The impact of a one-session-phonetic training on the improvement of non-native speakers' pronunciation of English

Amaury Flávio Silva

Technology College of Jacareí (FATEC Jacareí) - São Paulo, Brazil
Rua Faria Lima, 155 – Jardim Santa Maria, Jacareí – SP, Brazil, Zip Code 12328-070
amaury.silva@fatec.sp.gov.br

Abstract

Due to the difficulties L2[1] learners face regarding pronunciation, we conducted an experiment to find out if the participants of a one-session phonetic training would present any sign of improvement in their speech a week after the session. In order to evaluate their improvement, it was checked if the interword phonetic phenomena resyllabification, blending and hiding could be found in the subjects' speech. Furthermore, intraword-level pronunciation was also investigated. The findings have shown that betterment related to the presence of resyllabification occurred to all the subjects, but improvement to the other phenomena studied happened heterogeneously.

1. Introduction

Until the end of the 21st century, there was a limited number of studies regarding pronunciation (Derwing and Munro, 2005). This negligence is attributed to the fact that pronunciation was considered an aspect of language learning that could be naturally acquired through the learning process. However, since 2005 this viewpoint has been changing inasmuch as several studies, conferences, and articles about L2 pronunciation started to arise (Thomas and Derwing, 2014).

Despite the fact the importance of L2 pronunciation has become more evident, there are still L2 students, teachers and researchers who consider pronunciation teaching as being unnecessary as they reckon it can be learnt through exposure.

We regard pronunciation instruction as an essential part of the L2 teaching process. Its essential character becomes more evident when L2 learners, in spite of being studying the L2 for many years, still struggle to correctly pronounce the L2-language sounds, especially the ones that are not part of their L1 inventory systems. Nonetheless, we do not believe that achieving native-like pronunciation is necessary in that one's pronunciation being intelligible enough not to cause misunderstandings or hamper the flow of communication is what should be expected.

Owing to our belief that pronunciation instruction should be part and parcel of L2 language learning, we decided to carry out a study that aims to check the benefits of a one-session-pronunciation training in the improvement of the pronunciation of a group of subjects, Brazilian learners of English as a foreign language.

With regard to this one-session training, we hypothesize that there may be some kind of improvement in the subjects' pronunciation, but more sessions will be necessary to address all the pronunciation problems they may have. Moreover, the less proficient the students are, the higher will be the number of sessions necessary to help them deal with their pronunciation problems.

The dataset used during the training session was based on a study developed by Silva (2021) in which he studied examples of coarticulatory effects that we also incorporated in our pronunciation instruction session.

2. Goal of the paper

This paper, whose goal is to investigate the efficacy of a one-session phonetic training to enhance the participants' performance in pronunciation tasks also aims to provide a guideline that L2 teachers could use to assist their students improve. Furthermore, we hope that researchers could use the methods here applied to carry out new experiments in this area.

3. Theoretical Background

The increasing number of pronunciation-related studies since 2005 reveals the importance that pronunciation instruction has in the L2 learning process. Not only does it allow learners to become more confident when they speak, it also improves speech intelligibility as it helps to avoid misunderstandings.

Due to the importance of pronunciation, Thomas and Derwing (2014) wrote an article in which they evaluated 75 L2 pronunciation studies, most of which affirm that there was some kind of improvement in the speakers' pronunciation due to the training they took. The authors point out that diverging results take place owing to a few factors such as 'learner individual differences, goals and foci of instruction, type and duration of instructional input and assessment procedures' (p.1).

Most of the 75 studies focused on the achievement of native-like pronunciation by the learner and consisted in the use of computer-assisted tools. Moreover, the studies aimed at teaching the pronunciation of individual segments instead of teaching suprasegmental features, which would involve, for instance, resyllabification, prosodic boundaries, word stress, intonation, and speech rate.

In order to teach the pronunciation of segments, most of the time, the learners were engaged in activities that

[1] We use the term 'L2' to refer to the teaching of English as a foreign and as a second language.

required them to read texts aloud, instead of producing spontaneous speech.

When it comes to the quality of a pronunciation study, Thomas and Derwing (2014) mention a few features they should have. Firstly, they express their belief that pronunciation instruction should focus on 'helping students become more understandable' (p. 2). From this principle, they point out that an ideal pronunciation study should be able to give plenty of information on the subjects, have enough data that could be used to carry out statistical analyses, have a control group, and should not be limited to reading aloud tasks, i.e., it should also include spontaneous speech samples. Finally, it should include delayed assessment to verify the lasting effect of the pronunciation instruction.

With regard to qualitative analyses, they should encompass aspects such as motivation, type of interactions in the L2 and even social influences (Thomas and Derwing, 2014).

The training input of the studies surveyed, which was either classroom instruction or computer assisted pronunciation training, ranged from the manipulation of segments (Wang, 2002; Lee, 2009) to providing students with speech samples produced by native speakers so students could listen to them and compare them with their own productions (Gonzales-Bueno, 1997; Guilloteau, 1997, Weinberg and Knoerr, 2003; Lord, 2005; Pearson et al., 2011).

The learners' performances were evaluated by human listeners in 79 per cent of the studies and the other 21 per cent were evaluated using acoustic analyses.

The majority of the pronunciation training studies reviewed by Thomson and Derwing (2014) lacked explicit theoretical background so that the pronunciation training was solely based on the researchers' own experience. In our training, we considered the research about reduction phenomena led by Silva (2016, 2021), the findings on coarticulation conducted by Browman and Goldstein (1986, 1989), and the work developed by Vroomen and Gelder (1999) about resyllabification. We will be discussing this theoretical background later on in this section.

One important aspect that was not clear in the studies was the procedure taken during the training sessions (training input). The lack of clarity in the methodological procedures prevent other teachers and researchers from replicating the steps used in the studies in their own classes or research. Therefore, detailed methodological procedure is necessary 'for the benefit of other researchers and teachers' (Thomson and Derwing , 2014, p. 11).

The research on pronunciation training by Thomson and Derwing (2014) revealed that most of the participants showed some kind of improvement after the training. Nonetheless, the majority of studies only focused on the instruction of single sounds such as the contrast of /iː/ and /ɪ/. Should the studies be on several segmental and suprasegmental features, more time would be necessary so the learners could present significant improvement.

Another issue that questions the efficacy of the studies is whether or not the assessment used in them would reflect in the improvement of intelligibility when language is used in real-life contexts. For such issue to be solved, the studies should focus on 'more intelligible, as opposed to less-accented speech … (and) include a variety of assessment tasks' (Thomson and Derwing, 2014, p. 13-14). Furthermore, the authors state that evaluating the efficacy of the studies in a naturalistic fashion would take years, instead of weeks or months.

We believe that any research should depart from a well-established theoretical standpoint. Hence, since in our analyses we focused on the influence adjacent intra or interword segments have on one another, we turned to the studies developed by Browman and Goldstein (1986, 1989) on coarticulation.

According to Browman and Goldstein (1986, 1989), adjacent segments may be subjected to the phenomena called blending and hiding. Blending occurs when adjacent segments share the same articulator so that they cannot be produced without disturbance in their constriction location. An example of this phenomenon takes place when the segments [t] and [ð] from the context 'I want that' have to be produced one after the other. In this context, the constriction location of either segment may be disturbed as they are both characterized by a tongue tip gesture. Thus, the canonical production of the alveolar plosive and the interdental fricative may be realized as an approximant and as a dental fricative respectively.

Hiding occurs when adjacent segments do not share the same articulator so that the production of the first segment is overlapped by the production of the second one. Such phenomenon may occur when the segments [t] and [b] from the context 'I can't buy it' have to be produced one after the other. When this happens, the gesture of mouth closure to produce the bilabial consonant 'hides' the burst that would be caused by the release of the alveolar plosive.

Being aware of how these phenomena work allows speakers to reduce articulatory effort when they speak as the excursion of the articulators is decreased. The reduction in articulatory effort was studied by Silva (2016, 2021). In his investigations, he noticed that reduction is a strategy commonly used by native speakers and which can be characterized by the replacement of a segment that calls for high excursion of the articulators by one that does not (low-hierarchy reduction). Reduction can be also characterized by a segment deletion (high-hierarchy reduction).

Another phenomenon that causes reduction in articulatory effort is the one called resyllabification. It happens when 'consonants are attached to syllables other than those from which they originally came' (Vroomen and Gelder, 1999, p.413). An example of this phenomenon is the sentence 'you can evaluate this' in which the consonant /n/ of the word 'can' is coarticulated with the vowel /ɪ/ of the word 'evaluate.' This process contributes to maintain the speech flow as the speaker does not need to add a pause between adjacent words.

The analyses carried out in this study as well as the concepts explained during the training session were based on the phenomena blending and hiding (Browman and Goldstein, 1986-1989), reduction (Silva 2016, 2021), and resyllabification (Vroomen and Gelder, 1999).

4. Methods

In this section, we will describe details related to the subjects that participated in the study, the research dataset, the acoustic inspection and the training session.

4.1 Subjects

In order to conduct the analysis, we had the participation of four subjects, native speakers of Brazilian Portuguese (three males and one female), who study English as a foreign language. The subject 'English' is part of the Technological course the subjects were taking and all of them were enrolled in the same class, taking the third semester. It is important to point out that English is offered throughout the duration of the course, six semesters, and, despite the fact all the students were in the same class, their proficiency level was not the same.

The four participants will be referred to as subjects, 'S,' in this investigation.

4.2 Research dataset

The research dataset, table 1, is an extract from the program Actors' Studio (season 12, episode 13, released on July 2006) that was sent to the subjects so they would have to record and send it to the trainer before the training session. After the session, they would record it once more and send it to the trainer again so their improvement could be analyzed. We would like to point out that in our experiment, we asked the subjects to use their own smartphones or computers to record the dataset. This was done as they could not come to college to record it in its sound laboratory due to the restrictions related to the COVID-19 pandemic.

The same text was used in the pre and post-training phase as we aimed to analyze whether or not improvement could be observed in the second recording in terms of the group of words we selected that encompass the phenomena described in tables 2-4.

This dataset was also selected by Silva (2021) on his investigation about coarticulatory phenomena analysis.

It's funny, you know, someone comes into your life at a certain time and that's one of the great things that happens on Earth is you're mysteriously guided towards these people that you get to dance with, you know. And I thought "How great is that", he's kind of, like, I don't want to say an angel to her, but he's someone who needs as much as he's prepared to offer, and he has seen a lot of life, and he's not a typical lawyer-type.

Table 1: Research Dataset

Using the dataset above, we selected fragments in which the phenomena resyllabification, blending and hiding could take place. Furthermore, we also analyzed the pronunciation of a group of words that the students mispronounced in the pre-training recording.

The phenomenon resyllabification was investigated in 11 contexts, which we will present in the next table.

Contexts	Phonemes involved
'comes into'	/z/and /ɪ/
'at a certain'	/t̬/and /ə/
'one of'	/n/and /ə/
'on Earth'	/n/and /ɜr/
'and I'	/d/and /aɪ/
'great is'	/t̬/and /ɪ/
'kind of'	/d/and /ə/
'an angel'	/n/and /eɪ/
'as much as'	/tʃ/and /ə/
'seen a'	/n/and /ə/
'lot of life'	/t̬/and /ə/

Table 2: Resyllabification phenomenon

With regard to the phenomena blending and hiding, we analyzed eight contexts, presented in the next table.

Contexts	Phonemes involved
'certain time'	/n/and /t/
'great things'	/t/and /θ/
'guided towards'	/d/and /t/
'get to dance'	/t/and /t/
'prepared to'	/d/and /t/
'typical lawyer'	/l/and /l/
'these people'	/z/and /p/
'I don't want'	/t/and /w/

Table 3: Blending and hiding phenomena

Lastly, when it comes to word-level pronunciation, the words presented in the table below were investigated.

Words	Pronunciation errors found
'someone'	Phoneme substitution and insertion of a phoneme
'certain'	Phoneme substitution and word stress
'mysteriously'	Phoneme substitution and word stress
'towards'	Phoneme substitution
'thought'	Phoneme substitution
'offer'	Phoneme substitution and word stress
'lawyer'	Phoneme substitution and word stress

Table 4: Word-level pronunciation

4.3 Phonetic inspections

The phonetic inspection was carried out with the use of the free software PRAAT, version 6.0.39, developed by Paul Boersma and David Weenik (2018), from the Institute of Phonetic Science of the University of Amsterdam.

The inspections were based on the observation of the waveform, the broadband spectrogram, the fundamental frequency and the intensity of the phonemic segments.

4.4 Training Session

The training took place in a single 50-minute session of an online class. It was recorded so that the subjects could revisit it as many times as they wanted in order to review the concepts explained.

The training session the subjects participated was provided by the researcher of this work.

At the beginning of the training, which took place after the first recording of the dataset was sent by the subjects, the original recording of the dataset was played, and the corresponding script was projected on the screen for the subjects to follow it. The recording was played three times.

After that, the concept of resyllabification was explained and the first context where such phenomenon occurred according to table 2, 'comes into', was presented to the subjects (the orthography along with the recording). The context was played three times.

The subjects were asked to pay close attention to the recording of the context as they would have to repeat it afterwards. If they could not repeat it, the trainer would repeat the context himself at least three more times in order to assist the subjects grasp what and how they should say it.

Before moving on to the next context, the original recording was played one more time and the subjects were asked to repeat it. Not until all the subjects were able to repeat the context intelligibly, would the trainer teach the next context.

The procedure described above was followed to teach the other contexts including resyllabification, blending and hiding phenomena. Word-level pronunciation instruction followed the steps related to playing the recording three times before repetition. However, after analyzing the first recording, we reckoned the need to teach word stress and phoneme pronunciation.

It is important to point out that we did not use technical terms during the training as our focus was simply on improving their pronunciation.

When it comes to the difficulty the subjects presented to pronounce a word or group of words, the trainer noticed that it was necessary to teach the articulation of some phonemes, especially the ones not present in the subjects' L1inventory system. After the instruction of the articulation of such phonemes, improvements could be observed in their pronunciation.

The subjects, after the training session, had access to the original recording of the dataset and to a version recorded by the trainer, which was produced with a slower speech rate so that it could be helpful to less proficient subjects. These recordings were tools the subjects could use to improve their pronunciation before making the second recording that had to be sent within a week.

Once all the subjects had sent their recordings, we started the data analysis, whose results are presented in the next section.

5. Data analyses

The analyses in this chapter will feature figures that contain the waveform, spectrogram, segmentation, and spelling of a selection of the contexts investigated. However, at the end of section 5.1, a table with a summary of all the contexts investigated during the pre-training phase is provided and one at the end of section 5.2 with all the contexts analyzed in the post-training phase is available.

We reckon it is important to point out that the subjects reported that they recorded the dataset several times and that they sent us the version they judged to be the best.

5.1. Pre-training analyses

In this section, we will present the analyses that refer to the pre-training recordings. The first one refers to the context 'and I,' resyllabification.

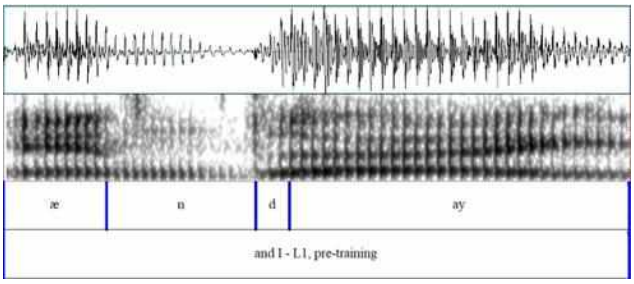

Figure 1: Production of 'and I' by L1, pre-training

Through the analysis of the broadband spectrogram and its corresponding waveform above, we can infer that there was no pause between the production of the adjacent segments [d] and [ay] so the phenomenon resyllabification was observed.

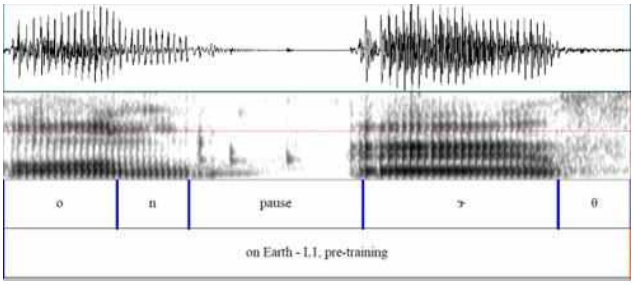

Figure 2: On Earth – S1, pre-training – Figure shows pause between the words 'on' and "Earth'

In the production of 'on Earth,' figure above, there was a pause between the segments [n] and [ɜr] so that the phenomenon resyllabification did not take place.

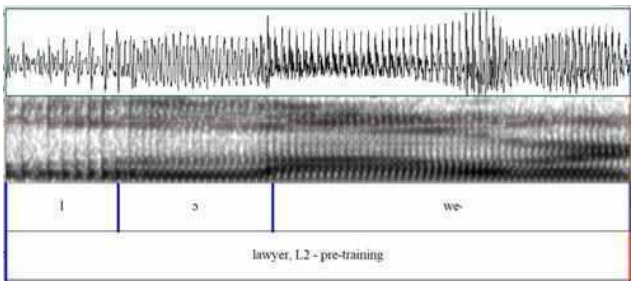

Figure 3: Production of 'lawyer' by S2

The figure above, which presents acoustic information, shows that the subject mispronounced the word lawyer in that [ləwər/] was produced instead of [lɔɪər/].

A summary of the data analyses that refer to the pre-training recordings is presented in the table below. The word or group of words written indicates that the phenomenon in the corresponding column was observed in their production.

Pre-training			
Subjects	**Resyllabification** (group of words in which the phenomenon was observed)	**Blending/Hiding** (group of words in which the phenomena were observed)	**Word-level pronunciation** (mispronounced words)
S1	'and I' 'great is' 'kind of' 'lot of life'	'guided towards' 'typical lawyer' 'these people' 'I don't want to'	All the words were mispronounced except 'someone' and 'offer'
S2	'at a certain' 'one of' 'on Earth' 'and I' 'kind of' 'lot of'	'great things' 'guided towards' 'get to dance' 'prepared to' 'typical lawyer' 'these people'	'mysteriously' 'thought' 'lawyer'
S3	'comes into' 'at a certain' 'one of' 'on Earth' 'and I' 'great is'	'certain time' 'get to dance' 'prepared to' 'these people'	'mysteriously' 'thought' 'lawyer'
S4	All the contexts except 'and I'	All the contexts except 'certain time'	'thought' 'offer' 'lawyer'

Table 5: Data analyses concerning the pre-training recordings

5.2. Post-training analysis

In this section, we will present the analyses that refer to the post-training recordings. The first one refers to the context 'on Earth,' resyllabification.

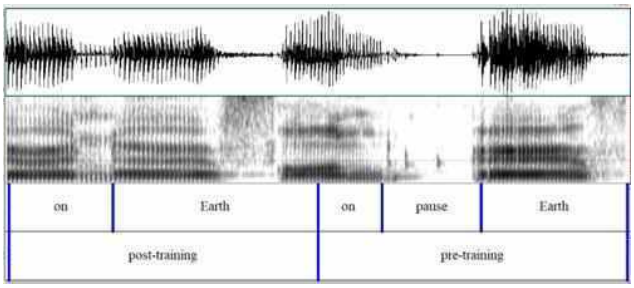

Figure 4: Concatenated productions of 'on Earth' by S1. Post-training left and pre-training right.

The concatenated productions of 'on Earth' by S1, presented in the figure above, demonstrate that the phenomenon resyllabification was observed in the post-training recording, but not in the pre-training recording. This fact is confirmed by the absence of pause between the segments [n] and [ɜr] in the post-training phase that did not occur in the pre-training phase as a pause is present in the spectrogram.

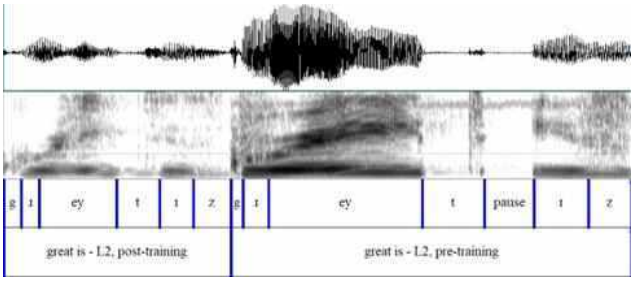

Figure 5: Concatenated productions of the post and pre-training versions of 'great is' by S2

As shown in the analysis of the context 'on Earth,' figure 4, in the context 'great is' by S2, figure above, the phenomenon resyllabification was observed in the post-training recording, but not in the pre-training one.

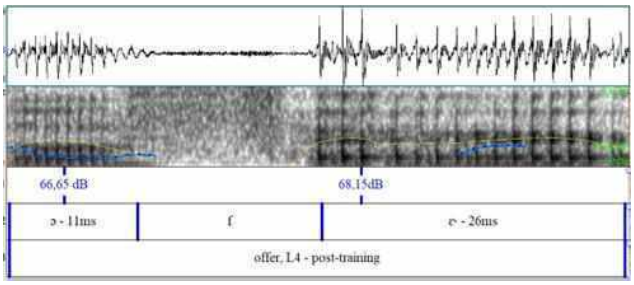

Figure 6: Production of the word 'offer' by S4

The analysis of the production of the context 'offer,' produced by S4, shows that the word stress was placed on the syllable '-fer' instead of the syllable 'of-', which is where the correct stress for the word 'offer' should occur. The word stress on the syllable '-fer' can be confirmed not only by the higher duration of the segment [ɛr], but also the higher intensity of this segment in comparison to the segment [ɔ]. What's more, S4 used the segment [ɛr] instead of /ər/ in the second syllable.

A summary of the data analyses that refer to the post-training recordings are presented in the table below. The word or group of words written indicates that the phenomenon in the corresponding column was observed in their production.

Post-training			
Subjects	**Resyllabification** (group of words in which the phenomenon was observed)	**Blending/Hiding** (group of words in which the phenomena were observed)	**Word-level pronunciation** (mispronounced words)
S1	All the contexts	'certain time' 'great things' 'get to dance' 'typical lawyer' 'these people' 'I don't want to'	All the words were mispronounced except 'someone' and 'offer'
S2	All the contexts	'certain time' 'great things' 'prepared to' 'these people'	'someone' 'mysteriously' 'thought' 'lawyer'
S3	'comes into' 'at a certain' 'one of' 'on Earth' 'and I' 'great is' 'kind of' 'a lot of life'	'certain time' 'great things' 'get to dance' 'prepared to'	'someone' 'mysteriously' 'thought'
S4	All the contexts except 'and I'	All the contexts	'thought' 'offer'

Table 6: Data analyses concerning the post-training recordings

6. Discussion

The analyses have shown that the one-session phonetic training was useful to help the subjects improve their pronunciation with regard to the resyllabification phenomenon. Nevertheless, no homogeneous improvement was observed in terms of the remaining phenomena analyzed.

The observed improvement in the resyllabification feature in the production of S1 and S2 was characterized by the use of this strategy in the production of all the contexts analyzed in the post-training recording, fact not observed in the pre-training one. S3 also demonstrated improvement in the use of this strategy in that it was used in two more contexts in the post-training recording. No improvement was observed in terms of resyllabification for S4, but they had already presented excellent performance of this strategy as there was only one context where it was not applied.

The presence of the phenomena blending, and hiding was found in the production of S1 in most of the contexts and in all the contexts produced by S4 in the post-training recording. Such phenomena were noticed in fewer contexts in the production of S2 and in the same number in the production of S3 in the post-training recording.

With regard to the last feature analyzed after the post-training session, word-level pronunciation, no improvements were observed in the production of S1, S2 made one more mistake and S3 improved the production of the word 'lawyer' but mispronounced a word he had produced correctly in the pre-training session, 'someone'. S4 improved the production of the word 'lawyer' but continued mispronouncing the words 'thought' and 'offer'.

Our findings have revealed different levels of improvement in the subjects' performance so that S1 is the one who presented the most improvement. S2 and S3's performances betterment was limited to the presence of the resyllabification phenomenon. S4 is the most proficient subject who presented only a few mistakes in the pre-training recording and was able to use the phenomena blending and hiding in all the contexts and to improve the pronunciation of a word after training.

The hypothesis we presented at the beginning of our work was confirmed as the subjects' pronunciation was somehow improved, but more sessions are necessary to address certain pronunciation problems such as word-level pronunciation and the phenomena hiding and blending.

In future studies, we could ask the subjects to report on the time they have dedicated to study and practice the pronunciation concepts studied during the training session. Furthermore, we could ask judges to evaluate the students' performance before and after the training session to find out if a perceptual betterment in their pronunciation was clear, i.e., if the level of intelligibility was enhanced.

We believe vehemently that, although the number of participants was not adequate through a quantitative standpoint as our aim was to conduct a qualitative investigation, the study has shown that improvement did occur, bringing to light the importance of phonetic instruction. Moreover, we expect that the procedure we used during the training session was clear enough so the study could be replicated by other researchers.

Lastly, we hope to continue our investigation by providing the subjects with more training sessions, evaluate them at least five months after the first training session and have more participants so we could carry out statistical analysis.

7. Reference

Paul Boersma, David Weenink. 2018. Praat: doing phonetics by computer, version 6.0.39. Available at: <http://www.praat.org>. Access on: 2 Dec. 2018

Catherine Browman, Louis Goldstein. 1986. Towards an articulatory phonology. Phonology, v. 3, pages. 219-252.

_____. Articulatory gestures as phonological units. Phonology, v. 6, 1989, pages 201-251.

Tracey M. Derwing, M, Murray J. Munro.2005. Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly* 16/1: pages 71-77.

Manuela Gonzales-Bueno. 1997. The effects of formal instruction on the acquisition of Spanish stop consonants. *Contemporary Perspectives on the Acquisition of Spanish* 2: pages 57-75.

Nancy Clarke Guilloteau. 1997. *Modification of phonetic categories in French as a second language: Experimental studies with conventional and computer-based intervention methods*. Unpublished Ph. D. thesis. University of Texas at Austin.

Lee Ji-Yeon. 2009. *The effects of pronunciation instruction using duration manipulation in the acquisition of English vowel sounds by pre-service Korean EFL teachers*. Unpublished Ph.D. thesis, University of Kansas.

Gillian Lord. 2005. (How) can we teach foreign language pronunciation? On the Effects of a Spanish Phonetics Course. *Hispania*, 88/3: pages 557-567.

Pamela Pearson, Lucy Pickering, Rachel DaSilva. 2011. The impact of computer assisted pronunciation training on the improvement of Vietnamese learner production of English syllable margins, In: Levis J. and LeVelle K. (eds). *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*, Iowa State University, pages 169-180.

Amaury F. Silva. 2021. Coarticulatory phenomena analysis in English based on the articulatory phonology. São Paulo. *CBTecLe* v.1, n.1.

_____.2016. *Percepção de reduções em inglês como L2*. Unpublished Ph.D. thesis, PUC-SP.

Ron I. Thomson, Tracey M. Derwing. 2014. The effectiveness of L2 pronunciation Instruction: a narrative review. Oxford, Oxford University Press.

Jean Vroomen, Beatrice De Gelder. 1999. Lexical access of resyllabified words: evidence from phoneme monitoring. Memory & cognition, 27(3), pages 413–421.

Xinchun Wang. 2002. *Training Mandarin and Cantonese speakers to identify English vowel contrasts: long term retention and elects on production*. Unpublished Ph.D. thesis, Simon Fraser University.

Alysse Weinberg, Hélène Knoerr. 2003. Learning French pronunciation: Audiocassettes or multimedia? *CALICO Journal*, 20/2: pages 315-336.