READING LIPS AND LEARNING SOUNDS: THE EFFECT OF VISUAL CUE SALIENCY ON PHONOLOGICAL PRODUCTION IN A SECOND-LANGUAGE

Brent Taylor Eisenbarth

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of the Arts in the Linguistics in the College of Arts and Sciences.

Chapel Hill
2018

Approved by:

Elliott Moreton

Jennifer L. Smith

Lucia Binotti

ABSTRACT

Brent Eisenbarth: Reading lips and learning sounds: The effect of visual cue saliency on phonological production in a second-language
(Under the direction of A. Elliott Moreton)

While L2 production is widely understood to show traces of L1 interference and general cue availability, the role of visual cue strength on L2 production is previously unstudied. As such, this experiment compares the production of 10 L2 Italian speakers' intervocalic /p/ - /pp/ and /k/ - /kk/ contrasts, to see if the more visible labial pair is produced more distinctly than its less visible velar counterpart. Participants read an adapted excerpt of Antoine de Saint-Exupéry's *The Little Prince* and a subsequent word list. After compensating for following vowel effects within a subset, intermediate and advanced participants were found to produce more distinct labial than velar pairs; an effect most evident in advanced participants.

Furthermore, this thesis discusses a potential asymmetry between the behavior of visible and non-visible gestures produced across different places of articulation, suggesting further research into the potential influence of varying visual cue strengths in L2 production.

To my mother, Kim Rose Eisenbarth.

ACKNOWLEDGEMENTS

There are many people to whom I hold gratitude, but foremost I want to express my gratitude towards my advisor Elliott Moreton. I deeply appreciate his support of the project and feel privileged to have participated in his course, LING 526, Second-Language Phonetics and Phonology, which led me to develop an interest in this research topic. I also send my thanks to Jen Smith and Lucia Binotti, the other two members of my committee, for their support and instruction. I am grateful to have held a role in both the Linguistics and Romance Studies departments and wish to express gratitude to the many other professors whom I consider influential in my linguistics educational experience, including but not limited to Bruno Estigarribia, Misha Becker, David Mora-Morin and Paul Roberge.

I also like to thank the members of the P-side group for their feedback on my project, and to my linguistics cohort for their advice and suggestions. I especially want to thank Katya Pertsova for suggesting that this experiment may warrant exploration with possible future perceptual studies. I also send thanks to the Odum Institute for assistance with statistical analysis, and to Michele Cammelli for meeting with me to look over the source text.

Lastly, I am grateful to my friends and family for supporting me as I worked through the thesis process. It would be incomplete to end a letter of thanks without acknowledging my parents, brothers, and close friends who have supported me throughout my educational career. The research of a university reflects the instructional and analytical efforts of the entire community. All errors within this project are my own, but all the fruits

of the project grow from the support and labor that other researchers, professors, and friends have invested in me and this project.

TABLE OF CONTENTS

LIST OF TABLES

1 Introduction

1.1 Study Goals and Implementation

This thesis explores the impact of visual input on L2 phonological learning. Particularly, this experiment tests whether L2 Italian speakers distinguish labial geminate and singleton voiceless stops more than their velar counterparts in production, with the goal of seeing whether the visibility of the articulators correlates to the production accuracy of articulatory gestures not present in the L1.

While visual cues' interaction with speech perception has been primarily studied in the context of immediate and short-term perceptual confusions, such as the McGurk effect, the data in a study on L2 Spanish spirantization by Menke and Face (2009) demonstrates an unexplained trend in which Spanish L2 speakers produce the labial spirant more accurately than the velar spirant. The effect appears stronger in speaker populations with more experience and fluency; for example, doctoral candidates of Spanish produce spirant /b/ 13.9% more accurately than spirant /g/.

The McGurk Effect suggests that the labial position is advantaged as more visually salient than other places of articulation. As such, if visual cues provide useful information to L2 learners, than the labial position's additional visual information may facilitate the learning of this production, so that within a natural class defined by manner of articulation, L2 learners would produce the gesture more accurately at the labial position than other places of articulation.

In this experiment, 2 graduate, 4 advanced and 4 intermediate learners of Italian read an adapted version of Antoine de Saint-Exupéry's *The Little Prince* in Italian and a word list of 25 words. All tokens are voiceless stops that pertain to one of four types: /p/, /pː/, /k/, or /kː/. 60 tokens, or 15 of each type, were collected from the short story task and 20 tokens, or 5 from each type, were collected from word list task to test whether labial singleton and geminate tokens are more distinct from each other than in the velar condition.

While overall the short story task and word-list task is ambiguous, when a subset controlled for the following vowels' backness is selected, the labial pair is significantly more distinct than the velar pair amongst intermediate and advanced speakers in the short story task. Literature on vowel backness and place of articulation suggests that the subset measurements are the most accurate.

This paper explores how the saliency of visual cues may influence L2 phonological production. Further study on this topic with similar L2 articulatory gestures realized in different places of articulation across speech may be useful for better understanding the mechanisms related to this phenomena. This research needs to be further conducted both with visible and non-visible phonetic patterns to what degree the features individual visibility may or may not condition this pattern; although, data from L1 English L2 Spanish speakers learning of Spanish VOT with voiceless tokens suggests that the difference in accuracy based on place of articulation found here and in the Menke and Face paper does not appear with non-visible cues (Reeder, 1998). More broadly, further study on visual saliency and possible impacts on L2 acquisition may provide further insight into the relationship between visual input and L2 phonological production, the role of input in L2 learning, and the role of visual cues in speech perception.

2 Theoretical Background

2.1 Introduction

Since the discovery of the McGurk effect, visual input has been shown to influence speech perception. This chapter establishes the theoretical framework behind this experiment with L2 Italian gemination, exploring the current knowledge on visual cues and their saliencies, the aforementioned Menke and Face (2009) study on Spanish spirantization, Italian gemination and the visibility of gemination.

2.2 McGurk Effect

In 1976, McGurk and McDonald discovered “a categorical change in auditory perception influenced by incongruent visual speech, resulting in a single percept of hearing as something other than the sound.” (Tiippana, 2014) Speakers process visual information alongside auditory information, so that when the information strongly conflicts speakers often perceive something distinct from both the original audio or visual input, be it a compromise of the two sounds in place of articulation, or an amalgamation of the segments.

In the original experiment, McGurk and McDonald (1976) presented 21 preschool, 28 elementary school, and 54 adult participants with speech tokens that consisted of audiovisual clips of a speaker producing labial and velar stops. A control group faced away from the screen hearing only the auditory portions of the tape, whereas the experimental group faced the screen and was exposed to mismatched audio and visual stimuli simultaneously. There were 100 disyllabic tokens, consisting of voiced and voiceless, labial and velar stops, and participants were asked to repeat back the sound they perceived after they heard and watched the clips.

The researchers grouped the responses of participants in the audiovisual experimental condition based on whether their perception concurred with the auditory stimuli, reflected the visual stimuli, concatenated the auditory and visual stimuli, or fused the auditory and visual stimuli into a distinct phoneme. Only a small number of responses, with no more than 6% per audiovisual token in the adult conditions, were classified as other. Below, the most common responses to the information are shown.

Table 1: McGurk Effect Responses

	Auditory Stimuli	Visual Stimuli	Majority (Plurality) Response	% of Adults
1	[ba.ba]	[ga.ga]	Fusion: [da.da]	98%
2	[ga.ga]	[ba.ba]	Combination: [bag.ba], [ga.ba], [ba.ga]	59%
3	[pa.pa]	[ka.ka]	Fusion: [ta.ta]	81%
4	[ka.ka]	[pa.pa]	Combination: [pak.pa], [ka.pa], [pa.ka]	44%

Adapted from McGurk and McDonald (1976)

Beyond clearly establishing that visual cues impact phonological processing, this data demonstrates that the visual saliency of the labial position. Tokens 2 and 4 consist of visual labial stimuli with velar auditory stimuli, whereas tokens 1 and 3 contain visual velar stimuli and auditory labial stimuli. The saliency of the visual cue in the visual labial stimuli (2 and 4) is strong enough that the majority of the time it overrides the audio velar cues to replace or insert a segment faithful to the visual cue's place of articulation. In contrast, the visual audio saliency of the visual velar labial stimuli is more equally matched, so that participants generally compromise when perceiving the place of articulation. A visual labial cue causes participants to override place of articulation information from their audio cues to insert or replace segments with labial place information, whereas a visual velar cue doesn't fully override the audio cues place of articulation and results in compromise. Visual cues from stops in the labial position are stronger than those in a velar position.

Token 4 (an audio velar, visual labial token) is the only token to have a plurality response rather than a majority response. In this case, it is notable that none of the adults chose a fused response (i.e. [ta.ta]) typical to the visual velar type tokens. Rather, the reason it didn't reach a majority is because more adults perceived solely by its auditory ([ka.ka]) or visual ([pa.pa]) stimuli in this condition that in the other conditions.

2.3 Visibility of Gemination

Visual cues are relevant beyond influencing the perception of place of articulation; particularly, speakers can see gemination. In 2017, Arai et al. tested the ability of Japanese native speakers to use visual input to distinguish singleton /s/ and /t/ from their geminate counterparts by observing their perception of different audiovisual cues. As our current project focuses on geminate stops, our focus lies with their findings on the singleton and geminate /t/, although both tokens yielded similar results. Beyond duration, formant transitions serve as an important cue in helping Japanese native speakers distinguish singleton and geminate tokens, to the extent that Japanese learners sometimes produce tokens, that while long, still sound like singletons to native speakers because of incorrect formant transitions (Yanagisawa and Arai, 2015a).

In this study, the researchers posited three types of audio cues and three types of visual cues. One set of audio cues had normative formant "transitions" (A_T) so they would sound like /atta/ and /assa/ to native speakers, whereas the other audio cues were flat (A_F); both visible in Figure 1. Without the distinctive transitions, A_F tokens are understood as /a/ and /ta/ spoken successively with a short break.

Figure 1: Arai et al (2017) Testing Tokens Spectrograph

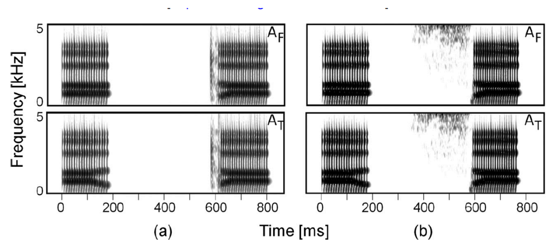


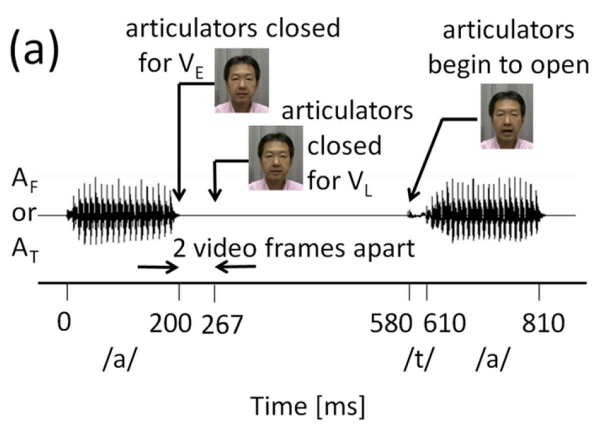
Fig. 1. Spectrograms of the A_F and A_T stimulus sounds. (a) /t/-experiment and (b) /s/-experiment.

From Arai et al (2017)

As such, the duration of audio content in the audio cues is the same, but what differs is whether the vowel formants contextualize the token as one geminate phoneme or as singleton preceding a pause. Lastly, in one trial of the experiment, the audio is absent from the audiovisual tokens (A_0).

The visual cues, similarly, either corresponded with geminate or singleton production. A video of a speaker was displayed, and if his mouth closed early, than this suggests a geminate consonant (V_E), whereas if he closed late, then this cues with a singleton consonant (V_L). Similarly, the visual cue could also be absent (V_0). Below is a diagram from Arai e al. 2017 of their testing stimuli.

Figure 2: Arai et al (2017) Token Design



From Arai et al (2017)

Each experiment consisted of three trials, and the t-experiment was administered before the s-experiment. In all trials, participants were asked to sit at an audiovisual monitor and headphone set, and select with a keystroke whether the token they perceived was a geminate or singleton. Each trial consisted of 20 tokens. In the first trial, only the audio conditions were played (A_TV_0 and A_FV_0). Then, in the second trial, the tokens mixed all the audio-visual combinations, so that the visual and audio cues could both suggest the same consonantal length or they could suggest contradictory length. As such, this trial included concurring tokens, A_TV_E (audio geminate, visual geminate) and A_FV_L (audio singleton, visual singleton), and contrasting tokens, A_FV_E (audio singleton, visual geminate) and A_TV_L (audio geminate, visual singleton). Lastly, a third trail consisted of the visual cues, A_0V_E, and A_0V_L.

Both the T- and S- experiments show that consonant length is a visible cue and the T- experiment additionally suggests that geminates are more visibly salient than singletons. In both experiments A_0 visual cases, participants can understand whether a token is singleton or geminate as accurately as they can in the cases in which the audio and visual information is consistent (A_TV_E, A_FV_L). This shows that consonantal length exists as a visible cue. Furthermore, while the mixed conditions pattern most accordingly to the audio condition, so that audio-geminate visual-singleton A_TV_L tokens patterns with other audio geminates and so that audio-singleton visual-geminate A_FV_E tokens pattern with audio singletons, there is still a significant difference between A_TV_L and A_FV_E's behavior. Visual cues can impede the perception of audio cues, as evidenced in the McGurk effect, and they seem to do so in the case of the A_FV_E tokens. In short, this suggests that geminates are more visually salient than singletons.

Figure 3: Arai et al (2017) Experimental Results

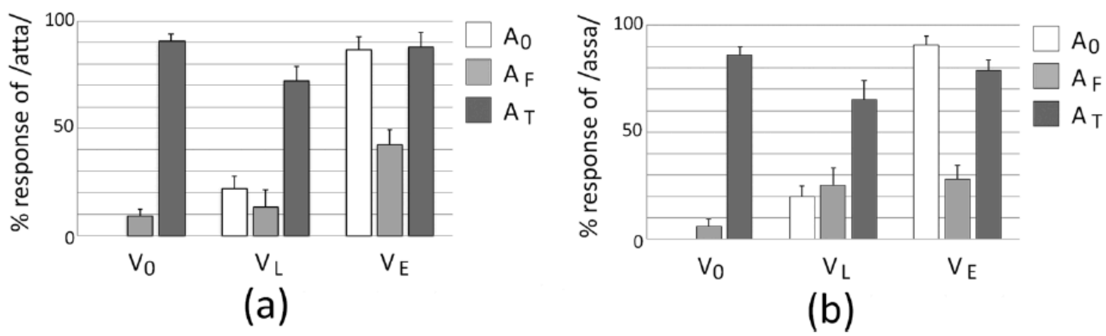


Fig. 4. Experimental results of (a) the /t/-experiment and (b) the /s/-experiment.

From Arai et al 2017

Before discussing the A_FV_E's behavior in the T condition, it is worth noting some differences between the T- and S- condition. The visual cues of the T- condition have 2 frames difference between the early and late closure in the visual condition, whereas the S- condition only had 1 a frame difference. Secondly, the T- condition's audio tokens are potentially less ambiguous and more natural than the S- condition's tokens. Arai et al inserted a silent interval of 140ms into the S- condition audio tokens, both because geminate fricatives sometimes have silence in real speech, and because it would have been unrealistic to follow a /a/ with no formant transitions by and the frication noise of a singleton /s/ immediately. While these are logical modifications to S- experiment, stop geminates are inherently silent during the closure and the T- experiment thus required no altercations to the natural tokens. A geminate /tː/ will always contain silence, removing the variability found in a geminate /sː/.

In the T experiment, 42.7% of A_FV_E tokens were listed as geminate, despite the auditory information denoting a singleton, whereas 72% of A_TV_E tokens were marked as geminate, which Arai et al report (2017) as more comparable to the pure audio and pure video conditions. (2017)[1] Furthermore, they report that visual gemination significantly impeded audio cues in phonological perception, whereas the visual singleton was not strong enough to significantly interrupt the geminate audio perception. Length is available as a visual cue per the visual-only A_0 experiments, but the mixed conditions furthermore suggest that geminates are more visually salient than singletons, which seems intuitively logical. A long closure provides a more observable and unique visual cue than a short one.

2.4 Place of Articulation and Spanish L2 Spirantization

In 2009, Menke and Face studied the development of Spanish L2 learners in their second-year of college Spanish, as graduating majors, and as doctoral students, to explore how students learn an allophonic pattern regarding Spanish voiced stops. Spanish voiced stops /b, d, g/ become spirantized intervocalically in all dialects, and English L1 learners frequently struggle to produce the allophones in their proper environments.

While their object was to discover whether more advanced speakers showed improved production of this allophony, their data incidentally also shows that all participants produced labial consonants better than velar consonants as shown in table 2. While the values were originally reported as the percentage of tokens produced correctly, these values have been converted to odds ratios here for better comparison. While they do not discuss this margin in their paper, it cannot be explained by frequency, as /b/ and /g/ have relatively the same frequency in Spanish at 2.1% and 1.4% respectively (Piñeros, 2009). /d/ actually has much greater frequency than the /b/ and /g/ at 4.0%, but, except for the fourth semester students, students produce less accurately than /b/ and more accurately than /g/. Labial and velars generally understood are marked compared to coronals, so these

[1] Thus, the "accuracy", if the auditory cue were to be understood as the correct response, of the $A_F V_E$ is only 57.3%. Here the auditory cue is singleton. The "accuracy" of the $A_T V_E$ tokens is 72%. In this case, the audio cue is geminate.

phonemes should have similar markedness (Lombardi, 2002). If it is true that visual influence holds a key role in the acquisition of L2 phonology, then this would predict that learners acquire /b/'s allophony more accurately than /g/'s because of additional visual information.

Table 2: Accuracy with Spanish Spirants by Proficiency Level

Odds of correct production	/b/	/d/	/g/	Odds ratio, /b/ vs. /g/
4th Semester	0.68	0.40	0.51	1.32
Major Students	2.11	1.53	1.31	1.60
Doctoral Students	7.06	3.74	2.88	2.46

Adapted from Menke and Face (2009)

It is furthermore interesting that the gap between /b/ and /g/ spirantization grows in doctoral students, even though accuracy improves in all phonemes. This seems to support the idea that there is a distinct interface between visual input and learning specific to L2 phonology acquisition. Frequency does not explain the accuracy of these three sounds, as /d/ is the most frequent, consisting of 4% of phonemes, and markedness also fails to make an accurate prediction, as /d/ is likely the least marked of the stops as a coronal articulation.

2.5 Gemination Cross-Linguistically

This experiment uses voiceless singleton stops rather than voiced singleton stops for a few reasons. Typically, though not exclusively, voiceless geminate stops are more common than their voiced counterparts, so that if a language has voiced geminate stops they generally have voiceless geminate stops (Blevins, 2008). Furthermore, as noted by Kawahara 2005, many claim that it may take more effort to produce a voiced geminate obstruent than a voiceless one due to difficulty of maintaining aerodynamic conditions required to sustain voicing throughout the closure. Furthermore, while the waveform served as the basis for measurement (see 3.1), spectrogram information is still useful when reading the waveform for stop closure. Esposito (1999) notes that voiced stops do not always have as sharp a transition on the spectrogram. As they may be more inherent cross-linguistically and due to ease of measurement, this experiment employs voiceless geminate stops in its test of impact of visual cues on a second language.

2.6 Structure of Italian Geminates

Standard Italian has a consonantal length contrast unique to the Italo-Romance branch of the Romance languages (Stevens, 2011). In Standard Italian, length is contrastive for all consonants intervocalically, except for five long consonants (Payne, 2006).

Table 3: Italian Consonant Inventory

Singleton Phonemes:	/p, b, t, d, k, g, t͡ʃ, d͡ʒ, s, f, v, n, m, l/
Geminate Phonemes:	/p, b, t, d, k, g, t͡ʃ, d͡ʒ, s, f, v, n, m, l/
Inherently Long Sing.:	/r, ʃ, ɲ, t͡ʃ, d͡z/

Adapted from Payne (2006)

Listeners distinguish singletons and geminates primarily by their length, but the V1 of the proceeding vowel may serve a weak as an additional correlate (Esposito, 1999). While research suggests that normally Italian speakers cannot identify gemination based off of the proceeding vowel alone, the V1s after a geminate are normally 25% shorter than a V1 following a singleton (Esposito, 1999).

In Italian, gemination occurs intervocalically, as geminates are often analyzed as holding a mora of the coda position of one syllable and onset of the following one simultaneously. There can be slight variation in the geminate length based-off of the following vowel, but the ratio of geminate to non-geminates generally falls just under 2 times the length, though Esposito et al find /k/ and /kː/ can both be individually shorter than /p/ and /p:/ in front of /a/ and /i/ in terms of absolute values. Payne (2006) finds that word-stress, intonational prominence, and the sonority of the consonant can slightly influence absolute duration. These features can alter the absolute values of gemination, but the general relative correlate is relatively consistent.

There is some dialectical variation found within Italian gemination. Some northern dialects use preaspiration to compensate for the length partially, but length of closure duration is still the primary correlate (Payne, 2006).

2.7 Project Aims and Hypotheses

As aforementioned, this project seeks to explore whether L2 phonologies show the influence of visual cues in the acquisition of the production of L2 phonological gestures, which we can refer to as a Visual Cue Saliency L2 Effect. This background section has established that Spanish spirantization allophony is better learnt in the labial position than in the velar position, that gemination is visible, and that the labial position is visually salient. This section also discusses the nature of Italian gemination.

Participants will read a short story in Italian and a word list. If visual input facilitates the learning of phonological processes, then participants will demonstrate more accurate elongation of geminate /p/ than of geminate /k/. This will be evident in the distinctiveness of labial and velar Italian geminates from their singular counterparts, as more accurate speakers will produce fewer geminates as singletons in their production.

2.7.1 Hypothesis 1

As such, my first hypothesis is that the ratio that consists of both the ratio of the voiceless labial geminates to singletons and the ratio of voiceless velar geminates to singletons will be greater than one. The ratio of ratios was expressed as a difference of logarithms to preserve maximum integrity of the data (see the statistical analysis in section

4.1), but this is conceptually similar to claiming that the difference in the ratio of /pː/ to /p/ to the ratio of /kː/ to /k/ will be a positive number. I anticipate /pː/ to be produced better, and thus longer than /kː/. Hypothesis one is a measure of relative distinctness and claims that the labial pair will be more distinct than the velar one.

2.7.2 Hypothesis 2

If hypothesis 1 is true, then I expect that the pattern found in the Menke and Face (2009) paper might be repeated amongst these three groups. In other words, I would expect the difference between the labial geminate singleton distinction (henceforth, the P contrast) and its velar counterpart (the K contrast) to become greater in more advanced speakers.

3 Methodology

3.1 Introduction

Now that the projects goals and hypotheses have been established, this section will detail the experimental design, measurement process, and token design. Participants were asked to complete 3 tasks in the following order: answer a questionnaire, read a short story and read a word list. After the experiment was conducted, the researcher selected a subset of the short story tokens, because the experimental design had not properly accounted for the influence of the following vowel in stop closure length, as discussed in 3.4.3. The guided reading was designed so that participants would not know that gemination specifically was the topic of study, to discourage careful reading regarding these phonemes.

3.2 Participants

The participants in this study are Italian L2 learners with between 2 and 10 years of formal study in the language. Among participants, the mean length of study was 3.95 years of Italian, and the median length was 5.5 years. Participants belong to one of three categories, defined relative to how many courses they taken beyond the fourth class in Chapel Hill's undergraduate language-learning sequence (Italian 204). The group of intermediate students had taken through Italian 204, advanced students had taken at least 4 major level Italian courses beyond Italian 204, and the last group consisted of graduate students, who that spring were finishing, at least their second year of graduate study.

Participants were all college-aged adults attending the University of North Carolina at Chapel Hill, who were currently enrolled in Italian coursework. There are 10 participants total in the results, with 4 participants in the beginner and intermediate groups and 2 in the advanced group. Two participants were excluded: one as he was a heritage speaker of Italian who acquired some Italian as a L1 speaker, and the other because his recording was of low quality.[2]

No participant reported knowledge of a language that included a singleton-geminate contrast, and all spoke English as a first language. The languages they have studied beyond Italian are listed in order of proficiency. The languages they have studied are reported as follows:

Table 4: Linguistic Experience

Part. #	*Exp. Group*	*# of Classes*[3]	*Yrs. Studied*	*Travel*	*Family*	*Other L2 Competencies, arranged from most to least fluent*
1	I.	4	4	None	None	Latin
2	I.	5	3.5	3 months	None	Spanish
3	B.	2	1	3 months	None	French
4	I.	5	3	5 months	None	Spanish, Latin, Purhepecha, Coptic
5	B.	0	2	3 days	None	Spanish, Swedish,[4] German, Turkish, Russian, French
6	I.	6	3	1 weeks	Yes, rel.	French
7	B.	3	4	3 months	Yes, rel.	
8	B.	1	3	None	Yes, rel.	Spanish
9	A.	18	10	4 years	Yes, spouse	Spanish, Latin, German, Fante
10	A.	11	6	1 year	None	French
Average:		*5.5*	*3.95*	*Travel: 0.62 years, ~ 7.5 months*		

---

[2] In this case, I was trying to see if the participant could send their own recording electronically, in order to increase the number of graduate students All participants in the study's results are Chapel Hill students, who were recorded in the Linguistics departments' professional sound booth.

[3] Number of Classes taken past 204 (not inclusive).

[4] Swedish does have contrastive gemination, as noted after the study. To the best of my knowledge, participant's behavior was indistinguishable from the other beginner students. Their study of Swedish was not advanced.

3.3 Questionnaire Design

First, participants responded to a questionnaire with questions about their language experience.[5] Questions pertained to their formal study of Italian, travel to Italy, and knowledge of other language amongst other things. Most importantly, participants reported how long they had studied Italian, and how many Italian-language major-level courses they have completed. Relevant data is compiled in Table 4, above.

3.4 Short Story Task

3.4.1 Short Story Methods

After completing the questionnaire, participants narrated an adapted Italian version of the first three chapters of *The Little Prince*, written by Antoine de Saint-Exupéry, available in appendix A.2. I adapted the translation, editing sentences and words, so that there were 60 tokens total embedded into the reading. Native Italian speakers proofread the edited work. There were 15 tokens of each type: 15 /p:/, 15 /p/, 15 /k:/, and 15 /k/ respectively. The tokens are available below, but further information on token design and balance will be covered in section 3.4.2.

Table 5: Short Story Task Tokens

1	s	k	*magnifico*	31	l	k	*bloccato*
2	l	k	*eccovi*	32	l	p	*grappolo*
3	l	k	*bracca*	33	l	p	*ceppo*
4	l	k	*bocca*	34	s	p	*stupore*
5	s	k	*mastica*	35	s	k	*pecora*
6	s	p	*dopo*	36	s	k	*trasecolo*
7	s	p	*capolavoro*	37	l	p	*apparizione*
8	l	p	*cappello*	38	l	p	*eppure*
9	s	p	*capo*	39	l	p	*appariva*
10	s	k	*placava*	40	s	k	*fatica*
11	s	p	*scopo*	41	l	k	*secco*
12	s	k	*applicassi*	42	s	p	*ripeté*
13	l	p	*mappe*	43	l	k	*traboccante*

[5] The questionnaire is available in Appendix A.1. The experiment consent form and other testing materials are available throughout Appendix A.

14	s	k	*logica*	44	s	k	*pericolo*
15	l	k	*sacco*	45	s	k	*ricordai*
16	s	p	*capiscono*	46	l	k	*piccolo*
17	s	k	*affaticano*	47	l	p	*groppo*
18	l	k	*d'occhio*	48	l	p	*appuntò*
19	s	p	*sapere*	49	s	k	*amico*
20	s	p	*l'opinione*	50	l	k	*vecchia*
21	s	p	*aperta*	51	s	p	*concepire*
22	s	p	*capire*	52	s	k	*poco*
23	l	p	*cappuccio*	53	l	p	*scoppio*
24	l	k	*peccato*	54	s	p	*riposo*
25	s	k	*politica*	55	s	p	*canapa*
26	l	p	*troppo*	56	s	p	*proposta*
27	l	k	*staccate*	57	l	p	*dappertutto*
28	l	k	*siccome*	58	s	k	*fico*
29	l	k	*meccanico*	59	l	p	*pioppo*
30	l	k	*acqua*	60	l	p	*scappare*

All of the participants were tested in the sound booth the Phonetics Lab at UNC in Dey 103. The recordings were captured at 44100 Hz in Praat. Participants were permitted to take breaks as needed, and were given short breaks of around 1 to 2 minutes at the beginning of the second and third chapter so that the researcher could start a new recording file; although, one intermediate student and one advanced student continued reading and thus have different breaks between the recordings of the reading section. Advanced students took around 10-15 minutes to complete this reading, whereas beginner students normally needed around 20-25 minutes.

3.4.2 Short Story Token Design and Balance

Tokens were chosen as they fit into the source text, and with attention to the vowel context, their relation to the stress in the word, and the number of syllables. While different places of articulation share a duration ratio target, there are a number of environmental factors that can influence the length of geminates. The following table shows summarizes the token balance, and a complete inventory of the tokens as marked for token balance is available in Appendix B.

Table 6: Token Balance Summary Chart

Categories:	Stressed V After	A. Syllable Length	V After Height	V After Backness
/pː/	4/15	2.866	8 Mid.	8 Back
/kː/	7/15	2.666	5 Mid.	5 Back
/p/	8/15	3.133	10 Mid.	7 Back
/k/	4/15	3.066	8 Mid.	8 Back

Geminate length may be affected by being followed by a stressed syllable, the word length and vowel quality (Payne, 2005, Esposito, 1999). The design goal is certainly to minimize the influence of these factors, but when impossible, the tokens are chosen so that these environmental factors support the distinctiveness of the K contrast. As both hypotheses of the experiment center around the idea that the P contrast will be more distinct than the K contrast, by having the tokens lean towards K contrast's distinctiveness when perfect balance is impossible, this helps protect against false positives.

With word length, both /pː/ and /p/ occur in longer words than /kː/ and /k/, which means that the P contrast is more likely to have shorter tokens than the K contrast. By having longer tokens in the geminate condition, the K-condition is biased towards being more distinct than the P-condition, as gemination is produced more than singletons. The K contrast is similarly advantaged in word stress. Being in a stressed syllable seems to elongate geminates (in environments where a stressed vowel follows the consonant), and /kː/ and /p/ are in more stressed syllables than /k/ and /pː/. While the geminates show more difference than singletons, /kː/'s length is still accentuated by a stress bias, supporting the

K-condition's distinctiveness. Vowel balance proved to be the most problematic element of the design, so a subsection of the short story task's data was analyzed as a subset controlled for vowel quality. This subset will be discussed here.

3.4.3 O-subset Design

After the experiment was conducted the researcher realized that previous research (Esposito, 1999), found that out of the various environmental aspects the quality of the following vowel most dramatically alters the absolute duration of the preceding stop. While the relative duration is roughly the same among vowels, without ensuring that there are an equal number of tokens across the four token types with each of the following vowel, the distribution of vowel balance of the experiment can bias the average durations of tokens across the four token types. This proves to be the clearest problem in this experimental design, and while the experimental balance's focus was originally concentrated on ensuring similar numbers of mid, low, and high vowels in each singleton-geminate group, the literature makes it clear that backness and height clearly must be prioritized together in the experimental design over other environmental influences discussed in 3.4.2.

Given that the following vowel quality seems to affects absolute stop length much more dramatically than other variation in the experiment (Esposito, 1999), the experimental conclusions would have been incomplete without analyzing a subset of the data controlled for the following vowel. The O-subset, as explained here, will serve to either confirm the experimental findings of the short-story task, or remove the influence of the following vowel in experimental design.

The O-subset consists of 26 tokens that follow /o/ or /ɔ/ in the short-story condition. No other vowels had sufficient tokens in all four conditions to be processed reliably: but the O-subset has 5 to 8 tokens for each type (/pː/, /p/, /kː/, and /k/). No other subsets are reputable, with some possible subsets having only 1 token or lacking a token for a certain token type. Containing nearly half of the short story list, the O-subset it is large and controlled so it can ensure that vowel misbalance among the types did not interfere in the results analysis. Esposito (1999) shows that /p/, /p:/, /k/ and /k:/ behave similarly to one another in front of /a/ and /u/ (the paper studied tokens proceeded by /i/, /a/, and /u/). Furthermore, the subset is not biased towards the P-condition's distinctness in word length or stress, as visible in table 16 in appendix B.2. While the subset was selected after conducting the experiment, its concordance or dissonance with the overall results is necessary information to account for behavior related to vowel backness and ensure the validity of the results of the short story and word list sections as a whole.

3.5 Word List Task

3.5.1 Word List Methods

After reading the short story, participants read a word list. The words were presented as a Google Slides presentation on a laptop computer screen, such as:

There were 25 words in all, of which 20 were measured as tokens, and 5 were distractor terms, which served only to mute slightly that gemination was under examination. The tokens in this study are near minimal pairs; although, this mandated the use of archaic or regional terms and placenames. The words are listed below:

Table 7: Word List Task Tokens

Capo	Hat	Teca	Display
Dopo	*After*	Seco	*With him (ar.)*
Cipolla	*Onion*	Reco	*I bring*
Tropo	*Too much*	Picone	*Placename*
Bepo	*A little machine*	Cechino	*Blind (dim.)*
Cappio	*Loop*	Tecca	*Little Machine*
Doppio	*Double*	Secco	*Dry*
Cippo	*Memorial / boundary stone*	Recco	*Placename*
Troppo	*Too much*	Piccone	*Pick*
Beppo	*Name*	Cecchino	*Sniper*

There were five additional words that served as filler: *bene, caffè, posto, luce, oro.* The slides were placed in a random order for each participant, using a website called Random.org.[6] Participants were instructed to click through the slides and read each recording. They were instructed only to repeat a word if they felt they had mispronounced it in such a way that they would normally correct themselves in a classroom setting, and the last iteration would be taken for measurement.

3.5.2 Word List Token Design and Balance

By and large, each measurement reflects a participant's first attempt at production. The tokens are well balanced in this task, but there are slight discrepancies in vowel backness. However, this task is much smaller than the short story task, and subject to greater variation as a result. Additionally as aforementioned, participants do seem to become aware of the focus on gemination during the task.

[6] Participant 8 word-list slides were not randomized due to experimenter error. They show no obvious outlier effects.

3.6 Measurement Process

After all participants were recorded, the experimenter randomly ordered the recordings and measured them according to this order. The reading passage consisted of three chapters (each recorded as a separate tape) and the word list was one tape, so that each participant was recorded on four tapes. As such, by randomizing the measuring order, of the four portions of each participant's production, the experimenter ensured that none of the chapters or participant groups were privileged above the other in terms of measurement, the concern being that later measurements might be more accurate than early measurements. The closure of the geminate and singleton tokens was measured in Praat (2018), as calculated from the nearest zero crossing to the left of the last glottal pulse on the waveform to the burst of the stop. The following details the measurement procedure.

3.6.1 Measurement Procedure

The following section outlines our measurement procedures. All tokens were measured at least twice to ensure the measurement's concurrence to these procedures.

1) File Organization

   1.a Open the sound file and add one band as a .txtgrid under Annotate.

   1.b Filter the sound file with a stop band filter, from 0 - 100. Set the smoothness to 10, and then press 'ok' to conserve the original sound file.

2) Locating the Beginning of the Selection

   2.a Find the target stop closure, and zoom into about .8 seconds so that most of the word is contained.

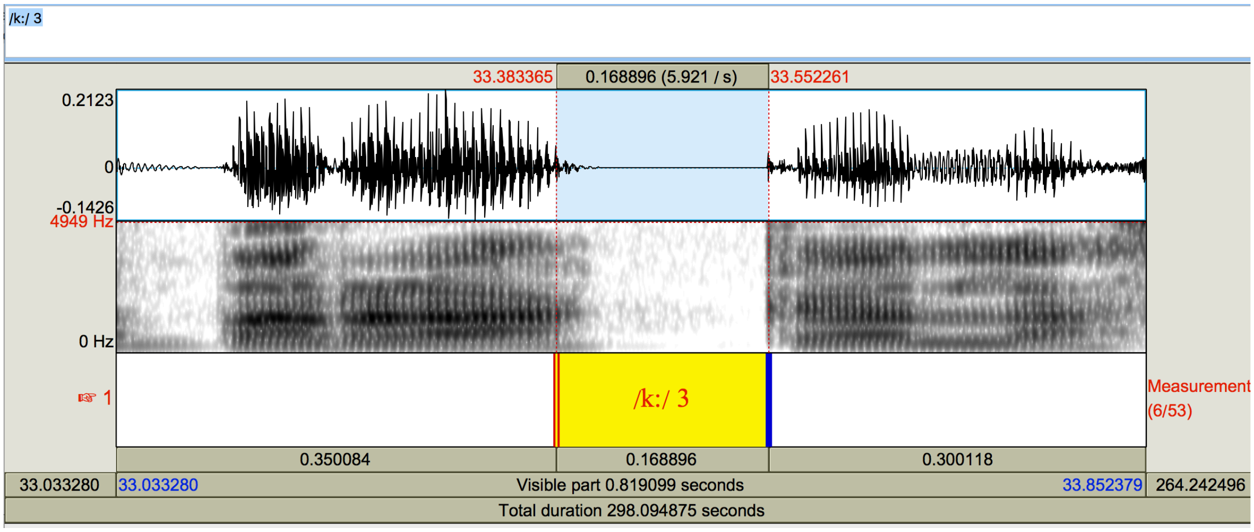


Image 1[7]**, the visible section here is .819 seconds.**

2.b Zoom into to 0.1 – 0.2 seconds near the closure to determine what the last glottal pulses with clear vowel formants may be.

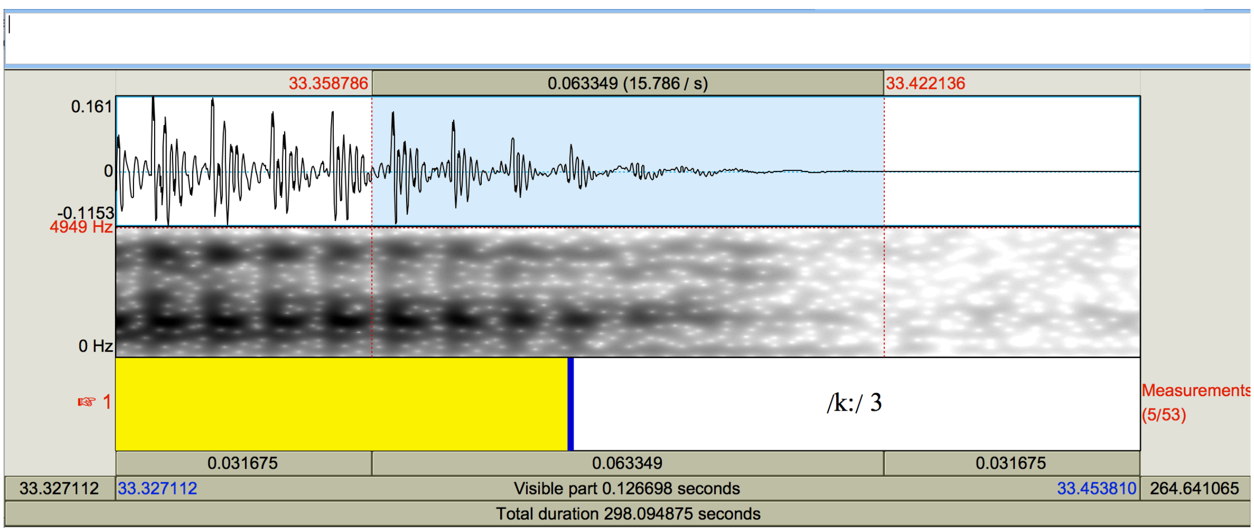


Image 2, selected portion corresponds to **Image 3**.

2.c To find the beginning of the closure measurement, insert a boundary on the nearest zero crossing to the left of the highest peak within the last repetition of the cycle as seen on the waveform. The last repetition on the waveform is defined as the one where:

[7] Images 1 – 4 are taken from token #3, "bocca", (/bok:a/) from speaker 4. Reading Section

2.c.1 The cycle on waveform maintains the same recognizable shape as earlier cycles in the middle of the vowel, so that the three waves with the highest peaks within this glottal pulse are also the three waves with the highest peaks in a prominent cycle in the middle of the vowel.

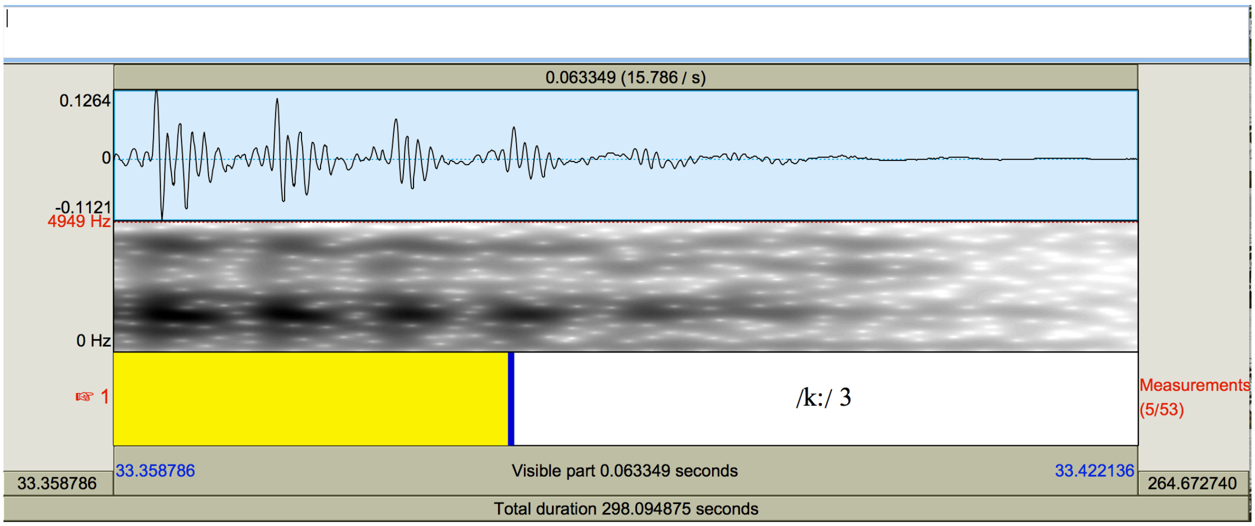


Image 3

2.d Use Praat commands to find the zero crossing described in **2.c**, set a boundary at this location.

3) Locating the Release

3.a. Use the waveform to identify the release of the articulators after the closure.

3.a.1 If there is an irregularity in the stop burst (for example, a click occurs before the burst), note this in an irregularity log.

3.b From here, locate the zero crossing at the first wave of the release on the waveform with Praat's functionality, and insert a boundary here.

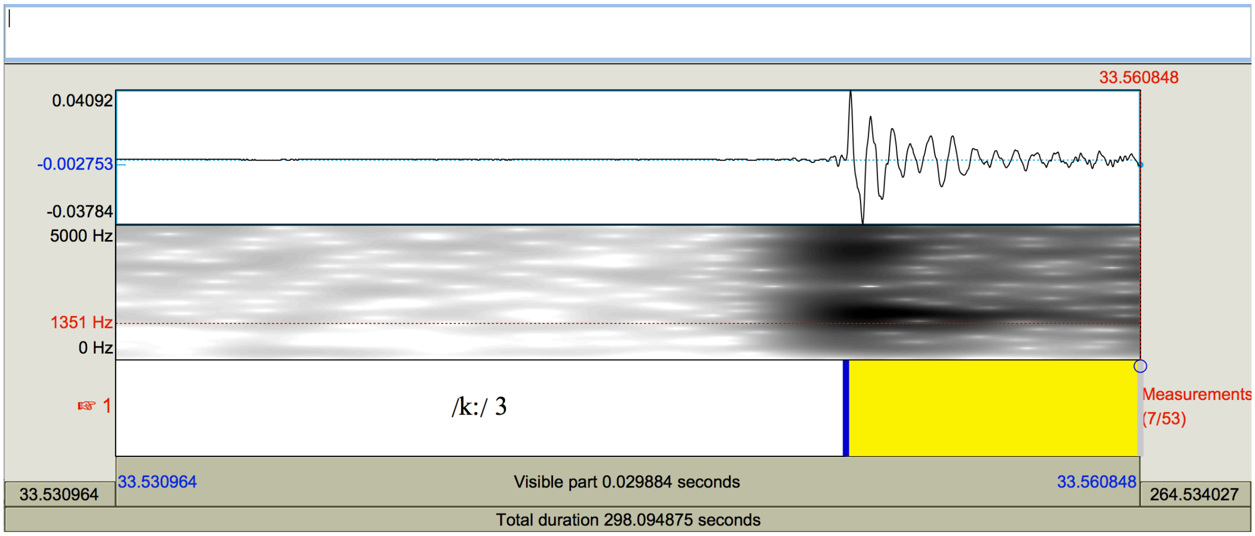


Image 4

4) Verification of Measurement and Removal of Inadequate Tokens

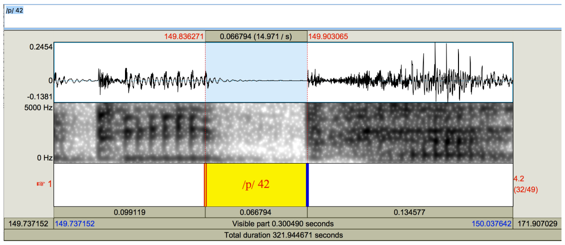
4.a If the stop consonant is produced incorrectly, then remove it from the count. Further details are available in section 3.3.3. Log irregularities for consistency.

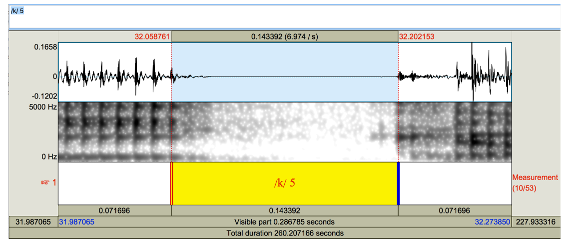
4.b Apply labels in between the boundaries on the text grid to mark the token number and the target phoneme.

After all tokens were passed over again to verify their correspondence to the conditions set forth in above, I used Dr. Katherine Crosswhite duration-logger Praat script available online at http://phonetics.linguistics.ucla.edu/facilities/acoustic/duration_logger.txt to extract the duration measurements. This logger places the durations of the labeled segments into a Text Edit file, which was then processed in Excel.

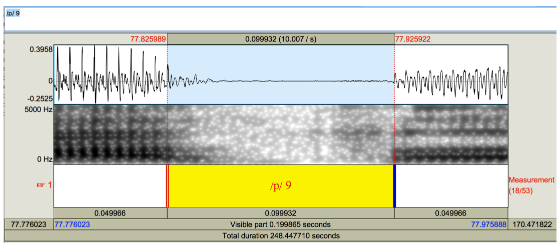
3.6.2 Exemplars

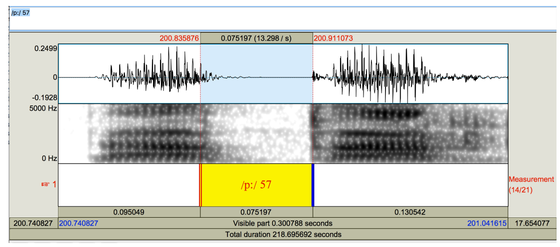
Here we present exemplar measurements of tokens as seen from each of the three groups of participants.



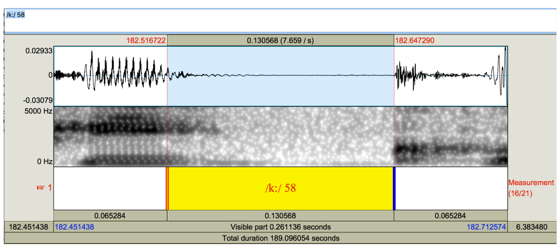


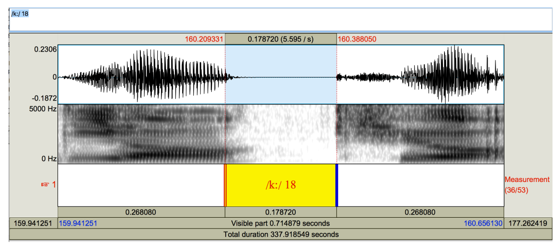
Beginners, Participant 4: 42, 6: 5





Intermediate, Participant 3: 9 5: 57





Advanced, Participant 13: 52 12: 18

3.6.3 Handling of Abnormalities

Abnormalities were logged and marked. Abnormalities in which the token was produced in a form other than a stop (a few singletons were spirantized) were discarded. All tokens were intervocalic, but in a few instances participants would metathesize the word so that the token was no longer intervocalic, so these tokens were excluded. In two instances background noise obscured the token. 25 tokens were excluded in total, or 3.125% of the total 800 tokens.

3.7 Hypothesis and Experimental Predictions

As introduced in 2.7, we will test two hypotheses here in the two experimental tasks and the reading list's O-subset. Our first hypothesis proposes that the P-condition is more distinct than the K-condition. Practically speaking, this means that we expect the ratio of the ratio of the P-condition and the ratio of the K-condition to be greater than 1. If it is true, then hypothesis 2 becomes relevant. The second hypothesis suggests that this distinction between P and K is stronger in more advanced participant groups, so that the ratio of ratios will be bigger in these groups. The statistical model and mathematical process for testing the hypotheses will be discussed in the following section, 4.1.

4 Results

4.1 Statistical Model

The data was as analyzed using a log-linear generalized model with a gamma-distributed error term. 12 groups are defined in the statistical analysis based on the place of articulation (p or k), the length (long or short) and experience (beginner, intermediate, and advanced) so that design matrix can be understood so that the log of duration is equal to [β_1(KLB)+ β_2(KLI)+ β_3(KLA)+ β_4(KSB)+ β_5(KSI)+ β_6(KSA)+ β_7(PLB)+ β_8(PLI)+ β_9(PLA)+ β_{10}(PSB)+ β_{11}(PSI)+ β_{12}(PSA)]. The variance accounted for multiple observations per subject, and the analysis was conducted in SAS PROC GENMOD.

As aforementioned, the first hypothesis proposed that the ratio of the singleton and geminate /p/ would be greater than the ratio of singleton and geminate /k/, so that log(sing. p) - log(gem. p) - log(sing. k) + log(gem. k) > 0, if (sing. p / gem. p) / (sing. k/gem. k) > 1.

The second hypothesis purports that the size of the ratio calculated to test hypothesis 1 is greater in higher-level proficiency groups than lower-level groups out of the three proficiency groups: beginner, intermediate, and advanced.

In the results section we report the ratios of ratios of the P- and K-conditions per task and per group here as "(PL/PS)/(KL/KS) Ratios." We also report the average duration of each token, /pː/, /p/, /kː/ and /k/, per level. Lastly, we report the ratio between the advanced and beginner conditions ratios of the P- and K-condition as A/B ratios per type to better support discussions on the developmental trends of the four different phonemes. If any of the results seem to comply with hypothesis 1 and 2, the A/B ratios allow us to identify which token types changed to form the ratio patterns hypothesized. These values are discussed for both the short story and word list task, as well as for the O-subset.

4.2 Results

4.2.1 Short story Task

The short story condition had 15 tokens of each type (/p:/, /p/, /k:/, /k/), and there were 4 participants in the beginning and intermediate groups, and 2 in the advanced. Plot 1 shows the average durations in seconds of each token type per group. The ratio of the distinctiveness of geminate and single /p/ to that of geminate and singleton /k/ is found in Table 8 and in Plot 1, as well individual A/B ratios per type that show how different participants produced each type of token in each of the four categories.

Plot 1: Short Story Task Results

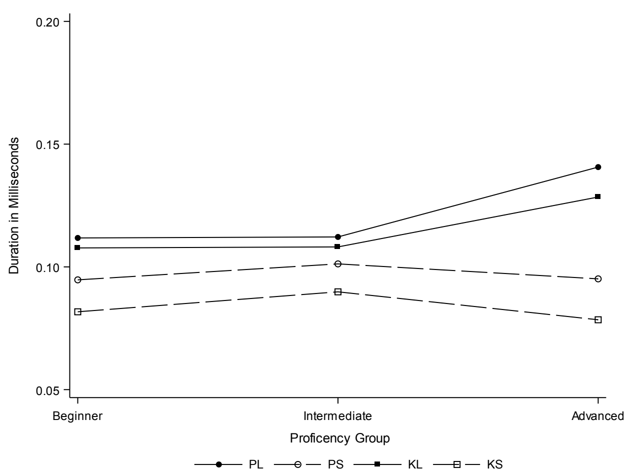


Table 8: Short Story Task Results

Label: Short story task	*Mean Estimate*	*MeanLowerCL*	*MeanUpperCL*	*Chi-Square*	*Pr > ChiSq*
Average Durations in Seconds:					
KL B	0.1078	0.0926	0.1254		
KL I	0.1080	0.0787	0.1483		
KL A	0.1286	0.0946	0.1748		

Label: Short story task	*Mean Estimate*	*MeanLowerCL*	*MeanUpperCL*	*Chi-Square*	*Pr > ChiSq*
KS B	0.0817	0.0759	0.0879		
KS I	0.0898	0.0641	0.1258		
KS A	0.0786	0.0712	0.0867		
PL B	0.1120	0.0968	0.1295		
PL I	0.1124	0.0899	0.1404		
PL A	0.1405	0.1198	0.1647		
PS B	0.0950	0.0800	0.1127		
PS I	0.1013	0.0755	0.1358		
PS A	0.0950	0.0929	0.0972		
(PL/PS)/(KL/KS) Ratios:					
Beginner: (PL/PS)/(KL/KS)	0.8938	0.7910	1.0099	3.24	0.0717
Intermediate: (PL/PS)/(KL/KS)	0.9224	0.8417	1.0108	2.99	0.0838
Advanced: (PL/PS)/(KL/KS)	0.9033	0.6903	1.1821	0.55	0.4587
A/B Ratios per Type:					
KL: Advanced/Beginner	1.1936	0.8475	1.6810	1.03	0.3111
KS: Advanced/Beginner	0.9621	0.8510	1.0876	0.38	0.5366
PL: Advanced/Beginner	1.2547	1.0115	1.5565	4.26	0.0390
PS: Advanced/Beginner	1.0007	0.8417	1.1898	0.00	0.9933

If /pː/ and /p/ are longer than /kː/ and /k/ in production, then we expect their (PL/PS)/(KL/KS) ratio of ratios to be greater than 1. Per the beginner, intermediate, and advanced (PL/PS)/(KL/KS) ratios Mean Lower CL and Mean Upper CL, we can see that the error ranges of all three groups in this condition fall on both sides of 1, so the short story task is inconclusive.

It is also useful to examine the ratios of the advanced tokens to the beginner's tokens per type, available under "A/B ratios per type." After accounting for error, if the number is greater than 1, then it means that the average length of the advanced tokens in the numerator is longer, whereas if the number is less than 1, then it means that the beginner condition was longer than the advanced condition. The short story task shows that /kː/ and /pː/ do seem to be significantly longer in advanced L2 learners compared to beginning L2 learners, so that learners are learning gemination. In contrast, the /k/ and /s/ do not evidence a significant difference between the groups.

4.2.2 Word List Task

The word-list task contains less data than the reading-list condition with 5 tokens for each type and tested the same participants as in the reading list. Plot 2 and Table 9 summarize the experimental results.

Plot 2: Word List Task Results

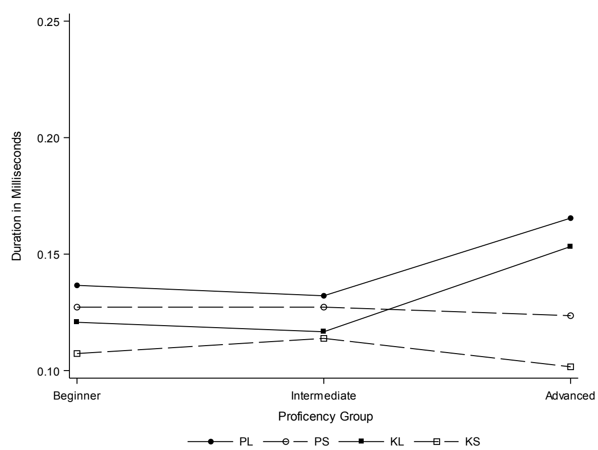


Table 9: Word List Task Results

Label	*Mean Estimate*	*MeanLowerCL*	*MeanUpperCL*	*Chi-Square*	*Pr > ChiSq*
Average Durations in Seconds:					
KL B	0.1477	0.1360	0.1603		
KL I	0.1563	0.0903	0.2704		
KL A	0.1964	0.1957	0.1972		
KS B	0.1383	0.1154	0.1657		
KS I	0.1494	0.0961	0.2322		
KS A	0.1242	0.0874	0.1765		

Label	Mean Estimate	MeanLowerCL	MeanUpperCL	Chi-Square	Pr > ChiSq
PL B	0.1849	0.1625	0.2105		
PL I	0.1842	0.1200	0.2827		
PL A	0.2243	0.2104	0.2391		
PS B	0.1570	0.1373	0.1796		
PS I	0.1607	0.1123	0.2302		
PS A	0.1469	0.1002	0.2155		
(PL/PS)/(KL/KS) Ratios:					
Beginner: (PL/PS)/(KL/KS)	1.1028	0.9348	1.3010	1.35	0.2459
Intermediate: (PL/PS)/(KL/KS)	1.0952	0.9838	1.2194	2.76	0.0967
Advanced: (PL/PS)/(KL/KS)	0.9656	0.8744	1.0664	0.48	0.4894
A/B Ratios per Type:					
KL: Advanced/Beginner	1.3301	1.2251	1.4441	46.23	<.0001
KS: Advanced/Beginner	0.8983	0.6051	1.3335	0.28	0.5946
PL: Advanced/Beginner	1.2131	1.0501	1.4015	6.88	0.0087
PS: Advanced/Beginner	0.9357	0.6236	1.4040	0.10	0.7481

We are less equipped to draw a conclusion from the word list task than from short story task, as there are fewer tokens in the word-list task. Furthermore, as mentioned in 3.5.2, participants also became aware that gemination was the focus of the study as they progressed through the randomized tokens of this task, whereas they were unaware during the short story task. All tokens' average durations were longer in the word list task than in the short story task; perhaps, this supports the idea that participants used more careful speech on the word list. Nonetheless, like the short story task, all three (PL/PS)/(KL/KS) values are inconclusive due to the error, and as to the P and K-conditions' ratios. Additionally, per the A/B ratios per type we can see that the geminates do seem longer in advanced populations, whereas the singletons do not evidence significant change.

4.2.3 O-subset Results

The O-subset proves to be the most interesting out of the three analyses of data, consisting of 26 tokens from the short story task. All tokens consist of the target sound followed by /o/ or /ɔ/, and Plot 3 and Table 10 describe the experimental results in detail.

Plot 3: O-subset Results

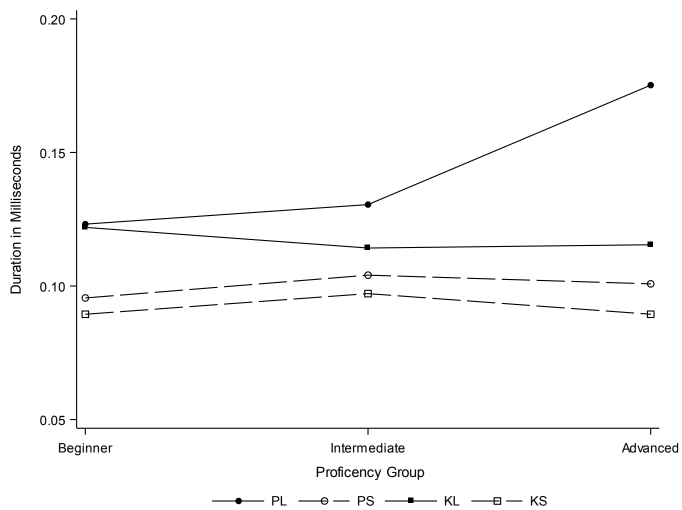


Table 10: O-subset Results

Label: O-subset	*Mean Estimate*	*MeanLowerCL*	*MeanUpperCL*	*Chi-Square*	*Pr > ChiSq*
Average Durations in seconds:					
KL B	0.1220	0.0940	0.1585		
KL I	0.1142	0.0844	0.1545		
KL A	0.1155	0.0978	0.1364		
KS B	0.0897	0.0798	0.1008		
KS I	0.0972	0.0690	0.1370		
KS A	0.0895	0.0780	0.1027		
PL B	0.1233	0.1003	0.1515		
PL I	0.1305	0.1050	0.1622		
PL A	0.1754	0.1585	0.1941		
PS B	0.0957	0.0776	0.1180		

Label: O-subset	*Mean Estimate*	*MeanLowerCL*	*MeanUpperCL*	*Chi-Square*	*Pr > ChiSq*
PS I	0.1042	0.0767	0.1414		
PS A	0.1007	0.0996	0.1018		
(PL/PS)/(KL/KS) Ratios:					
Beginner: (PL/PS)/(KL/KS)	0.9466	0.8265	1.0842	0.63	0.4283
Intermediate: (PL/PS)/(KL/KS)	1.0666	1.0316	1.1028	14.35	0.0002
Advanced: (PL/PS)/(KL/KS)	1.3493	1.1138	1.6345	9.37	0.0022
A/B Ratios per Type:					
KL: Advanced/Beginner	0.9467	0.6945	1.2905	0.12	0.7290
KS: Advanced/Beginner	0.9978	0.8329	1.1954	0.00	0.9813
PL: Advanced/Beginner	1.4226	1.1307	1.7898	9.05	0.0026
PS: Advanced/Beginner	1.0520	0.8527	1.2979	0.22	0.6363

While the short story and word lists task show inconclusive results concerning the ratio of the P- and K-conditions, the O-subset shows significant results in the intermediate and advanced categories of the (PL/PS)/(KL/KS) ratios, because the Mean Lower CL is greater than 1. Furthermore, the Mean Lower CL of the advanced group is greater than the Mean Higher CL of the intermediate group, so that there is no overlap between the groups. This shows that the advanced group has indisputably longer (PL/PS)/(KL/KS) ratio than the intermediate group. Furthermore, the A/B per token type values show that only /pː/ production has a statistically significant change between the most advanced group and the beginner group. The error ratio (Mean Lower CL and Mean Upper CL) of all other phonemes are on both sides of 1.

4.3 Hypotheses and Statistical Results

In responding to the hypotheses, the (PL/PS)/(KL/KS) ratios are the most useful information. Table 11 summaries these ratios below.

Table 11: (PL/PS)/(KL/KS) Ratio Results Summary Chart

Word List	*Mean Estimate*	*Mean Lower CL*	*Mean Upper CL*
Short Story Task			
Beginner: (PL/PS)/(KL/KS)	0.8938	0.7910	1.0099
Intermediate: (PL/PS)/(KL/KS)	0.9224	0.8417	1.0108
Advanced: (PL/PS)/(KL/KS)	0.9033	0.6903	1.1821
Word List			
Beginner: (PL/PS)/(KL/KS)	1.1028	0.9348	1.3010
Intermediate: (PL/PS)/(KL/KS)	1.0952	0.9838	1.2194
Advanced: (PL/PS)/(KL/KS)	0.9656	0.8744	1.0664
O-subset			
Beginner: (PL/PS)/(KL/KS)	0.9466	0.8265	1.0842
Intermediate: (PL/PS)/(KL/KS)	**1.0666**	**1.0316**	1.1028
Advanced: (PL/PS)/(KL/KS)	**1.3493**	**1.1138**	1.6345

4.3.1 Hypothesis 1: /pː/ and /p/ will be more distinct than /kː/ and /k/

The first hypothesis seems to be verified by the results of this study when controlled for vowel context; however further experimentation may be useful. The O-subset shows that when controlled for vowel-context, the P-condition (/pː/ and /p/) is more distinct than the K-condition (/kː/ and /k/). In the overall short story task and word list task, the results are ambiguous as shown in table 11.

4.3.2 Hypothesis 2: The trend observed in hypothesis 1 will become more prominent in advanced student categories

Hypothesis 1 was built from hypothesis 2, so that it was anticipated that labial geminate and singleton sound contrast would be more distinct from the velar counterpart in more advanced speakers. While hypothesis 1 proposes an observable effect, hypothesis 2 suggests that this effect should be visible as a trend at different stages of L2 acquisition. This is visible in O-subset, as the ratio of ratios in the intermediate group's low and high range is 1.0316 and 1.1028ms respectively, whereas the advanced groups' answers range between 1.1138 and 1.6345ms. There is no overlap between the values of the two groups. That said, the advanced group had two participants while the intermediate group had four, so the intermediate group should have statistically stronger data than the advanced group. Although the tests administered here support hypothesis 2, it may perhaps be wisest to confirm with further experimentation as to the results of this trend over time.

4.4 Statistical Trends

The A/B ratios (advanced-to-beginner) per type show that the geminate /p/ in the O-subset was the only type of token to become longer between beginner and advanced groups in a statistically significant way. In other words, the statistical response to hypothesis 1 discussed above seems to be related to increasing the length of the labial geminate stop, and not due to changes in length between any of the other stops.

The A/B ratios of short stops also are enlightening, as all tasks report very large ranges with these token types. The ratio difference between the productions of these stops by speakers ranges from about .6 to 1.35, so while we say there is no statistically significant change due to the ratio being possibly both greater than or less than 1, there is a lot of variation within their production. The O-subset shows much less range in these categories, with the range being approximately .8 to 1.25. This leads to two suggestions: first, that the production short stops may be more susceptible to environmental factors, which makes sense given that long consonants can often resist changes that short consonants do not undergo, and that in this experiment the most expected change was that participants would learn to produce longer geminates. While the short-story task is inconclusive, in the careful speech of the word-list task, the A/B ratios of /kː/ and /pː/ show that participants do show that advanced participants produce noticeably longer geminate consonants than beginner participants.

5 Discussion

5.1 Introduction to Experimental Discussion

The results of this experiment are consistent with the theory that higher visual saliency of a phone corresponds to higher accuracy in L2 production than sounds of the same natural class at other places of articulation. That said, this alone does not exclusively implicate that our cue visibility theory, referred to in section 2.7 as a Visual Cue Saliency L2 Effect, is the cause. As such, in 5.2.2, we juxtapose the distinct production across place of articulation in studies with visible cues with the indistinct behavior across place of articulation found in an additional study entailing L2 production of non-visible cues (Reeder, 1998). Visibility, of course, is prerequisite to the visibility differences that may influence L2 sound production. This asymmetry between the production of visible and non-visible cues across place of articulation supports this claim that the differing visual saliency of cues is responsible for differences in the L2 production of visible gestures across place of articulation.

Beyond this key juxtaposition, this section continues to explore whether the scope of L2-influencers can be expanded to include visual cue saliency. This would suggest that L2 learning is partially facilitated by passive reinforcement of phones available through visual cues. In this section, our discussion of experimental results entails conversation about factors influencing experimental conclusions, proposed methods for future experimentation on visual cue saliency's effect in L2 production, and discussion of the potential impacts of this finding on our understandings of second language grammars, speech perception, and even pedagogy.

5.2 Experimental Conclusions

As discussed in the results, the data from the O-subset shows that intermediate and advanced participants produce the P-condition more distinctly than the K-condition. Similar to the data of the Menke and Face (2009) paper, the second hypothesis also concurs with the data in the O-subset, as L2 Italian advanced participants produce a significantly greater distinction than intermediate participants. The effect of visual input captured in this experiment grows more noticeable in more advanced populations in accord with the second hypothesis, even though one might have anticipated that advanced speakers show less external influence in their L2 phonologies.

5.2.1 Validity of O-subset's Behavior

Although the O-subset was selected after the experiment was conducted, I believe that it best explains what is happening in the data. Geminate duration may vary depending on the following vowel, but it is proportionally relatively similar across the places of articulation, so that based off of the results of the short-story and word-lists tasks, the (PL/PS)/(KL/KS) ratio of ratios of the O-subset should have been just as ambiguous as the data of the other two tasks. However, not only is the O-subset, which contains around half of the tokens, produced with a more distinct P-condition than K-condition, but more advanced participants produce a greater distinction. The experimental tokens were not balanced for vowel quality, and other subsets are hard to produce because the tokens proceeded by /a/ are by-and-large /k/ type tokens and the other vowels are not individually well represented in the experiment. Originally, the experimental design was focused on balancing other characteristics such as height and stress and not vowel quality (height and backness together).

While future experimentation will be useful, as discussed below, the O-subset's behavior can only be understood as distinct from that of the other tokens through this explanation. Furthermore, the framework of the theory that visual cue salience causes the difference in production between labial and velar counterparts in the O-subset and the Menke and Face (2009) finds further justification, actually, from an L2 acquisition pattern regarding place of articulation and a non-visible cue.

5.2.2 Spanish L2 Deaspiration and Non-Visible Cues

If the O-subset pattern was deviated purely from the influence of labial visual saliency, as the Visual Cue Saliency L2 Effect proposes, then it follows that non-visible articulations should be produced without bias across places of articulation. Indeed, VOT differences are likely not visibly salient due to the role of glottis in voice onset timing, and the Reeder (1998) findings on L1 English L2 Spanish production of Spanish voiceless stops does seem to concord with our prediction that non-visually salient productions lack a labial preference because /p/, /t/ and /k/ are all equally non-visible. More advanced participants produce similar aspiration proportionally across all places of articulation.

L1 English learners learn to deaspirate voiceless stops when learning Spanish. Reeder (1998) explores the production of these tokens by these learners at 4 different levels, with 10 participants in each level and a control group of 5 native speakers. Level 1 and 2 speakers are enrolled in third- and first-semester Spanish courses respectively, level 3 are students enrolled in graduate-level Spanish coursework, and level 4 speakers are Spanish L2 teaching faculty. As such, Level 1 and 2 speakers, are very early in their L2 acquisition. The experience groups in Reeder (1998) align nicely with our experiment and with Menke and Face (2009), where level 2 parallels our beginner groups, level 3 parallels our intermediate groups, and level 4 best parallels our advanced groups. Table 12 reports the mean VOT of each phoneme per group.

Table 12: Mean VOT per Phoneme per Group

	/p/	/t/	/k/
Level 1 (Novice)	54ms	53ms	73ms
Level 2	51ms	55ms	74ms
Level 3	36ms	36ms	60ms
Level 4 (Advanced)	29ms	29ms	49ms
Native Speaker (Control)	17ms	17ms	28ms

Adapted from Reeder (1998)

In table 12, the native speaker group shows us that /p/ and /t/ have slightly different targets than /k/; however, /p/, /t/ and /k/ do not all have the same target VOT, as seen in the native speaker productions above. In order to understand this data, we have to look proportionally at the relationship between the target VOT and the L2 productions.

Table 13 juxtaposes the anticipated length of /k/, based off of native speaker productions, to its observed length, so that the anticipated length of /k/, X, is calculated proportionally from the ratio of the observed values of /p/ and /t/ in both the Native Speaker group and the L2 production group. X is equal to the observed length of the native speaker groups' /k/, times the quotient of the observed length of /p/ and /t/ in the L2 speaker group and that of the native speaker. In other words:

X = [Native Speaker /k/] * ([L2 /p ~ t/] ÷ [Native Speaker /p ~ t/])

Table 13: Expected VOT of /k/ Proportionately Calculated from Native Productions

	/p/ + /t/ Base	*Anticipated* /k/ (X)	Measured /k/
Level 1 (Novice)	53 ~ 54ms	*Ant.* 87.29 ~ 88.94ms	*Actual* 73ms
Level 2	51 ~ 55ms	*Ant.* 84 ~ 90.58ms	*Actual* 74ms
Level 3	36ms	*Anticipated* 59.29ms	*Actual* 60ms
Level 4 (Advanced)	29ms	*Anticipated* 47.76ms	*Actual* 49ms
Native Speaker (Control)	17ms	—	28ms

When the target production is taken into account, and the L2 productions are understood as proportional to the target the actual production of level 3 and level 4 speakers of these tokens is actually consistent between /p ~ t/ and /k/. In other words, while advanced participants produce longer /k/ tokens, they produce them with the same proportion to native speech that regulates their production of /p ~ t/. The measured-values of the beginner Level 1 and 2 participants are not consistent; however, beginner participants may also be less important in understanding how the availability of visual cues may impact L2 production. The O-subset of our study was also significant only in intermediate and advanced groups, and the effect witnessed in Menke and Face (2009) was also more significant in these. In Reeder (1996) intermediate and advanced L2 Spanish learners acquire deaspiration, a non-visible cue showing no bias across place of articulation towards the labial position.

5.2.3 Synthesis of Data on Visible and Non-Visible Cues

Here lies a key asymmetry for our conclusions. The Reeder (1996) deaspiration data suggests that non-visible cues indeed are produced without bias across place of articulation, while Menke and Face (2009) and our O-subset suggest that visual cues correlate with L2 production so that visual saliency boosts the production of the labial counterpart. The availability of labial over velar visual cues is seen where the articulatory cue is visible, and non-visible cues do not confirm to this pattern. This of course suggests that differing visual cue saliency may be the cause of this pattern and is consistent with the predictions of a possible Visual Cue Saliency L2 Effect.

5.3 Future Experimental Design and Repeatability

The current experimental findings can be strengthed both by repeating the present experiment and by designing similar, new experiments that test L2 learner's production of foreign phonological patterns that are both accompanied by a salient visual cue and produced at different places of articulation.

5.3.1 Future Repetitions of this Study

As discussed, token balance proved to be the primary source of obstruction in this experiment, and in future studies should be well controlled, weighted over other factors such as word stress and length. The O-subset allowed this experiment to control for this within a subset, however further repetitions of this experiment should have an equal number of tokens within each type that are followed by the same vowel. This would allow further descriptive insight onto the relationship between the following vowel and gemination, as Esposito (1999) has explored.

Also, while intonation and reading speed may be a concern, it is useful to compare the behavior of the short story and word list tasks. All four token types are dramatically longer than they are in the short story, and this gives good reason to believe that the short story (from which the O-subset is pulled) shows more realistic behavior. In addition to the length of word-list tokens suggesting that participants are employing careful reading, it was evident to the researcher that participants became aware that geminate-singleton contrasts were being specifically tested as the progressed through the word list. While there are drawbacks to the rigid structure of the short story, it has a more naturalistic design, so it is perhaps advisable to use as many tokens as possible (balanced for the following-vowel) than to use a word list.

Lastly, this study isolates the phonetic pattern of gemination from its context within the production of the word. If visibility of the phonetic pattern predicts its accuracy with which it is produced in an L2, then perhaps the vowel context of the geminate raises or lowers the visibility of the geminate. Depending on whether the underlying form of the word reflects this effect evidenced in production, we may expect more distinction in the production of geminate-singleton minimal pairs with low-vowels around the singleton or geminate than we might in one surrounded by two high-vowels. This could be incorporated as an additional step into a perception-production experiment as discussed in section 5.4.2.

5.3.2 Variations of this Study in Different Languages

Our results lead us to anticipate that the labial position will be generally privileged over the velar position when the L2 phonological pattern is visible, so that in L2 productive phonologies, phonetic patterns found in various places of articulation will be the most accurate in the labial position. I would be interesting to see if Japanese L2 speakers exhibit the same behavior with gemination, to study spirantization cross-linguistically in other languages (similar to the Menke and Face study), to examine the production of vowel allophony in languages with patterns shared between rounded and unrounded vowels, and to test “invisible” phonetic patterns that an L2 learner may have to learn across places of articulation, such as nasalization or voicing / VOT contrasts.

5.4 Further Questions

5.4.1 Visual and Auditory Saliency

Beyond the primary theory of the Visual Cue Saliency L2 Effect, the research conducted in this study raises questions about how the varying visibility of binary features may affect production. Upon what other dimensions, besides place of articulation, could cue visibility be a meaningful factor that impacts L2 production, or even typological patterning? As discussed in section 2.3, the Arai et al (2017) experiment suggests that gemination may be one such dimension, so that geminates are more visibly salient than singletons. The behavior of the visible geminate audio singleton tokens (A_FV_E tokens) suggest that visual geminacy is stronger than an auditory singleton because the visual geminacy causes significant perceptual interference, whereas the visible singleton audio geminate condition (A_TV_L tokens) shows that visual singletons are weak compared to auditory gemination. In the A_TV_L tokens the visual singletons can only cause comparably negligible interference, if any at all, to the audio geminate token.

This leads to further questions on the comparative strength of auditory and visual cues, and theoretically as to whether this imbalance in the strength of visual cues would have visible effect on L2 production over time as seen in differing acquisition and production of singleton and geminate tokens. While singletons certainly do not need to be "acquired" as foreign phonetic pattern as geminates often need to be by an L2 learner, the asymmetry between the strength of these visual cues, in light of how this misbalance between places of articulation seems to affect acquisition, leads us to wonder how other imbalances in visual cue strength between the presence and absence of a feature may lead to diverging behavior between the two forms in L2 production. In this context of this experiment, the study served primarily to show that visual gemination cues do disrupt audiovisual perception and that visual gemination cues are meaningful to phonological processing, but the question as how differing visual and auditory cue strengths, especially of binary features, can affect L2 phonological production and acquisition remains.

Conducting a version of the McGurk experiment with geminate and singleton tokens across different places of articulation may perhaps provide further insight into understanding the cross-section of visual- and audio-cue strength with place of articulation. Miller and Nicely (1955) tested participants to create confusion matrices of 16 English consonants (including all the stop consonants) found that place of articulation was the most frequently confused natural class over confusion in the manners of articulation. If place of articulation is a weak cue audibly, then not only is it curious as to the extent to which visual geminate cues could interrupt audio geminate cues given the elongation of the audio cue, but this also inquires to the role of visual cues in language processing and comprehension.

While the research presented hear is focused on how visual and auditory cues impact L2 acquisition, it is fathomable that there could be relationships between cue strength and assimilatory tendencies cross-linguistically. For example, many languages exhibit some form of place neutralization in the coda. Furthermore, coda place-assimilation happens informally and incidentally in causal speech frequently in languages such as English, in part because the audio cue is even weaker in a coda position (Tavabi et al 2009). By understanding the relative strengths of visual and auditory cues in comparison to one another, it may be possible to make predictions as to what types of place assimilations are most likely to occur.

5.4.2 Perceptual Studies

The primary finding that visual-saliency seems to affect the acquisition of visually noticable phonetic patterns across different places of articulation prompts further questioning into the underlying phonological perceptions of L2 speakers. Does this acquired pattern only exist in production, or to what degree does this L2 production correlate with perception? To what degree does L2 perception and do L2 speakers' underlying forms reflect the type of access that they have to different cues?

One potentially could conduct an experiment similar to the Arai et al experiment, testing second-language learners of a language with a singleton-geminate labial contrast at their accuracy in distinguishing normative (not-mixed) audiovisual, auditory, and visual, singleton and geminate tokens. The tokens would need to be real minimal pairs in order to try and understand underlying representation. One would anticipate that if L2 perception reflects production, in this case, participants would perceive /pː/ more distinctly from /p/ (P-condition) with greater accuracy than /kː/ and /k/ (K-condition) in all the visual and audiovisual categories. Furthermore, if the underlying forms were somehow clearer due to the visual information, we would anticipate participants to perceive the P-condition more accurately than the K distinction even in the audio condition. The visual cues are not available for the perception of this condition, so if participants are more accurate in the P-condition than the K-condition of this experiment, it suggests that the quality of their underlying forms reflects the availability of the visual cues available in perception.

5.4.3 Pedagogical Questions

Like all L2 production research, this research is useful in its interface with pedagogical research and questions. Particularly this experiment asks how instructors can help bridge the gap in production between the labial and velar positions. While posing specific pedagogical experiments in detail goes beyond the scope of this thesis, pedagogical work could explore how explicit extended production practice with less visible phonemes / allophones helps students produce both places of articulation more accurately. Practice makes perfect, but perhaps weighted practice would be more efficent.

This research may also interface well with research on metalinguistic knowledge in production. Do advanced students with a background in philology or linguistics produce tokens better than other students? While this experiment proposes that learners process visual cues at a subconscious level rather than explicitly, studies such as Wremble (2013) have found that metalinguistic knowledge is useful learning phoneme distinctions and allophonic patterns. Perhaps the incorporation of light non-theoretical phonetics information and even diagrams that mimic and explain the motion in less-visible places of articulation may help students better visualize and understand how their articulators should be shaped during sound production. While this experiment and theory deals with implicit visual information, research on the role of explict visual information during language instruction may assist in bridging any gaps left over from less visable places of articulation.

5.5 Conclusions

In conclusion, the primary finding of this experiment is that the strength of visual cues at different places of articulation may influence the accuracy with which L2 learners acquire phonetic patterns. Graduating Italian students (intermediate level) and Italian graduate students (advanced level) produce the P-condition much more distinctly than they produce the K-condition in a reading passage in the O-subset. In other words, when we control for vowel context, it seems that participants produce a more distinct labial geminate-singleton contrast than its velar counterpart.

Theoretically, we support this claim by pointing to the McGurk Effect, discussing the visibility of geminates, and discussing both Spanish L2 spirantization (Menke and Face 2009), a visible pattern that conforms to this pattern of the proposed Visual Cue Saliency L2 Effect, and Spanish L2 deaspiration of voiceless stops (Reeder, 1998), a non-visible pattern that generally behaves, proportionally, the same regardless of place of articulation. The non-visual pattern of deaspiration contextualizes this effect.

That said, before this theory of a Visual Cue Saliency L2 Effect can be purported more strongly, it needs to be further verified. It would be useful to retest the behavior of Italian geminates in a sample well-controlled for the following vowel's context like the O-subset, and further visible and non-visible phonetic patterns across different places of articulation need to be tested for their conformity or rejection of the aforementioned pattern. It would be useful to test visible spirantization, gemination, and the non-visible aspiration, patterns in other L2 languages. Beyond these it would be useful to test phonetic patterns, such as other voicing distinctions and labialization, to see if they conform to this pattern.

Lastly, this research suggests several related research questions. Variation in L2 production leads us to ask if there is corresponding variation in L2 perception and in the underlying representation of these phonemes based on the strength of visual cues from that place of articulation. Beyond place of articulation, we ask if there other dimensions with which differences in the strength of visual and auditory cues leave marks on a second language. Whenever research on L2 acquistion finds differing behaviors between the target language and learners, the research raises interesting pedagogical questions for instructors, who seek to pedagogically bridge the gap between the target and the actual production. Research on visual cue availability and L2 acquisition may prove to hold many interesting questions for current researchers.

Appendix A: Testing Materials

A.1: Questionnaire:

How many years have you studied Italian?

Have you traveled to Italy? If so, for how long?

Do you have family that speaks Italian? If so, how many hours a week do you speak Italian with them? When did you start speaking Italian with them?

What other languages do you have experience with?

How many major-level (>300 level) Italian courses have you taken, where Italian is the language of instruction?

Do you have any regional preferences or dialect in your Italian? Do you have any connection with any particular sub-region of Italy or the Italian-speaking world?

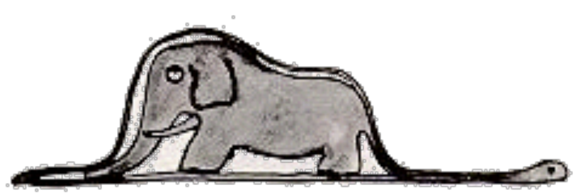
A.2: Elicitation Text

Quando avevo sei anni, in un libro sulle foreste primordiali, intitolato "Storie vissute della natura", vidi un magnifico disegno. Rappresentava un serpente boa nell'atto di inghiottire un animale. Eccovi la copia del disegno.

C'era scritto: "Appena bracca la sua preda, il boa la mette in bocca tutta intera, e non la mastica. Dopo di che non riesce più a muoversi e dorme durante i sei mesi che la digestione richiede". E si fermerebbe. Il movimento sarebbe energia sprecata.

Meditai a lungo sulle avventure della giungla. È a mia volta riuscii a tracciare il mio primo disegno. Il mio disegno numero uno. Era così:

Mostrai il mio capolavoro alle persone grandi, domandando se il disegno li spaventava. Ma mi risposero: "Spaventare? Perché mai, uno dovrebbe essere spaventato da un cappello?" Il mio disegno non era il disegno di un capo di vestiario. Era il disegno di un boa feroce che placava la fame con un elefante. Allo scopo che vedessero chiaramente che cos'era, disegnai l'interno del boa. Bisogna sempre spiegargliele le cose, ai grandi. Il mio disegno numero due si presentava così:



Questa volta mi risposero che lasciassi da parte i boa, sia di fuori che di dentro, è che mi applicassi invece alle mappe di geografia, alla storia, all'aritmetica e alla analisi logica. Fu così che a sei anni io rinunziai a quella che avrebbe potuto essere la mia gloriosa carriera di pittore. Il fallimento del mio disegno numero uno e del mio disegno numero due mi aveva lasciato con le pive nel sacco. I grandi non capiscono mai niente da soli e i bambini si affaticano a spiegargli tutto ogni volta. Allora scelsi un'altra carriera e imparai a pilotare gli aeroplani. Ho volato un po' sopra tutto il mondo: è veramente la geografia mi e stata molto utile. A colpo d'occhio posso distinguere la Cina dall'Arizona, e se uno si perde nella notte, questo sapere è di grande aiuto.

Ho incontrato molte persone importanti nella mia vita, ho vissuto a lungo in mezzo ai grandi. Li ho conosciuti intimamente, li ho osservati proprio da vicino. Ma l'opinione che avevo di loro non è molto migliorata.

Quando ne incontravo uno che mi sembrava di mente aperta, tentavo l'esperimento del mio disegno numero uno, che ho sempre conservato. Cercavo di capire così se era veramente una persona comprensiva. Ma, chiunque fosse, uomo o donna, mi rispondeva: "É un cappuccio". Che peccato! È allora non parlavo di boa, di foreste primitive, di stelle. Mi mettevo al suo livello. Gli parlavo di bridge, di golf, di politica, di cravatte. E lui era tutto soddisfatto di avere incontrato un uomo tanto sensibile. Non ero troppo immaginativo.

Così ho trascorso la mia vita solo, senza nessuno cui poter parlare, fino a sei anni fa quando ebbi un incidente col mio aeroplano, nel deserto del Sahara. Le parti dell'aeroplano si erano staccate ed erano sparse da tutte le parti. Qualche cosa si era rotto nel motore, e siccome non avevo con me né un meccanico, né dei passeggeri, mi accinsi da solo a cercare di riparare il guasto.

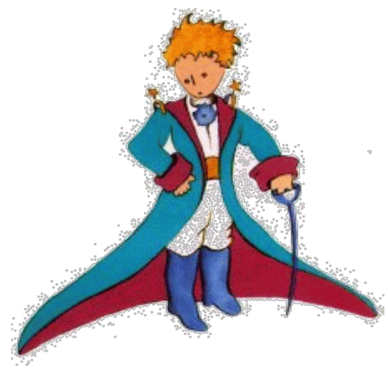
Era una questione di vita o di morte, perché avevo acqua da bere soltanto per una settimana, ed ero bloccato sotto un grappolo de stelle. La prima notte, dormii sulla sabbia, a mille miglia da qualsiasi abitazione umana. Ero più isolato che un marinaio abbandonato in mezzo all'oceano, su una zattera, dopo un naufragio. Mi sono seduto su un ceppo a pensare.

Potete immaginare il mio stupore di essere svegliato all'alba da una strana vocetta: "Mi disegni, per favore, una pecora?"

"Cosa?"

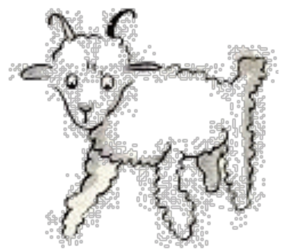
"Disegnami una pecora".

Mi trasecolo. Balzai in piedi come fossi stato colpito da un fulmine. Mi strofinai gli occhi più volte guardandomi attentamente intorno. E vidi una straordinaria personcina che mi stava esaminando con grande serietà. Qui potete vedere il miglior ritratto che riuscii a fare di lui, più tardi. Ma il mio disegno e molto meno affascinante del modello.



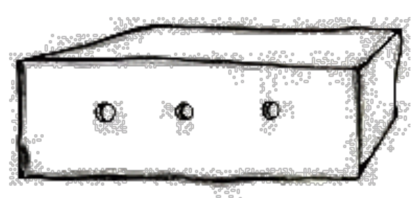
La colpa non è mia, pero ho mai imparato a disegnare altro che serpenti boa dal di fuori o serpenti boa dal di dentro.
Ora guardavo fisso l'improvvisa apparizione con gli occhi fuori dall'orbita per lo stupore. Dovete pensare che mi trovavo a mille miglia da una qualsiasi regione abitata, eppure il mio ometto non appariva smarrito in mezzo alle sabbie, né tramortito per la fatica, o per la fame, o per la sete, o per la paura. Niente di lui mi dava l'impressione di un bambino sperduto nel deserto secco, a mille miglia da qualsiasi abitazione umana. Quando finalmente potei parlare gli domandai: "Ma che cosa fai qui?"
Come tutta risposta, egli ripeté lentamente come si trattasse di cosa di molta importanza: "Per piacere, disegnami una pecora! "
Quando una situazione è così traboccante di mistero, non si osa disubbidire. Per assurdo che mi sembrasse, a mille miglia da ogni abitazione umana, e in pericolo di morte, tirai fuori dalla tasca un foglietto di carta e la penna stilografica. Ma poi ricordai che i miei studi si erano concentrati sulla geografia, sulla storia, sull'aritmetica e sulla grammatica e gli dissi, un po' di malumore, che non sapevo disegnare. Mi rispose: "Non importa. Disegnami una pecora! "Non avevo mai disegnato una pecora e allora feci per lui uno di quei due disegni che avevo fatto tante volte: quello del boa dal di fuori; e fui sorpreso di sentirmi rispondere: "No, no, no! Non voglio l'elefante dentro al boa. Il boa è molto pericoloso e l'elefante molto ingombrante. Dove vivo io tutto è molto piccolo. Ho bisogno di una pecora: disegnami una pecora".
Con un groppo in gola, sapevo che dovevo tentare. Feci il disegno.

Mi guardò attentamente, e poi appuntò. "No! Questa pecora è malaticcia. Fammene un'altra".



Feci un altro disegno.

Il mio amico mi sorrise gentilmente, con indulgenza. "Lo puoi vedere da te", disse, "che questa non è una pecora. È un ariete. Ha le corna". Rifeci il disegno una terza volta, ma fu rifiutato come i tre precedenti." Questa è troppo vecchia. Voglio una pecora che possa vivere a lungo".
Questa volta la mia pazienza era esaurita, avevo fretta di rimettere a posto il mio motore. Buttai giù un quarto disegno. E tirai fuori questa spiegazione: "Questa e soltanto la sua cassetta. La pecora che volevi sta dentro". Fui molto sorpreso di vedere il viso del mio piccolo giudice illuminarsi:



"Questo e proprio quello che speravo. Pensi che questa pecora dovrà avere una gran quantità d'erba?"
"Perché?"
"Perché dove vivo io, tutto e molto piccolo!"
"Ci sarà certamente abbastanza erba per lei, e molto piccola la pecora che ti ho data".
Si chinò sul disegno:
"Non così piccola che - oh guarda! - sì e messa a dormire…
"E fu così che feci la conoscenza del piccolo principe. E lui lo guardò.

Ci misi molto prima di concepire da dove venisse. Il principe, che mi faceva una domanda dopo l'altra, pareva che non sentisse mai le mie.
Sono state le parole dette per caso, che poco a poco, mi hanno rivelato tutto. Così, quando vide per la prima volta il mio aeroplano (non lo disegnerò perché sarebbe troppo complicato per me), mi domandò:
"Che cos'e questa cosa?"
"Non è una cosa… vola. È un aeroplano. È il mio aeroplano".
Ero molto fiero di fargli sapere che volavo.
Allora grido: "Come? Sei caduto dal cielo!"
"Si", risposi modestamente.
"Ah! Questa è buffa!"
E il piccolo principe scoppio in una bella risata che mi irritò. Voglio che le mie disgrazie siano prese sul serio. Poi riprese:
"Allora anche tu vieni dal cielo! Di quale pianeta sei?"
Intravidi una luce, nel mistero della sua presenza, e lo interrogai bruscamente:
"Tu vieni dunque da un altro pianeta?"

Ma non mi rispose. Scrollò gentilmente il capo osservando l'aeroplano.

"Certo che su quello non puoi venire da molto lontano!"

e si sprofondò in una lunga meditazione. Poi, tirando fuori dalla tasca la mia pecora, sprofondo nella contemplazione del suo tesoro.

Voi potete bene immaginare come io fossi incuriosito da quella mezza confidenza su "gli altri pianeti". Cercai dunque di tirargli fuori qualche altra cosa:

"Da dove vieni, ometto? Dov'è la tua casa? Dove vuoi portare la mia pecora?"

Mi rispose dopo un riposo meditativo: "Quello che c'è di buono, è che la cassetta che mi hai dato, le servirà da casa per la notte".

"Certo. E se sei buono ti darò pure una corda di canapa per legare la pecora durante il giorno. E un paletto."

La mia proposta scandalizzò il piccolo principe.

"Legarla? Che buffa idea!"

"Ma se non la leghi, andrà in giro è si perderà…"

Il mio amico mi guardò come se fossi sciocco.

"Ma dove vuoi che vada!"

"Dappertutto. Il mio pianeta è microscopico! C'è solo la mia casa, un albero di fico e un pioppo."

Una pecora non può scappare.

A.3: Consent Form

Study: Second Language Phonological Learning

This study explores how learners acquire pronunciation in second-languages. During this study, you will be asked to listen to an Italian speaker pronounce several words. Your goal is to learn how to pronounce them to the best of your ability. You can practice saying the word after hearing the speaker present the word, and at the end we will record you reading the words. These words are common Italian words that may be covered in an introductory Italian class. All data will be anonymous, only the researchers will have access to the recordings, and participation is optional. Consent to participate in this study can be withdrawn at any time. There are no known risks and the only the benefit is advancing the cause of science.

In this study, you will learn a series of words from an Italian speaker. He will repeat each word twice, and you can pause the video to practice the word if you wish. At the end, you will be asked to read back the list of words to the best of your ability.

Signature: _______________________________________ Date: ____________

Appendix B: Token Information

B.1 Token Balance Extended Summary Chart

Table 14: Token Balance Extended Summary Chart

	/kː/	/pː/	/k/	/p/
V After Height				
Low	8	3	7	1
Mid	5	8	8	10
High	2	4	0	4
V After Backness				
Back	5	8	8	7
Mid	8	3	7	1
Front	2	4	0	7
V After Quality				
a	8	3	7	1
e	0	3	0	3
i	2	1	0	4
o	5	5	8	7
u	0	3	0	0
Back				
Stress After	7	4	4	8
Stress Before	8	7	6	3
Syllable Count	2.666666667	2.866666667	3.066666667	3.133333333

B.2 Full token Balance Chart

Table 15: Full Token Balance Chart

Token #	Length	P. of A.	Token	V After Height	V After Backness	V Aft Quality	Stress	Syllable Count
/k:/								
2	l	k	*eccovi*	Mid	Back	o	after	3
3	l	k	*bracca*	Low	Mid	a	before	2
4	l	k	*bocca*	Low	Mid	a	before	2
15	l	k	*sacco*	Mid	Back	o	before	2
18	l	k	*d'occhio*	High	Front	i	before	2
24	l	k	*peccato*	Low	Mid	a	after	3
27	l	k	*staccate*	Low	Mid	a	after	4
28	l	k	*siccome*	Mid	Back	o	after	3
29	l	k	*meccanico*	Low	Mid	a	after	4
30	l	k	*acqua*	Low	Mid	a	before	2
31	l	k	*bloccato*	Low	Mid	a	after	3
41	l	k	*secco*	Mid	Back	o	before	2
43	l	k	*traboccante*	Low	Mid	a	after	4
46	l	k	*piccolo*	Mid	Back	o	before	2
50	l	k	*vecchio*	High	Front	i	before	2
/p:/								
8	l	p	*cappello*	Mid	Front	e	after	3
13	l	p	*mappe*	Mid	Front	e	before	2
23	l	p	*cappuccio*	High	Back	u	after	3
26	l	p	*troppo*	Mid	Back	o	before	2
32	l	p	*grappolo*	Mid	Back	o	before	3
33	l	p	*ceppo*	Mid	Back	o	before	2
37	l	p	*apparizione*	Low	Mid	a	n/a	5
38	l	p	*eppure*	High	Back	u	after	3
39	l	p	*appariva*	Low	Mid	a	n/a	4
47	l	p	*groppo*	Mid	Back	o	before	2
48	l	p	*appuntò*	High	Back	u	n/a	3
53	l	p	*scoppio*	High	Front	i	before	2
57	l	p	*dappertutto*	Mid	Front	e	n/a	4
59	l	p	*pioppo*	Mid	Back	o	before	2
60	l	p	*scappare*	Low	Mid	a	after	3
/k/								
1	s	k	*magnifico*	Mid	Back	o	n/a	4
5	s	k	*mastica*	Low	Mid	a	n/a	3
10	s	k	*placava*	Low	Mid	a	after	3
12	s	k	*applicassi*	Low	Mid	a	after	4

14	s	k	*logica*	Low	Mid	a	n/a	3
17	s	k	*affaticano*	Low	Mid	a	before	5
25	s	k	*politica*	Low	Mid	a	n/a	3
35	s	k	*pecora*	Mid	Back	o	before	3
36	s	k	*trasecolo*	Mid	Back	o	after	2
40	s	k	*fatica*	Low	Mid	a	before	3
44	s	k	*pericolo*	Mid	Back	o	before	4
45	s	k	*ricordai*	Mid	Back	o	n/a	3
49	s	k	*amico*	Mid	Back	o	after	2
52	s	k	*poco*	Mid	Back	o	before	2
58	s	k	*fico*	Mid	Back	o	before	2
/p/								
6	s	p	*dopo*	Mid	Back	o	before	2
7	s	p	*capolovoro*	Mid	Back	o	n/a	5
9	s	p	*capo*	Mid	Back	o	before	2
11	s	p	*scopo*	Mid	Back	o	before	2
16	s	p	*capiscono*	High	Front	i	after	4
19	s	p	*sapere*	Mid	Front	e	after	3
20	s	p	*l'opinione*	High	Front	i	n/a	4
21	s	p	*aperta*	Mid	Front	e	after	3
22	s	p	*capire*	High	Front	i	after	3
34	s	p	*stupore*	Mid	Back	o	after	3
42	s	p	*ripeté*	Mid	Front	e	n/a	3
51	s	p	*concepire*	High	Front	i	after	4
54	s	p	*riposo*	Mid	Back	o	after	3
55	s	p	*canapa*	Low	Mid	a	n/a	3
56	s	p	*proposta*	Mid	Back	o	after	3

Table 16: O-subset Balance Chart

Token #	Length	P. of A.	Token	V Aft Height	V Aft Backness	V Aft Quality	Stress	# of Syls.
2	l	k	eccovi	Mid	Back	o	after	3
15	l	k	sacco	Mid	Back	o	before	2
28	l	k	siccome	Mid	Back	o	after	3
41	l	k	secco	Mid	Back	o	before	2
46	l	k	piccolo	Mid	Back	o	before	2
						Tokens Aft: 2		
26	l	p	troppo	Mid	Back	o	before	2
32	l	p	grappolo	Mid	Back	o	before	3
33	l	p	ceppo	Mid	Back	o	before	2
47	l	p	groppo	Mid	Back	o	before	2
59	l	p	pioppo	Mid	Back	o	before	2

						Tokens Aft: 0		
1	s	k	magnifico	Mid	Back	o	n/a	4
35	s	k	pecora	Mid	Back	o	before	3
36	s	k	trasecolo	Mid	Back	o	after	2
44	s	k	pericolo	Mid	Back	o	before	4
45	s	k	ricordai	Mid	Back	o	n/a	3
49	s	k	amico	Mid	Back	o	after	2
52	s	k	poco	Mid	Back	o	before	2
58	s	k	fico	Mid	Back	o	before	2
						Tokens Aft: 2		
6	s	p	dopo	Mid	Back	o	before	2
7	s	p	capolovoro	Mid	Back	o	n/a	5
9	s	p	capo	Mid	Back	o	before	2
11	s	p	scopo	Mid	Back	o	before	2
34	s	p	stupore	Mid	Back	o	after	3
54	s	p	riposo	Mid	Back	o	after	3
56	s	p	proposta	Mid	Back	o	after	3
						Tokens Aft: 3		

REFERENCES

Arai, T., Iwagami, E., & Yanagisawa, E. (2017). Seeing closing gesture of articulators affects speech perception of geminate consonants. *The Journal of the Acoustical Society of America, 141*(3). doi:10.1121/1.4978343

Blevins, Juliette (2008). *Explaining Diversity in Geminate Consonant Inventories.* URL: https://www.eva.mpg.de/lingua/conference/08_springschool/pdf/course_materials/blevins_evening_lecture.pdf (Visited: 3/12/18)

Boersma, Paul & Weenink, David (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.31, retrieved 22 August 2017 from http://www.praat.org/

Esposito A. and Di Benedetto, M. G (1999). Acoustical and perceptual study of gemination in Italian stops. *The Journal of the Acoustic Society of America*, *106*, p 2051-2062.

Face, T. L., & Menke, M. R. (2009). Acquisition of the Spanish Voiced Spirants by Second Language Learners. In J. Collentine, M. García, B. Lafford, & F. M. Marín (Eds.), *Selected Proceedings of the 11th Hispanic Linguistics Symposium* (pp. 39-52). Somerville, MA: Cascadilla.

Kawahara, Shigeto. (2005). Voicing and geminacy in Japanese: An acoustic and perceptual study. In K. Flack and S. Kawahara (eds.) *University of Massachusetts Occasional Papers in Linguistics 31: Papers in Experimental Phonetics and Phonology.* pp. 87-120.

Lombardi, L. (2002). Coronal epenthesis and markedness. *Phonology, 19*(2). pp. 219-251. doi:10.1017/S0952675702004323

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748. doi:10.1038/264746a0

Miller, G. A., & Nicely, P. E. (1955). An Analysis of Perceptual Confusions Among Some English Consonants. *The Journal of the Acoustical Society of America, 27*(2), 338-352. doi:10.1121/1.1907526

Piñeros, Carlos Eduardo (2009) *Estructura de los sonidos del español;* Upper Saddle River, NJ, Pearson Education Inc. ISBN-13: 978-0-13-194437-4

Stevens, M. (2011). Consonant length in Italian: Gemination, degemination and preaspiration. In S. M. Alvord (Ed.), Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology (pp. 21-32). Somerville: Cascadilla Proceedings Project.

Payne, E. (2005). Phonetic variation in Italian consonant gemination. *Journal of the International Phonetic Association, 35*(2), 153-181. doi:10.1017/S0025100305002240

Reeder, J. T. (1998). English Speakers' Acquisition of Voiceless Stops and Trills in L2 Spanish. *Texas Papers in Foreign Language Education, 3*(3), 101-118.

Tavabi, K., Elling, L., Dobel, C., Pantev, C., & Zwitserlood, P. (2009). Effects of Place of Articulation Changes on Auditory Neural Activity: A Magnetoencephalography Study. *PLoS ONE, 4*(2). doi:10.1371/journal.pone.0004452

Tiippana, K. (2014). What is the McGurk effect? *Frontiers in Psychology, 5*, 1-3. doi:10.3389/fpsyg.2014.00725

Wrembel, Magdalena. (2013). 'Metalinguistic awareness in third language phonological acquisition'. 119-143.

Yanagisawa, E., and Arai, T. (2015a). "Effects of formant transition and intensity damping of preceding vowel off-glide on perception of Japanese geminate consonant Sokuon," J. Acoust. Soc. Jpn. 71 (10), 505–515.