N. A., Huss, D., Dickison, D., & Anderson, J. R. (2008). The acquisition of robust and

  flexible cognitive skills. *Journal of Experimental Psychology: General 137*(3), 548-565.

  doi: 10.1037/0096-3445.137.3.548.

Tarone, E. (2015). Online Foreign Language Education: What Are the Proficiency Outcomes?

  The Modern Language Journal, *99*(2), 392-393.

Thomas, E. R. (2011). *Sociophonetics: An introduction*. New York, NY: Palgrave Macmillan.

Thomson, R. (2013). Computer Assisted Pronunciation Training: Targeting Second Language

  Vowel Perception Improves Pronunciation. *CALICO Journal, 28*(3), 744-765. Retrieved

  4 October, 2015, from

  http://www.equinoxpub.com/journals/index.php/CALICO/article/view/22985

Timmis, I. (2002). Native-speaker norms and International English: a classroom view. *ELT Journal*, *56*(3), 240-249.

Tozcu, A. & Coady, J. (2004). Successful Learning of Frequent Vocabulary through CALL also Benefits Reading Comprehension and Speed. *Computer Assisted Language Learning, 17*(5), 473-495. doi: 10.1080/0958822042000319674

Tranel, B. (1987). *The Sounds of French: An Introduction*. Cambridge, New York: Cambridge University Press.

VanPatten, B. & Williams, J. (2014). *Theories in Second Language Acquisition: An Introduction* (2nd ed.). New York, NY: Routledge.

Warschauer, M., & Grimes, D. (2007). Audience, authorship, and artifact: The emergent semiotics of Web 2.0. *Annual Review of Applied Linguistics*, *27*, 1-23.

Wilson, J. T. S. (2006). *Anxiety in Learning English as a Foreign Language: Its Associations with Student Variables, with Overall Proficiency, and with Performance on an Oral Test: Tesis Doctoral*. Editorial de la Universidad de Granada.

Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, and interpretation*. Thousand Oaks, Calif: Sage Publications.

Woodrow, L. (2006). Anxiety and Speaking English as a Second Language. *RELC Journal, 37*(3), 308-328. doi: 10.1177/0033688206071315

Yamazaki, K. (2014). Toward Integrative CALL: A Progressive Outlook On the History, Trends, and Issues of CALL. *The Tapestry Journal, 6*(1).

Yeni-Komshian, G. H., Caramaza, A., & Preston, M. S. (1977). A study of voicing in Lebanese Arabic. *Journal of Phonetics, 5*(1), 35-48.

Yin, R. K. (2003). *Case study research: Design and methods*. Thousand Oaks, Calif: Sage Publications.

Yip, F. W. M, & Kwan, A. C. M. (2006). Online vocabulary games as a tool for teaching and learning English vocabulary. *Educational Media International, 43*(3), 233-249. Doi:

10.1080/09523980600641445

Young, S. S. C., & Wang, Y. (2014). The Game Embedded CALL System to Facilitate English

Vocabulary Acquisition and Pronunciation. *Journal of Educational Technology &*

Society, 17(3), 239-251.

Zielinski, B. (2006). The intelligibility cocktail: An interaction between speaker and listener

ingredients. *Prospect*, *21*, 22-45.

Zhang, H., Song, W. & Burston, J. (2011). Reexamining the Effectiveness of Vocabulary

Learning via Mobile Phones. *The Turkish Online Journal of Educational Technology*

10(3), 203-214. Retrieved from

http://search.proquest.com.argo.library.okstate.edu/docview/964176435?accountid=4117

Zlatin, M. A. (1974). Voicing contrast: Perceptual and productive voice onset time characteristics

of adults. *The Journal of the Acoustical Society of America*, *56*(3), 981-994.

APPENDICES

Appendix A: Pretest and Posttest (with legend)

Please read the following words (22) twice each:
(3) Geminates (*Duolingo*): donna, nello, latte
(3) Geminates (non *Duolingo*): gonna, della, gatto
(3) VOT /p, t, k/ (*Duolingo*): pesce, cucino, torta
(3) VOT /p, t, k/ (non *Duolingo*): pera, carta, tela
(5) Distractors (*Duolingo*): **carota**, **cena**, siete, uomo, manzo
(5) Distractors (non *Duolingo*): divano, abete, **sole**, nuvola, **data**

Please read the following sentences (18) twice each:

Duolingo	Non *Duolingo*	Distractors
Voi mangiate il pollo	Io mangio il pasto	Tu bevi l'acqua
Loro sono donne	Lui ha il testo	E' un libro
Io mangio un biscotto	L'uomo ha il codice	Lei non è alta
Io mangio il pane	Io mangio il tonno	Lui beve
Lei mangia la torta	Lui batte il libro	Un uovo per favore
Il cuoco ha la mela	Noi siamo belli	
	La carota è nel piatto	

Please read the following passage once:
Al ristorante, il **locale** è pieno. Il cameriere, un uomo, è alto. Porta al **tavolo** il **pollo**, il **pesce**, le **carote**, il pane ed i **piatti**. Io ordino il succo, voi bevete acqua. La carne è buona. Alla fine, mangiamo un dolce al cioccolato.

149

Appendix B: Pre-interview

Demographics
1. Gender
2. Age
3. Level of Education
4. Place of birth

Language Background
1. Do you speak any Italian?
2. Do you have any experiences learning Italian?
3. Do you have any travel experience(s) to Italy?
     a. If yes, please briefly describe your experience(s).
4. Do you speak Spanish?
     a. If yes, how would you rate your writing/reading/speaking/listening/pronunciation in Spanish from 1-9, with 1 being the lowest score and 9 being the highest?
     b. How did you learn it?
5. Do you speak any other language? If yes, how did you learn it?
6. Are you in the process of studying any other language? If yes, how are you learning it?

Motivation and expectations
1. What would you say is your motivation to join this study?
2. What do you expect from this study?

Perceptions on language learning
1. How do you feel people best learn a language?
     a. How do you feel you best learn a language?

Perceptions on technology for learning
1. In what ways do you use technology in your daily life?
2. Have you used technology for learning in general?
     a. If so, how?
3. What are your feelings towards the use of technology for learning?

Perceptions on technology for language learning
1. What do you think about technology for language learning (ex. Rosetta Stone)?
2. Have you ever used technology for language learning?
     a. Can you give an example?
     b. Was it successful?
3. Why do you think it was/wasn't successful?

Perceptions on *Duolingo*
1. Are you familiar with *Duolingo*?
2. Have you ever used it before?
     a. For what language did you use it?
     b. Did it help you improve?
     c. Why do you think it did/didn't?

Appendix C: Questionnaire

1. How much did you use *Duolingo* this week?

2. On a scale from 1 to 9, 1 being "I have improved a lot" and 9 being "I haven't improved at all", how much would you say you have improved this week?

1 I have improved a lot	2	3	4	5	6	7	8	9 I haven't improved at all

3. Which activity(ies) has been the most helpful to you this week?

4. Which activity(ies) has been the least helpful to you this week?

5. What things do you feel you are struggling the most with this week?

6. What things do you feel are easiest this week?

7. Do you have any other comments?

Appendix D: Post-interview

Motivation and expectations
1. At the beginning of the study you said that ___. Did this study meet your expectations?
    1.1. In which way did it meet them? /Why did it not meet them?

Perceptions on language learning
1. At the beginning of the study you said that ___. How do you feel *Duolingo* fit with your language learning preferences?

Perceptions on technology for learning
1. At the beginning of the study you said that ___. What do you think about technology for learning at the end of this study?
    1.1 Has this been a successful experience with technology for learning?
         1.1.1   Why yes/no?

Perceptions on technology for language learning
1. At the beginning of the study you said that ___. What do you think about technology for language learning at the end of this study?
    1.1 Has this been a successful experience with technology for language learning?
         1.1.1   Why yes/no?
         1.1.2   If yes, which part of the experience has been the most successful? Why?

Perceptions on *Duolingo*
1. At the beginning of the study you said that ___. What do you think about *Duolingo* at the end of this study?
2. Do you think you were successful at improving your Italian?
3. Which activity do you feel has improved your pronunciation skills the most?
4. Which activity do you feel has improved your pronunciation skills the least?
5. Which activity do you feel was irrelevant to your improvement in pronunciation?
6. Did you feel like the software was easy to use?
    6.1 If yes, which aspects were easy?
    6.2 If no, which aspects did you find hard to use?

Language information:
1. What thing do you feel you struggled the most with in terms of pronunciation (ex. a specific sound, stress, rhythm, etc)?
2. What thing do you feel you was the easiest in terms of pronunciation?

Appendix E: Rater Scales

You will listen to sentences that were read by a variety of speakers. The sentences might at times seem odd, as they have no context. Do not worry about this.

- Please rate the comprehensibility of the speaker on the first scale provided here. The scale is from 1 to 9, with "1" indicating that the speaker is extremely easy to understand and "9" indicating that the speaker is impossible to understand.

- Please rate the accentedness of the speaker on the second scale provided here. The scale is from 1 to 9, with "1" indicating that the speaker has no foreign accent and "9" indicating that the speaker has a very strong foreign accent. The first 2 items are given as practice to get you used to the task.

The first 2 items are given as practice to get you used to the task.

AUDIO FILE HERE

Comprehensibility

1 Extremely easy to understand	2	3	4	5	6	7	8	9 Impossible to understand

Accentedness

1 No foreign accent	2	3	4	5	6	7	8	9 Very strong foreign accent

VITA

Micol Martinelli

Candidate for the Degree of

Master of Arts

Thesis: EFFECTIVENESS OF ONLINE LANGUAGE SOFTWARE (DUOLINGO) ON ITALIAN PRONUNCIATION FEATURES: A CASE STUDY.

Major Field: English

Biographical:

      Education:

      Completed the requirements for the Master of Arts in English (Linguistics) at Oklahoma State University, Stillwater, Oklahoma in July, 2016.

      Completed the requirements for the Bachelor of Arts in Psychology and English at University of Hawaii at Manoa, Honolulu, Hawaii in 2009.

      Experience:

- ❖   GTA/GRA at Oklahoma State University, 2014-2016
- ❖   English Instructor, The Bridge Institute, Italy, 2012-2014
- ❖   Translator, The Bridge Institute and Freelance, 2010-2014

      Professional Memberships:

      Linguistics Society of America (LSA)