Gradience and Variability of Intervocalic /s/ Voicing in Highland Ecuadorian Spanish

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Christina Marie García

Graduate Program in Spanish and Portuguese

The Ohio State University

2015

Dissertation Committee:

Dr. Terrell A. Morgan, Advisor

Dr. Rebeka Campos-Astorkiza, Advisor

Dr. Scott Schwenter

Dr. Anna M. Babel
Copyright by

Christina Marie García

2015
ABSTRACT

This dissertation focuses on Highland Ecuadorian Spanish (HES), more specifically in the city of Loja, exploring how linguistic and social factors interact to condition a sound change in progress. While Lojano Spanish has many distinctive features, I examine intervocalic /s/ voicing, that is the variable pronunciation of /s/ as [s] or [z] (for example [losamiɣos] ~ [lozamiɣos] ‘the friends’). This feature has been attested in a few modern dialects of Spanish, most prominently in HES (cf. Robinson 1979) and Catalanian Spanish (McKinnon 2012, Davidson 2014). The present work provides a comprehensive look at this feature in Lojano Spanish by investigating both production and perception.

First, I analyze the production of this feature by carrying out a detailed acoustic analysis of 31 recordings with native Lojanos. These recordings were collected during two fieldwork trips spent in Loja in 2010 and 2013. The recordings include both sociolinguistic interviews, as well as a reading task. I measured over 2,969 tokens of /s/ for percent voicing and examined the influence of the following factors: word position, stress, speech rate, preceding/following vowel, and participants’ age and gender. Voicing is analyzed as both a continuous dependent variable (percent voicing) as well as a categorical variable (voicing category). Following Campos-Astorkiza (2014), voicing category is considered a tripartite distinction: voiceless (0-20%), partially voiced (20-
90%) and fully voiced (100%). Overall, the production results show that /s/ voicing in Lojano Spanish is not categorical, but rather is a gradient, variable process. Increasingly voiced realizations are more likely in faster speech; in word final and initial contexts as opposed to medial; when /s/ is between unstressed syllables; before non-high vowels; and in the speech of younger participants and male participants.

Second, I explore the perception of /s/ voicing in Lojano Spanish by conducting an online experiment. Twenty-four native speakers of HES participated in two tasks in the online experiment, the design of which was adapted from Boomershine et al. (2008). In the first, a similarity rating task, participants heard pairs of audio files and rated these pairs on a scale of 1 very similar to 6 very different. The second was a traditional discrimination task in which the participants heard these same pairs and decided if the two tokens were the same or different (Liberman et al. 1957). For the stimuli, I recorded a native speaker of Lojano Spanish saying sequences of words that pertain to the three word positions (medial: asa, initial: la saca, final: las ata), producing voiced and voiceless variants of each. The pairs of audio files either had the same type of voicing, “identity” pairs ([asa] vs. [asa]) or different voicing ([aza] vs. [asa]), “difference” pairs. The results show that the listeners hear difference pairs as different more frequently than identity pairs; however, the robustness of this effect depends on word position, with the most robust effect observed in final position, followed by initial position, and the least robust effect in medial position. This demonstrates a connection between production and perception since it is precisely the environments in which Lojanos voice /s/ more (word-finally and initially) that they perceive most readily the difference between [s] and [z].
ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance and support of many people. First and foremost, I want to thank my advisors, Dr. Terrell Morgan and Dr. Rebeka Campos-Astorkiza. Terrell has been my constant guidance literally from day one and recognized my potential as a scholar before I even knew what the word “scholar” meant. He set an outstanding example for me of what it means to be a good professor, in every sense of the word, and I will never be able to repay the opportunities he has bestowed upon me during my time at Ohio State and the incredible experience that has come from these opportunities. Thank you for believing in me and always being in my corner when I needed it.

Rebeka has been a steady source of inspiration and support for me, both in research and in the classroom. I greatly appreciate her extreme dedication to my work, which has manifested itself in everything from teaching me the ins and outs of acoustic analysis to helping me think of stimuli, and much, much more. Her consistency and commitment has been one of the main forces driving me to finish my dissertation and it will continue to motivate me in the future. Thank you for always making time to meet with me, even from afar.

I would also like to thank my committee members, Dr. Scott Schwenter and Dr. Anna Babel, for bringing their valuable perspective to my work over the years. I am very
grateful that Scott awakened the variationist spirit in me so long ago, gave me the opportunity to present at my very first conference, and for allowing me to crash advisee Panera outings to talk over stats, look at abstracts, discuss ideas, and the like. I am thrilled to have a fellow andinista in Anna on my committee and appreciate that she always challenges me to see beyond the data. Additionally, I thank the faculty in the Linguistics Department who challenged me to develop into the scholar I am today: Judith Tonhauser, Cynthia Clopper, and Kathryn Campell-Kibler. I am particularly indebted to Kathryn for showing me the ways of all types of sociolinguistic research and teaching me much of what I know about R.

The collection of the data in this dissertation took place during two trips to Loja, which were made possible by different sources of funding. The first trip was funded by an International Affairs Grant from the OSU Office of International Affairs and the Bulkatin-Pardo Graduate Award. The second trip to Loja was funded by a Tinker Foundation Field Research Grant. I acknowledge as well the transcription help of Wendy Feliz, and the expert statistics consulting of Elizabeth Petraglia, Allen Po-hsu, and Sophie Nguyen.

I would also like to recognize my amazing colleagues and friends. Hannah Washington and Mary Beaton have been with me in this adventure since the very first day of classes and are the two best dissertation buddies anyone could ask for. Hannah made sure I was always on track with just about everything, Mary gave me all the supportive hugs I all too often needed, and they both made this whole process a lot more enjoyable and helped to keep me sane. I also thank Abby Walker for being the best collaborator on the planet and putting so much hard work and ingenuity into our joint
endeavors. ¡Mano a mano para siempre! To those who paved the road ahead of me, I am so fortunate to have had your guidance and advise. I am grateful to Jenny Barajas for her constant positive attitude, giving me the encouragement I needed even at the most challenging times. I have always looked up to Sonia Barnes and Mary Johnson and their work has both inspired and informed my own. This dissertation quite honestly might not have been written without the vision of and many discussions with Whitney Chappell, and I am extremely appreciative that she first pointed out this linguistic feature to me. I also appreciate the others who brightened my existence at Hagerty Hall, Meghan Dabkowski with her laugh that can be heard down the hall, and Karen Lopez-Alonso with her infectious smile.

Además quiero agradecer a todos los participantes y las personas que colaboraron con el reclutamiento de este estudio. Gracias por compartir sus historias y sus vidas conmigo. A las familias Samaniego y García, estoy perennemente agradecida por su constante apoyo, cariño, y hospitalidad inmensa. Me hicieron enamorarme de Loja y siempre llevaré conmigo una parte de esta hermosa ciudad.

I would also like to recognize the never-ending love and support of my family. I am thankful to my parents, Dan and Sandy, for always believing in me no matter what I wanted to do in life, and to my brother, Kevin, for providing the soundtrack to my dissertation and encouraging me all along. Finally, quiero reconocerles a mi esposo, Santiago, y mi perrito, Rocco. Sin la constante compañía, y besos peludos, de Rocco durante el proceso de escribir la tesis, estoy segura de que me habría vuelta loca. A Santiago, mi fundación en todo, le agradezco eternamente los sacrificios que ha hecho por apoyarme, por siempre creer en mí, y por su paciencia infinita. Sin tu devoción y
ánimo, nada de eso hubiera sido posible y tampoco tendría importancia en absoluto. Te amo con todo mi corazón.
VITA

2009…………………………………………B.A. Spanish and Linguistics, The Ohio State University

2011…………………………………………M.A. Hispanic Linguistics, The Ohio State University

Publications


Fields of Study

Major: Spanish and Portuguese
TABLE OF CONTENTS

ABSTRACT ........................................................................................................ ii

ACKNOWLEDGEMENTS .............................................................................. iv

VITA ................................................................................................................. viii

TABLE OF CONTENTS .................................................................................... ix

LIST OF TABLES ........................................................................................... xii

LIST OF FIGURES .......................................................................................... xiii

CHAPTER 1: INTRODUCTION ............................................................................. 1
  1.1 Description of the phenomenon ................................................................. 1
  1.2 Overview and goals ................................................................................... 3
  1.3 Research questions ................................................................................... 6

CHAPTER 2: BACKGROUND .............................................................................. 8
  2.1 Historical developments of the region ....................................................... 8
    2.1.1 Historical background of the Ecuadorian Andes .................................. 8
    2.1.2 Loja then and now ............................................................................. 10
  2.2 Highland Ecuadorian Spanish (HES) ......................................................... 15
  2.3 Spanish sibilants ....................................................................................... 18
    2.3.1 History ............................................................................................... 18
    2.3.2 Current /s/ phenomena ...................................................................... 20
  2.4 Intervocalic /s/ voicing ........................................................................... 22
    2.4.1 In other varieties ............................................................................... 22
    2.4.2 Intervocalic /s/ in Highland Ecuadorian Spanish .................................. 23
    2.4.3 Previous methodologies ................................................................. 25
    2.4.4 Linguistic factors examined ............................................................. 31
    2.4.5 Social factors examined ................................................................. 36
    2.4.6 Origins of intervocalic /s/ voicing ...................................................... 43
  2.5 Perception ................................................................................................ 47
    2.5.1 Categorical and allophonic perception .............................................. 48
    2.5.2 Perception of fricative voicing ........................................................... 53
  2.6 Research questions .................................................................................. 59
5.1.3 Conditioning factors .................................................................................................................. 153
  5.1.3.1 Speech rate.......................................................................................................................... 154
  5.1.3.2 Stress................................................................................................................................... 155
  5.1.3.3 Preceding/following vowel ............................................................................................... 157
  5.1.3.4 Position within the word ...................................................................................................... 158
  5.1.3.5 Gender ............................................................................................................................... 160
  5.1.3.6 Age ...................................................................................................................................... 161
  5.1.4 Theoretical implications of production data ........................................................................... 161
    5.1.4.1 Previous formal analyses ................................................................................................. 161
    5.1.4.2 Proposed analysis ............................................................................................................. 169
5.2 Perception of intervocalic /s/ voicing ......................................................................................... 182
  5.2.1 Summary of perception results ............................................................................................ 183
  5.2.2 Effect of word position .......................................................................................................... 184
  5.2.3 Comparison to previous studies ........................................................................................... 186
5.3 Conclusions ................................................................................................................................ 188
  5.3.1 Status of intervocalic /s/ in Lojano Spanish ........................................................................... 188
  5.3.2 Contributions ......................................................................................................................... 190
  5.4 Future research ......................................................................................................................... 193

REFERENCES .................................................................................................................................... 197

APPENDIX A: QUESTION USED IN THE INTERVIEWS ................................................................. 210

APPENDIX B: LIST OF STIMULI IN READING TASK ............................................................... 211

APPENDIX C: STIMULI USED IN THE PERCEPTION EXPERIMENT ..................................... 214
LIST OF TABLES

Table 1: Target words used in reading task .......................................................... 66
Table 2: Means for percent voicing ...................................................................... 77
Table 3: Distribution of tokens by voicing category ............................................... 78
Table 4: Results for read data, continuous analysis ............................................. 80
Table 5: Results for read data, categorical analysis ............................................. 84
Table 6: Voicing category by stress; read data ...................................................... 87
Table 7: Results for interview data, continuous analysis .................................... 94
Table 8: Results for interview data, categorical analysis ................................... 98
Table 9: Voicing category by stress; interview data ............................................. 101
Table 10: Voicing category by stress; interview data .......................................... 113
Table 11: Summary of significant predictors in production results ..................... 114
Table 12: Mean response of training stimuli, Blocks 1-3 .................................. 125
Table 13: Responses for training stimuli, Blocks 4-6 ......................................... 126
Table 14: Results for Block 1, similarity rating task of medial position ............. 127
Table 15: Results for Block 2, similarity rating task of initial position .............. 129
Table 16: Results for Block 3, similarity rating task of final position ............... 131
Table 17: Results for Block 4, discrimination task for medial position ............ 134
Table 18: Results for Block 5, discrimination task for initial position ............... 135
Table 19: Results for Block 6, discrimination task for final position ............... 135
Table 20: Summary of significant predictors in production results .................. 154
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regional Map of Ecuador</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>Map of Ecuador with linguistic regions according to Lipski (1994)</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>Historical development of Spanish sibilants</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>Bar graph showing participants’ age and gender</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>Read production of <em>proceso</em> from participant F10 – 27% voiced</td>
<td>68</td>
</tr>
<tr>
<td>6</td>
<td>Read production of <em>caña</em> from participant M7 – 100% voiced</td>
<td>68</td>
</tr>
<tr>
<td>7</td>
<td>Histogram of percent voicing</td>
<td>73</td>
</tr>
<tr>
<td>8</td>
<td>Histogram of percent voicing for read data and interview data</td>
<td>75</td>
</tr>
<tr>
<td>9</td>
<td>Standard deviation of percent voicing in read data and interview data</td>
<td>75</td>
</tr>
<tr>
<td>10</td>
<td>Scatterplot of percent voicing versus speech rate; read data</td>
<td>81</td>
</tr>
<tr>
<td>11</td>
<td>Voicing category by gender; read data</td>
<td>85</td>
</tr>
<tr>
<td>12</td>
<td>Voicing category by speech rate; read data</td>
<td>86</td>
</tr>
<tr>
<td>13</td>
<td>Conditional inference tree for read data</td>
<td>91</td>
</tr>
<tr>
<td>14</td>
<td>Scatterplot of percent voicing versus speech rate; interview data</td>
<td>95</td>
</tr>
<tr>
<td>15</td>
<td>Voicing category by speech rate; interview data</td>
<td>99</td>
</tr>
<tr>
<td>16</td>
<td>Voicing category by word position; interview data</td>
<td>100</td>
</tr>
<tr>
<td>17</td>
<td>Voicing category by following vowel; interview data</td>
<td>102</td>
</tr>
<tr>
<td>18</td>
<td>Conditional inference tree for interview data</td>
<td>111</td>
</tr>
<tr>
<td>19</td>
<td>Speaker effect in read data based on categorical model</td>
<td>115</td>
</tr>
<tr>
<td>20</td>
<td>Speaker effect for interview data based on categorical model</td>
<td>116</td>
</tr>
<tr>
<td>21</td>
<td>Speech rate (syllables per second) by percent voicing, by speaker</td>
<td>111</td>
</tr>
<tr>
<td>22</td>
<td>Screenshot of the online survey, similarity rating task in Blocks 1-3</td>
<td>119</td>
</tr>
<tr>
<td>23</td>
<td>Screenshot of the online survey, discrimination task in Blocks 4-6</td>
<td>120</td>
</tr>
<tr>
<td>24</td>
<td>Percentage of responses for Block 1, difference pairs and identity pairs</td>
<td>128</td>
</tr>
<tr>
<td>25</td>
<td>Percentage of responses for Block 2, difference pairs and identity pairs</td>
<td>130</td>
</tr>
<tr>
<td>26</td>
<td>Percentage of responses for Block 3, difference pairs and identity pairs</td>
<td>132</td>
</tr>
<tr>
<td>27</td>
<td>Range of responses by participant, Blocks 1-3</td>
<td>141</td>
</tr>
<tr>
<td>28</td>
<td>Responses by participant, Blocks 4-6</td>
<td>143</td>
</tr>
<tr>
<td>29</td>
<td>Histogram of percent voicing</td>
<td>150</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION

1.1 Description of the phenomenon

The variation found in modern Spanish in the production of /s/ is of considerable interest to Hispanic Linguistics given that the realization of this sound is a key shibboleth for differentiating dialects of Spanish (cf. Lipski 1994), among linguists and non-specialists alike. In fact, sibilants in general have long been a topic of concern to Hispanic linguists since the development of these sounds has traced the evolution of Spanish.

Old Spanish had six sibilants, which included a phonemic voicing distinction between /s/ and /z/ (cf. Penny 1991), contrastive in intervocalic position ([pasa] passa ‘he/she passes’ vs. [kaza] casa ‘house’). During the transition to Modern Spanish, the voicing contrast in the six sibilants was lost, merging /s/ and /z/ to a single phoneme: /s/. This phonological change sets Spanish apart from most other Romance languages, in which the phonemic status of /s/ and /z/ is maintained (Recasens 2002). In Modern Spanish, [z] is found only as an allophone of /s/, whereby /s/ voices before voiced consonants (e.g. desde ‘from’ [dezəðe]). According to the traditional phonological rules of Modern Spanish (cf. Morgan 2010), intervocalic /s/ is pronounced as voiceless (los amigos ‘the friends’ [losamiɣos]).
Contrary to the traditional description of the distribution of [s] and [z] in Modern Spanish, previous studies have found that a few present-day dialects exhibit voicing of intervocalic /s/, both within and across words. In these dialects, intervocalic /s/ may be produced as [z], as in example (1).

(1) Standard Pronunciation Dialect with voicing

<table>
<thead>
<tr>
<th>English</th>
<th>[Spanish]</th>
<th>[voiced]</th>
</tr>
</thead>
<tbody>
<tr>
<td>la sopa ‘the soup’</td>
<td>[la.so.pa]</td>
<td>[la.zo.pa]</td>
</tr>
<tr>
<td>casa ‘house’</td>
<td>[ka.sa]</td>
<td>[ka.za]</td>
</tr>
<tr>
<td>los otros ‘the others’</td>
<td>[lo.so.tros]</td>
<td>[lo.zo.tros]</td>
</tr>
</tbody>
</table>

This feature has been attested in the Spanish of the following places: Catalonia (McKinnon 2012, Davidson 2014), Madrid (Torreira & Ernestus 2012, Hualde & Prieto 2014), Central Spain (Torreblanca 1986), and Highland Ecuador (Robinson 1979, Lipski 1989, Calle 2010, Chappell 2011, Strycharczuk et al. 2013). The specific realization of intervocalic /s/ varies greatly between these dialects and previous studies have differed in whether considering this process to be categorical or gradient. While some studies claim that intervocalic /s/ voicing is a categorical process (Robinson 1979, Lipski 1989), modern studies using acoustic analysis have shown that it may be gradient in nature (cf. Strycharczuk et al. 2013), and dependent on both linguistic and social factors, such as position within a word and speaker gender. This debate on whether intervocalic /s/ voicing is a categorical or gradient process can inform the larger question of how phonological processes are realized phonetically.
1.2 Overview and goals

The goal of this dissertation is to examine the production and perception of intervocalic /s/ voicing in the Spanish spoken in Loja, Ecuador. As will be described in detail later, the Spanish spoken in the city of Loja (Lojano Spanish) is one of the sub-dialects of Highland Ecuadorian Spanish (HES). Studies on this variety are few and far between and it is unclear what the realization of intervocalic /s/ is in Lojano Spanish. Based on interviews with two male Lojanos, Robinson (1979) concludes that intervocalic /s/ voicing is not present in this dialect. On the other hand, Chappell (2011) provides anecdotal evidence from one male Lojano speaker who variably voices /s/ in intervocalic position. Given these conflicting reports and the small number of speakers observed, more production data from Loja is needed to verify whether intervocalic /s/ voicing is a feature of this dialect.

Beyond filling a gap, Lojano Spanish is noteworthy because of its development and present-day situation. Until relatively recently, the city of Loja was all but isolated from other Highland towns due to the surrounding mountainous terrain and poor road construction. Thus, Lojano Spanish has remained relatively conservative and has not been in significant contact with other varieties of HES until recently. Furthermore, while there are several indigenous languages spoken in the province of Loja, their presence and influence in the city of Loja has been minimal, while language contact has been more of a factor in other areas of the Highlnds. The present-day sociolinguistic situation is in transition in Loja. In the past decade, the road connecting Loja to other cities in Ecuador has been drastically improved, and so communication and exchange with these cities has increased. While it has not been empirically documented, it is likely that Lojano Spanish
is changing due to this increased communication. Given the presence of intervocalic /s/ voicing in other HES varieties, studying this feature can help us not only to describe Lojano Spanish, but can also help us understand the role of linguistic features in this community which finds itself at a social crossroads.

Despite the fact that there are several production studies on intervocalic /s/ voicing, how this feature is perceived by native speakers is unclear. Robinson (1979) claims that speakers of HES can distinguish between would-be minimal pairs such as ha sido ‘it has been’ [asiðo] and has ido ‘you have gone’ [aziðo], but this has not been corroborated by empirical evidence. More broadly, studies on the perception of fricatives in Spanish are scarce and have not been able to provide clear support for whether native speakers can perceive the difference between [s] and [z] (Borzone de Manrique & Massone 1981, Widdison 1995). Thus, while linguists are able to note production differences between [s] and [z], it is unknown whether naïve native speakers of HES can distinguish [s] and [z] in intervocalic position or whether they perceive these two sounds as one and the same.

This dissertation provides data in two main areas: production and allophonic perception. The production data comes from sociolinguistic interviews that I collected during December 2010 and July/August 2013. I recorded interviews with each of the 31 participants in two speech styles: interview (semi-spontaneous) and read speech. The participants are all native Lojanos and the sample is balanced for age and gender. Almost 3,000 tokens of intervocalic /s/ were extracted from these interviews. Each token of intervocalic /s/ was analyzed acoustically in WaveSurfer (Sjölander & Beskow 2010) and measured for percent voicing by dividing the total voicing during the fricative by the total
fricative duration. A series of statistical models were fit to the data to examine the effect of the following factors: position within the word, speech rate, preceding/following vowel, stress, and participant age and gender. Separate analyses were run for interview and read speech, and two analyses were run in each case: one with percent voicing as the continuous dependent measure and one with voicing category (unvoiced, partially voiced, fully voiced) as the dependent variable.

There are several areas of research that are relevant to the production results. As mentioned in section 1.1, these results add to the growing body of evidence coming out of laboratory phonology that shows the line between phonetics and phonology is not quite as clear-cut as was once thought. The framework that I use to discuss this interplay is Articulatory Phonology (Browman & Goldstein 1989). In this framework, the most basic unit of speech is the articulatory gesture and many phonological processes can be explained by the reduction in magnitude or overlap of gestures. Using this model, I relate intervocalic /s/ voicing to other reduction processes that occur in Spanish, such as the voicing of intervocalic voiceless stops (Hualde et al. 2011). Ultimately, based on the results for gender and age, I determine that intervocalic /s/ voicing represents a change in progress in Lojano Spanish, and furthermore I claim that the gradient and variable nature of this feature is best explained with a gestural analysis that also takes into account the principles of sound change.

In the realm of perception, I provide data from an online experiment looking at allophonic perception of [s] and [z] in intervocalic position. This online experiment consists of a series of similarity rating and discrimination tasks, and the stimuli were recorded by a native male Lojano. In the similarity rating tasks, the participants hear two
audio files (i.e. [asa]-[aza]) and must respond on a scale from 1 ‘very similar’ to 6 ‘very different’ based on how they hear these audio files. In the discrimination tasks, the participants hear the same pairs and must decide whether the two audio files sound the “same” or “different”. A few recent studies have shown that listeners can at times differentiate two allophones of their language to some degree (Boomershine et al. 2008, Johnson & Babel 2010), and my results confirm that this is the case for the allophones [s] and [z] in Lojano Spanish, while also showing that this perception is dependent on linguistic context. In this way, there is a connection between production and perception since Lojanos hear the difference between [s] and [z] most robustly in the environments where they voice intervocalic /s/ the most.

1.3 Research questions

The present study addresses the following research questions:

- To what degree is intervocalic /s/ voicing present in Lojano Spanish? How does the realization of intervocalic /s/ in this dialect compare to other HES varieties?
- Which linguistic and social factors influence voicing in this dialect?
- Can the gestural approach of Articulatory Phonology adequately explain intervocalic /s/ voicing in Lojano Spanish?
- Do native Lojanos hear the difference between [s] and [z] in intervocalic position? To what degree?
- How can the challenge of examining allophonic perception be overcome and what type of task is best suited to this type of perception?
• How does the perception of [s] and [z] in Lojano Spanish compare to the perception of other sounds in Spanish?

In the chapters that follow I will address each of the aforementioned aspects of intervocalic /s/ voicing in Lojano Spanish. Chapter 2 provides background on the sociohistorical situation of the Ecuadorian Andean region, Highland Ecuadorian Spanish, Spanish sibilants, the phenomenon of interest, and a review of previous methodologies. Chapter 3 focuses on the production study, with a detailed explanation of the methodology, data analysis performed, results, and conclusions. In Chapter 4, I discuss the methodology and results of the online experiment that examined allophonic perception. Finally, in Chapter 5, I conclude with a discussion of the results, a few remarks about the overall status of this variable, the contributions of this dissertation, and areas of future research.
CHAPTER 2: BACKGROUND

2.1 Historical developments of the region

2.1.1 Historical background of the Ecuadorian Andes

According to historical anecdotes, the first contact between the native population in Ecuador and the Spaniards took place in 1527 (De la Torre & Stiffler 2008), when explorers led by Francisco Pizarro reached the coast of what is now Ecuador. This presence was heightened in 1533 and 1534 when Sebastián de Belalcázar, the lieutenant of Francisco Pizarro, campaigned in the northern Andean region of Ecuador, which resulted in the foundation of San Francisco de Quito towards the end of 1534. Quito (and subsequently Ecuador) continued to gain importance as the Spanish colonies developed, with the establishment of the seat of the Royal Audiencia of Quito in 1563. Many years later, after much struggle for independence from Spain, Ecuador finally became a nation in 1830.

Before the arrival of the Spaniards, the native population of Ecuador was quite diverse. According to De la Torre & Stiffler (2008), there were numerous indigenous groups that inhabited Ecuador prior to the Incan Conquest. These indigenous groups, including the Cañari and Shuar, among others, were relatively isolated, and as far as we know, there was no unifying governmental structure. The Incan Conquest began in the
1400s and by roughly 1480, Tupac Yupanqui had conquered the southern provinces of Ecuador. Later, his son, Huayna Capac would go on to push their reign further north into the Andes, leaving many costal and lowland indigenous communities untouched. Given that the Spanish arrived right on the heels of this conquest, the Incas were never able to fully establish a centralized government or uniform language or religion in Ecuador, as had been the case in Peru. Thus, while Kichwa¹ is presently the most widely spoken native language in Ecuador (Instituto Nacional de Estadística y Censos), Spanish and Kichwa have been in contact with many other native languages throughout the history of Ecuador.

Geographically, four major regions characterize Ecuador, as seen in Figure 1 below: Coast, Highlands/Andes, Amazon, and the Galapagos Islands. Each region has its own history and customs. Guayaquil, the largest city in terms of population, sits on the southern coast and is known as the economical center of the country. Ecuador’s capital, Quito, which is in the Northern Highlands, is the second largest city in the country and represents the political center. While there are very few cities of significant size in the Amazonian region, this area is of utmost economical importance given its oil production. The focus of the present research is the Highland region, which reaches from the northernmost province of Carchi to the province of Loja in the south. While the map in Figure 1 shows each of the provincial capitals for the Highlands, anecdotally the cities considered to be most important are Quito, Cuenca, and Loja. With a population of half a million inhabitants, Cuenca is known for its well-preserved historical center, which was

¹ Kichwa (in Spanish, Quichua) is the traditional name of the variant of Quechua spoken in Ecuador.
declared a World Heritage Site by the United Nations Educational, Scientific and Cultural Organization in 1999 (UNESCO, 1999). Farther south, the city of Loja, with a population of just over 180,000 (INEC, 2010), is recognized as the *cuna de artistas* ‘cradle of artists’, boasting many important musicians and writers. Given that Loja is the main site of the fieldwork for the present study, I will turn now to the history of this city in particular.

![Regional Map of Ecuador](image)

**Figure 1**: Regional Map of Ecuador (Guia Turística Ecuatoriana, 2015)

### 2.1.2 Loja then and now

Before the arrival of the Spaniards, the area that is now the province of Loja was
home to diverse indigenous populations, including the Palta, Malacatos, and Cañari people (Paladines 2001). From what scholars have been able to gather, the indigenous groups of this region were not unified, but rather functioned more or less independently. After an initial settlement in 1546, Alonso de Mercadillo founded the city of Loja in its present-day location in 1548 (Jaramillo 1955). He named it after his hometown in Spain, which lies within the southern region of Andalusia on the coast of the Mediterranean Sea. Juan de Salinas, the first governor of the region surrounding Loja, was from Álava (Jaramillo 1955), in the Basque region of Spain. Upon examining further information about the settlers to all of the Spanish colonies, it is clear that the largest number of them came from Andalusia. Boyd-Bowman (1976a: 585) documents the immigration of almost 55,000 of these settlers from 1493 to 1600 and attests that 36.9% of them came from Andalusia. He goes on to report that large numbers of immigrants to America also came from Extremadura (16.4%), New Castile (15.6%), and Old Castile (14.0%), which are in central and northwestern Spain. These figures are quite similar for the Spaniards that came specifically to Quito (Boyd-Bowman1976b: 88-89). Alvar (1996) confirms this, stating that most of the Spaniards that came to colonize Ecuador were from Andalusia or Castile. More specifically for Loja, Jaramillo (1955) adds that the names of colonial towns surrounding Loja likely divulge where the early settlers were from: Valladolid and Santa María de Nieva (Castile and Leon), and Loyola (Basque Country). Additionally, analysis of common last names in Loja shows that there was likely a considerable Basque population among its settlers, as well as Sephardic population (Ordóñez Chiriboga 2005). In conclusion, these sources indicate that is likely that the first Lojanos had quite diverse origins.
Beyond the history of Loja, it is of utmost consequence to consider the other factors that have shaped and molded Lojano society and its people. Arguably the most crucial factor is the isolation that the city has experienced since its foundation. This isolation was clearly originally caused by geography. As the city of Loja is sandwiched in a valley in the middle of the Andes Mountains, travel to and communication with other cities in Ecuador has been limited. Paladines (2001) highlights this isolation, stating that Loja has always been far from the economic center of Guayaquil and the political center of Quito. He concludes therefore that Loja has had to be self-sufficient and confront problems without outside support.

Apart from the extreme isolation, Loja has also been characterized by being a transitional zone, the border between one place and another. Paladines (2001:15) describes Loja in this way as a “encrucijada territorial e histórica” [territorial and historical crossroads]. This is evidenced both in North-South direction as well as West-East, as can be appreciated in the map in Figure 1. Loja is what separates the rest of Ecuador from Peru, the last post before crossing the border. Thus, being self-sufficient has often included being able to defend the border during times of conflict with Peru, while also time engaging in mercantile and cultural exchange with Peru. Additionally, while firmly planted in the highlands, Loja and its people have always had one arm extended to the coast and the other to the Amazon, as both regions are in close proximity to Loja. The transitional nature is even seen in the vegetation of the province of Loja, which is a mix of diverse ecosystems, including both dry and mountainous forests and agricultural land.

Paladines (2001) claims that these geographical factors have had many social
effects. First of all, provincial life in Loja was prolonged due to its isolation. While many of the other cities in Ecuador progressed and developed rapidly, Loja remained quite conservative in nature for many years. Furthermore, isolation has had a profound impact on the character of its population. Paladines (2001: 61) describes Lojanos as open, frank, and yet reserved. Interestingly, however, open-mindedness is not traditionally something that is associated with reserved-ness. It seems, perhaps, that since Loja is one of the most important cultural centers of Ecuador (Paladines 2006), intellectual and artistic (especially musical) propensity has caused this duality of both reserved-ness and openness. Beyond mere openness, Lojanos are known for being explorers and pioneers, populating many remote areas of the country, and migrating abroad in modern times.

The marginality experienced by the city has also led to an extremely strong regional identity, which can be found in everyday conversations. For instance, when asked where he or she is from, a Lojano might first answer, “Soy Lojano/a” [I am Lojano], and then clarify, “Soy ecuatoriano/a” [I am Ecuadorian]. Part of this regional Lojano identity is a differentiation from other Ecuadorians and even other highlanders. As Paladines, a native Lojano, states: “hablamos diferente, pensamos diferente, comemos diferente, vivimos y morimos diferente” [We speak differently, think differently, eat differently, live and die differently] (2001: 62). In part, this sense of uniqueness is likely the result of Loja being a transitional area. Loja is not just highland, coastal, or Amazonian, but rather experiences the multiplicity of being the junction of all these things at once. Finally, Lojanos’ distinctive and instantly recognizable accent also serves to set them apart from other Ecuadorians and plays a big role in their sense of identity (Paladines 2001: 120).
One final characteristic to note about Loja is the distinct relationship the city has had with the indigenous populations of the area. Palacios Alcaine (2005) cites that in comparison to the northern Andes, Loja is as a region that favors “castellanization” and negative attitudes towards Kichwa and other indigenous languages. Boyd-Bowman (1953) confirms this, stating that there are numerous exclusively native communities in the northern Highlands (at least at the time), whereas the indigenous people are integrated more with the “white” population in Loja. In fact, the town of Saraguro boasts one of the few significant indigenous populations within the province of Loja.

The city of Loja today is ripe for sociological and linguistic study because it finds itself at a crossroads between the remoteness of the past and the interconnectedness of the future. While it was once very difficult to traverse the roads between Loja and other neighboring cities, and indeed Loja was not connected by highway to other highland cities until the mid-twentieth century (Toscano Mateus 1953: 15), that is no longer the case. According to an expenditure report by the Ministerio de Transporte y Obras Públicas (MTOP), over $520 million USD was spent on public works, most importantly highways, in the province of Loja during the time period of 2007-2014 (MTOP 2014). This represents roughly 7% of the total expenditures by the MTOP and the third highest of any province in Ecuador, topped only by the provinces of Manabí on the coast, and Pichincha, the province where the capital city Quito lies.

The effect of the rapid improvement of highways in Loja has been profound. Lojanos can now travel much more freely and easily than in the past, which has increased overall communication between Loja and other highland towns. To cite one relevant example, in the 1960s it took roughly 10 hours to get from Loja to Cuenca, its closest
sizeable neighbor to the north. In the early 2000s, this travel time was reduced to around 5 hours, and today it takes a mere 2.5 hours to get between the two cities. Along with these infrastructural changes have come societal changes. As will be discussed in more detail later, generally young Lojanos now orient outward and aspire to study, work, and live in cities such as Quito, Cuenca, and Guayaquil. Thus, the future for Loja undoubtedly will hold many changes, both social and linguistic in nature.

2.2 Highland Ecuadorian Spanish (HES)

Given the strong sense of regionalism within highland Ecuador, we might wonder if the language of these regions is unified under the term “Highland Ecuadorian Spanish”. I believe there is some unification when Ecuador is considered as a whole. In juxtaposition against the costeños ‘coastal-dwellers’ of the Pacific Coast of Ecuador, serranos ‘highlanders’ do share certain cultural aspects, including gastronomy and clothing, among other things. While someone from Loja and Cuenca will point out the differences between their hometowns, it is also likely that they would feel more akin to each other than to someone from the coast.

This is mirrored linguistically – Toscano Mateus (1953), Boyd-Bowman (1953), Lipski (1994), and Aguirre (2000) all affirm that the three major dialect regions in Ecuador are Coast, Highlands, and Amazon. The three main sounds that distinguish Coastal Ecuadorian Spanish and Highland Ecuadorian Spanish (HES) are /ɣ/-/ʎ/, /r/ and /s/. While Coastal varieties exhibit yeismo (/ɣ/ and /ʎ/ have merged to one phoneme), Highland varieties have mostly conserved lleismo (maintenance of /ɣ/ and /ʎ/ as separate
phonemes). As such, the words *malla* ‘net/mesh’ and *maya* ‘Mayan’ would be homophonous on the Coast ([ma.ja]), while they are distinguished in the Highlands:

*malla* [ma.ʎa] and *maya* [ma.ja] (although see the discussion in the next paragraph). As for the phoneme /r/, it is realized as an alveolar trill on the Coast, while it is assibilated in some Highland varieties. Finally, syllable-final /s/ is weakened in Coastal Ecuadorian Spanish (e.g. *costa* ‘cost’ [koh.ta]), while it is usually fully pronounced in HES (e.g. [kos.ta]). In addition to these sounds, there are certain lexical items (for instance borrowings from Kichwa) that are used exclusively in HES, as well as certain grammatical structures (e.g. *dar* + gerund, see Haboud (1998) for a description of HES contact features). As for the Amazonian region, Lipski (1994) warns that a true “Amazonian Ecuadorian Spanish” has yet to emerge since most of the population, at the time of his writing, were transplants from neighboring Highland provinces. While it has not been documented empirically, I believe the same to be true of the Galapagos Islands given that its population is a melting pot of migrants from all parts of mainland Ecuador.

Despite the similarities to be found within the Highland region, Lipski (1994: 248-249) further clarifies that there are important differences between sub-dialects of HES. He demarcates four subdivisions within the Highlands, as seen in Figure 2: extreme north (Carchi), Central Highlands (from Imbabura to Chimborazo, including Quito), Cañar/Azuay (including Cuenca), and Loja. While there is very little existing information on the Spanish of Carchi, the other three sub-dialects of HES have been descriptively studied. Gómez (2003) asserts that delateralization of /ʎ/ and assibilation of /ɾ/ determine the two major isoglosses of the Ecuadorian highlands. In the first case, Loja and Cuenca
are known for preserving the palatal lateral production of /ʎ/ (malla [ma.ʎa]), while this sound is realized as a pre-palatal sibilant fricative in the northern region surrounding Quito (malla [ma.ʒa]). In the second case, Loja is the only Highland province noted to have trill /r/, while this sound is assibilated in the rest of HES, most notably Cuenca, and to a lesser extent Quito. As will be seen later, the realization of intervocalic /s/ also varies within the Highlands. Consequently, it is important to keep in mind the implication of regionalism within the Andes, both socially and linguistically.

Figure 2: Map of Ecuador with linguistic regions according to Lipski (1994), adapted from http://www.luventicus.org/mapas/ecuador/esmeraldas.html

In contrast to traditional descriptions, my own observations in Loja lead me to believe that Lojano Spanish is undergoing change, whereby yeísmo and assibilation of /r/ can now be heard in the speech of the younger generations.
2.3 Spanish sibilants

Spanish sibilants have long been a significant topic of debate within the world of Hispanic Linguistics since their development traces history from medieval Spanish to the dialects of Spanish spoken today. In medieval Spanish, the sibilants went through the processes of deaffrication, devoicing and, from there, into further changes based on the dialect (Penny 1991). In present day, /s/ voicing has resurged in a few dialects of Peninsular and Latin American Spanish (Robinson 1979, Torreblanca 1986, Lipski 1989, among others).

2.3.1 History

As described in Penny (1991), Old Spanish had seven sibilants, six of which are of interest here: /ts/, /dz/, /s̱/, /ẕ/, /ʃ/, /ʒ/. The historical change that affected these sibilants, which was complete “by the end of the Middle Ages” (p. 99), was the deaffrication of /ts/ and /dz/ to dental fricatives /s̱/ and /ẕ/. The second change was the devoicing of /ẕ/, /ẕ/, and /ʒ/, which merged these sounds with their voiceless counterparts, resulting in a system with three sibilants distinguished only by place of articulation: dental /s̱/, apico-alveolar /s̱/ and pre-palatal /ʃ/. According to Penny, this change took place in the late sixteenth century (p. 99). From there, the process of what Widdison (1997: 256) calls “distancing” converted /s̱/ to /θ/ and /ʃ/ to /χ/. While these sounds remained voiceless, the point of articulation changed from dento-alveolar /s̱/ to interdental /θ/, and from pre-
palatal /ʃ/ to velar /x/, resulting in a three-way distinction between /θ/, /ʃ/ and /x/. In Andalusia and America, this final change had a different development, ultimately resulting in a two-way distinction between /s/ and /x/.³

Penny (1991: 101-103) presents two hypotheses for the evolution of Andalusian and American Spanish. One possibility is that the affricates merged with the dental sibilants (this dialect being devoid of alveolar sibilants) before devoicing took place. The other analysis shows alveolar and dental sibilants merging after devoicing. Regardless of how the change took place, the eventual result in these dialects of Spanish is one (alveolar) sibilant. Figure 3 shows a summary of these changes.

Figure 3: Historical development of Spanish sibilants

³ In Andalusia and some parts of Latin America, this phoneme is realized as a glottal fricative [h].
2.3.2 Current /s/ phenomena

In modern-day Spanish, /s/ is still subject to considerable variation, and therefore is of substantial interest to linguists. Two main threads of research on /s/ have formed: /s/ weakening and voicing assimilation of /s/. The first, /s/ weakening, is arguably the most studied variable in Hispanic Linguistics, "además de ser uno de los fenómenos más comentados al nivel del público no especialista" (‘besides being one of the phenomena most likely to be commented on by non-specialists’) (Lipski 1983: 242). This process can be described as the variable production of /s/ as [h] (aspiration) or a phonetic zero [∅] (deletion), as in example (2).

(2) Standard Pronunciation Aspiration Deletion

costa ‘the coast’ [kos.ta] [koh.ta] [ko.ta]

This variable is subject to conditioning by numerous linguistic factors, such as syllable position, as well as social factors, such as dialect and age. While /s/ weakening is traditionally thought to occur exclusively in syllable-final position (as in the example above), more recent studies have also found syllable-initial weakening in some Spanish varieties (Brown & Torres Cacoullos 2003, among others). For linguists and non-linguists alike, /s/ weakening is one of the key shibboleths to figuring out where someone is from within the Spanish speaking world (Lipski 1994), and Hammond (2001) estimates that roughly 50% of modern Spanish speaking varieties are characterized by /s/ weakening. This variable has been studied in production for the following dialects, among others: Dominican (Alba 2000), New Mexican (Brown 2005), Venezuelan (Calles
1986), Chilean (Cid-Hazard 2003), Cuban (Terrell 1979, Hammond 1980, Dohotaru 2004), Puerto Rican (Poplack 1981, López Morales 1983, Figueroa 2000), Colombian (Lafford 1986, File-Muriel 2009, File-Muriel & Brown 2011), Argentine (Terrell 1978), Andalusian (Gerfen 2002), Canary Islands (Samper Padilla 1990), as well as comparative works (Terrell 1981, Brown 2009). Although [h] and  symbol are the most common variants found, recent studies have also shown that another possible realization is a glottal stop [ʔ], resulting in más alto ‘taller’ [maʔ.al.to] (see Chappell 2013 for Nicaraguan Spanish).

Perception studies on this variation are less common; however, scholars have begun to investigate the effect of /s/ on dialect recognition (Boomershine 2006), how /s/ weakening is categorized (Bishop 2007, Schmidt 2013), and what social associations are tied to /s/ weakening (Mack 2009, Walker et al. 2014).

Apart from /s/ weakening, recent linguistic studies have also investigated the voicing assimilation that /s/ undergoes before a voiced consonant. Traditionally, it was assumed that /s/ voices to [z] before voiced consonants in accordance with a categorical phonological process, as in desde ‘from’ [dezðe]. Recent studies have shown that voicing assimilation before voiced consonants is gradient in nature (Romero 1999, Schmidt & Willis 2011, A. García 2013, Campos-Astorkiza 2014). These studies demonstrate that the voicing of /s/ before voiced consonants is dependent on such factors as the manner and place of articulation of the following consonant, phrasal position, and speech rate. Thus, voicing assimilation seems to be a gradient process in that varies with speech rate, as well as following consonant.

Both of these lines of research are important for the present work because, as will
be discussed later, intervocalic /s/ voicing could be viewed both as /s/ weakening, in that the /s/ in question in shortened, and as voicing assimilation, since the /s/ is adopting the voicing of the surrounding vowels. Furthermore, these studies offer a backdrop for other processes that affect /s/ and a point of comparison for the variation found with intervocalic /s/.

2.4 Intervocalic /s/ voicing

Though intervocalic /s/ voicing was originally thought to be a feature unique to Ecuadorian Spanish, more and more studies have shown that it is present to some extent in quite a few modern Spanish dialects. The phenomenon can be described as the categorical or variable realization of /s/ as [z] when it is between vowels in word-initial (la sopa ‘the soup’), medial (casa ‘house’), or final position (los otros ‘the others’), as shown in example (3), repeated from section 1.1. I will first give an overview of non-Ecuadorian dialects where intervocalic /s/ voicing has been attested and then focus on what has been documented for the dialect of interest.

(3)  Standard Pronunciation  Dialect with voicing

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>la sopa</td>
<td>[la.so.pa]</td>
<td>[la.zo.pa]</td>
</tr>
<tr>
<td>casa</td>
<td>[ka.sa]</td>
<td>[ka.za]</td>
</tr>
<tr>
<td>los otros</td>
<td>[lo.so.tros]</td>
<td>[lo.zo.tros]</td>
</tr>
</tbody>
</table>

2.4.1 In other varieties

Outside of Ecuador, this feature has been described for the Spanish of the following
places: Catalonia (McKinnon 2012, Davidson 2014), Madrid (Torreira & Ernestus 2012, Hualde & Prieto 2014), Central Spain (Torreblanca 1986), and Bogotá (A. García 2013). Looking at Catalan-Spanish bilinguals, McKinnon (2012) and Davidson (2014) both find variable voicing in word-final /s/, while word-medial /s/ is rarely voiced. For Madrid Spanish, Torreira & Ernestus (2012) find some voicing in all positions, but more so word-finally. Similarly, Hualde & Prieto (2014), who also examine Madrid Spanish, find more voicing in word-final position, although quite a bit of voicing word-initially as well. Torreblanca (1986) details that he has heard voicing of intervocalic /s/ due to articulatory relaxation in Toledo, Ávila, and Cáceres. Beyond Spain, A. García (2013) looks at the speech of 15 speakers originally from Bogotá, some currently living in Milwaukee, finding overall low average rates of voicing. These studies will be explored in more detail in sections 2.4.3-2.4.5. While not explicitly investigating this feature, a few other studies have documented intervocalic /s/ in Asturian Spanish (Bárkányi 2014), Mexican Spanish (Schmidt & Willis 2011), and Buenos Aires Spanish (Rohena-Madrazo 2011). Most notably, among Schmidt & Willis’ (2011) participants from Mexico City, a male speaker systematically produced tokens of intervocalic /s/ with majority voicing.

2.4.2 Intervocalic /s/ in Highland Ecuadorian Spanish

Now turning to Ecuador, the reports of intervocalic /s/ voicing go back to the first studies of Ecuadorian Spanish. All descriptions of this feature agree that it is characteristic of HES, while not being found in Coastal Ecuadorian dialects. Toscano Mateus (1953: 79) provides the first documentation, stating: ‘La s final de palabra (aspirada en la Costa), cuando precede a una palabra que comienza por vocal, se
pronuncia sonora en Quito y la Sierra (loz hombres, como en francés les hommes)” [The
s at the end of the word (aspirated on the Coast), when it comes before a word that starts
with a vowel, it is voiced in Quito and the Highlands (loz hombres, like in French les
hommes)]. While he makes no mention of the precise location within the Highlands in
this general statement, the author goes on to specify that /s/ is also voiced sporadically in
prefixes such as des-hilar ‘to fray’ and des-herbar ‘to weed’ in Cuencano Spanish.
Noting that [z] is found in some Spanish borrowings in Kichwa and that /z/ is a phoneme
of Ecuadorian Kichwa (p. 23), Toscano Mateus seems to hint that a substrate influence
might be involved.

This early report was not confirmed empirically until many years later. Robinson
(1979) interviewed two male speakers from each of the highland provinces and provides
impressionistic evidence that leads him to conclude there are three major dialect regions
according to the realization of intervocalic /s/. For the Northern-Central highlands
(including Quito), he claims there is categorical voicing before word boundaries. In
Cuenca (South-Central highlands), /s/ voices categorically in this same context as well as
before prefixal morpheme boundaries, such as those mentioned previously. Finally,
Robinson claims that intervocalic /s/ is never voiced in the southern-most province of
Loja and the northern-most province of Carchi. These findings were later supported by
other impressionistic studies of Quiteño Spanish (Lipski 1989) and Cuencano Spanish
(Calle 2010), as well as the general description of Ecuadorian Spanish in Aguirre (2000).

More quantitative studies have called into question the categorical nature of this
phenomenon. Concerning Quiteño Spanish, Chappell (2011) and Strycharsczuk et al.

24
both find that voicing of intervocalic /s/ may not be completely categorical, at least for some speakers. Chappell (2011) shows that the voiced variant is found in 91% of cases of word-final /s/, while only occurring sporadically in word-medial and initial contexts. She concludes that voicing of /s/ at a word boundary is not categorical as Robinson (1979) and Lipski (1989) had predicted. Inspecting individual differences more closely, Strycharczuk et al. (2013) show two patterns of voicing in Quiteño Spanish: some speakers voice intervocalic /s/ gradiently, while others voice categorically, but optionally. Given the results of these recent instrumental studies, the situation in Loja might be different than earlier impressionistic descriptions would predict.

2.4.3 Previous methodologies

Taking into consideration the distinct dialects examined for intervocalic /s/ voicing, it is vital to consider the different methodologies that have been utilized and the advantages and disadvantages of each. The first concern evident from the literature review is the nature of the data. Several of these studies have used exclusively interview or spontaneous data (Robinson 1979, Torreblanca 1986, Lipski 1989, Chappell 2011, Torreira & Ernestus 2012, Hualde & Prieto 2014). The advantage of spontaneous-like speech is that it is assumed to be more representative of how speakers talk on a daily basis. In the traditional Labovian notion of “attention paid to speech” (Labov 1971), the assertion is that participants are less likely to pay attention to the phenomenon of interest in interview data, in which they answer a series of questions about themselves, and therefore will not modify their production. Some of the disadvantages of this type of data, 

---

4 See also Chapter 6 of Strycharczuk (2012) for a more thorough description of this study.
however, include that the quality of recording is often comprised due to the more naturalistic environment, and that it is difficult to ensure participants will produce sufficient tokens in the linguistic contexts of interest. Other studies of intervocalic /s/ voicing have used exclusively read data (Strycharczuk et al. 2013, Davidson 2014). While this type of data can range in how naturalistic it is, in general read speech is less natural than interview or spontaneous speech. Thus, some scholars have questioned the conclusions drawn from read speech since it might not be entirely indicative of how participants speak day to day. One advantage of read data, on the other hand, is that it is possible to control how many tokens per context will be elicited and therefore the data will be better balanced. These two poles represent the struggles of the field of sociophonetics, which has to straddle the methodological aims of both sociolinguistics and phonetics (Foulkes et al. 2012, Di Paolo & Yaeger-Dror 2011, among others).

Since advantages and disadvantages can be found on both sides of this issue, other scholars investigating intervocalic /s/ have decided to use both read and interview data (Calle 2010, McKinnon 2012, A. Garcia 2013). In this way, the researcher can ensure sufficient tokens of each context are elicited, but also compare the participants’ behavior in more naturalistic speech, in some sense to “check” the patterns found in the read data. Beyond being a methodological compromise, using both types of speech can also allow the researcher to look at factors that vary between read and spontaneous speech, such as speech rate.

Once the data has been collected, the issue turns to data analysis. This is where the question of categorical versus gradient comes to the forefront. It is important to distinguish between different definitions of categorical and gradient since these terms are
used in different ways in phonetics, phonology, and sociolinguistics. “Categorical” can either mean that voicing happens all the time whenever there is an /s/ in the conditioning environment or that there are only two variants to consider: voiceless and voiced, with nothing in between. On the other hand, “variable” is most often used to describe cases where voicing does not apply all of the time in the conditioning environment, and “gradient” is used when the realization of voicing varies along a continuum of 0% to 100% voicing, although “variable” and “gradient” are at times used interchangeably. Often the chosen data analysis has dictated whether intervocalic /s/ voicing is seen as categorical or gradient/variable. For instance, those studies that use impressionistic, auditory coding (Robinson 1979, Torreblanca 1986, Lipski 1989, Calle 2010) have no choice but to use the binary variants [s] and [z] because the researcher chooses a variant based on what they hear. In the case of Robinson (1979) and Lipski (1989), we can assume that these authors also wish to invoke the other sense of “categorical” since they state things like “/s/ always voices at a word boundary”.

More recent studies that use acoustic measurements have had to decide whether to treat intervocalic /s/ voicing as a binary or continuous phenomenon. I will discuss here the various acoustic measurements that have been used, and leave the issue of binary versus continuous analysis for Chapter 3. Studies looking at intervocalic /s/ voicing and the production of /s/ more broadly have used a wide array of acoustic measurements as dependent variables: fricative duration, voicing duration, percent voicing, uninterrupted voicing, percent frames voiced (voicing report), and several distinct intensity measures. Although preceding vowel duration has been shown to correlate with fricative voicing, it has not been used in studies of intervocalic /s/ in Spanish.
Fricative duration and some measure of voicing are the two most widely used acoustic measurements in studies on the voicing of /s/ in any environment. Two studies of intervocalic /s/ have considered fricative duration: Strycharczuk et al. (2013) and Hualde & Prieto (2014). Both find an inverse relationship between fricative duration and percent voicing, and thus consider contexts with shorter fricative durations as favoring voicing. The various acoustic measurements of glottal vibration have all been used in studies on intervocalic /s/ voicing and are at times contested in terms of their validity. A few scholars have mentioned that acoustic parameters are not always reliably indicative of vocal cord activity and have preferred the use of articulatory data, such as EGG (Smith 1997, Romero 1999). This type of data is difficult and expensive to collect, however, and Recasens & Mira (2012) conclude that acoustic parameters correlate more with glottal vibrations than EGG activity does. For these reasons, most previous studies have relied on acoustic measures. Chappell (2011) and Torreira & Ernestus (2012) both classify glottal vibrations as uninterrupted versus interrupted voicing by examining the voicing bar and Praat pitch tracking, and only consider completely uninterrupted voicing as voiced. Like auditory analysis, this method forces intervocalic /s/ voicing to be considered categorically even if it might actually be gradient in nature. Nevertheless, this may be the most reliable measure of glottal vibration in some cases, for instance where there is significant background noise that prevents the researcher from utilizing a more precise measure.

Voicing duration, percent voicing, and percent frames voiced have been used in order to quantify /s/ voicing. Percent voicing (also called voicing ratio) is calculated by first measuring (usually by hand) the fricative duration and voicing duration, and then
dividing the latter by the former. While voicing duration goes into the calculation of percent voicing, Recasens & Mira (2012) claim that it is preferable to use percent voicing rather than the raw voicing duration since the former is in some sense a normalized measurement of voicing. For example, 50 ms of voicing is considerable if the total fricative duration is only 60 ms, but that same amount of voicing would only be one-fourth of a token that has a fricative duration of 200 ms. Thus, using percent voicing allows for a more uniform picture of voicing across different length tokens. Strycharczuk et al. (2013) use both voicing duration and percent voicing, and percent voicing has also been used by McKinnon (2012), Davidson (2014), A. García (2013) and Schmidt & Willis (2011). In these studies, as well as Campos-Astorkiza (2014) on pre-consonantal /s/, voicing duration is judged as the portion of the fricative that exhibits periodicity in the waveform and a voicing bar in the spectrogram. Modern technology also offers an instrumental method of measuring percent voicing – Bárányi (2014) and Hualde & Prieto (2014) use the “voicing report” feature in Praat to determine the percent of frames voiced in a given token of /s/.

Finally, two recent studies have used measurements of intensity to quantify the production of intervocalic /s/. Strycharczuk et al. (2013) use intensity difference since it has been shown to illustrate gradient voicing differences: “Mean intensity was measured for the low frequency portion of the fricative (filtered from 0 to 900Hz) and for the unfiltered fricative. The total intensity value was then subtracted for the low-frequency intensity to calculate the intensity difference” (p. 140). Smaller differences in intensity are associated with voicing because voiced fricatives have a high intensity low-frequency portion. Torreira & Ernestus (2012) measure low-band intensity dip duration and high
band intensity difference. The latter is the same measurement used in Strycharczuk et al. (2013), while the former is measured by taking the duration of the portion of the vowel-fricative-vowel sequence where the energy dips into the lower frequencies. Since this dip in intensity usually corresponds to the articulation of the /s/, the authors use low-band intensity dip duration as an estimate of fricative duration. These intensity measures are used to provide another, at times more precise measure, of voicing, as well as to approximate segmental duration in cases where it is not possible to measure the duration of /s/.

Gradoville (2011) provides an evaluation of all of the aforementioned instrumental measurements as well as harmonicity and center of gravity in the context of voicing of palatal sibilants in Argentine Spanish. He compares each of the measurements to auditory coding of [ʒ] and [ʃ] in order to see which matches up the best with the results of impressionistic coding, assuming a better match is a more “valid” technique. Gradoville concludes that the pulse-based analysis offered by the Praat voicing report is the most reliable instrumental measurement of fricative voicing in Spanish.

Given the methodologies of these previous studies, I believe the most accurate measurement of intervocalic /s/ voicing for my data is to calculate percent voicing by hand. The issue with instrumental measurements, such as the voicing report, is that these automated reports can be fooled by “phantom” glottal pulses, which in reality are the result of background noise. This is of particular concern for my data as it was collected in the field as opposed to in a laboratory setting. While no measurement is perfect, measuring percent voicing offers a reliable way of examining the potential gradience of
intervocalic /s/ voicing. This technique does not assume a priori that the variation will fall neatly into the two categories of voiceless and voiced. Additionally, as Recasens & Mira (2012) mention, percent voicing offers a somewhat normalized measurement since the relative duration of /s/ that a speaker may have is factored out. A detailed description of how percent voicing was measured for the present study can be found in section 3.2.3.

2.4.4 Linguistic factors examined

In addition to different dependent variables, past studies on intervocalic /s/ voicing have also considered different independent predictors. These linguistic factors include: position within the word, word class, word frequency, preceding/following segment, stress, speech rate, and type of speech. The factors examined and which are found to be significant depend on the dialect at hand. In the discussion of these factors, I will talk about which “favor voicing” or show “more voicing”, which can mean fully voiced tokens or tokens with higher percent voicing, since as we saw studies differ in their dependent measures.⁵

As already evidenced, the factor most studied for intervocalic /s/ voicing is position within the word. While all studies have included tokens of word-final and word-medial /s/, tokens of word-initial /s/ are not always included. Nonetheless, most studies have found more voicing (or voicing exclusively) in word-final context (Robinson 1979, Lipski 1989, Calle 2010, Chappell 2011, Torreira & Ernestus 2012, McKinnon 2012,

---

⁵ Where relevant I will mention the results of other acoustic measurements (besides measures of glottal vibrations), although in general scholars have found correlations between the measure of “voicing” they use and other measurements such as duration and intensity.
Strycharczuk et al. 2013, Davidson 2014), while only one study finds more voicing in word-medial context (A. García 2013). Hualde & Prieto (2014) also find the most voicing in word-final context; however, they do not find a significant difference between final and initial contexts. The theories on why position within a word might play a role will be discussed later. While all of the studies of Peninsular and Ecuadorian varieties conclude that intervocalic /s/ voicing is present, A. García (2013) asserts based on her results that this feature is not present in Highland Colombian Spanish since the rates of intervocalic /s/ voicing she finds are quite low in comparison to other studies.

Previous results for word class, word frequency, and morphological status of /s/ are somewhat contradictory. Chappell (2011) finds more voicing in determiners, adjectives, nouns, and complementizers than in verbs, participles, and adverbs. In addition, she finds more voicing in frequent as compared to infrequent words. On the other hand, Torreira & Ernestus (2012) find no effect of morphological status of the morpheme containing the /s/ (whether it is a plural marker, for instance) or word frequency on voicing. Davidson (2014: 6) incorporates word class into the independent factor of prosodic boundary, where “prosodic boundary was coded as either weaker for /s/ tokens in a determiner/preposition + noun phrase (e.g. las aguas ‘the waters’) or stronger for /s/ tokens in a noun/verb + adjective/adverb phrase (e.g. aguas ácidas ‘acidic waters’).” He finds that voicing is favored across a stronger prosodic boundary for two groups of participants, but does not find any effect of morphological status. In line with usage-based theories, if intervocalic /s/ voicing is a change in progress, we might expect more voicing in frequent words since “those words that are used more often have more opportunity to be affected by phonetic processes” (Bybee 2001: 11). So far the few
studies looking at morphological, lexical, and frequency effects have garnered inconclusive results.

Since most studies using read data have kept constant the surrounding linguistic context, it has not been included as an independent variable. Among the studies using spontaneous data, however, Chappell (2011) and Torreira & Ernestus (2012) consider the effect of preceding and following vowel. Although Torreira & Ernestus (2012) find no effect of either factor, Chappell’s (2011) Quiteño results show that voicing is favored before a low vowel and disfavored before a mid or high vowel. Chappell warns, however, that this result must be taken with a grain of salt since following segment is likely to interact with position in the word. Given that in Spanish nouns and adjectives the last vowel tends to be an –a or –o, it is likely that the following vowel results are simply repeating the pattern already seen, that voicing is favored word-finally. Preceding segment did not reach significance in her data.

Despite Lipski’s (1989: 50-53) claims that intervocalic /s/ voicing is not dependent on stress or speech rate in Quiteño Spanish, recent studies have illustrated an effect of these two factors for other dialects. Among those that take stress into consideration, Torreira & Ernestus (2012) and Davidson (2014) indicate some effect of stress for Madrid and Catalan Spanish, respectively, though Chappell (2011) finds it is not a significant predictor. Torreira & Ernestus (2012) code stress according to whether /s/ was in the onset of a stressed syllable or not, assuming resyllabification in the case of word-final /s/. While these authors do not find stress to be a significant predictor of voicing, they do show that /s/ exhibits longer lower-band intensity dips in stressed syllables. Utilizing a different coding schema for stress, Davidson (2014) considers the stress of the
preceding and following vowel (stressed or unstressed) of each token as two separate factor groups. In his data, voiced tokens are favored between unstressed vowels, although this result is only seen for one of the four participant groups (the Catalan-dominant group).

All studies of intervocalic /s/ that have taken into account speech rate have found a direct correlation between speech rate and voicing: as speech rate increases, voicing of intervocalic /s/ increases. For Madrid Spanish, estimated speech rate, measured continuously by calculating the duration of the three phones preceding/following the target /s/, is highly correlated with voicing, lower-band dip duration, and high-band intensity difference, in each case showing more reduction at higher speech rates (Torreira & Ernestus 2012). As one of the two independent predictors considered, A. García (2013) shows for her Bogotá sample that the average percent voicing of /s/ in a word list increases in slow to normal to fast speech conditions. Finally, directly contradicting Lipski (1989), Strycharczuk et al. (2013) find an effect of speech rate in their Quiteño Spanish data. Though this effect at first seems to be across the board, a closer look at individual tendencies shows two distinct patterns for the interaction of voicing and speech rate. Four of Strycharczuk et al.’s participants exhibit very little difference in percent voicing rates between fast and normal speech, while the other three participants exhibit much higher percent voicing rates in fast speech. The authors take this to mean that the first group of participants is voicing intervocalic /s/ categorically (independent of speech rate) and the second group is voicing intervocalic /s/ gradiently (dependent on speech rate). The results of these studies are also in line with research on /s/ weakening, which has shown that speech rate is the number one predictor of /s/ production in the
Spanish of Cali (File-Muriel & Brown 2011) and Barranquilla (File-Muriel 2012). Furthermore, A. García (2013) provides evidence that voicing assimilation of /s/ before voiced consonants is dependent on speech rate. Since all of these are processes affecting /s/, it is clear that speech rate is of utmost importance in the overall production of this sound, but also that it is important to include a precise measure of speech rate (instead of as “fast” vs. “slow”) so that the results can be more reliably compared between studies.

The last linguistic factor investigated in previous research is type of speech. A. García (2013) looks at what she calls “formality” of speech by incorporating both a reading passage and interview questions. Unlike most sociolinguistic studies that use interviews, in her study, participants were posed hypothetical questions on a computer screen. Her use of particular hypothetical questions is a clever way to ensure that the desired tokens of /s/ are produced. However, the fact that participants were talking to a computer screen may have caused their speech to be somewhat unnatural. Even though she did find an effect of speech rate, A. García does not find different rates of voicing between the passage and interview data.

McKinnon (2012) compares data from sociolinguistic interviews to that of a reading paragraph in his study on Catalan Spanish: the interview data favors voicing and the read data disfavors voicing. Under the Labovian notion of “attention paid to speech” (cf. Labov 1971), it is assumed that speakers are likely to pay more attention to their speech in more formal tasks and thus might “modify” their speech in these tasks. Given McKinnon’s results, a traditional approach might conclude that speakers are conscious of intervocalic /s/ voicing and therefore mitigate their pronunciation in the read data to voice less. Nonetheless, since many studies (File-Muriel & Brown 2011, Torreira
& Ernestus 2012, File-Muriel 2012, A. García 2013) have found speech rate to be such an important predictor of /s/ production, any conclusions based on attention paid to speech must be questioned. It is difficult to separate the fact that people speak faster in informal tasks from the possibility that they might be style shifting, especially if an independent measurement of speech rate is not included, as is the case with McKinnon’s study.

Previous studies have shown that a complex interaction of linguistic factors condition the realization of intervocalic /s/. Most important for the present study are those that examine sub-dialects of HES. In these studies, the factors that seem to be most important are position within a word and speech rate, although it is difficult to draw broad generalizations due to the vast methodological differences in previous treatments of HES. Based on the results of studies of other dialects, it would also be useful to consider factors such as stress of surrounding vowels and type of speech.

2.4.5 Social factors examined

Although the research on intervocalic /s/ voicing is still quite recent, some studies have also looked at various social factors. These social factors include: regional dialect, gender, age, speaker, and level of bilingualism. While not all of these factors will be relevant in each sociolinguistic situation, it is important to give thought to how they could be applied to /s/ voicing in Lojano Spanish. As evident from previous discussion, rates of intervocalic /s/ voicing have been shown to vary from country to country, but also within the sub-dialects of HES. While not all of his conclusions were borne out by modern acoustic analysis, it is possible that the three dialect regions put forth by Robinson (1979) are still valid with some modifications. Starting in the north, Chappell (2011) and
Strycharczuk et al. (2013) have shown overall high rates of voicing in word-final position in Quiteño Spanish, while there is relatively little voicing in word-medial or initial position. Although acoustic data is still missing for Cuenca, Calle (2010) corroborates Robinson’s initial claim and finds that Cuencanos do systematically voice only in word-final position. Unfortunately, she does not consider words where /s/ appears before a prefixal morpheme boundary. Calle (2010) is also the only study to look at two sub-dialects simultaneously. Her sample included six speakers from Quito and six speakers from Cuenca of both genders and different ages. Calle finds voicing to be more consistent in the Cuencano participants, while there is quite a bit of variation among the Quiteño participants. Though this could suggest a difference between the two dialects, this conclusion must be questioned given the small number of participants in her sample and the diverse residential background of the Quiteños. Nevertheless, taken all together, previous studies on HES show important regional differences.

Since its inception, sociolinguistics has continuously put gender at the forefront, and studies in many distinct communities have found overwhelmingly similar patterns in regards to the role of participants’ gender. In general, under stable sociolinguistic stratification men use more nonstandard forms, in change from above women favor prestige forms, and in change from below women are the innovators (Cheshire 2002). While this pattern has been found time and time again, Cheshire (2002) warns that the situation is likely not so simplistic, and that trying to find one explanation for this gender pattern is impossible because each community has a vastly different sociocultural situation. In addition, Foulkes & Docherty (2006) point out that it is often very difficult to separate the effect of non-learned, non-arbitrary factors, such as vocal tract length,
from the effect of arbitrary, learned behaviors. A few previous studies on intervocalic /s/
voicing have considered the effect of participant gender and have arrived at contradictory
results. In Chappell’s (2011) analysis of Quiteño radio data, gender was not selected as a
significant predictor. On the other hand, McKinnon (2012) finds that Catalan-Spanish
bilingual females slightly favor voicing (with a factor weight of .53), while their male
counterparts slightly disfavor voicing (.48). Torreira & Ernestus (2012) state in the
introduction of their article that they consider participant gender as a predictor, but it does
not appear in the list of predictors in any subsequent analyses, and so the effect of gender
in Madrileño Spanish is unknown. Finally, Strycharczuk et al. (2013) find no effect of
gender in the production patterns of their Quiteño participants.

In addition to gender patterns, the influence of age on linguistic variation can give
us important information about the status of a feature. Bailey (2002) overviews what has
been found for age in past variationist studies. He distinguishes between those studies
using real-time and apparent-time evidence. Real-time data consists of two or more
samples of the same community separated by a significant time period, with the goal of
observing change between the two samples. Although real-time data provides the best
time comparison, it relies on the existence of a previous sample or else the researcher has
to sample a given community twice and wait many years to publish their findings. When
a previous sample is used, the methodology and sampling techniques may be vastly
different between the two samples. Even in the case where the same researcher collects
data twice, there may have been major demographic changes in the community in the
intervening time period.

Given these methodological difficulties, far more sociolinguistic studies have
relied on apparent-time data. The central assumption of the apparent-time construct is that by comparing the speech of participants of different age groups, we can get a snapshot of the changes that have been taking place in the speech community. Crucially, this also assumes that people’s speech is solidified in their early 20’s and that their speech patterns do not change after that point. The clear advantage of using apparent-time data is that the researcher does not have to wait to collect another sample of data and changes are examined in the same exact community since it only involves one sample. However, apparent-time studies have been criticized due to age grading: some changes are associated with particular times of life and repeat themselves from one generation to the next, and so people do at times exhibit changes throughout their lifespan. Bailey (2002) cites several longitudinal studies (following the same participants) that have shown this to be the case. These studies usually invoke the concept of the linguistic marketplace, whereby speech patterns are more conservative during working years than earlier or later in life.

Despite these critiques, Bailey et al. (1991) compare results from apparent and real time studies of the same variable and find overwhelmingly that apparent-time conclusions do match what is found for corresponding real time studies. However, it is important to note that the studies reviewed include very large, representative samples. Thus, Bailey (2002) concludes that the use of apparent-time data is warranted as a surrogate for real-time data but only for adult participants in large, representative samples. Calle (2010) is the only previous /s/ voicing study to consider the effect of age. She finds that younger Quiteños do not voice as much as older ones and concludes that perhaps this is indicative of a change in progress in Quiteño Spanish. Since she only interviewed six
participants from Quito, there are only two participants per age group, and so this is not the large, representative sample we would hope for in order to look at apparent-time.

Whether or not speaker is included as a random effect, almost all previous studies have found considerable interspeaker variation. As detailed before, Strycharczuk et al. (2013) find two different patterns of voicing for Quiteño Spanish based on speaker. A. García (2013) admits substantial interspeaker variation among her fifteen Bogotano participants, and shows that only four speakers reach rates of 50% voicing or more in any context. Similarly, Davidson (2014) infers that the trend of higher rates of voicing among his Spanish dominant Catalanian participants is being led by a few participants in that group. Finally, Torreira & Ernestus (2012) conclude that there is a significant random effect of speaker in the analysis of all three of their dependent variables (voicing, low-band intensity dip duration, and high band intensity difference). Clearly, we have to be even more concerned with the effect of speaker when the sample size is small, since then the production of one participant could drastically bias the overall patterns.

The final social factor that has been looked at for intervocalic /s/ voicing is language contact. McKinnon (2012) and Davidson (2014) consider the influence of Catalan on Catalanian Spanish. Both authors do so by examining the influence of participants’ level of bilingualism and residence. Given that word-final /s/ assimilates in voicing to following vowels in Catalan, the hypothesis is that increased experience or contact with Catalan will lead to higher rates of intervocalic /s/ voicing in Spanish. For McKinnon (2012), this is defined through two independent factors. He coded the language use of his sixteen bilinguals according to their reports of whether they use Spanish more, Catalan more, or both languages equally. Additionally, he coded
“residence” as inside the city of Barcelona or outside the city of Barcelona, assuming there is more Spanish presence inside the city given the influx of immigrants from other Spanish speaking countries and areas of Spain. On the other hand, Davidson (2014) combines these two factors together to create four social groups for his sample of twenty female, middle class bilinguals. Group A are Catalan-dominant participants living outside the city of Barcelona, Group B are Catalan dominant within the city, Group C are Spanish dominant outside the city, and Group D are Spanish dominant within the city. He assumes that Group A has the most exposure to Catalan, Groups B and C are equal in terms of exposure, and Group D has the least Catalan exposure.

The results for social groups in Catalonian Spanish, however, are not as expected. In McKinnon’s (2012) sample, “Spanish Dominant” disfavors voicing, “Both Languages Equally” favors voicing, and “Catalan Dominant” neither favors nor disfavors. Additionally, participants who live inside the city of Barcelona voice more than those who live in the greater province of Barcelona (outside the city limits). Davidson (2014) finds the highest rates of word-final [z] in Group A (42%) and Group D (36%). While the first result is expected given that Group A has the most exposure to Catalan, the high rates of voicing in Group D, Spanish dominant participants within Barcelona, are puzzling. Since he also looks at /l/ velarization, another Catalan contact feature, Davidson compares the rates of /s/ voicing to those of /l/ velarization. While Group D shows very low rates of /l/ velarization, a few participants in this group show high rates of intervocalic [z]. Therefore, he hypothesizes that /s/ voicing may be below the level of consciousness since it is influencing the speech of Spanish dominant speakers. On the other hand, Catalonians are likely more aware of /l/ velarization since the Spanish
dominants avoid using it in their speech. This goes with previous studies that Davidson cites that claim /l/ velarization is a stereotype of Spanish in contact with Catalan.

With regards to position, McKinnon (2012) and Davidson (2014) find an effect of Catalan contact: word-final /s/ voicing is present while word-medial voicing is almost non-existent. Both authors attribute this to the fact that Catalan has a phonemic voicing contrast between /s/ and /z/ word-medially, but not finally. Their bilingual participants do not voice word-medial /s/ when speaking Spanish because that is precisely the context where they can distinguish [s] and [z]. Torreira & Ernestus (2012) and Hualde & Prieto (2014) also find more word-final voicing for the “non-contact” variety of Madrid Spanish, but both find much more voicing in word-medial and intial contexts than either McKinnon (2012) or Davidson (2014) do, providing support to the contact hypothesis.

Given the results of both Catalanian and Madrileño Spanish, it is clear that Catalan contact is not the sole cause of intervocalic /s/ voicing, but that it is likely to influence the factors that condition /s/ voicing.

Of the social factors examined, those that are most relevant to the present study are age and gender. If intervocalic /s/ voicing is indeed coming into Lojano Spanish, then we should see a clear pattern among different age groups in Loja. As mentioned previously in this section, the apparent-time construct can be used to examine linguistic change, based on the assumption that younger generations will be adopting innovative features while older generations will be set in their conservative ways of talking. The interaction between age and gender can illustrate even clearer patterns of what is going on in the society and certain patterns have surfaced time and time again, as was discussed previously. In the case of Lojano Spanish, if intervocalic /s/ voicing is a prestige feature
and the variation pattern follows previous studies, then the young women should be leading this innovation. Again, it is essential to take into account the size of the sample when drawing conclusions based on age and gender and also to consider the effect that individual speakers may have on broader trends. Finally, while contact with another language is clearly of importance to the realization of this feature in Catalan Spanish, it carries less weight for the present study given that most people living in the city of Loja are monolingual Spanish speakers. Instead, we might consider contact between dialects in that Loja has been increasingly exposed to Cuencano and Quiteño Spanish, which could result in influence from these dialects where intervocalic /s/ voicing is prevalent.

### 2.4.6 Origins of intervocalic /s/ voicing

One of the issues to consider in relation to intervocalic /s/ voicing is whether this phenomenon is a remnant of the voicing contrast found in medieval Spanish or whether it is an independent development. While this is not the focus of the present study, it has been an important consideration in many of the analyses of intervocalic /s/ voicing, and thus will be reviewed here. There are three hypotheses about the origins of modern /s/ voicing in HES: (1) it is an archaism that has remained from Old Spanish, (2) it has resulted from contact with Kichwa, (3) it is a modern development, the result of phonetic forces.

The first of these three hypotheses maintains that modern /s/ voicing in HES is a remnant of the voicing contrast previously found in Old Spanish, as described in section 2.3.1. Robinson (1979) is the main proponent of this theory, showing that modern voicing in Cuencano Spanish occurs in every position that it did in Old Spanish, except for word-
medial intervocalic position, where Cuencano Spanish does not voice. Thus, he seems to imply that in HES the devoicing process took root in word-medial position, where /s/ and /z/ contrasted phonemically in Old Spanish, but did not take place in the other intervocalic contexts, leaving [z] before a morpheme or word boundary.

The second hypothesis looks to contact with Kichwa to explain modern voicing. Toscano Mateus (1953: 23) mentions, “El quichua ecuatoriano presenta rastros de la antigua s sonora en viejos préstamos, como cazarana (casarse) y cazuna (hacer caso). Hay que tener en cuenta que la s sonora es también un fonema quichua” [Ecuadorian Kichwa presents vestiges of the old voiced s in old borrowings, such as cazarana (to get married) and cazuna (to pay attention to). One has to take into consideration that the voiced s is also a phoneme of Kichwa]. He goes on to state (p. 29), “En el español vulgar de la Sierra han penetrado fonemas quichuas: se pronuncia, por ejemplo, la s sonora en puzu” [In the vernacular Spanish of the Highlands, Kichwa phonemes have penetrated: one pronounces, for example, the voiced s in puzu]. Although he does not explicitly state this, Toscano Mateus seems to imply here that there could be substrate influence. Rosenblat (1967: 130) rejects this claim, suggesting that voicing in medieval Spanish must have influenced voicing of quechuismos since /z/ did not originally exist in Quechua, though he does not provide any proof from historical documents for this latter assertion. Cassano (1974) later maintains Ronsenblat’s position since /s/ voicing was known anecdotally to occur in varieties other than Ecuadorian Spanish. Apart from Spanish borrowings into Kichwa and vice versa, Muysken (1985) also notes the presence of [z] in some Spanish verbal roots in Media Lengua, a bilingual mixed language spoken most notably in Salcedo, Ecuador. These verbal roots include azi (hacer ‘to do/make’).
and *dizi* (*decir* ‘to say’), among others. Finally, in my own study of Kichwa, I have noted that Quindi Pichisaca (2010) lists /z/ as a phoneme of Austro-Kichwa, but only gives the following examples for words with /z/: *zampu* ‘of mixed race’, *zipi* ‘chapped skin’, *kuzu* ‘worm’, *guzu* ‘swamp/bog’, *zupu* ‘bump’ (p. 18). Thus, it is hard to give much weight to the Kichwa contact hypothesis because /z/ does not appear to be a productive phoneme in the language.

Finally, other scholars have invoked phonetic forces to support the idea that intervocalic /s/ voicing is a modern innovation (Torreblanca 1986, Widdison 1997, Bradley & Delforge 2006, among others). Torreblanca (1986) references data from many different dialects of Spanish to discuss in each case whether modern intervocalic /s/ voicing is a relic from medieval Spanish or a modern innovation. Overwhelmingly, Torreblanca decides that most cases of modern intervocalic /s/ voicing are due to articulatory relaxation and are not remnants of medieval sibilant voicing, because the distribution of [z] in these modern dialects is different than that of medieval Spanish. He argues that a few dialects, Judeo Spanish for example, clearly exhibit the medieval voicing distinction, since their sibilant voicing distribution is parallel to that of medieval Spanish. In the case of Ecuador, Torreblanca assesses the information provided by Toscano Mateus (1953) and Robinson (1979) and decides that the information is too contradictory to be able to make a claim about whether Highland Ecuadorian /s/ voicing is an archaism or also due to articulatory relaxation.

Widdison (1997) provides additional phonetic reasoning for the sibilant patterns

---

6 This denomination is used to refer to the provinces of Azuay, Loja, Zamora-Chinchipe, and El Oro. Cuenca is the capital of the province of Azuay and the city of Loja is the capital of the province of Loja.
observed in Spanish, including the historical devoicing examined in section 2.3.1, the modern devoicing of /ʒ/ in Argentine Spanish, and the modern voicing of intervocalic /s/.

Through the examination of sibilant patterns in other languages and from the perception data that he collected, Widdison is able to show how the aforementioned sibilant (de)voicing processes may have their origin in the effects of co-articulation and the nature of phonetic gestures. Widdison’s perception experiment, which takes advantage of the variation between [ʒ] and [ʃ] in Porteño Spanish, shows that voiceless fricatives may be perceived as voiced when the duration is “maximally abbreviated” (p. 261). With these results, he hypothesizes that duration may have served as the primary distinguisher between voiced and voiceless sibilants in Medieval Spanish, and thus for a time compensated for the phonetic devoicing that occurred. In the case of intervocalic /s/ voicing, the opposite argument explains why a “greatly abbreviated syllable-final [s]” (p. 262) may spontaneously voice even when it is not followed by a voiced consonant.

Torreblanca (1986) and Widdison (1997) both show that there are certain articulatory, acoustic, and perceptual factors to take into consideration when analyzing voicing in sibilants.

Going along with the idea that modern /s/ voicing is an innovation, a few scholars have noted that it may be a feature that is spreading throughout sub-dialects of HES, and as such we should find varying rates within the Highlands. Chappell (2011) mentions preliminary observations of a male Lojano speaker in which she finds voiced tokens, but not to the extent that she found for Quiteño data. Chappell suggests that perhaps intervocalic /s/ voicing is encroaching on Lojano Spanish due to the increased contact
with other HES varieties (e.g. Cuencano). Since Strycharczuk et al. (2013) find that some Quiteño speakers voice categorically while others do not, they hypothesize that intervocalic /s/ voicing might be on the way to becoming a categorical process. While they do not provide data from Loja, these authors seems to suggest that other HES varieties might be further behind Quiteño Spanish on this trajectory.

Despite the recent increase in studies looking at intervocalic /s/ voicing, there are still several areas of research that should be examined. Most importantly, with only a passing mention in Robinson (1979) and Chappell (2011), the status of this feature in Lojano Spanish is still unknown. It remains to be seen whether intervocalic /s/ in Lojano patterns with other Highland varieties, and furthermore which linguistic and social factors influence voicing in this variety. Past studies provide important insights into which techniques can be used to investigate intervocalic /s/ voicing in Lojano Spanish and which factors could prove to be important.

2.5 Perception

Given that intervocalic /s/ voicing is present to different extents in the sub-dialects of HES, one might also wonder if speakers of HES are aware of this variable. It is one thing for linguists to notice this feature and measure it, but this does not address whether speakers hear the different variants. The question that arises is, are naïve native speakers of HES able to perceptually differentiate between [s] and [z] in intervocalic position? In this section, I will address different ways in which this question can be answered.
2.5.1 Categorical and allophonic perception

There are many ways of testing naïve native speakers’ perception, and each has its own advantages and disadvantages depending on the research questions. One extremely popular paradigm is that of categorical perception. Liberman et al. (1957) originally designed the categorical perception task, and since then it has been applied to many different phenomena in many languages (cf. Pisoni & Lazarus 1974, Logan et al. 1991, Martínez-Celdrán 1993, Schneider et al. 2003, among many others). The basic design involves two parts: an identification task and a discrimination task. In an identification task, participants are presented with a series of stimuli and are usually asked to circle or write down what they hear. Most often, identification tasks are forced-choice in that participants are given the possible options they are to choose from. There are two main types of discrimination tasks. In an AX discrimination task, participants hear two stimuli and have to decide if they are the same (the same stimulus played twice) or different. Critically, the participants hear pairs where the correct answer is same and other pairs where the correct answer is different. In an ABX discrimination task, participants hear three stimuli and have to decide if the last one they hear (X) is more similar to the first one (A) or the second one (B). In this type of task, stimulus X is identical to either A or B.

Categorical perception tasks have most often been used to investigate the role of one or more acoustic cues on participants’ perception, and to do this, researchers usually employ manipulated stimuli. For instance, in their original experiment, Liberman et al. (1957) synthesized tokens of /b, d, g/ that varied in the direction and length of the second

---

7 Because forced-choice is the most common type of identification task, the reader can assume any identification task referred to is of this type unless otherwise specified.
formant transition, creating a 14-point stimulus scale. For the identification task, participants in this study heard these synthesized tokens and had to decide if they heard ‘b’, ‘d’ or ‘g’. The ABX discrimination tasks that the authors used varied in whether they were one-step, two-step, or three-step. In a one-step task, participants hear stimuli A and B that are adjacent on the stimulus scale – for example, one stimulus in which the second formant transition is 300 ms and another stimulus in which the second formant transition is 290 ms. In a two or three-step task, however, participants hear two stimuli that are two steps away on the stimulus scale (300 ms and 280 ms) or three steps away (300 ms and 270 ms). Of course, the researcher is responsible for deciding what the relevant “step” is based on what acoustic property is being manipulated. The results of the identification and discrimination tasks are then compared to see if the participants break up the stimulus scale into categories and where the boundary between these categories falls. Liberman et al. (1957) conclude that in general their participants exhibit categorical perception since the crossover points in the stimulus scale between categories /b/-/d/ and /d/-/g/ from the identification task correspond to the points at which the participants most easily discriminated tokens.

Martínez Celdrán (1993) carried out a similar categorical perception task for Spanish [β, p, pp], manipulating only the closure duration, and concludes that “kaba” is perceived when there is 26.4-61.6 ms of closure duration, “kapa” for 70.4-140.8 ms, and “kappa” for 149.6+ ms of closure duration. He finds the same crossover pattern that is indicative of categorical perception. For instance, the participants in his study judged

---

8 These are not the actual stimuli that Liberman et al. (1957) used. I have decided to simplify for the sake of clarity in describing the task.
stimuli as “kaba” up until 61.6 ms of closure duration and from there on started judging stimuli with higher closure durations as “kapa”. The participants also responded “different” when hearing A and X of 61.6 ms and 70.4 ms more consistently because these two stimuli belong to two different categories (61.6 = “b”, 70.4 = “p”). On the other hand, when hearing two stimuli of the same category, for instance tokens with 26.4 and 61.6 ms closure durations, participants did not respond “different” as frequently.

While the traditional design includes identification and discrimination tasks, other scholars have used one or the other to look at the influence or weight of acoustic cues in the perception of contrasts. Cole & Cooper (1975) and Stevens et al. (1992) use identification tasks to investigate the influence of fricative duration, vowel duration, voicing, and transitions on the perception of voiced and voiceless fricatives in English. The advantage of the identification task, whether used alone or in tandem with a discrimination task, is that it allows the researcher to explore the effect of one or more acoustic cues and it usually provides straightforward, easily interpretable results. The disadvantage is that it relies on orthography or some written way for participants to show their identification of each token they hear.

Although most of the studies already cited have been concerned with the perception of phonemes, it is important to note that this is not the only type of perception. Just because participants do not exhibit categorical perception, it does not mean that they do not hear some type of difference between stimuli. This point has been taken up by studies looking at allophonic perception. In these studies, the use of identification tasks has been impossible because allophones are usually not distinguished orthographically. Two such studies (Boomershine et al. 2008, Johnson & Babel 2010) are important here for their
methods even though they do not consider fricative voicing.

Boomershine et al. (2008) look at the perception of [d], [ɾ] and [ð] in Spanish and English. The authors are interested in how allophony versus contrast affects speech perception: the contrast [ɾ] - [ð] is phonemic and can be found in minimal pairs in both Spanish and English, while [d] - [ɾ] is phonemic in Spanish but cannot be found in minimal pairs since the two sounds occur in distinct environments, and [d] – [ð] is phonemic in English and also present in minimal pairs. To get at this research question, they use a similarity rating task and a speeded AX discrimination task. The stimuli for the first two experiments were VCV sequences recorded in English where the C was [d], [ɾ] or [ð] and V was [a], [i] or [u]. In Experiment 1, a similarity rating task, Spanish and English speaking participants heard pairs of these stimuli and had to rate them on a scale from (1) very similar to (5) very different. The results of normalized ratings show that [d]-[ɾ] are rated more different for Spanish participants, [d]–[ð] more different for English participants, and the [ɾ]-[ð] ratings do not differ between the two participant groups.⁹ Experiment 2, a speeded AX discrimination task, followed the same basic design as described before, except that participants only heard 100 ms of silence between trials in order to encourage them to respond quickly. Their reaction times were recorded. The underlying assumption is that reaction times for the stimulus pairs that are actually different (where A is a different stimulus than X) can be taken as a measure of perceptual distance. Slower reaction times are indicative of smaller perceptual distance while faster reaction times correlate with larger perceptual distance. The normalized reaction times

⁹ Boomershine et al. (2008) do not find any effect of which vowel was in the VCV sequence.
exhibit the same pattern of results seen in Experiment 1, showing that similar language-specific effects emerge in both off-line (similarity rating) and on-line tasks (discrimination). Given that there was a potential confound with the stimuli being recorded in English, Boomershine et al. then replicate Experiments 1 and 2 with VCV sequences recorded in Greek, which has a contrast between all three sounds, and find very similar results. Based on these results, the authors conclude that the phonological relationship between two sounds in a language does influence how they are perceived. It is interesting that even though English speakers heard more of a difference in the [d]–[ð] pair, Spanish speakers still heard some difference between these sounds which are allophones in Spanish.

Johnson & Babel (2010) use an identical methodology to examine the perception of [f, θ, s, ʃ, x, h] in English and Dutch. While English lacks [x], Dutch does not have [θ] and [ʃ] is phonologically ambiguous. The stimuli for both experiments consisted of English VCV syllables with each of the six fricatives and either [a], [i] or [u]. In the similarity rating task, as compared to English listeners, Dutch listeners rated [s]-[ʃ] as more similar with all vowel pairings and [s]-[θ] and [ʃ]-[θ] pairs as more similar for some vowel pairings. There was no difference between Dutch and English for pairs with [x], [f] or [h]. Contrastively, the results of the speeded AX discrimination task using the same stimuli show no significant effect of the participants’ language. Their results show that it is not just phonemic or allophonic status that affects perception. Since [s] and [ʃ] only contrast in a few contexts in Dutch, they were judged as more similar by Dutch speakers
than by English speakers. It is noteworthy that despite replicating the methods of Boomershine et al. (2008), there is no effect of native language in the speeded discrimination task. Johnson & Babel mention that this may be due to the fact that they encouraged participants to respond as quickly as possible, and thus obtained overall lower reaction times than the previous study.

Clearly, these two types of tasks can be useful for looking at allophonic perception. The main advantage of the similarity rating task is that it does not rely on orthography. Participants can indicate their responses of more or less similar without having to identify the sounds they hear, which perhaps would be an impossible task in the case of allophones. Of course, because it is not an identification task, similarity ratings can only tell us what participants hear as more or less similar in the stimuli and not if they hear two categories. As with the similarity rating task, discrimination tasks do not rely on orthography but cannot be used on their own as proof for categorical perception. In addition, speeded discrimination tasks require sophisticated technological equipment since reaction times are recorded. With any perception task, it is vital to consider what the aims of the experiment are since each of the tasks described here is more or less appropriate depending on the research question.

2.5.2 Perception of fricative voicing

Of the methodologies examined in the last section, identification tasks have most often been used to investigate the perception of fricative voicing. Studies looking at voiced and voiceless fricatives are more common in English, but there are a few studies that offer some preliminary conclusions on how fricative voicing is perceived in Spanish.
Past studies have shown there are many acoustic cues to voicing: fricative duration, vowel duration, location and duration of voicing within the fricative, and fricative-vowel transition. Given the breadth of studies in English, I will focus on those that look at syllable-initial fricatives since that is the position I am interested in. Also, I will review only results of English [s] and [z], and not other fricatives, for the sake of brevity.

Cole & Cooper (1975) present results on two series of perception experiments comparing syllable-initial voiced and voiceless affricates and fricatives in English. Both experiments included an identification task in which university students had to circle which syllable they heard, for instance ‘SA’ or ‘ZA’. In the first experiment, participants heard stimuli of natural CV tokens of [sa] with variable fricative and vowel durations. The original [sa] token was shortened in increments of 40 ms to produce a six-step scale: 240 (original), 200, 160, 120, 80, and 40 ms. Fricative portions were removed from just before the vowel since there are more cues in the beginning of a fricative. The following vowel was either 100, 200 or 300 ms long. The results from the first experiment show that [sa] stimuli with 80 and 40 ms fricative portions are perceived voiced the majority of the time regardless of the length of the vowel. There was a significant interaction between fricative duration and vowel duration, whereby a longer vowel slightly increases the number of voiced responses, but no independent effect of vowel duration was obtained. The results of the second experiment are not relevant for the present study, and for this reason they are not reported here. Based on the results from the first experiment, Cole & Cooper conclude that fricative duration is a better cue to voicing in syllable-initial [s], contrary to what had been found for syllable-final position, where vowel duration plays much more of a role (Raphael 1972).
In a similar study, Jongman (1989) aims to determine the minimal length of frication noise required in order to properly identify English fricatives. The stimuli for this study were CV syllables with following [i, a, u] that were then manipulated to include the full CV syllable, the full frication, or 70, 60, 50, 40, 30, 20 ms of frication. Whereas Cole & Cooper (1975) always presented participants with a CV syllable, this study also includes stimuli of isolated frication. Twelve university students participated in an identification task in which they were instructed to write down the sound they heard (from a list of options). For the identification of [s] and [z], there was no effect of following vowel in the full CV condition and very little effect of frication duration on voicing perception in the rest of the stimuli. For instance, even in the 20 ms condition, [s] was perceived voiceless 92% of the time, although [z] was only perceived voiced 68% of the time in this condition. Given the confusion rates of [s] and [z], Jongman concludes that 30 ms of [z] and 50+ ms of [s] are necessary for correct identification. Fricative duration is not found as important in this study; however, it is possible that participants were using a different mode (i.e. not speech) of perception since they heard the fricative portions in isolation.

The last study for English, Stevens et al. (1992), provides evidence that other factors in addition to vowel and fricative duration can affect voicing perception. They present the results of an acoustic study as well as four perception experiments, of which Experiments 2-4 (all identification tasks) are relevant here. In Experiment 2, VCV tokens were synthesized to contain [s] of 70, 85, and 100 ms duration. In addition, glottal vibration of different amplitudes and durations was put at the beginning or end of the fricative. There is a complex interaction between these characteristics, but overall they
find that shorter fricative duration, longer glottal excitation, and higher voicing amplitude lead to more voiced responses, and that glottal excitation at the beginning of the fricative has more of an impact. In Experiment 3, VCV tokens were synthesized with different locations and durations of glottal excitation, and different lengths of F1 transitions. More voiced responses are obtained when the voicing is placed at the edges of the fricative as opposed to in the middle. The effect of F1 transitions is minimal, but a full F1 transition can sway the responses towards voiced even when no voicing is present if the fricative is quite short (70 ms long). Finally, Experiment 4 looks at the perception of naturally produced [asa] and [aza] in which portions of the middle of the fricative were taken out to produce 70, 85, and 100 ms long fricatives. The results show a slight trend in the expected direction, shorter tokens are perceived as more voiced, but overall the participants were extremely accurate: judging [asa] as [s] and [aza] as [z] over 90% of the time regardless of duration. Stevens et al. conclude that the portion of the fricative in which there is no glottal vibration is most important in the identification of voiced and voiceless fricatives – the portion of the fricative in which there is no glottal vibration must exceed 60 ms in order to be heard as voiceless.

In summary, we can see that the most important acoustic cues to differentiating syllable-initial [s] and [z] in English are fricative duration and glottal vibration, with vowel duration and F1 transitions as secondary cues. The differences seen between the results of these studies are likely due in part to methodological disparities; for example, Cole & Cooper (1975) consider post-pausal position while Stevens et al. (1992) consider intervocalic position. It is important to note that none of these studies has proven categorical perception of [s] and [z] in English, as they do not include discrimination
tasks. Studies in Spanish have also looked at the effect of fricative duration and glottal vibration on perception, but in general have been less thorough.

Widdison (1995, 1996, 1997) reports on one perception experiment concerning palatal fricatives and a second experiment that looks at the perception of /s/. In the first, a female speaker of Buenos Aires Spanish was recorded producing phonetically voiceless [ʃ], and this original 165 ms token was abbreviated to 125, 85, and 45 ms by removing 40 ms portions from the middle of the fricative. The participants, Spanish speakers from various countries studying at an American university, were instructed to choose between <SH> and <ZH> for the stimuli they heard. Since some participants were not familiar with Argentine Spanish, Widdison carried out an extensive training portion in which he instructed them that <SH> is indicative of Buenos Aires pronunciation while <ZH> is characteristic of other Southern Cone dialects. Eight participants scored better than 75% on this pretest and moved on to the experimental identification test. The results show that 165 and 125 ms tokens are perceived as majority <SH> and 85 and 45 ms tokens are perceived as majority <ZH>. Given the extensive training and that the participants were not judging their native dialect, it is unclear how natural these results are. In Widdison’s second experiment, phonetically trained participants rated voicing on a scale of 1 to 5 for tokens of ‘los otros’ that had been manipulated for duration and intensity. Regrettably, there was very little consensus between the participants and thus the results of this second experiment are inconclusive.

\[10\]

It is unclear whether any participants in this study are actually from Buenos Aires, but if this is the case, they would likely know that both [ʃ] and [ʒ] can be heard in Buenos Aires Spanish, a potential confound.
Blecua & Rost (2013) find a variety of productions of intervocalic /f/ in casual speech that include: voiceless fricative [f], partially voiced fricative [f̬], voiced fricative [v], and voiced approximant [β]. In addition to production data, these authors also carried out a perception experiment to see whether this variation in production gives rise to perceptual variation. For the stimuli, nonsense (C)V_V sequences were taken from their spontaneous corpus. These tokens contained one of the four allophones of /f/ or an allophone of /b/. Upon hearing the stimuli, phonetically trained Spanish participants had to choose between orthographic “ifo”, “ivo”, and “ibo”. Blecua & Rost note that they decided to use phonetically trained participants to ensure they could distinguish “ivo” since [v] is not a phoneme of Spanish. The majority of participants heard fully voiced variants of /f/ as “ivo” or “ibo” and partially voiced and voiceless [f] as “ifo”. The authors conclude that listeners can interpret intervocalic /f/ as something different than what the speaker intended. Though this study provides some evidence that speakers can potentially hear sub-phonemic differences, there are a few issues with the design. As with the previous study, it is unclear here whether the results are representative of what naïve native speakers would hear since the participants were phonetically trained. In addition, the stimuli are likely to have differed in more than just voicing, for instance vowel duration, which could affect the results.

Finally, Gradoville (2011) also provides some pertinent insights despite not having carried out a perception experiment. This author had two native and two non-native linguistically trained judges impressionistically code Buenos Aires palatal sibilants as [ʃ] or [ʒ]. He notes, “Something that is interesting to note here is that tokens with as little as
40% voicing (60% voiceless) are perceived as voiced” (p. 69). Gradoville goes on to state that he believes glottal vibrations at the fricative-vowel edges are most important in identifying it as [ʒ]. If this were in fact the case, it would show that what Stevens et al. (1992) found for location of glottal vibration is also true for Spanish fricatives. Still, it is essential to take into account fricative duration in addition to glottal vibrations, especially since Martínez Celdrán (1993) shows that perception of voicing in Spanish bilabial stops is dependent on closure duration and not glottal vibrations. Taking all the evidence together, it appears that both fricative duration and glottal activity affect the perception of fricative voicing in Spanish, but this remains to be confirmed for /s/, and indeed it remains to be confirmed whether native speakers hear the distinction [s]-[z] at all.

2.6 Research questions

The preceding review of previous literature illuminates many areas of research that deserve more careful attention. First and foremost, while there are production studies of both Quiteño and Cuencano Spanish, no study has systematically examined intervocalic /s/ voicing (or any other feature) in the Spanish spoken in Loja. As such, the research questions for the production data in the present study are two-fold. The first is whether there is some degree of intervocalic /s/ voicing present in the Spanish spoken in Loja, Ecuador. If this feature is indeed present in Lojano Spanish, the second question is which linguistic and social factors influence intervocalic /s/ voicing in this dialect. Beyond filling a gap in the literature, production data looking at these two questions can inform a larger question that has arisen in phonetics and phonology. Many recent studies in
laboratory phonology have shown that production processes traditionally thought to be categorical are gradient in nature when examined with acoustic measures. Given the aforementioned debate as to whether intervocalic /s/ voicing is a categorical or gradient process, the instrumental assessment of this feature lends itself to be part of this larger discussion in laboratory phonology. Going along with recent studies, my aim is to show that the division between phonetics and phonology is not entirely clear-cut, and that phonetic and phonological factors can both inform our understanding of processes such as intervocalic /s/ voicing. Furthermore, I will show that considering the interplay of phonetic/phonological and social factors provides a broader, more complete analysis of this feature.

Turning to perception, no study to my knowledge has provided empirical evidence for the perceptual status of [s] and [z] in HES, although Robinson (1979) claims impressionistically that Cuencano speakers can and do distinguish pairs such as ha sido ‘it has been’ [a.si.ðo] and has ido ‘you have gone’ [a.zi.ðo]. In fact, previous studies on perception of Spanish fricative voicing in general are few and far between. The question here is whether naïve native speakers of Lojano Spanish are able to distinguish [s] and [z] in intervocalic context. The investigation of this research question has greater implications for the study of allophonic perception. Recent studies have shown that listeners may be able to perceptually distinguish allophones in their native language (Boomershine et al. 2008, Johnson & Babel 2010), but more evidence is needed to confirm the relationship between allophonic or phonemic status and perception.
CHAPTER 3: PRODUCTION

3.1 Introduction

As described in section 2.4, intervocalic /s/ voicing has been attested in several modern dialects of Spanish, and previous studies have found a variety of linguistic and social factors relevant to this linguistic phenomenon. Studies on HES and other dialects have confirmed that the most influential linguistic factors are position within a word, speech rate, and stress. Depending on the dialect, such social factors as level of bilingualism and gender have also been found to affect intervocalic /s/ voicing. The production data described in this chapter aims to determine to what extent intervocalic /s/ voicing is present in Lojano Spanish, and which linguistic and social factors have an effect on its realization. In section 3.2, I will describe the methodology utilized to elicit production data, including participants, interview schedule, stimuli, acoustic analysis, and independent variables. For section 3.3, I will focus on the results of the production data, including a description of the statistical analyses, with the results divided by continuous and categorical analyses. Finally, in section 3.4, I will conclude with a summary of the production results.
3.2 Methodology

3.2.1 Participants

In order to get at the research questions previously mentioned, I interviewed thirty-five native speakers of Lojano Spanish during December 2010 and August 2013. I recruited participants who had lived the majority of their lives in the city of Loja, using the “Friend of a Friend” technique (Milroy 1980). I tried to obtain a sample that was balanced for gender as well as age. Older participants were hard to come by, and so the sample is a little biased towards the younger age groups. Of the thirty-five interviews that I collected during the two trips, I ended up with thirty-one usable interviews, sixteen with males and fifteen with females. The youngest participant is 20 years old, the oldest is 68 years old, and the average age in the sample is 38 years old. These participant characteristics can be seen in Figure 4, where each age group is broken up by female (red) and male (blue) participants. As we can see, the ages are relatively balanced for gender, with the exception of the aforementioned issue with locating older participants.
Overwhelmingly, the participants are middle or upper middle class and highly educated. Eighteen of the participants were getting a bachelor’s degree or had already obtained a bachelor’s degree. Nine had obtained graduate or professional degrees (in addition to bachelor’s degrees). Only two had obtained no higher degree than a high school diploma and another two have primary school educations. Given the relative homogeneity of the sample, it is not representative of all social strata of Loja. Most of the participants have taken English at some point in their life, but none have lived in the United States for a significant amount of time or consider themselves fluent in English.

All of the 20-30 year-old participants are college students or research assistants, except one 29 year-old who is an entrepreneur. The occupation of the remaining participants varies greatly and includes: small business owner, janitor, government employee,
agronomist, professor, radiologist, housewife, TV show host, surgeon, designer, project manager, and retired judge.

3.2.2 Interview schedule and token extraction

Each interview lasted between one and two hours and consisted of five tasks. All interviews were digitally recorded on a Zoom H2 Handy Portable Stereo Recorder. The interviews are part of a larger project on Lojano Spanish and thus the present analysis is only concerned with two of the five tasks. In the first task, the questionnaire, the participants were asked orally a series of background questions about themselves, in addition to questions about the city of Loja and how it has changed during their lifetime. A list of these questions can be found in Appendix A. The second task, the reading task, consisted of the participants reading sentences from PowerPoint slides as I clicked through the slides. The remaining tasks are not used in this analysis.

For each participant, sixty tokens where extracted from the questionnaire period. Where possible, these tokens were taken from the participants’ answer to the question, “What is the city of Loja like?” This portion of the questionnaire was chosen because it was the last question in the questionnaire period, and thus the participants had already become relatively comfortable with the interview situation and the interviewer. Of these sixty tokens, there were twenty (the first twenty available) in each of the word positions described in section 2.4: word-initial, medial and final. For the purpose of this analysis, the tokens that came from the questionnaire will be denoted “interview”.

The reading task was designed to elicit tokens of /s/ in different intervocalic contexts. Besides the tokens with /s/, there were also fillers and sentences that elicited
/ʎ/, /ʃ/ and /r/. The reading stimuli contained 32 target words with intervocalic /s/ that were each uttered twice, producing 64 randomized tokens for each participant. These target words were distributed across four intervocalic contexts: word medial at a morpheme boundary (biʃabuelo ‘great-grandfather), word medial not a morpheme boundary (caʃa ‘house’), word-final (has ido ‘you have gone’) and word-initial (ha ʃido ‘it has been’). The target words contained target /s/ where the following vowel was stressed or unstressed, although stress was ultimately coded in a different manner, as will be described later. A complete list of the target words can be seen in Table 1. The target words were presented to the participants in carrier sentences and the participants were asked to read each sentence in a natural, relaxed style as if they were talking to a friend. The complete list of stimuli can be found in Appendix B. For each participant, 60 interview and 64 read tokens were analyzed. A few tokens had to be excluded from the analysis due to pauses, unstressed vowel reduction, mispronunciations, and poor recording quality or background noise. In total, 2,969 tokens of intervocalic /s/ were analyzed from the 31 speakers.
Table 1: Target words used in reading task

3.2.3 Acoustic analysis

After extracting the tokens, a series of acoustic measures were taken from each token using the synchronized spectrograms and waveforms in Wavesurfer (Sjölander & Beskow 2010). As discussed in section 2.4.3, studies on /s/ in Spanish, and on sibilants more generally, have used many different acoustic and articulatory measures to analyze voicing. Given the results of these studies, I believe that overall the best acoustic measure of glottal vibrations is to calculate percent voicing by hand. Accordingly, the dependent measure I use here is percent voicing. To calculate percent voicing, I first measured the duration of the fricative in each instance of /s/. Since all of the tokens were intervocalic, I judged the beginning of the fricative by the disappearance of the formant structure of the preceding vowel, change in the waveform, and the presence of frication in the high frequencies. Similarly, the end of the fricative was judged by the onset of the formant structure of the following vowel, a change in the waveform from the transition of the fricative to the vowel, and the disappearance of frication in the high frequencies. I then
measured the duration of voicing of each token. For the tokens that were not fully voiced, I took a measurement of the voicing at the beginning of the fricative and added that to the voicing at the end of the fricative in order to obtain the total voicing. In the case of tokens that were not fully voiced, it was much more common that there was a little voicing at both edges than just voicing at the beginning or end of the fricative. Voicing was judged using two cues: periodicity of the waveform and voicing bar in the spectrogram. Finally, I divided the total voicing by the total duration of the fricative to get the percent voicing. Any tokens that were unclear in terms of total duration or duration of voicing were checked with another phonetician.

Figure 5 and Figure 6 serve to exemplify how the acoustic analysis was done. The boxes in both figures highlight the voiced portion of the fricative. Figure 5 shows the production of the word *proceso* ‘process’ from a female speaker (F10), taken from the reading task. This token is 88 milliseconds long and 27% voiced. At the beginning and the end of the fricative, we can see a small portion of periodic wave, which is accompanied by a voicing bar. Figure 6 shows the production of the word *casa* ‘house’ from a male speaker (M7), taken from the reading task. This token is 68 milliseconds long and 100% voiced. As evidenced by the periodic waveform and continuous voicing bar, the entirety of this token is voiced.
Figure 5: Read production of *proceso* from participant F10 – 27% voiced

Figure 6: Read production of *casa* from participant M7 – 100% voiced
### 3.2.4 Independent variables

The independent linguistic factors taken into consideration are: position within a word, preceding vowel, following vowel, stress, and local speech rate. Position within a word is coded as word-final, word-medial at a prefixal morpheme boundary, other word-medial, or word-initial. These four categories were later collapsed into three based on the results and token distributions: word-final, word-medial, and word-initial. There was no significant difference between word-medial at a prefixal morpheme boundary and other word-medial in either type of data, which justifies the decision to collapse them into the same category. A recent study has called into question whether word-final prevocalic /s/ undergoes resyllabification in HES (Robinson 2012). Thus, in order to avoid having to decide whether word-final /s/ is part of the preceding or following syllable, the factor “stress” is coded according to the stress of the preceding and following vowels, producing the following options: unstressed-unstressed (los abuelos ‘the grandparents’), unstressed-stressed (las águilas ‘the eagles’), stressed-unstressed (autobús azul ‘blue bus’), and stressed-stressed (manatí únicos ‘unique manatees’). Stress is coded in the same manner for all word positions, although it is important to note that stressed-stressed tokens are impossible in medial position given that words in Spanish can only carry one primary stress. Preceding and following vowel are coded as the vowel that immediately precedes or follows the /s/ in question, with the following options: [i], [e], [a], [o], or [u]. In the case of diphthongs, the glide closest to /s/ was coded as its nuclear vowel equivalent. Thus, in a token such as la suegra ‘the mother-in-law’, the following vowel would be coded as [u]. Similarly, for a token such as habéis amado ‘you all have loved’, the preceding vowel would be coded as [i].
The local speech rate for each token was measured in syllables per second, combining the techniques of File-Muriel and Brown (2011) and Kendall (2009). File-Muriel and Brown (2011) calculate the local speech rate as the phonemes per second in the three word sequence surrounding a token of /s/, that is, including the word before and after the target word. According to Kendall’s (2009) review of speech rate, it seems that most scholars prefer to use syllables per second. Thus, I measured syllables per second of the three-word sequence. To do this, I first took the duration of the sequence from the word before the target word to the word after the target word, for example, *todos los amigos* ‘all of the friends’. If there was a pause before or after the target word, then this sequence consisted of only two words instead of three. I then divided the number of syllables in this sequence by the duration in seconds to get the local speech rate at the time of the token. Position within a word, preceding vowel, following vowel, and stress are categorical, while local speech rate is a continuous variable. The two social factors considered here are age and gender (male or female) of participant. Age is considered as a continuous factor.

3.3 Results

3.3.1 Statistical analysis

Returning to the issue of data analysis, Rohena-Madrazo (2011), Schmidt & Willis (2011), Strycharczuk et al. (2013), A. García (2013), and Hualde & Prieto (2014) choose to keep percent voicing as a continuous dependent variable. One problem with this method of analysis is that percent voicing does not usually exhibit a normal distribution, which violates the assumptions of linear regression, as well as other
statistical tests. Rohena-Madrazo (2011) evades this problem by performing an arcsine transformation of his data in order to obtain a normal distribution. Other studies use percent voicing to derive a non-continuous variable and create different voicing categories. These studies that have employed a non-continuous analysis have “cut up the pie” in different ways, so to speak. While the binary analysis presented in Schmidt & Willis (2011) and that of Davidson (2014) both compare “majority” voiced versus voiceless tokens, Campos-Astorkiza (2014) and McKinnon (2012) use a tripartite categorization: voiceless (0-20%), partially voiced (20-90%) and fully voiced (100%). Furthermore, McKinnon (2012) carries out two binary analyses, one collapsing voiceless and partially voiced tokens and the other collapsing partially and fully voiced tokens, and does not find noteworthy differences between the results of the two analyses. Campos-Astorkiza (2014) justifies her three categories by showing that the majority of tokens of /s/ before a voiceless consonant are less than 20% voiced. However, the question of what categories to use is a difficult one since this decision can at times seem arbitrary.

Given the inconsistencies among previous studies as well as the distribution of my data, I consider percent voicing as both a continuous and categorical dependent variable. As File-Muriel (2012) points out, many previous studies on Spanish /s/ have unnecessarily forced what could be gradient processes to be categorical by using binary categories (aspirated versus sibilant, for instance). Using both continuous and categorical analyses will allow me to more completely examine intervocalic /s/ voicing in Lojano Spanish without assuming anything about the nature of this process, whether it is gradient or categorical.
Figure 7 shows a histogram of the distribution of percent voicing in my data. While there is a normal (although skewed) distribution of tokens below 90% voicing, there is a clear break between these tokens and the fully voiced tokens. In fact, there were no tokens between 90 and 99% voicing. Since overall the tokens are not normally distributed and instead exhibit two peaks, I do not employ linear regression, but rather regression with an inflated beta distribution for the continuous analyses. This type of regression allows for inflated values for 0 and 1 as compared to the rest of the scale, which is the case for percent voicing in my data. Additionally, the inflated beta distribution allows for a non-normal distribution of the data in between 0 and 1 because it is based on a beta distribution instead of a normal distribution, which solves the previous issue that other scholars have had with using linear regression. The continuous analyses were performed using mixed effects in R (R Core Team 2014) with the package gamlss (Stasinopoulos et al. 2015), which accounts for the distribution of the data by fitting four parameters to the data: mu, sigma, nu, and tau. The first, mu, shows the main effects of the model. The second parameter, sigma, which is based on the standard deviation, allows us to see which independent factors exhibit the most variation. Finally, nu demonstrates under what conditions a token is more likely to be at 0 than in between 0 and 1, and tau in turn demonstrates under what conditions a token is more likely to be at 1 than in between 0 and 1. The inflated beta distribution is most appropriate for my data because it is specifically designed to work with percentages. The inflated beta models were fit to the data using an iterative process, and the best model was determined by comparing AIC scores and residual plots between models.
Following Campos-Astorkiza (2014), I consider three categories in the categorical analyses: unvoiced (0-20%), partially voiced (20-90%) and fully voiced (100%). For these analyses, I will use ordinal logistic regression (also known as proportional odds modeling) to look at the distribution of the three voicing categories. Ordinal logistic regression allows for more than two categories and it assumes that there is a structure to the categories. That is, this test assumes that the three categories are hierarchically ordered: unvoiced < partially voiced < fully voiced. Using this test will allow me to see, for instance, what the odds of being in category 1 (unvoiced) are compared to categories 2 and 3 (partially/fully voiced). This analysis improves upon previous methodologies because it does not employ statistical methods that force variation into binary categories. Furthermore, ordinal logistic regression is more appropriate than standard multinomial logistic regression because it includes the information about the hierarchical order of the
data. The ordinal models were fit to the data using mixed effects with *clmm2* in the *ordinal* package (Christensen 2011) in *R*. These models were fit to the data using an iterative process, and nested models were compared using *ANOVA*.

The data from the read and interview modes were subjected to separate statistical analyses due to the following reasons. First, in both continuous and categorical analyses of the whole dataset, there is a very robust effect of mode, whereby there is more voicing in the interview data than in the read data. Additionally, it was determined that the distribution of percent voicing varied notably between the read and interview data. Figure 8 shows a histogram of the distribution of percent voicing in the read data (left) and the interview data (right). Finally, as seen in Figure 9, the amount of variance is much higher in the interview data than in the read data, and thus it is statistically unsound to combine the two sets of data into one analysis. It is likely that there is more variation in the interview data simply because the participants were speaking freely and not confined to the task of reading set sentences. Thus, the participants’ speech flows naturally with the ebb and flow of the conversation in the interview data, as opposed to being relatively constrained in the read data. This method of analysis improves upon previous studies by considering the effects of both mode and speech rate, without giving preference to one over the other, and by considering speech rate as a continuous variable.
Figure 8: Histogram of percent voicing for read data (left, N=1582) and interview data (right, N=1387)

Figure 9: Box plot of percent voicing in read data (left) and interview data (right)
For all of the aforementioned reasons, separate analyses were performed for read and interview data in order to examine whether different factors influence intervocalic /s/ voicing in each type of data. Mixed models were fit to the data using the independent factors described previously: position within the word, stress, preceding vowel, following vowel, speech rate, age, and gender. Speaker was also included as a random effect. One of the assumptions of regression analysis is independence – that each data point is an independent observation. My data violates this assumption since I have multiple data points for each speaker. Thus, including speaker as a random effect will allow me to ensure that the patterns seen for independent predictors go above and beyond what we would expect simply based on individual variation. Interactions between independent variables were included where relevant in all analyses, including interactions between social and linguistics factors, but no significant interactions were found. In total, four analyses were fit to the data:

1. Read data – Inflated beta regression – continuous percent voicing (N=1582); section 3.3.3
2. Read data – Ordinal logistic regression – voicing categories (N=1582); section 3.3.4
3. Interview data – Inflated beta regression – continuous percent voicing (N=1387); section 3.3.5
4. Interview data – Ordinal logistic regression – voicing categories (N=1387); section 3.3.6
3.3.2 Overall results

Before examining the effect of individual factors on /s/ voicing, it is important to look at the overall patterns. Table 2 shows the mean percent voicing by type of data and position within the word. What is first most evident from this table is that the average percent voicing is higher in interview data than in read data. Furthermore, it is already apparent that there is an effect of position within the word. The percent voicing is higher in word-final and initial position than word-medially, with word-final position showing the highest percent voicing of all positions. Crucially, percent voicing does not reach 100% in any given position in either type of speech.

<table>
<thead>
<tr>
<th></th>
<th>Word-Initial</th>
<th>Word-Medial</th>
<th>Word-Final</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Read Data</strong></td>
<td>33.5</td>
<td>30.9</td>
<td>44.1</td>
<td>36.2</td>
</tr>
<tr>
<td><strong>Interview Data</strong></td>
<td>45.9</td>
<td>40.7</td>
<td>50.5</td>
<td>45.7</td>
</tr>
</tbody>
</table>

Table 2: Means for percent voicing (N=2,969)

The overall patterns are further illuminated once we consider the voicing categories. Table 3 below shows the distribution of tokens in each voicing category by type of speech. As a reminder, the categories used are: unvoiced (0-20%), partially voiced (20-90%) and fully voiced (100%). Mirroring the pattern seen for mean percent voicing, there are more fully voiced tokens in interview than in read data. On the other hand, there are more unvoiced tokens in read than in interview data. Setting aside methodological differences for the moment, it is already clear that the general pattern seen here for Lojano Spanish is vastly different than the results upheld by previous studies of HES, since voicing is found in all three positions and is variable in each position.
### Table 3: Distribution of tokens by voicing category (N=2,969)

<table>
<thead>
<tr>
<th></th>
<th>Unvoiced</th>
<th>Partially Voiced</th>
<th>Voiced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Read Data</strong></td>
<td>675 (42.7%)</td>
<td>691 (43.7%)</td>
<td>216 (13.6%)</td>
<td>1,582</td>
</tr>
<tr>
<td><strong>Interview Data</strong></td>
<td>447 (32.2%)</td>
<td>568 (41.0%)</td>
<td>372 (26.8%)</td>
<td>1,387</td>
</tr>
</tbody>
</table>

3.3.3 Read data – continuous analysis

Now that the overall patterns have been examined, the results of each of the statistical analyses will be discussed in turn. First, I will focus on the results of the read portion of the data, which is comprised of 1,582 tokens of intervocalic /s/. As a reminder, in this task, each participant was instructed to read sentences aloud from PowerPoint slides. The continuous analysis of this dataset will be examined first, followed by the categorical analysis. The best model for the continuous analysis of the read data with percent voicing as the dependent variable, using inflated beta regression, includes the following factors as main effects: gender, speech rate, stress, preceding vowel, and following vowel. In addition to these main effects, there is a significant random effect of speaker. The discussion of interspeaker variation will be saved for section 3.3.9. In the read continuous analysis, position within the word and age are not selected as significant independent predictors.

The results of this inflated beta model are seen in Table 4. As mentioned in section 3.3.1, inflated beta regression fits four parameters to the data. The estimates seen in Table 4 are only for the main effects of the principal parameter ($\mu$). The results of the other three parameters will be discussed later. In the first column of Table 4, positive estimates indicate higher percent voicing, while negative estimates indicate lower percent...
voicing. The second and third columns include the standard error (SE) and t-values, respectively. Finally, the last column shows the relative p-value. For categorical predictors, the reference level is listed in parentheses after the name of the predictor.

For social factors, Table 4 demonstrates that gender is a significant predictor of percent voicing in the read data. Male participants have significantly higher percent voicing than female participants. Four linguistic factors are found significant for the continuous read analysis: speech rate, stress, preceding vowel, and following vowel. There is a significant, direct relationship between speech rate and voicing: as speech rate increases, percent voicing increases. This relationship is further evidenced by the scatterplot in Figure 9, in which a correlation line is fit between percent voicing and speech rate. While the direction of this correlation is clear, the scatterplot also demonstrates that those tokens that are fully voiced (1.0) occur at most speech rates, although the density of these tokens is greatest at a speech rate of about 6 syllables per second. This will be discussed in more detail in the next section with the results of the categorical read analysis.
<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>t value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-1.83</td>
<td>0.09</td>
<td>-21.09</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Gender (reference level is Female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.48</td>
<td>0.03</td>
<td>14.58</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Speech Rate</td>
<td>0.10</td>
<td>0.01</td>
<td>7.70</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Stress (reference level is Unstressed-Unstressed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stressed-Unstressed</td>
<td>-0.19</td>
<td>0.05</td>
<td>-3.82</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Unstressed-Stressed</td>
<td>-0.48</td>
<td>0.04</td>
<td>-12.20</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Stressed-Stressed</td>
<td>-0.53</td>
<td>0.10</td>
<td>-5.41</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Preceding (reference level is [a])</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[e]</td>
<td>0.08</td>
<td>0.05</td>
<td>1.85</td>
<td>0.07</td>
</tr>
<tr>
<td>[o]</td>
<td>0.17</td>
<td>0.05</td>
<td>3.54</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>[i]</td>
<td>0.18</td>
<td>0.06</td>
<td>2.74</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Following (reference level is [a])</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[i]</td>
<td>-0.05</td>
<td>0.05</td>
<td>-1.03</td>
<td>0.30</td>
</tr>
<tr>
<td>[o]</td>
<td>0.04</td>
<td>0.04</td>
<td>1.00</td>
<td>0.32</td>
</tr>
<tr>
<td>[e]</td>
<td>0.10</td>
<td>0.05</td>
<td>2.13</td>
<td>&lt;0.04</td>
</tr>
<tr>
<td>[u]</td>
<td>0.34</td>
<td>0.12</td>
<td>2.91</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

AIC: -1354.55

Table 4: Results for read data, continuous analysis (N=1582)
Table 4 demonstrates that percent voicing is significantly higher when the stress pattern is unstressed-unstressed in comparison to the other three stress patterns. Releveling this factor shows that there is a three-way distinction for stress: stressed-stressed, unstressed-stress < stressed-unstressed < unstressed-unstressed. That is, while the lowest percent voicing is found in stressed-stressed tokens, there is not a significant difference between these and unstressed-stressed tokens. The next highest percent voicing is found in stressed-unstressed tokens, which is significantly different than...
unstressed-unstressed. Thus, it is not just the stress of the preceding or following vowel that influences percent voicing in the read data, but rather the combination of the two.

In comparison to the effects just described, the effect of preceding and following vowel on voicing category is much less robust, as evidenced by the fact that the model only gets marginally worse when either factor is taken out. For preceding vowel (of which there are no cases of [u]), percent voicing increases if the /s/ is preceded by [o] or [i] in comparison to [a]. Releveling shows that the only significant vowel comparisons are [a]-[o] and [a]-[i]. For following vowel, percent voicing increases if the /s/ is followed by [e, u] as compared to [a], with [i, o] showing percent voicing not significantly different than [a]. Releveling shows many non-significant vowel comparisons: [e]-[o], [o-a], [o-i], and [i-a].

The results for the other three parameters, while not as central to the analysis as $\mu$, can illuminate further patterns in the data. For $\sigma$, which represents the factors that have a higher standard deviation, speech rate and stress are selected as significant. This means that there is more variability in percent voicing in the read data as speech rate increases, and also there is more variability in percent voicing for unstressed-unstressed tokens than for the other stress patterns. In the case of the third parameter, $\nu$, no predictors are selected as significant and thus it is merely a constant in the model. Thus, there are no conditions under which a token in the read data is more likely to be at 0% voicing than in the range between 0 and 100%. Finally, for $\tau$, which shows the conditions under which a token is more likely to be 100% voiced than in the range in between 0 and 100%, speech rate, word position, stress, and following vowel are selected as significant. As speech rate increases, and in final and initial positions (as compared to
medial), it is more likely that a token will be 100% voiced than in the range between 0 and 100%. Additionally, it is more likely that a token will be 100% voiced when surrounded by unstressed vowels, and less likely when the following vowel is [a] or [i].

3.3.4 Read data – categorical analysis

This section will focus on the categorical analysis of the read data, using ordinal logistic regression, and the comparison of the two read analyses will be discussed in the summary in the subsequent section. As a reminder, in the categorical analyses, the dependent measure is a ternary voicing category: voiceless (0-20%), partially voiced (20-90%), and fully voiced (100%). The best model for the categorical analysis of the read data with voicing category as the dependent variable includes the following factors as main effects: gender, speech rate, stress, preceding, and following vowel. Speech rate is considered as a continuous predictor, while all others are categorical predictors. In addition to these main effects, there is a significant random effect of speaker, which will be discussed later. In the read categorical analysis, age and position within the word are not selected as significant independent predictors.

The results of this ordinal logistic model are seen in Table 5. As mentioned in section 3.3.1, this type of regression calculates the odds of being in a progressively higher category, assuming there is a hierarchical order to the categories: unvoiced < partially voiced < fully voiced. The model first calculates the odds of being in the unvoiced category as compared to partially/fully voiced, and then calculates the odds of being in the partially voiced category as compared to fully voiced. Thus, the estimates in the first column of Table 5 are a combination of these two proportional odds calculations, with
positive numbers indicating increased odds of being in a significantly higher category, whether that be in the category jump from unvoiced to partially voiced, or from partially to fully voiced. On the other hand, negative estimates indicate increased odds of being in a significantly lower category, again either in the category jump from fully to partially voiced, or from partially voiced to unvoiced. The second column shows the standard error (SE), the third the z-value, and the last column shows the relative p-value. As before, the reference level for categorical predictors is listed in parentheses after the name of the independent variable.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>z value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (reference level is Female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.92</td>
<td>0.45</td>
<td>4.25</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Speech Rate</td>
<td>0.37</td>
<td>0.06</td>
<td>6.48</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Stress (reference level is Unstressed-Unstressed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stressed-Unstressed</td>
<td>-0.78</td>
<td>0.17</td>
<td>-4.64</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Stressed-Stressed</td>
<td>-1.51</td>
<td>0.36</td>
<td>-4.21</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Unstressed-Stressed</td>
<td>-1.55</td>
<td>0.14</td>
<td>-11.14</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Preceding (reference level is [a])</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[e]</td>
<td>0.19</td>
<td>0.16</td>
<td>1.22</td>
<td>0.22</td>
</tr>
<tr>
<td>[i]</td>
<td>0.30</td>
<td>0.21</td>
<td>1.45</td>
<td>0.15</td>
</tr>
<tr>
<td>[o]</td>
<td>0.49</td>
<td>0.17</td>
<td>2.94</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Following (reference level is [a])</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[i]</td>
<td>-0.07</td>
<td>0.18</td>
<td>-0.40</td>
<td>0.69</td>
</tr>
<tr>
<td>[o]</td>
<td>0.29</td>
<td>0.15</td>
<td>1.92</td>
<td>0.05</td>
</tr>
<tr>
<td>[e]</td>
<td>0.41</td>
<td>0.16</td>
<td>2.52</td>
<td>&lt;0.02</td>
</tr>
<tr>
<td>[u]</td>
<td>1.59</td>
<td>0.40</td>
<td>4.01</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Table 5: Results for read data, categorical analysis (N=1582)

For social factors, only gender is selected significant in the read categorical analysis. The positive estimate in Table 5 demonstrates that being male increases the
odds of producing a significantly higher voicing category. The specific effect of gender is further illustrated in Figure 10, which shows how many tokens are in each of the three voicing categories for female and male participants. Even though the read data is relatively well balanced for gender (N=777 for female, N=805 for male), male participants have about one-third more partially voiced tokens (N=417) than female participants (N=274). Additionally, male participants have exactly five times more fully voiced tokens in the read data (N=180) than female participants (N=36).

Figure 11: Voicing category by gender; read data (N=1582)

For linguistic factors, we see an effect of speech rate, stress, preceding vowel, and following vowel on voicing category in the read data. In the case of speech rate, as speech rate increases, the odds of being in a higher voicing category significantly
increase. Figure 12 shows a scatterplot of each token within the three voicing categories (1=Unvoiced, 2=Partially, 3=Voiced) by the speech rate of that token. It seems from this figure that the speech rate is most different for the fully voiced tokens. The distribution of both unvoiced and partially voiced tokens is densest from 4 to 8 syllables per second, while for voiced tokens, it is densest from 6 to 9 syllables per second. This figure also demonstrates there is less variation in speech rate for voiced tokens that for those in the other two voicing categories.

Figure 12: Voicing category by speech rate; read data (N=1582)
The effect for stress in Table 5 demonstrates that the odds of being in a higher voicing category decrease significantly in all stress patterns as compared to unstressed-unstressed. That is, the highest rates of voicing are found when /s/ is both preceded and followed by unstressed vowels. After releveling stress, it is clear that there is a three-way stress distinction since there is no significant difference between unstressed-stressed and stressed-stressed. The odds of being in a higher voicing category increase in the following cline: unstressed-stressed, stressed-stressed < stressed-unstressed < unstressed-unstressed. Interestingly, the raw counts and percentages in Table 6 show that most of this effect is likely being driven by the distribution of voiced tokens. While the ratio of partially voiced tokens by stress does not follow the pattern just described, the percentage of voiced tokens increases in the predicted order, with unstressed-stressed having the least and unstressed-unstressed having the most. Thus, stress in the read data seems to matter most when comparing unvoiced/partially voiced to voiced tokens.

<table>
<thead>
<tr>
<th></th>
<th>Unvoiced</th>
<th>Partially Voiced</th>
<th>Voiced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>unstressed-stressed</td>
<td>52% (327)</td>
<td>42% (260)</td>
<td>6% (40)</td>
<td>627</td>
</tr>
<tr>
<td>stressed-stressed</td>
<td>63% (42)</td>
<td>30% (20)</td>
<td>7% (5)</td>
<td>67</td>
</tr>
<tr>
<td>stressed-unstressed</td>
<td>39% (131)</td>
<td>46% (158)</td>
<td>15% (51)</td>
<td>340</td>
</tr>
<tr>
<td>unstressed-unstressed</td>
<td>32% (175)</td>
<td>36% (253)</td>
<td>22% (120)</td>
<td>548</td>
</tr>
</tbody>
</table>

Table 6: Voicing category by stress; read data (N=1582)
In comparison to the effects just described, the effect of preceding and following vowel on voicing category is much less robust, as evidenced by the fact that once again the model only gets marginally worse when either factor is taken out. Nevertheless, the particular effects seen for preceding and following vowel in the read categorical analysis mirror those for the read continuous analysis. For preceding vowel (of which there are no cases of [u]), the odds of being in a higher voicing category increase if the /s/ is preceded by [o] in comparison to [a]. Releveling confirms that the only significant difference is between [o] and [a] and no other vowel pairing. For following vowel, the odds of being in a higher voicing category increase if the /s/ is followed by [e, u] as compared to [a], with [i, o] showing probabilities not significantly different than [a]. Releveling shows many insignificant vowel comparisons: [e]-[o], [o]-[a], [o]-[i], and [i]-[a], which also calls into question the importance of this result. Furthermore, crosstabs between preceding/following vowel and other factors show that the results for vowel are likely a product of the target words chosen, and these words are not balanced when considering the cross section of position and stress.

3.3.5 Read data – summary

Overall, the continuous and categorical analyses for the read data are quite similar. The independent factors selected as significant for both analyses are: gender, speech rate, stress, preceding vowel, and following vowel. Of these, the most important effects are seen in gender, speech rate, and stress. In both analyses, significantly higher rates of voicing are correlated with male speech, faster speech, and unstressed-unstressed pattern. The factors not selected significant in either read analysis are age and word
position. As will be seen later, these factors are significant in the interview data. Therefore, it is likely that the amount of variation in the read portion is simply not great enough in order to see the effect of age and word position. This is supported by the boxplot that was seen in Figure 9, which shows a much larger range of voicing on average for the interview data. In both read analyses, there is a significant random effect of speaker. Interspeaker variation for the read data will be discussed in section 3.3.9.

As mentioned previously, interactions between independent variables were tested where relevant and no significant interactions were found. To take a closer look at potential interactions, the conditional inference tree seen in Figure 13 was generated using percent voicing as the dependent measure.\textsuperscript{11} Tagliamonte & Baayen (2012) allege that since regression analysis and conditional inference trees provide us with different information, they should be used in tandem in sociolinguistic research. The nodes of the tree in Figure 13 show how the independent variables break down to influence percent voicing, which is seen in the gray bars for each node at the bottom of the tree. The best way to understand the tree is to read it from top to bottom. For instance, to explain the leftmost nodes, we could say that for males, in stressed-unstressed/unstressed-unstressed tokens at lower speech rates, participants under 37 years old voice more than participants over 37. Additionally, the factors that are towards the top of the tree are relatively more important since they have an influence over more nodes and thus a larger portion of the data. Comparatively, the factors that are towards the bottom of the tree can be considered less important since they only affect a small subset of the data.

\textsuperscript{11} It is important to note that conditional inference trees are based on normal linear regression and so their results are not as reliable as those of the inflated beta analyses.
The first split we see in the tree is for gender, showing the relatively high importance of this factor. Since this is the first split, we are able to compare the production of male (left half of the tree) and female participants (right half of the tree). While some of the specific effects differ for males and females, in general the same factors are important for both genders, and the direction of effects is similar and matches the results seen previously in continuous and categorical analyses. In both left and right branches, the next split is for stress, separating stress patterns with more voicing (stressed-unstressed/unstressed-unstressed) from those with less voicing (stressed-stressed/unstressed-stressed). For males, speech rate is important in all stress patterns, while for females, it is only important in stressed-stressed/unstressed-stressed tokens. Nevertheless, the effect of speech rate is the same: voicing increases in higher speech rates. For males, the threshold for speech rate is between 6 and 7 syllables per second, while a female must speak at more than 8 syllables per second for speech rate to affect her rates of voicing.

While age is not significant for the read data, the tree allows us to see the trends for different ages of participants. For both males and females, age only appears in the node for a particular stress pattern. The general trend shows more voicing for younger participants, although the split between “younger” and “older” is higher for males (37 years old) than for females (20 years old). Additionally, we can see that word position interacts with several other factors, whereby word position is only important in stressed-unstressed or unstressed-unstressed tokens, for male participants whose speech rates are higher than 6 syllables per second. This interaction would be more noteworthy if word position were part of the best model for the read data; however, it is not a significant
predictor in either read analysis. Finally, it is interesting that neither preceding nor
following vowel appears in this conditional inference tree, suggesting that neither interact
with any other independent factor.
Figure 13: Conditional inference tree for read data (N=1582)
3.3.6 Interview data – continuous analysis

Now I will turn to the results of the interview portion of the data, which is comprised of 1,387 tokens of intervocalic /s/. As a reminder, in this task, the participants were engaged in conversation with me about themselves and the city of Loja. The continuous analysis of this dataset will be examined first, followed by the categorical analysis. The best model for the continuous analysis of the interview data with percent voicing as the dependent variable, using inflated beta regression, includes the following factors as main effects: gender, age, speech rate, word position, stress, and following vowel. Again, age and speech rate are considered as continuous predictors, while all others are categorical predictors. In addition to these main effects, there is a significant random effect of speaker. In the interview continuous analysis, preceding vowel is not selected as a significant independent predictor. The results of the best inflated beta model for the interview data are seen in Table 7. As before, the estimates seen in Table 7 are only for the main effects of the principal parameter ($\mu$). In the first column of Table 7, positive estimates indicate higher percent voicing, while negative estimates indicate lower percent voicing.

Table 7 demonstrates that both gender and age are significant predictors of percent voicing in the interview data. As was seen for the read portion of the data, male participants have significantly higher percent voicing in the interview data than female participants, although the effect size is somewhat smaller in interview data. There is an inverse relationship between age and percent voicing, whereby younger participants have significantly higher percent voicing in the interview data than the older participants. There is no significant interaction between age and gender for the interview data.
<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>t value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-1.21</td>
<td>0.14</td>
<td>-8.62</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Gender (reference level is Female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.19</td>
<td>0.05</td>
<td>4.10</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age</td>
<td>-0.008</td>
<td>0.002</td>
<td>-4.65</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Speech Rate</td>
<td>0.08</td>
<td>0.01</td>
<td>6.52</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Position (reference level is Word Medial)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word Final</td>
<td>0.23</td>
<td>0.06</td>
<td>3.48</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Word Initial</td>
<td>0.26</td>
<td>0.06</td>
<td>4.07</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Stress (reference level is Unstressed-Unstressed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stressed-Unstressed</td>
<td>0.06</td>
<td>0.07</td>
<td>0.81</td>
<td>0.42</td>
</tr>
<tr>
<td>Stressed-Stressed</td>
<td>-0.16</td>
<td>0.11</td>
<td>-1.47</td>
<td>0.14</td>
</tr>
<tr>
<td>Unstressed-Stressed</td>
<td>-0.20</td>
<td>0.06</td>
<td>-3.18</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Following (reference level is [a])</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[e]</td>
<td>-0.02</td>
<td>0.07</td>
<td>-0.27</td>
<td>0.78</td>
</tr>
<tr>
<td>[o]</td>
<td>-0.06</td>
<td>0.09</td>
<td>-0.70</td>
<td>0.48</td>
</tr>
<tr>
<td>[i]</td>
<td>-0.14</td>
<td>0.07</td>
<td>-2.06</td>
<td>&lt;0.04</td>
</tr>
<tr>
<td>[u]</td>
<td>-0.33</td>
<td>0.10</td>
<td>-3.16</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

$AIC = 836.03$

Table 7: Results for interview data, continuous analysis (N=1387)

The effect of speech rate in the interview data is also similar to what was seen previously. As evidenced in Table 7, and further corroborated by the scatterplot in Figure 14, as speech rate increases in the interview data, percent voicing increases. Figure 14 is quite similar to the scatterplot seen previously in Figure 9, however we see that on average speech rate is higher in the interview data and varies more. Again, fully voiced (1.0) tokens can be found along the whole range of speech rate.
For position within the word, the estimates in Table 7 indicate that there is significantly higher percent voicing in final and initial positions, as compared to medial position. While the highest percent voicing is found in initial position, there is no significant difference between initial and final position. In regards to stress in the interview data, there is significantly higher percent voicing in unstressed-unstressed than in unstressed-stressed; however, the comparison between unstressed-unstressed and the other two stress patterns is not significant. In fact, stressed-stressed is not significantly different than any other stress pattern, which may be due to the small number of tokens in this pattern in the interview data (N=73). On the other hand, releveling shows
significantly higher percent voicing in stressed-unstressed than in unstressed-stressed. As for the read data, the stress of both preceding and following vowel seems to matter here.

In the case of the interview data, the results for following vowel are more informative because there is more variation in the vowel patterns in this portion of the data since the tokens were not predetermined as with the read data. Interestingly, the significant patterns that arise with following vowel for the interview data fall into natural classes. Table 7 shows significantly lower percent voicing for high vowels ([i, u]) than for non-high vowels ([a, e, o]). The exact ordering of following vowel from lowest to highest percent voicing is: [u] < [i] < [o] < [e] < [a]. However, there is no significant difference between [u] and [i], or between [o], [e], and [a]. Curiously, there is also no significant difference between [i] and [o], showing that these two vowels are on the periphery of their respective vowel classes in terms of voicing. Thus, overall non-high vowels correlate with higher percent voicing and high vowels correlate with lower percent voicing.

There are some interesting patterns to be seen upon examining the other three parameters. For sigma, the parameter responsible for standard deviation, word position and stress are selected as significant. As compared to medial position, in the interview data there is the same amount of variability in final position, and more variability in initial position. For stress, the least amount of variability is found in stressed-unstressed and unstressed-stressed tokens, while those in stressed-stressed and unstressed-unstressed exhibit higher variability. The best model for nu includes the factors age, speech rate, and word position. This shows that a token is more likely to fall at 0% voiced than in between 0 and 100% for older speakers, at lower speech rates, and in initial position. Finally, the
fourth parameter \textit{tau}, demonstrates very similar effects as the main parameter. A token is most likely to fall at 100\% voiced than in between 0 and 100\% at higher speech rates, in final position, when surrounded by unstressed vowels, and when the following vowel is \{a, e, o\}.

\section*{3.3.7 Interview data – categorical analysis}

In this section, I will examine the categorical analysis of the interview data, using ordinal logistic regression. The best model for the categorical analysis of the read data with voicing category as the dependent variable includes the following factors as main effects: gender, age, speech rate, word position, stress, and following vowel. In addition to these main effects, there is a significant random effect of speaker. In the interview categorical analysis, preceding vowel is not selected as a significant independent predictor. The results of this ordinal logistic model are seen in Table 8. As before, positive numbers indicate increased odds of being in a significantly higher category, while negative estimates indicate increased odds of being in a significantly lower category.

Gender and age are both significant in the interview categorical analysis and the expected effects are seen. Male participants have increased odds of being in a higher voicing category, and as age increases, the odds of being in a higher voicing category decrease. The specific effect of these two factors will be discussed further in section 3.3.8. The effect of speech rate in the interview categorical analysis is also in the same direction as previously: as speech rate increases, the odds of being in a higher voicing
category increase. Figure 15, the distribution of voicing categories by speech rate, further
demonstrates that it is the voiced tokens that on average occur at higher speech rates.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>z value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender (reference level is Female)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.09</td>
<td>0.34</td>
<td>3.22</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Age</td>
<td>-0.04</td>
<td>0.01</td>
<td>-3.12</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Speech Rate</td>
<td>0.25</td>
<td>0.03</td>
<td>7.95</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Position (reference level is Word Medial)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word Initial</td>
<td>0.32</td>
<td>0.15</td>
<td>2.15</td>
<td>&lt;0.04</td>
</tr>
<tr>
<td>Word Final</td>
<td>0.52</td>
<td>0.16</td>
<td>3.34</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Stress (reference level is Unstressed-Unstressed)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stressed-Unstressed</td>
<td>-0.09</td>
<td>0.16</td>
<td>-0.56</td>
<td>0.57</td>
</tr>
<tr>
<td>Unstressed-Stressed</td>
<td>-0.49</td>
<td>0.14</td>
<td>-3.43</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Stressed-Stressed</td>
<td>-0.50</td>
<td>0.27</td>
<td>-1.87</td>
<td>0.06</td>
</tr>
<tr>
<td><strong>Following (reference level is [a])</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[o]</td>
<td>0.12</td>
<td>0.21</td>
<td>0.58</td>
<td>0.57</td>
</tr>
<tr>
<td>[e]</td>
<td>-0.08</td>
<td>0.17</td>
<td>-0.46</td>
<td>0.64</td>
</tr>
<tr>
<td>[i]</td>
<td>-0.46</td>
<td>0.16</td>
<td>-2.83</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>[u]</td>
<td>-0.60</td>
<td>0.24</td>
<td>-2.53</td>
<td>&lt;0.02</td>
</tr>
</tbody>
</table>

Table 8: Results for interview data, categorical analysis (N=1387)
For position within the word, Table 8 shows that the odds of being in a higher voicing category increase in initial and final position, when compared to medial position. The odds are highest in final position, although there is no significant difference between final and initial positions. Figure 15 shows the relative distributions by word position for the interview data. Here it becomes clear that the effect of word position is being driven principally by the fully voiced tokens. While the number of partially voiced tokens is very similar between medial (194) and initial (200) positions, and actually drops for final position (174), the number of fully voiced tokens steadily increases from medial (112), to initial (125), to final (135) position.
The effects in the interview categorical analysis for stress and following vowel are identical to what was seen for the interview continuous analysis. The results in Table 8 show that the odds of being in a higher voicing category increase in unstressed-unstressed and stressed-unstressed tokens, which are not significantly different from each other, in comparison to lower odds with unstressed-stressed tokens. As before, there is no significant difference between stressed-stressed and any other stress pattern. The relative token distributions by category and stress in Table 9 make it clear that the voiced tokens are driving this effect. While the proportion of partially voiced tokens is similar in all stress patterns, the proportion of voiced tokens in stressed-unstressed and unstressed-unstressed tokens is notably higher. The distributions in Table 9 suggest that in the
interview data, the effect of stress may be mostly due to following stress, given that stressed-stressed and unstressed-stressed behave similarly, as do stressed-unstressed and unstressed-unstressed. However, this is a tentative assumption due the small number of tokens in the stressed-stressed condition, and the fact that there are no stressed-stressed tokens in medial position. Thus, the rates of voicing in these tokens may actually be artificially high given that they all come from final or initial position, which were shown to exhibit higher voicing than medial position.

<table>
<thead>
<tr>
<th></th>
<th>Unvoiced</th>
<th>Partially Voiced</th>
<th>Voiced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>stressed-stressed</td>
<td>41% (30)</td>
<td>40% (29)</td>
<td>19% (14)</td>
<td>73</td>
</tr>
<tr>
<td>unstressed-stressed</td>
<td>38% (176)</td>
<td>42% (190)</td>
<td>20% (89)</td>
<td>455</td>
</tr>
<tr>
<td>stressed-unstressed</td>
<td>27% (96)</td>
<td>43% (151)</td>
<td>30% (107)</td>
<td>354</td>
</tr>
<tr>
<td>unstressed-unstressed</td>
<td>29% (145)</td>
<td>39% (198)</td>
<td>32% (162)</td>
<td>505</td>
</tr>
</tbody>
</table>

Table 9: Voicing category by stress; interview data (N=1387)

The model results demonstrate that the odds of being in a higher category increase for following [a, e, o], while they are significantly lower for following [i, u], again following the non-high versus high vowels divide. The particular order of the vowels from lowest to highest voicing is a bit different than before: [u] < [i] < [e] < [a] < [o]. Nevertheless, the overall pattern is the same as with the interview continuous analysis, and in this case there is a significant difference between each non-high/high vowel pairing. Figure 17 illustrates this overall pattern from least voicing (after [u]) to most voicing (after [o]). In the case of [u] and [i], the voiced tokens represent less than one-
fourth of the total tokens, while the voiced tokens are closer to one-third of all tokens for [e, a, o].

Figure 17: Number of tokens in each voicing category by following vowel; interview data (N=1387)

3.3.8 Interview data – summary

The results of the continuous and categorical analyses of the interview data exhibit the same significant predictors: gender, age, speech rate, word position, stress, and following vowel. Significantly higher voicing is found for male participants and for younger participants. For speech rate, percent voicing or voicing category increases as speech rate increases. In the case of word position, the continuous analysis indicates the highest relative percent voicing in initial position, while the categorical analysis shows
that the greatest number of fully voiced tokens occurs in final position.\textsuperscript{12} In comparison to final position, a large number of tokens in initial position are partially voiced, which makes the estimate higher for initial position in the continuous analysis. For stress, in both analyses there is significantly more voicing when /s/ is followed by an unstressed vowel as compared to a stressed vowel. Both analyses demonstrate that voicing increases when /s/ is followed by a non-high vowel as compared to a high vowel. There is a significant random effect of speaker in both interview analyses. Interspeaker variation in the interview data will be discussed in the section 3.3.9. Finally, preceding vowel is not selected as significant for either interview analysis.

Although no significant interactions were found in the interview analyses, the conditional inference tree in Figure 18 was generated in order to look more closely at the intersection of factors such as age and gender. Since the dependent measure for this tree is percent voicing, the gray bar for each node shows the relative percent voicing for that combination of independent variables. Whereas in the read data the first split was for gender, in this tree for the interview data the first split is for speech rate, separating tokens that fall above or below 7.7 syllables per second.

Gender is the second split in the tree and appears for both left and right branches of speech rate. Age is the third split on both sides and for both males and females, except in the case of males at higher speech rates. The relative break of the age continuum, however, is different for males and females. While males are divided into older and younger than 60 years old, the breakoff point for females is 20 years old, showing that

\textsuperscript{12} This is corroborated by the results for \textit{tau} in the continuous analysis, which indicate that a token is more likely to fall at 100\% voicing if it is in final position as compared to initial position, although the difference between the two positions is not significant.
while young and middle-aged men voice at similar rates, younger females have a
different voicing pattern than middle-aged and older females. In both cases, there is
higher voicing in tokens from younger participants than older participants. For younger
females at relatively slower speech rates, there is another split for speech rate, illustrating
that young females are more likely to voice when speaking faster than roughly 5.8
syllables per second. This suggests that the speech rate threshold for voicing to kick in is
actually lower for younger females than for older females. That is, to voice at rates
comparable to younger females, an older female has to speak even faster.

The fact that stress and following vowel are far down on the tree and are only
relevant for five nodes of the tree confirms that these two factors are less important than
those already discussed. In the case of stress, for males under 60 years old, at relatively
slower speech rates, voicing is higher in stressed-stressed, stressed-unstressed, and
unstressed-unstressed tokens, but only if the speech rate is higher than 5.2 syllables per
second. On the other hand, following vowel is only relevant for females over 20 years old
at speech rates less than 7.7 syllables per second. For this portion of the data, voicing is
higher when /s/ is followed by [a, e, o] (non-high vowels) as compared to [i, u] (high
vowels).
Figure 18: Conditional inference tree for read data (N=1582)
3.3.9 Interspeaker variation

In all of the best models for both datasets, there is a significant random effect of speaker, indicating that there is a considerable amount of interspeaker variation, as is to be expected. Nevertheless, the fact that in some cases age and gender still come out significant shows that there are broader voicing patterns that go above and beyond the random variation of speaker. As Drager & Hay (2012) explain, the random intercept for each speaker indicates how much that individual diverges from the predicted trends of the statistical model. These authors advocate utilizing the information provided by random intercepts instead of just including speaker as a random effect because of the independence assumption, as sociolinguists tend to do. Looking at random intercepts for each speaker can allow me to see if any speakers are voicing more or less than their age and gender might predict. Examining interspeaker variation in more detail is also important given the results of Strycharczuk et al. (2013), who find that participants pattern in two groups when the effect of speech rate by speaker is considered.

Figure 19 shows the speaker effect for the read data. This figure is based on the categorical analysis of the read data. The random intercept for each individual speaker is plotted along the scale of -3 to 2, where zero is the expected voicing given a speaker’s age and gender. Thus, this graph allows us to see which participants diverge most from what their age and gender would predict: speakers who fall to the left side of the graph (negative numbers) voice less than predicted, while speakers who fall to the right side of the graph (positive numbers) voice more than predicted. Those speakers who fall in the middle and are very close to the zero line voice more or less as their age and gender would predict. The group of speakers that voice more than predicted in the read data
(with a random intercept of more than 1) includes four female and two male participants: F20, M1, F19, F5, M12, and F3.

Figure 20 shows the speaker effect for the interview data. This figure is based on the categorical analysis of the interview data. The random intercept for each individual speaker is plotted along the scale of -2 to 2. Again, positive numbers indicate that the speaker voices more than predicted and negative numbers indicate that the speaker voices less than predicted. The group of speakers that voice more than predicted in the interview data includes two female and three male participants (in order of random intercept): M5, F13, F3, M4, and M2. Interestingly, there is only one participant who has a relatively high random intercept in both datasets (F3). Besides this participant, those speakers who diverge from the predicted pattern and voice more than predicted are different between the read and interview data, showing that individual behavior in read and interview tasks is not necessary the same.

Individual speakers are likely to differ considerably in terms of speech rate, and so looking at speech rate by speaker can help to shed light on why certain speakers voice more than their age and gender would predict. A scatterplot of percent voicing by speech rate for each speaker, with read and interview data combined, is seen in Figure 21. For most of the speakers who diverge from expected patterns, their respective scatterplot exhibits a steep slope for the correlation line between voicing and speech rate. This means that for these speakers, for each increment in speech rate, the percent voicing increases more rapidly than for other speakers with less steep slopes. That is, the effect of speech rate on voicing is greater for these speakers. This could in part explain why these participants diverge from expected patterns. It is also important to note that only one
speaker (M15) exhibits an almost flat correlation line between percent voicing and speech rate and all other speakers in this sample exhibit a positive correlation.\textsuperscript{13} Thus, the result that voicing increases as speech rate increases truly is generalizable for all speakers.

\textsuperscript{13} The reason that M15 exhibits this pattern is likely because he did not elaborate much in his answers and spoke in a monotone voice, which does not allow for much variation in speech rate.
Figure 19: Speaker effect in read data based on categorical model (0 = expected voicing given age and gender)
Figure 20: Speaker effect in interview data based on categorical model (0 = expected voicing given age and gender)
Figure 21: Speech rate (syllables per second) for all data by percent voicing, by speaker, with correlation line fit
3.4 Summary of all production results

In Table 11, we can see a summary of the production results for each of the types of data and both continuous and categorical analyses. In this table, the box is checked for an independent factor if it was found to be significant in that particular dataset and analysis. The three factors that are most consistent overall are gender, speech rate, and stress. In all analyses, males voice significantly more and there is a direct relationship between speech rate and voicing, so that as speech rate increases, voicing increases. For stress, there is a three-way distinction in the read data and only a two-way distinction in the interview data. In the case of the read data, voicing increases along this scale: stressed-stressed/unstressed-stressed < stressed-unstressed < unstressed-unstressed. On the other hand, stressed-stressed is not significantly different than any other stress pattern in the interview data, resulting in the following scale: unstressed-stressed < stressed-unstressed/unstressed-unstressed. As mentioned in section 3.3.7, the behavior of stressed-stressed tokens may be a bit unusual given there are so few of them and they all come from final or initial position. Given that word position is also significant in the interview data, looking at the cross tabulations of word position and stress in Table 10 can help to further illuminate these patterns. As expected, the distributions by word position and stress are not even; however, with the exception of stressed-stressed in medial position, there are no cells with a very small number (less than 25) tokens. We do see that there are disproportionately more unstressed-unstressed tokens in final and initial position, which may slightly augment the overall rates of voicing found for unstressed-unstressed. There are also more medial tokens for unstressed-stressed and stressed-unstressed than in initial or final position. Nevertheless, the interaction between word position and stress is not
significant in the interview data, and thus the overall patterns already discussed are upheld. Overall, combining read and interview analyses, we can see that there is more voicing when /s/ is followed by an unstressed vowel.

<table>
<thead>
<tr>
<th></th>
<th>stressed-stressed</th>
<th>unstressed-stressed</th>
<th>stressed-unstressed</th>
<th>unstressed-unstressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final</td>
<td>28</td>
<td>51</td>
<td>122</td>
<td>219</td>
</tr>
<tr>
<td>Initial</td>
<td>45</td>
<td>172</td>
<td>27</td>
<td>229</td>
</tr>
<tr>
<td>Medial</td>
<td>0</td>
<td>232</td>
<td>205</td>
<td>57</td>
</tr>
</tbody>
</table>

Table 10: Voicing category by stress; interview data (N=1387)

Following vowel is also found to be significant in all analyses; however, the exact results differ by type of data. It was determined that the results for following vowel in the read data are likely due to the structure of the data itself, while in the interview data the results for following vowel follow natural classes, with more voicing before non-high vowels. Age and word position were only found to be significant in the interview data. While not significant factors in the read data, the trends are the same, whereby there is more voicing by younger participants and in final position. This suggests that there is simply not enough variation in the read data for these trends to be significant, while in the interview data the high amount of variability allows the effects of age and word position to be seen. Finally, preceding vowel is only significant for the read data, but as with following vowel, the results are most likely due to the structure of the target words. The reasoning behind the effects of these independent factors will be explored in Chapter 5.
Table 11: Summary of significant predictors in production results

<table>
<thead>
<tr>
<th></th>
<th>Read Data</th>
<th>Interview Data</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Continuous</td>
<td>Categorical</td>
<td>Categorical</td>
<td>Categorical</td>
</tr>
<tr>
<td>Gender</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Speech Rate</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Word Position</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Stress</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Preceding Vowel</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Following Vowel</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
CHAPTER 4: ALLOPHONIC PERCEPTION

4.1 Introduction

As was seen in Chapter 2, while the number of production studies of intervocalic /s/ voicing is growing, we still do not have a clear idea of the perceptual reality of [s] and [z] in Spanish. Although trained linguists can often easily note the production of [s] or [z], it is unclear whether naïve native speakers hear the difference between these sounds or not given the lack of phonemic distinction between these phones. In this chapter, I will discuss an online experiment that I conducted to test this perception. In section 4.2, I will detail the methodology for this experiment, including stimuli, participants, and experimental design. In section 4.3, I will detail the results of this experiment, and conclude in section 4.4 with a summary of these results.

4.2 Methodology

4.2.1 Experimental design

As mentioned previously, the main challenge in testing allophonic perception is to find a way to do so naturally without relying on orthography. One option would be to choose two arbitrary symbols and train participants to respond X when they hear [s] and Y when they hear [z], as Widdison (1995, 1996, 1997) did. However, this would probably bias the participants to hear differences. I also considered using the “minimal
pairs” that Robinson (1979) mentions. For example, I could play an audio stimulus of [aziðo] and ask participants to circle ‘ha sido’ or ‘has ido’. I do not believe this methodology would work with Lojano participants, since they have been shown to voice in all positions and thus could pronounce either ‘ha sido’ or ‘has ido’ as [aziðo].

Based on the literature review in section 2.5.1, I believe the methodology used by Boomershine et al. (2008) and Johnson & Babel (2010) adapts best to the perception of intervocalic /s/ voicing. Both of these studies, among others that they cite, have shown that similarity rating and discrimination tasks work well for allophonic perception, which is what we have in the case of [s] and [z] in Spanish.

Following Boomershine et al. (2008) and Johnson & Babel (2010), I designed an online experiment with two parts: a similarity rating task and an AX discrimination task. The stimuli consist of intervocalic [s] and [z] in words where [s]/[z] is in word-initial, medial or final position. By including all three positions, I cover all the locations where HES speakers could potentially hear this feature. For medial position, the stimuli are tokens of the word *asa* ‘he/she grills’. For initial and medial position, the stimuli are two word sequences: *la saca* ‘he/she takes it out’ and *las ata* ‘he/she ties them’. Accordingly, the full set of stimuli is: [asa], [aza], [lasaka], [lazaka], [lasata], and [lazata].

Boomershine et al. (2008) do not discuss whether it is important that the stimuli be real or pseudo words but they include both in their experiment: [aða] – *hada* vs. [iði] – nonce. Using real words is the only way to ensure (without having any text on screen corresponding to the stimuli) that the participants hear the stimuli as two words. This was also reinforced by the fact that participants were told they would hear sequences of two words in the case of *la saca* and *las ata* stimuli. It is assumed that there will be no
ambiguity as to which word the /s/ pertains to in the case of the two word-sequences *la saca and las ata since the sequences *las aca and *la sata are unattested in Spanish.

Various productions of these stimuli were recorded by a phonetically trained male Lojano Spanish speaker. I chose the experimental stimuli by examining the waveform and spectrogram for each recording. I chose ‘unvoiced’ tokens that had between 0-15% voicing, and ‘voiced’ tokens with very close to 100% voicing. As a reminder, the cutoff for unvoiced tokens in the production experiment was 20% voiced; see section 3.3.1 for more details on how this was determined. The advantage of using naturally recorded stimuli in this way is that no manipulation is necessary, however special measures must be taken to ensure the stimuli pairs are as identical in every way except for the consonant. Following Boomershine et al. (2008: 6), I carefully controlled for the vowels’ duration and intonation between tokens, and equated the peak amplitude for each of the tokens in order to control for amplitude across tokens. Before piloting, the stimuli were checked by two professional phoneticians to ensure that [s] versus [z] was the only audible difference between the two tokens of a pair. Acoustically speaking, the tokens of [s] and [z] differed in both duration and voicing as expected, since the fricatives in the original recordings were not manipulated. All experimental materials were piloted with three native non-linguists, three native linguists, and three non-native linguists, to make certain that the recording of the stimuli was successful and that the tasks were straightforward to complete.

14 The native speakers who piloted the stimuli were native speakers of Spanish, but not specifically of Lojano Spanish.
The similarity rating and discrimination tasks are presented to the participants via the online platform, SurveyGizmo (Vanek and McDaniel 2006). All instructions are given in written form via the online experiment. Before completing the experimental blocks, the participants were asked for their consent to participate in the study and completed a short background questionnaire, including gender, age, residential history, occupation, languages studied, and whether the participant was using headphones or speakers to complete the survey. The experimental portion of the survey is comprised of six blocks, consisting of three similarity rating tasks and three discrimination tasks, one for each word position. In the similarity rating tasks (Blocks #1-3), participants are instructed that they will hear a pair of words/pair of two word sequences and be asked to rate them on a scale of 1-6, where 1 is muy similar ‘very similar’ and 6 is muy diferente ‘very different’, as seen in the screenshot in Figure 22. At the beginning of each block, participants are given the chance to practice with six to eight VCV sequences before completing the experimental trials. The practice materials include stimuli where the answer is very clearly 1 or 6 to make sure that the participants understand how to use the scale. For example, one of the training stimuli for Block 1 consisted of [ama]-[ata], which would quite easily be judged as 6.
In the first of the three experimental similarity rating tasks, the participants hear pairs that differ only in the consonant and that were always physically different tokens even if they contain the same sounds. This ensures that participants are not responding to some other characteristic of the recordings when they respond that the two stimuli sound the same. For instance, a pair could consist of [aza] and [asa], or [aza] and then a different token of [aza]. Each pair is separated by one second of silence. The other two similarity rating tasks are identical in design but use the [lasaka]/[lazaka] (#2) and [lasata]/[lazata] pairs (#3).

In the discrimination tasks (Blocks 4-6), the participants hear the same pairs of stimuli they heard in Blocks 1-3 and are instructed to determine whether the pair they
heard were identical sounds/sequences or different, as seen in Figure 23. “Igual” ‘same’ would be a correct response if for example two tokens of [asa] were played, while “diferente” ‘different’ would be a correct response if stimuli differing in voicing were played ([asa]-[aza]). Block 4 includes the word-medial stimuli ([asa]/[aza]), Block 5 includes the word-initial stimuli ([la saka]/[la zaka]), and Block 6 includes the word-final stimuli ([la sata]/[la zata]). In all six blocks, the audio pair for each page plays automatically and the participants are not able to play the audio again. This encourages participants to answer each question quickly and rely on their spontaneous impressions. The participant must answer the question on the page before moving on to the next page.

Figure 23: Screenshot of the online survey, discrimination task in Blocks 4-6
In each of the tasks, the experimental stimuli are randomly presented with fillers that are similar but contain the sounds [d], [ð], [r], [n], and [ŋ]. There were always twice as many filler stimuli as target stimuli in hopes that the participants would not realize which were the sounds of interest. For Blocks 1 and 4 (word-medial position), the fillers are [ada], [aða], and [ara]. In Blocks 2 and 5 (word-initial position), the fillers are [ladaβa] and [laðaβa]. Finally, Blocks 3 and 6 (word-final position), the fillers are [tanaltō], [taŋaltō], [konoro], and [koŋoro]. A complete list of the stimuli for each block can be found in Appendix C. Blocks 1 and 4 included eight training items, twenty-two fillers, and twelve target stimuli. Blocks 2, 3, 5, and 6 included six training items, twenty fillers, and twelve target stimuli. The same randomized order of stimuli was used for all participants.

After completing the experimental blocks, the participants were asked to leave any comments they may have about what sort of difference they heard in the pairs and whether or not the speaker they heard could be someone from their community. Finally, participants have the option of leaving their email if they would like to be informed about the results of the study and they are thanked for their time.

---

15 Even if participants were able to correctly identify /s/ as the sound of interest, it is unlikely that they could alter their perception since they had to respond to each stimulus instantaneously.

16 The sequences [ada] and [aða] are both productions of hada ‘fairy’, while [ara] is that of ara ‘he/she plows’. [tanaltō] and [taŋaltō] correspond to tan alto ‘so tall’, while [konoro] and [koŋoro] correspond to con oro ‘with gold’.
### 4.2.2 Participants

Participants for the perception experiment were recruited via email and potential participants were sent a link that led them directly to the survey. In order to participate, the participants had to be at least 18 years old, be a native speaker of Spanish, have been born in Ecuador, and have not lived outside of Ecuador for more than 3 years. This was to ensure that the participants would have sufficient and recent contact with Highland Ecuadorian Spanish. Only those participants who have lived the majority of their life in Loja are used for the present analysis.

A total of 21 participants completed the entire survey (Blocks 1-6). Another 3 participants completed Blocks 1-3 and have been included in the results for the similarity rating tasks. The age of the participants ranges from 18 to 60 years old. There are eleven male participants and thirteen female participants. Their knowledge of English ranges from none at all to 20 years of studying English. The vast majority of participants have not studied any other language other than English. One participant studied French for one year and another participant studied German for three years. Fourteen participants used speakers to listen to the stimuli and the other ten participants used headphones. The results of the training stimuli were checked to ensure participants properly understood the task. No participants had to be excluded from the analysis since all performed as expected on the training materials, as will be seen in section 4.3.2.

### 4.3 Results

Responses were collected via SurveyGizmo and exported to Excel for analysis. Each token was coded for the block and type of token (filler, target, etc.).
discussion that follows, pairs that included two audios of the same production ([asa]-[asa]) will be referred to as “identity pairs”, while pairs that included two audios of different productions ([asa]-[aza]) will be referred to as “difference pairs”. Unlike Boomershine et al. (2008) and Johnson & Babel (2010), my survey did not record reaction times, and so I analyze the results in the traditional way described in section 2.5.1, which involves directly comparing participants’ responses to difference and identity pairs.

4.3.1 Data analysis

The data was separated by block and imported into R (R Core Development Team 2014) for analysis. For Blocks 1-3, a Welch Two Sample t-test was performed on the responses to target stimuli to determine whether there was a significant difference between mean response for identity and difference pairs. Given that the data in Blocks 1-3 is not normally distributed, the results of these t-tests were further corroborated by a permutation test for each block. In every block, the results of the permutation test matched those of the t-test, and thus only the p-values of the t-tests are reported here. In addition, a mixed linear model was fit to Blocks 1-3 combined with response as the dependent measure, block as an independent predictor, and participant as a random effect. This was done to examine whether Blocks 1-3 are significantly different from each other and to investigate the variability across participants. For Blocks 4-6, a two-tailed Pearson’s chi-squared test with Yates’s continuity correction was performed to test whether participants responded ‘diferente’ significantly more to difference pairs than to identity pairs. A mixed logistic model was fit to Blocks 4-6 combined to investigate the
effect of block and participant on the binary response variable. As mentioned previously, block in both parts of the experiment corresponds to word position, and thus an effect of block illustrates differences between the word positions: medial (Block 1/4), initial (Block 2/5), and final (Block 3/6).

4.3.2 Results of training blocks

Before looking at the experimental results, it is useful to consider the results of the training session in each block to ensure that participants understood the task at hand. As mentioned previously, among the training materials, pairs that are very clearly different were included in order to encourage participants to use the whole scale from 1 to 6. These clearly different pairs included a phonemic contrast: [ama]-[ana], [ama]-[ata], [ara]-[ala], [asa]-[ata] (Block 1/4); [lasaka]-[laβaka], [lalata]-[laβata] (Block 2/5); [marasul]-[malasul], [elexe]-[enexe] (Block 3/6). If the participants understand the task, we should expect responses close to 6 ‘muy diferente’ on Blocks 1-3, depending on how they are interpreting the scale. For Blocks 4-6, we should see mostly “different” responses to these training stimuli.

Table 12 shows the mean response for each of the training stimuli that include a clear phonemic contrast. Not listed in this table are the training stimuli that were also fillers or target stimuli in the experimental blocks, that is, the stimuli that include an allophonic pair. For the phonemic contrast training pairs in Table 12, the mean response ranges from 4.583 to 5.583. This is evidence that overall the participants understand the task because these means are relatively close to the expected response (6). Given that
these training stimuli represent the participants’ first time using the scale, it is likely that some of them did not judge these stimuli as “very different” because they were waiting to see what the range of “different” would be in order to use the whole scale appropriately. Nevertheless, the mean responses for these phonemic contrasts are quite high, which is good assurance that participants are not misunderstanding how to use the scale.¹⁷

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>ama-ana</td>
<td>4.583</td>
</tr>
<tr>
<td>mar azul-mal azul</td>
<td>4.667</td>
</tr>
<tr>
<td>el eje-en eje</td>
<td>5.083</td>
</tr>
<tr>
<td>la lata-la bata</td>
<td>5.125</td>
</tr>
<tr>
<td>la saca-la vaca</td>
<td>5.292</td>
</tr>
<tr>
<td>ara-ala</td>
<td>5.333</td>
</tr>
<tr>
<td>ama-ata</td>
<td>5.500</td>
</tr>
<tr>
<td>asa-ata</td>
<td>5.583</td>
</tr>
</tbody>
</table>

Table 12: Mean response of training stimuli, Blocks 1-3 (N=192)

The responses for training stimuli for Blocks 4-6 are displayed in Table 13.

Again, this table only includes training stimuli in which there is a phonemic contrast. As expected, the vast majority of responses is “different”. Out of 168 responses, there are

¹⁷ Interestingly, although not relevant to the present study, the training stimuli that receive the lowest responses are those in which the contrasting phonemes are of the same natural class. In the case of *ama-ana*, both [n] and [m] pertain to the class of nasals, and for *mar azul-mal azul*, [l] and [ɾ] form the natural class of liquids.
only five “same” responses. It is likely that these five responses are simply errors. Thus, participants clearly understand what they should do in the discrimination tasks in Blocks 4-6.

<table>
<thead>
<tr>
<th></th>
<th>Different</th>
<th>Same</th>
</tr>
</thead>
<tbody>
<tr>
<td>ama-ana</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>ama-ata</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>ara-ala</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>asa-ata</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>la saca-la vaca</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>la lata-la bata</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>mar azul-mal azul</td>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>el eje-en eje</td>
<td>19</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 13: Responses for training stimuli, Blocks 4-6 (N=168)

4.3.3 Similarity rating task

4.3.3.1 Block 1

In Block 1, participants heard tokens of ‘asa’: identity pairs of [asa]-[asa] and [aza]-[aza], and difference pairs of [aza]-[asa] and [asa]-[aza]. As mentioned previously, these target stimuli were mixed with filler pairs containing [ada], [aða], and [ara]. Participants saw the scale in Figure 22 and had to respond by clicking one of the numbers
on the scale and continuing to the next page. The participants only heard each audio pair once.

Table 14 shows the results for Block 1. The t-test shows that the mean response for identity pairs (1.604) is significantly lower than the mean response for difference pairs (2.073). As a reminder, the scale in this task goes from 1 ‘very similar’ to 6 ‘very different’. The overall low numbers for all stimuli in this block demonstrates that in general the pairs sounded very similar to each other, which is a clear contrast to the relatively high averages seen in Table 12 for the training stimuli. Nevertheless, there is a statistical difference between identity and difference pairs in which difference pairs are heard as more different (higher on the scale) than identity pairs.

<table>
<thead>
<tr>
<th></th>
<th>Difference pairs (N=192)</th>
<th>Identity Pairs (N=96)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>2.073</td>
<td>1.604</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>1.272</td>
<td>1.192</td>
</tr>
</tbody>
</table>

Table 14: Results for Block 1, similarity rating task of medial position (t = 3.0759, df = 201, p = 0.00239)

While the means in Table 14 show the general tendencies, it is also helpful to look at the frequency of each possible response to see the broader picture. Figure 24 illustrates the percentage of responses in Block 1 for the difference pairs (left, dark bars) and identity pairs (right, light bars). Each bar shows the proportion out of total responses for that particular response (1, 2, 3, 4, 5 or 6). For the difference pairs, most responses are concentrated in the 1-2 range, however roughly one-fourth of the responses are 3 or
higher. Thus, for some participants the difference pairs are heard as quite different, even maximally different (6 on the scale). For identity pairs in Block 1, there is much less variability in the responses, with 1 representing three-fourths of the responses to identity pairs. Participants do use the whole scale for the identity pairs; however, the percentage of responses higher than 2 for identity pairs is negligible. These two graphs also suggest that there is some variation in participants’ responses, which is greater for difference pairs than identity pairs, as is also evidenced by the standard deviations in Table 14. The difference between participants will be discussed in further detail in section 4.3.6.

![Graphs showing percentage of responses for Block 1, difference pairs (left, N=192) and identity pairs (right, N=96)](image.png)

Figure 24: Percentage of responses for Block 1, difference pairs (left, N=192) and identity pairs (right, N=96)

### 4.3.3.2 Block 2

The task for Block 2 was identical to Block 1, however participants now heard ‘la saca’ in identity pairs of [lasaka]-[lasaka] and [lazaka]-[lazaka], and difference pairs of [lazaka]-[lasaka] and [lasaka]-[lazaka], mixed with filler pairs containing [ladaβa] and
The results for Block 2 are seen in Table 15. As with Block 1, in Block 2 there is a statistically significant difference between identity and difference pairs. The responses for difference pairs are on average significantly higher than the responses for identity pairs, demonstrating that difference pairs are heard as more different. The p-value for the t-test done on Block 2 is quite small, showing this is a robust effect.

<table>
<thead>
<tr>
<th></th>
<th>Difference pairs (N=192)</th>
<th>Identity Pairs (N=96)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>2.333</td>
<td>1.573</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>1.363</td>
<td>0.971</td>
</tr>
</tbody>
</table>

Table 15: Results for Block 2, similarity rating task of initial position ($t = 5.4477$, $df = 253$, $p = 0.000000121$)

The proportion of responses seen in Figure 25 confirms the general tendencies seen for the mean response in Block 2. As before, the difference pairs are exhibited on the left (dark bars) and the identity pairs on the right (light bars). In comparison to Block 1 responses, for the difference pairs in Block 2, the highest concentration of responses is seen for 2 instead of 1, which is also evidenced by the higher mean response for difference pairs in Block 2. Nevertheless, the response for identity pairs in Block 2 is quite similar to that of Block 1, whereby roughly three-fourths of responses are at the lowest, “very similar”, value on the scale. As with Block 1, here in Block 2, there is more variability in the responses to difference pairs than identity pairs, in which there is more consistency between participants. This is confirmed by the higher standard deviation seen in Table 15 for difference pairs than for identity pairs.
Figure 25: Percentage of responses for Block 2, difference pairs (left, N=192) and identity pairs (right, N=96)

4.2.3.3 Block 3

The task for Block 3 was identical to that of Blocks 1 and 2. In Block 3, participants heard tokens of ‘las ata’ with identity pairs of [lasata]-[lasata] and [lazata]-[lazata], and difference pairs of [lazata]-[lasata] and [lasata]-[lazata], mixed with filler pairs containing [ladaβa] and [laðaβa]. Table 16 shows the results for Block 3. The results of the t-test exhibit a statistically significant difference between identity and difference pairs in Block 3. Difference pairs are heard as more different (higher on the scale) than identity pairs. As with Block 2, the $p$-value of the t-test is quite small, showing a robust effect of pair type in Block 3.
Table 16: Results for Block 3, similarity rating task of final position (t = 5.4898, df = 265, p = 0.0000000943)

<table>
<thead>
<tr>
<th></th>
<th>Difference pairs (N=190)</th>
<th>Identity Pairs (N=96)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.174</td>
<td>1.448</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.332</td>
<td>0.881</td>
</tr>
</tbody>
</table>

Figure 26 shows the percentage of responses for Block 3, again with the difference pairs in the left graph (dark bars) and the identity pairs in the right graph (light bars). As compared to the other blocks, the distribution for the difference pairs in Block 3 is most similar to that of Block 2. Again, the highest percentage of responses is seen for 2, which is followed closely by the percentage of responses for 1. The same pattern seen before is seen for the identity pairs in Block 3. Most responses are concentrated at 1 and there is a steady decline in percentage of responses as the scale increases from 1 to 6. Like the results for the other blocks, in Block 3 there is more variability for the difference pair responses and more consistency between participants for the identity pair responses.
4.2.3.4 Comparison Blocks 1-3

In Blocks 1-3, participants responded to audio pairs by selecting a corresponding number on the scale from 1 ‘very similar’ to 6 ‘very different’. In all three blocks, there is a significant difference between identity and difference pairs, in each case with significantly higher responses for difference pairs. This indicates that participants hear difference pairs as more different than identity pairs, suggesting they distinguish in some manner between [s] and [z] in word-medial (Block 1), word-initial (Block 2), and word-final position (Block 3). The difference between identity and difference pairs is most robust in word-final position as evidenced by the lowest $p$-value, which is followed closely by word-initial position. The difference is least evident in word-medial position, which has the highest $p$-value of the three blocks. Figures 24-26 also support this conclusion, showing that the most frequent response for medial position is 1, while the most frequent response for initial and final position is 2. The linear mixed model of
Blocks 1-3 combined shows that there is a significant difference between Block 1 and Block 2 ($p < 0.02$), but no significant difference between Block 1 and Block 3 or between Block 2 and Block 3. This indicates that there is a difference only between medial and initial position. There is also a significant random effect of participant, evidenced by the variance value for the random effect (0.4537). This will be discussed further in section 4.3.6.

4.3.4 Discrimination task

4.3.4.1 Block 4

In Block 4, participants heard the same stimuli as in Block 1: identity pairs of [asa]-[asa] and [aza]-[aza], and difference pairs of [aza]-[asa] and [asa]-[aza], mixed with filler pairs containing [ada], [aða], and [ara]. The screen displayed the page shown in Figure 23 and after hearing each audio pair, participants had to respond by clicking *igual* ‘same’ or *diferente* ‘different’ and continuing to the next page. The participants heard each audio pair only once.

Table 17 shows the distribution of ‘different’ and ‘same’ responses for Block 4. While only roughly ten percent of the responses to identity pairs are ‘different’, for difference pairs almost one-fourth of the responses are ‘different’. The chi-square test shows that this distribution is statistically significant. Thus, participants hear [s]-[z] in word-medial position more frequently as different than [s]-[s].
<table>
<thead>
<tr>
<th></th>
<th>Identity Pairs</th>
<th>Difference pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different</td>
<td>8 (9.5%)</td>
<td>37 (22.0%)</td>
</tr>
<tr>
<td>Same</td>
<td>76 (90.5%)</td>
<td>131 (78.0%)</td>
</tr>
<tr>
<td>Total</td>
<td>84</td>
<td>168</td>
</tr>
</tbody>
</table>

Table 17: Results for Block 4, discrimination task for medial position ($c^2 = 5.1435$, df = 1, $p = 0.02333$)

### 4.3.4.2 Block 5

The task for Block 5 was identical to that of Block 4. For Block 5, participants heard the same stimuli as in Block 2: identity pairs of [lasaka]-[lasaka] and [lazaka]-[lazaka], and difference pairs of [lazaka]-[lasaka] and [lasaka]-[lazaka], mixed with filler pairs containing [ladaβa] and [laðaβa]. The distribution of responses for Block 5 is seen in Table 18. In this block, the contrast between identity and difference pairs is even more marked. As in Block 4, identity pairs are heard as ‘different’ about ten percent of the time; however, difference pairs are heard as ‘different’ almost half of the time. As the $p$-value indicates, this distribution is highly significant; indicating that [s]-[z] in word-initial position is heard more frequently as different as compared to [s]-[s].
### Table 18: Results for Block 5, discrimination task for initial position ($\chi^2 = 34.1039$, df = 1, $p = 0.000000005225$)

<table>
<thead>
<tr>
<th></th>
<th>Identity Pairs</th>
<th>Difference pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Different</strong></td>
<td>8 (9.5%)</td>
<td>80 (47.6%)</td>
</tr>
<tr>
<td><strong>Same</strong></td>
<td>76 (90.5%)</td>
<td>88 (52.4%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>84</td>
<td>168</td>
</tr>
</tbody>
</table>

4.3.4.3 Block 6

The task for Block 6 was identical to that of Blocks 4 and 5. For Block 6, participants heard the same stimuli as in Block 3: identity pairs of [lasata]-[lasata] and [lazata]-[lazata], and difference pairs of [lazata]-[lasata] and [lasata]-[lazata], mixed with filler pairs containing [ladaβa] and [laðaβa]. Table 19 shows the distribution of responses for Block 6. In this block, identity pairs are heard as ‘different’ roughly 12% of the time. On the other hand, one-third of difference pair stimuli are heard as different. The chi-square test for this block demonstrates that this distribution is statistically significant. Thus, participants hear [s]-[z] in word-final position as different more consistently than they hear [s]-[s] as different.

### Table 19: Results for Block 6, discrimination task for final position ($\chi^2 = 12.2166$, df = 1, $p = 0.0004737$)

<table>
<thead>
<tr>
<th></th>
<th>Identity Pairs</th>
<th>Difference pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Different</strong></td>
<td>10 (11.9%)</td>
<td>56 (33.3%)</td>
</tr>
<tr>
<td><strong>Same</strong></td>
<td>74 (88.1%)</td>
<td>112 (66.7%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>84</td>
<td>168</td>
</tr>
</tbody>
</table>
4.3.4.4 Comparison Blocks 4-6

In Blocks 4-6, participants responded to audio pairs by selecting ‘different’ or ‘same’. In all three blocks, there is a significant difference between the distribution of responses to identity and difference pairs, in each case with more ‘different’ responses for difference pairs. Interestingly, the effect of word position in Blocks 4-6 is similar to what was seen for Block 1-3. The difference between identity and difference pairs is most robust in word-initial position (Block 5), followed by word-final position (Block 6), and finally word-medial position (Block 4). The logistic mixed model of Blocks 4-6 combined shows that there is a significant difference between Block 4 and Block 5 (p < 0.001), between Block 4 and Block 6 (p < 0.02), and between Block 5 and Block 6 (p < 0.01). This shows that the difference between the three word positions is significant. The variance value for the random effect of participant in this model is 0.03064, indicating that there is not a significant random effect of participant for Blocks 4-6.

4.3.5 Results of open-ended questions

As mentioned before, at the conclusion of the experimental blocks, the participants were asked to respond to three open-ended questions. A response was not required on these questions in order to submit the survey; however, the nineteen of the twenty-one participants who completed the entire survey elected to answer the open-ended questions. When reporting the responses to these questions, I will paraphrase what the participants wrote. The first question is: Cuando escuchó una diferencia en los pares, ¿cómo describiría esa diferencia? [When you heard a difference in the pairs, how would you describe this difference?]. While there were a variety of answers to this question,
eight participants mentioned some aspect of accent, tone, or velocity of speech. Since the same word in Spanish (‘acento’) can be used colloquially both for word accent (i.e. stress) and dialectal pronunciation, it is hard to know which meaning the participants were invoking, although given its co-occurrence with words like entonación ‘intonation’ and tono ‘tone’, it is likely the former. Other responses signaled ‘minimal difference’ (N=4), ‘clear difference’ (N=1), ‘difference in the consonants’ (N=3), and ‘bad pronunciation’ (N=1). No participant specifically mentioned /s/ in this commentary, although two participants mentioned hearing a difference in the /n/ and /l/ sounds. The most puzzling response comes from one participant who wrote “como el sonido de una avispa” [like the sound of a wasp]. While it is impossible to say for sure what the participant means, it is possible that the participant is referring to the buzz-like, friction noise of [z].

The second open-ended question is: ¿Le parece que el hablante que escuchó podría ser un miembro de su comunidad? ¿Por qué sí o por qué no? [Do you think the person you heard could be a member of your community? Why or why not?]. In response to this question, ten participants indicated ‘yes’, five responded ‘no’, and three said ‘maybe’. Of those who said the speaker could be a member of their community, there was a range of reasons as to why. These reasons included: ‘por su acento’ [because of his accent] (N=4), ‘por su pronunciación’ [because of his pronunciation] (N=3), ‘porque habla mal/arrastra’ [because he speaks poorly, drags his speech] (N=2), and ‘por la entonación y cantado’ [because of the intonation and sing-song quality]. It is interesting that one participant mentions the verb arrastar ‘to drag’, since this is normally used in everyday conversation to refer to the assibilation of trill /t/. Given that there were no
tokens of /r/ in the stimuli, it is probable that this participant is hinting at the pronunciation of /s/ as [z]. While only one participant mentions intonation for this question, which is one of the notable features of HES as mentioned in section 2.2, this might be due to the fact that the participants only heard sequences of one or two words in which intonational patterns are less evident. The five participants that responded ‘no’ to this question all indicated that the speaker’s pronunciation was somehow different from their own, without mentioning specific sounds. Two of the participants who responded ‘maybe’ erroneously thought there was more than one speaker in the experiment, and so indicated that some of the speakers could be members of their community and others could not. The other participant who responded ‘maybe’ to this question simply wrote that it was difficult to discern.

The final open-ended question allowed the participants to provide any other comments they may have: ¿Algún otro comentario sobre el estudio? [Any other comment about the study?]. Most participants left this question blank or filled in “no” or “none”. Three participants responded that the survey was somewhat boring, one participant commented that it seemed like there were two speakers, and another wrote “Se puede interpretar lo mismo, pero con una pequeña diferencia en el sonido.” [It can be interpreted the same, but with a small difference in the sound]. Finally, one participant suggested using full sentences to make dialectal differences more noticeable, and another participant praised the study saying, “Realmente interesante y muy bien desarrollado.” [Really interesting and very well designed].

Overall, the open-ended questions show that while the participants consistently heard a difference between [s] and [z] stimuli, they did not explicitly comment on this
difference. This could be due to at least two different things. First, given that the number of fillers was greater than the number of target stimuli, it is possible that participants genuinely did not know which sounds were of interest and thus did not “pick out” certain sounds to highlight when commenting on what they heard as different. It is also possible that the perceptual difference between [s] and [z] is under the radar for these participants. This is supported by the fact that only one participant explicitly mentioned the pronunciation of /s/ as a distinguishing feature of HES when asked during the sociolinguistic interview portion of the recordings to describe different dialects of Ecuador. If the participants, on the other hand, were more readily aware of this feature, we might expect that a few would comment on the online survey that they heard a difference in the pronunciation of las eses ‘the s’s’. Indeed this type of explicit comment has been found for the perception of other phonetic features that are more salient, such as /s/ aspiration (Walker et al. 2014), but this is not the case for intervocalic /s/ voicing with Lojano participants.  

4.3.6 Participant differences
Sections 4.2.3.4 and 4.3.4.4 presented the results of two mixed models that examine the random effect of participant in Blocks 1-3 and 4-6. For Blocks 1-3, the variance for the random effect of participant is 0.45, which indicates a significant random effect of participant. It is important to look deeper into how precisely the participants’ behavior differs on the similarity rating tasks. Figure 27 shows two graphs that illustrate

18 While there is certainly a continuum of awareness, and different types of awareness exist, it is clear that Lojanos are less explicitly aware of intervocalic /s/ voicing than many Caribbean speakers are of /s/ aspiration and deletion.
the response values for each participant in Blocks 1-3. The top graph includes the responses by participant for the difference pairs and the bottom graph is for the identity pairs. Comparing the two graphs makes it clear that there is much more variation between participants in the difference pairs than in the identity pairs. Thus, the significant random effect of participant in Blocks 1-3 is mostly due to the difference pairs. In the top graph, we can see that Participants 58, 114, 153, and 173 have a mean response of 3 or higher, which means that they hear the difference pairs on average as more different than the other participants. Participant 173 also has a relatively high mean response for identity pairs, showing that this participant perhaps is overall more attuned to subtle pronunciation differences.
Figure 27: Range of responses by participant for difference (top) and identity pairs (bottom), Blocks 1-3
For Blocks 4-6, the mixed logistic regression showed a variance for the random effect of participant of 0.03. The effect of participant for Blocks 4-6 is not therefore particularly worrisome; however, looking at individual differences can reveal whether the same participants who deviate from the overall patterns in Blocks 1-3 continue this behavior in Blocks 4-6. Figure 28 includes two graphs that demonstrate the percentage of ‘different/same’ responses by participant for Blocks 4-6. The dark bars represent the percentage of ‘different’ responses, while the lighter bars represent the percentage of ‘same’ responses. The top graph is for difference pairs and the bottom graph is for identity pairs. As a reminder, not all of the participants who completed Blocks 1-3 also completed Blocks 4-6, which is why the participant ID numbers are different between Figures 27 and 28. Again, there appears to be more variation between participants in their response to difference pairs than identity pairs. Looking at the top graph in Figure 28, Participants 58, 100, 112, 114, 178, and 183 responded “different” to difference pairs about 60% or more of the time. The behavior of Participants 58 and 114 is markedly different than the rest of the participants in all blocks, including both similarity rating and discrimination tasks, suggesting that perhaps their ears are more attuned to the difference between [s] and [z]. Both of these participants have studied English for several years, 7 and 13 years respectively, and so it is possible that the phonemic contrast of /s/ and /z/ in English is affecting their perception of these sounds in Spanish.
Figure 28: Responses by participant for difference (top) and identity pairs (bottom), Blocks 4-6
4.4 Summary of all perception results

In Blocks 1-3, the similarity rating tasks, participants rated difference pairs higher on the scale than identity pairs, showing that they hear difference pairs as more different. A series of t-tests shows that the difference in response between difference pairs and identity pairs in Blocks 1-3 is significant. The effect of pair type is most robust for word-final position, followed closely by word-initial position, and least robust in word-medial position. There is a significant difference in response between word-medial and word-initial position; however, no other position comparison reaches significance. In Blocks 4-6, the discrimination tasks, participants responded “different” to difference pairs more often than they did to identity pairs. A series of chi-square tests shows that the distribution of “same” and “different” responses in Blocks 4-6 is significantly different. The difference between identity and difference pairs is most robust for word-initial position, followed by word-final position, and least robust in word-medial position. There is a significant difference between each of the position comparisons for Blocks 4-6.

In all blocks, the difference between [s] and [z] is least readily heard in word-medial position. For word-initial and final positions, the results are different between the similarity rating and discrimination tasks. For the similarity rating tasks (Blocks 1-3), the difference between [s] and [z] is most readily heard in word-initial position, although there is no significant difference between this and word-final position. On the other hand, for the discrimination tasks (Blocks 4-6), the difference between [s] and [z] is most readily heard in word-final position, which is significantly different than all other positions. Considered together, these results suggest that the participants heard [s]-[z] most clearly in word-final position, and least clearly in word-medial position, at least for
these particular stimuli. While the random effect of participant is significant for Blocks 1-3, it is not significant for Blocks 4-6, indicating that participants exhibit more consistency in the discrimination tasks. This might be due to the fact that they only had two choices to choose from in the discrimination tasks, so there was less opportunity for individual differences. The allophonic perception results presented here will be discussed in further detail in Chapter 5.
CHAPTER 5: DISCUSSION AND CONCLUSION

5.1 Production of intervocalic /s/ voicing

5.1.1 Categorical, continuous, gradient, or variable?

In section 3.3.2, the overall voicing patterns for the 2,969 tokens of intervocalic /s/ were examined. A closer examination of these patterns can illuminate whether intervocalic /s/ voicing in Lojano Spanish is a continuous or categorical phenomenon. As a reminder, percent voicing was measured by dividing the duration of voicing during the /s/ by the total duration of the /s/. The histogram seen in Figure 7 showed that there are tokens all along the gamut of voicing from 0% to 100%, with the exception of 90-99% voicing in which no tokens are present. The average percent voicing for the read data is 36.2%, while for interview data the average reaches 45.7%. The average percent voicing by word position ranges from 30% to 50%. These average rates are substantially higher than what A. García (2013) finds for Highland Colombian Spanish: she concludes that intervocalic /s/ voicing is not a feature of that variety because percent voicing in her data ranges from 2% to 34%, depending on speech rate and formality. Breaking the data into the three voicing categories (unvoiced, partially voiced, fully voiced), there are 216 fully voiced tokens in the read data, representing 13.6% of the total read tokens. For the interview data, there are 372 fully voiced tokens, representing 26.8% of all interview tokens.
As previously mentioned in section 2.4.3, the terms “categorical”, “continuous”, “gradient”, and “variable” have been used to invoke different concepts in previous studies. Before deciding how to classify intervocalic /s/ voicing in Lojano Spanish, it is important to clearly distinguish these terms. There are at least two senses in which the term “categorical” can be used. First, it can refer to when there are only two discrete variants for the phonetic realization of a phoneme. For this feature, that would be completely unvoiced ([s]) versus completely voiced ([z]). Whether there are discrete categories or not may depend on the perception of the sounds. “Categorical” can also indicate that one variant occurs in a particular context 100% of the time. For example, if /s/ is realized as completely voiced whenever it is in word-final position, this may be considered categorical. Most often, when previous studies on intervocalic /s/ voicing use the term “categorical”, it is not specified which sense of the term the author wishes to invoke; however, the particular analysis indicates that it is likely both. For instance, to say that /s/ voices to [z] categorically in word-final position means that tokens of intervocalic /s/ that occur in word-final position are always 100% voiced. These different meanings of “categorical” do not in fact entail one another since a process could be categorical in that there are two discrete categories, but optional in that it can apply or not.

On the other hand, “gradient”, “continuous”, and “variable” have all been used to signify the opposite of categorical, but in what precise sense? Is a variable process necessarily also gradient? “Gradient” and “continuous” are most commonly synonyms and can be best understood as the realization of a particular sound along a continuum as opposed to as discrete variants. There are a myriad of cases of gradience in human
language. To cite one example, different vowels are phonetically realized along a continuum of F1, F2, and F3, even if perceptually they are discrete categories. Modern acoustic analysis allows phonetic features to be measured in a continuous fashion, whereas this was not the case in the past. For this reason, now more than ever the issue of gradience has come to the forefront in phonetics and phonology. “Variable” processes are those that do not occur 100% of the time. If a given process occurs in a particular linguistic environment 43% of the time and 76% of the time in another linguistic environment, this process could be thought of as variable. Thus, phonological rules that are thought to apply across the board, 100% of the time, are not considered variable. It is possible for a phonetic process to be both gradient and variable, but one characteristic does not entail the other. For instance, in the case of the sounds of interest, even if there are two discrete categories in the production, the process of intervocalic /s/ voicing would be variable if two-thirds of word-final tokens are completely voiced ([z]) versus unvoiced ([s]). For this precise reason, it is of utmost importance to take extreme caution when using the terms “categorical”, “continuous”, “gradient”, and “variable”.

Unfortunately, whether a process is best understood as gradient or involving discrete categories has often been predetermined by the investigator’s analysis. Before acoustic analysis was as readily used as it is today, phonetic and phonological analysis was done auditorily by listening for one variant or the other, for [s] or [z] in this case. Clearly, this presupposes that there are two (or more) discrete categories to listen for. As File-Muriel (2012), among others, points out, many previous studies on Spanish /s/ have unnecessarily forced what could be gradient processes to be categorical by using binary categories (aspirated versus sibilant, for instance, in the case of /s/ weakening). Thus, one
of the goals of this dissertation was to conduct the analysis in such a way that would not
decide a priori whether intervocalic /s/ voicing in this dialect is a categorical or gradient
process.

Based on the distribution of tokens in Figure 7, repeated here as Figure 29, I
conclude that the voicing of intervocalic /s/ in Lojano Spanish is a gradient, continuous
process. As mentioned and clearly evident from Figure 29, the range of percent voicing in
the dataset goes from 0% to 100%, with only a small break in between 90 and 99% voiced. This break in the distribution and lack of tokens 90-99% voiced interestingly has
also been found by other scholars who use a continuous measure of voicing (Campos-
Astorkiza 2014, among others). Nevertheless, besides the gap in tokens 90-99% voiced,
we can see there are tokens at all other values of percent voicing, suggesting gradience.
So what might we expect if voicing in this dialect were a categorical process? Certainly it
would be unreasonable to expect tokens only at 0% and 100% given what we know about
the anatomy and functionality of the vocal chords. We might expect, however, a more
clearly bimodal distribution for percent voicing in which there are two peaks close to 0%
and 100% voiced with some tokens immediately before and after these peaks. While the
distribution in Figure 29 is somewhat bimodal, there are many partially voiced tokens
that fall in the middle range (20%-90% voiced), that is, not close to either peak. In fact,
there are 691 partially voiced tokens in the read data and 568 in the interview data, which
represents more than 40% of all of the data. For these reasons, intervocalic /s/ voicing in
Lojano Spanish is best understood as a gradient process.
This brings us to the issue of variability. Does voicing in this dialect exhibit variation, or is it the case that certain contexts have some degree of voicing while others do not have any voicing at all? Even without considering the results of the statistical models, the percentages cited at the beginning of this section make it clear that intervocalic /s/ voicing in Lojano Spanish, in addition to being gradient, is also a variable process. High percent voicing or fully voiced tokens can occur or not occur in a particular social or linguistic context. Indeed the same speaker can, and this does occur in the dataset, voice an intervocalic /s/ in one production of a given word, and then minutes later produce the same word without voicing.\(^\text{19}\) Nonetheless, this variability is not random and is conditioned by both linguistic and social factors, as was seen in Chapter 3 and will

\(^{19}\) For instance, M7 produces the fully voiced token of \textit{casa} seen previously in Figure 6 roughly two minutes after producing a completely voiceless token of the word \textit{casa}. 150
be discussed in more detail in section 5.1.3. In conclusion, neither definition of “categorical” is appropriate for describing intervocalic /s/ voicing in Lojano Spanish, and instead it is best understood as a gradient and variable process.

5.1.2 Comparison to other varieties of HES

Now that the nature of this process is apparent, it is useful to compare Lojano voicing to what has been described for other subdialects of HES before turning to the conditioning factors. A more detailed description of the studies mentioned in this section can be found in section 2.4.2. One limitation in making these comparisons, however, is that methodologies have varied greatly from study to study. While Robinson (1979), Lipski (1989), and Calle (2010) rely on auditory analysis, Chappell (2011), Strycharczuk et al. (2013), and the present study use acoustic analysis. Despite using acoustic information, Chappell (2011) considers voicing as binary, and so the only results for HES that are truly comparable to the present results are those of Strycharczuk et al. (2013). Thus, these methodological differences must be kept in mind.

Starting with Cuencano Spanish, Robinson (1979) and Calle (2010) claim that intervocalic /s/ voicing in this variety occurs categorically after a prefix or word boundary. Since their analyses are impressionistic, it is safe to assume that they wish to invoke both senses of categorical; that is, that fully voiced tokens occur 100% of the time in the contexts listed. The only participant in either study to deviate from this pattern is a young female Cuencano in Calle’s (2010) sample. It should be noted that Calle (2010) does not consider word-initial /s/ and Robinson (1979) claims there is no voicing in this position. While these results have not yet been corroborated with acoustic data, we can
conclude preliminarily that there is little variation in the realization of intervocalic /s/ in Cuencano Spanish.

For Quiteño Spanish, traditional impressionistic reports have been followed up more recently with acoustic analyses. The earlier studies, Robinson (1979) and Lipski (1989), attest that intervocalic /s/ voices categorically in Quiteño Spanish in word-final position, and Lipski adds to this that /s/ also voices word-finally before a hesitation pause. On the other hand, Calle (2010) finds that some of her Quiteño participants follow this pattern, while others, the younger Quiteños, do not voice at all. However, these results must be taken with a grain of salt given that some of Calle’s participants had not lived their entire lives in Quito. The recent acoustic analyses of Quiteño Spanish have partially confirmed what these studies have claimed, while also suggesting that voicing in this dialect may be more complex than was originally assumed. Chappell (2011) concludes that voicing of /s/ at a word boundary in Quiteño Spanish is not categorical as previous studies had predicted given that she finds the voiced variant in 91%, that is not 100%, of the tokens of word-final /s/. Still, she finds only sporadic voicing in word-initial and medial position, and so overall the pattern of voicing is quite similar to what Robinson (1979) and Lipski (1989) suggest. Finally, Strycharczuk et al. (2013) find that four participants voice /s/ categorically, although optionally, while the other three voice /s/ gradiently. That is, some of their participants produce tokens only around 0 and 100% voiced, while their other participants produce the full continuum of percent voicing. Thus, there is some preliminary evidence for variable intervocalic /s/ voicing in Quiteño Spanish, although this should be confirmed with a larger sample.
Setting aside the methodological differences momentarily, it is evident that there is much more variation present in Lojano Spanish for this feature than what has been described for Cuencano or Quiteño Spanish. Voicing is found consistently in all word positions in Lojano Spanish. Furthermore, as Figure 21 demonstrated, all speakers in the present analysis voice intervocalic /s/ gradiently, producing tokens all along the continuum of 0% to 100% voiced. This directly contrasts to the results of Strycharczuk et al. (2013), since about half of their participants voice intervocalic /s/ categorically. Although additional fine-grained analyses are needed for Cuencano and Quiteño Spanish, it appears from the comparison of the existing studies that Lojano Spanish is unique in exhibiting variable and gradient voicing of intervocalic /s/. A possible reason for why there might be more variability in Lojano Spanish will be discussed in section 5.1.4.2.

5.1.3 Conditioning factors

In Chapter 3, four different statistical analyses of the realization of intervocalic /s/ were presented for Lojano Spanish. As a reminder, these analyses are based on 2,969 tokens of intervocalic /s/ taken from interviews with thirty-one natives of Loja, Ecuador. Based on the distinct distributions and levels of variance between the read and interview data, separate analyses were run for the two modes. In addition, both continuous and categorical analyses were conducted for read and interview datasets, resulting in four total analyses. The dependent measure for continuous analyses is percent voicing, while it is voicing category (unvoiced, partially voiced, fully voiced) for the categorical analyses. A summary of the results of these four analyses was seen in section 3.4 in Table 11, which is repeated here as Table 20.
As evidenced by Table 20, there is a complex interplay of social and linguistic factors that conditions the voicing of intervocalic /s/ voicing in Lojano Spanish. For social factors, gender is a significant predictor in all four analyses, while age is only significant for the interview data. Speech rate, stress, and following vowel also have a significant effect on voicing in all of the analyses. Finally, preceding vowel is only significant for the read analyses, while word position is only significant for the interview analyses. These factors will be discussed one by one in order to elucidate why they effect the realization of intervocalic /s/ and compare the results to previous studies.

**5.1.3.1 Speech rate**

In all four analyses, there is a significant effect of speech rate and the direction of the effect is the same. Speech rate is measured here as syllables per second for the three word sequence surrounding the target /s/. For continuous analyses, as speech rate increases, percent voicing increases. In the categorical analyses, the odds of being in a higher voicing category increase as speech rate increases. Thus, there is a direct
correlation between speech rate and voicing of intervocalic /s/, and the effect of speech rate is very robust in both read and interview speech, with $p$-values in all analyses of less than 0.001. The fact that speech rate is significant in both read and interview data shows that this effect is not limited to a particular mode of speech. Even in the read data where there is less variation for speech rate, the variation that does exist significantly affects the realization of /s/.

The effect of speech rate corroborates what previous studies on intervocalic /s/ voicing and /s/ aspiration have found. Torreira & Ernestus (2012), A. García (2013), and Strycharczuk et al. (2013) find a positive correlation between intervocalic /s/ voicing and speech rate, although for Strycharczuk et al. (2013), this correlation is only present for some participants. In addition, A. García (2013) shows that as speech rate increases, voicing assimilation of pre-consonantal /s/ increases. Furthermore, File-Muriel & Brown (2011) and File-Muriel (2012) show speech rate is the number one predictor of /s/ weakening in the Spanish of Cali, Colombia. Although these studies use different methods of quantifying speech rate, its presence across distinct /s/ phenomena and varieties is evidence of how important this factor is to the realization of /s/ in Spanish.

5.1.3.2 Stress

In addition to speech rate, stress of the surrounding vowels is a significant predictor of voicing in all four analyses. As a reminder, stress is coded as the stress of the preceding vowel plus the stress of the following vowel, producing the following options: unstressed-unstressed, unstressed-stressed, stressed-unstressed, and stressed-stressed. All of the stress patterns are possible for word-final and initial position, while stressed-
stressed is impossible in word-medial position. For examples of each category, refer to section 3.2.4.

While the overall effect of stress is similar in the read and interview data, the particular order of stress patterns in terms of voicing differs. In the read data, percent voicing and the odds of being in a higher voicing category increase according to the following scale: stressed-stressed/unstressed-stressed < stressed-unstressed < unstressed-unstressed. That is, the most voicing is found in the read data when /s/ is preceded and followed by an unstressed vowel. However, in the interview data, the scale from least to most voicing is as follows: unstressed-stressed < stressed-unstressed/unstressed-unstressed. Because of the uneven distribution of stressed-stressed tokens in the data, the results for this stress pattern are somewhat inconclusive in the interview data, although generally stressed-stressed shows similar voicing rates as unstressed-stressed.

The results for stress in part justify the coding scheme that was employed. At least in the read data, it is not the individual effect of the stress of the preceding or following vowel, but instead the cumulative effect of the stress of both vowels that conditions the amount of voicing. That is, there is a three-way distinction for stress: the least amount of voicing is found when the following vowel is stressed (regardless of preceding vowel), followed by more voicing in stressed-unstressed, and the most voicing in unstressed-unstressed. Taking the results of all of the data together, we can conclude that /s/ being preceded and followed by an unstressed vowel is the ideal condition for voicing.

Although previous studies have not always utilized the same coding schema for stress, their results are similar to those of the present study. Torreira & Ernestus (2012) code stress according to whether /s/ was in the onset of a stressed syllable or not, and find
that /s/ exhibits shorter lower-band intensity dips, which correlates with voicing, in unstressed syllables. Davidson (2014) codes stress the same way as the present study and finds that for the Catalan-dominant group of participants, voiced tokens are favored between unstressed vowels. The only other study to consider stress of surrounding vowels, Chappell (2011), does not find a significant effect of stress. Given the significant effect of both speech rate and stress, the results of the present study contradict directly Lipski’s (1989: 50-53) claim that intervocalic /s/ voicing is not dependent on stress or speech rate. Lipski made this claim in reference to Quiteño Spanish, and the results presented here show that it is unfounded for Lojano Spanish.

5.1.3.3 Preceding/following vowel

Preceding and following vowel were coded as the segment that precedes and follows the target /s/, with the following options: [a], [e], [i], [o], and [u]. Unlike speech rate and stress, the read and interview data differ notably when it comes to the effect of preceding and following vowel. Both are selected as significant in the analyses of the read data, whereas only following vowel is significant in the interview analyses. The vowel results for the read data were found to be of less importance given that only a few vowel comparisons were found significant, and the effects that were seen could be easily attributed to the structure of the data itself. For this reason, I will focus on the results for following vowel in the interview data, where more clear patterns are observed.

In both continuous and categorical interview analyses, there is higher percent voicing/higher voicing category for [a, e, o] than for [i, u]. These groupings of vowels fall neatly into two natural classes: non-high vowels [a, e, o] and high vowels [i, u]. For the
continuous analysis, the order of vowels from least to most voicing is as follows: \([u] < [i] < [o] < [e] < [a]\). For the categorical analysis, the odds of being in a higher voicing category increase along this cline: \([u] < [i] < [e] < [a] < [o]\). Even though the precise ordering of the vowels differs between the two analyses, the high vowel versus non-high vowel distinction is clear in both since there is no significant difference between \([i]\) and \([u]\), or between any combination of \([a]\), \([e]\), and \([o]\). Thus, there is significantly more voicing in the interview data when /s/ is followed by a non-high vowel than when it is followed by a high vowel.

The results for vowel are different from what previous studies have found. Torreira & Ernestus (2012) find no effect of either preceding or following vowel. Chappell (2011), on the other hand, shows that voicing is favored when /s/ is preceded by a low vowel (\([a]\)) as compared to a mid or high vowel (\([e, o, i, u]\)). The distinction in Chappell’s results seems to be between low and non-low vowels, while for the present study it is between high and non-high vowels. Nonetheless, the results correspond somewhat in that there is more voicing before \([a]\). The other studies of intervocalic /s/ have not looked at the effect of preceding or following vowel because they have used exclusively read or scripted data in which preceding and following segment were controlled for.

5.1.3.4 Position within the word

As mentioned in Chapter 2, position within the word has been central to the discussion of how intervocalic /s/ is realized in HES. Unsurprisingly, this factor is also a significant predictor in the present results, although the particular effect is different than
in other varieties of HES. In the continuous interview analysis, there is significantly higher percent voicing in word-final and initial position than in medial position. Word-initial represents the highest percent voicing; however, the difference between word-initial and final is not significant. The results for the categorical interview analysis are quite similar: the odds of being in a higher voicing category are significantly greater in word-final and initial position as compared to medial. The odds are greatest in word-final position, but the difference between word-final and initial is not significant. Combining the continuous and categorical results, we see that there is more voicing overall in word-final and initial position, and significantly less in word-medial position.

The fact that there are high rates of voicing in word-final position is not surprising given the results of previous studies. Robinson (1979), Lipski (1989), Calle (2010), Chappell (2011), Torreira & Ernestus (2012), McKinnon (2012), Strycharczuk et al. (2013), Davidson (2014), and Hualde & Prieto (2014) all find more voicing or voicing exclusively in word-final position. Indeed, the only study to find contrary results is A. García (2013), whose results show more voicing in word-medial position, although the overall low rates of voicing in her data lead her to conclude intervocalic /s/ voicing is not a feature of Highland Colombian Spanish. What is more surprising in my own study is that voicing in Lojano Spanish in word-initial position is as prevalent as word-final position. Unfortunately, not all previous studies consider voicing in word-initial position, and thus it is impossible to conjecture whether they would have found voicing in this environment or not. Of those that do consider word-initial position, most find very little voicing in this context (Calle 2010, Chappell 2011, McKinnon 2012), with the exception of Hualde & Prieto (2014), who find no statistical difference between voicing in final and
initial position in their Madrid Spanish data. Thus, the results for word position in Lojano Spanish represent somewhat of a departure from past studies on other HES varieties in that Lojanos not only mark the end of words with more voicing, but also the beginning of words. It is also important to keep in mind that while the rates of word-medial voicing are less than the other word positions in Lojano Spanish, the mean percent voicing for word-medial intervocalic /s/ still reaches about 40% in the interview data, and so /s/ voices notably more in word-medial position in Lojano Spanish than in other varieties of HES. In this regard, voicing across the three word positions is actually more similar in Lojano Spanish given the almost nonexistence of word-initial and medial voicing in other HES dialects.

5.1.3.5 Gender

The effect of gender is significant in all four analyses of the data, and the effect is quite robust given that the $p$-value for three of the four analyses is less than 0.001. In read and interview data, tokens of intervocalic /s/ produced by male participants are more likely to be in a higher voicing category, and produced with higher percent voicing, than those produced by female participants. This result departs from what past studies have found. While most previous studies have not found the effect of gender to be important (Chappell 2011, Strycharczuk et al. 2013, Torreira & Ernestus 2012), McKinnon (2012) finds that Catalan-Spanish bilingual females slightly favor voicing, while their male counterparts slightly disfavor voicing. On the other hand, preliminary observations from Costa Rican Spanish suggest that in this dialect males may voice intervocalic /s/ more than females (Chappell & García, ms.).
5.1.3.6 Age

While age does not reach significance in the read data, it is a significant predictor of voicing of intervocalic /s/ in the interview data. As a reminder, age is considered as a continuous predictor. In the continuous and categorical interview analyses, there is an inverse relationship between age and voicing. As age increases, the percent voicing or odds of being in a higher voicing category decrease. This means that younger participants voice significantly more than older participants. This effect is more robust in the continuous analysis than in the categorical analysis. The trends observed in the read data are the same, and thus it is likely that age does not reach significance in the read data simply because the amount of variation in the read data is less than the interview data.

Among previous studies of intervocalic /s/, only one has considered age as an independent variable. Calle (2010) finds that younger Quiteños do not voice as much as older ones and concludes that perhaps this is indicative of a change in progress in Quiteño Spanish. However, since she only interviewed six participants from Quito, there are only two participants per age group, and so the strength of this conclusion must be questioned. In any case, the trends for age observed in the present study for Lojano Spanish are the opposite of what Calle finds for Quiteño Spanish since younger Lojanos voice intervocalic /s/ more than older Lojanos.

5.1.4 Theoretical implications of production data

5.1.4.1 Previous formal analyses

Given the limited presence of intervocalic /s/ voicing in Modern Spanish, a few scholars have tried to account for it with formal phonological analyses. Most of these
analyses have been concerned with voicing in HES. I will review and critique how various scholars have dealt with this issue and how their analyses might fit my data, leaving a discussion for what I believe to be the most fruitful analysis for section 5.1.4.2.

Since all of the formal analyses of intervocalic /s/ voicing predate the first acoustic analysis of this feature (Chappell 2011), they are based on the impressionistic data of Robinson (1979) and Lipski (1989). After reviewing these analyses, we will see whether they can still be upheld given the HES data that has been advanced more recently. The first analysis put forth is that of Lipski (1989). As mentioned previously, the author asserts that /s/ voices categorically at the word boundary in Quiteño Spanish, both prevocally and before a hesitation pause. In a footnote (p. 51), Lipski expresses his reservations about Robinson’s (1979) claim that /s/ also voices before prefixal morpheme boundaries in Cuencano Spanish, but it does not seem as if he has any data from Cuenca to prove otherwise. The analysis that Lipski proposes is a derivational account that makes a connection between intervocalic /s/ voicing and other Spanish phenomena, most notably /s/ aspiration. Using a word delimitation rule, an unattached C-slot, i.e. the representation of the word boundary, is added on after word-final consonants, including /s/. In the next step of the derivation for HES, /s/ voices when it appears before this C-slot. Sometime before resyllabification, the C-slot is then eliminated, resulting in voicing of /s/ in intervocalic positions. The strength of this claim is that it can also account for voicing assimilation of word-final /s/ before a voiced consonant. It also brings together many word-final consonantal modifications, such as /n/ velarization, under a unified analysis. Like other scholars have pointed out (cf. Colina 2009), however, Lipski’s account requires another rule to “undo” this voicing in contexts where it is not
evidenced, such as word-finally before voiceless consonants. In addition, Bradley (2005) points out that Lipski’s analysis does not allow for variable voicing of /s/ before voiced consonants since it predicts fully voiced [z] in this context.

Taking advantage of the theoretical tools of Optimality Theory (OT; Prince & Smolensky 1993) and systemic contrast, Bradley (2005) and Bradley & Delforge (2006) offer another solution. The analysis presented in these two articles is essentially the same, only differing in the dialects that are covered. Bradley (2005) focuses solely on HES and includes an explanation for prefixal voicing in Cuencano Spanish. Bradley & Delforge (2006) provide a broader scope, including diachronic sibilant voicing patterns and modern day voicing in both HES and the Peninsular dialects mentioned in Torreblanca (1986); see section 2.4.1 for more details on (non-Catalonian) Peninsular intervocalic /s/ voicing. Given that the analysis is so similar between these two articles, I will combine the two and discuss them as one analysis, only mentioning them separately in cases where they differ.

Bradley and Delforge exploit the concept of underspecification and therefore assume a tripartite voicing distinction: phonologically contrastive [+voice] and [-voice], and targetless/neutral [0voice], represented as [S]. This underspecified [S] does not have its own voicing target and therefore is realized via interpolation between the voicing targets of the surrounding sounds. More specifically, they state, “The realization of neutral [S] depends on phonetic factors such as sibilant duration and intensity, stress, adjacency to major prosodic boundaries, speech register, and speaking rate” (Bradley & Delforge 2006: 9). The other essential tenet of this analysis is systemic contrast, which these authors borrow from Dispersion Theory (Flemming 1995). Intervocalic position
favors both internal and transitional cues of sibilants, allowing them to be most perceptually distinct in this environment. Based on this fact, they propose a systemic contrast preservation constraint, which specifies that two sibilants must contrast at least as much as they do in intervocalic position, preserving minimal degree of perceptual distinctiveness.

Within their OT analysis, Bradley & Delforge also make use of traditional input-output faithfulness constraints as well as two non-systemic markedness constraints, which dictate that syllable-initial sibilants be voiceless in the output and that no obstruent have a [voice] feature. The re-ranking of these constraints accounts for the diachronic changes in sibilant voicing from Medieval Spanish to Modern Spanish and also the /s/ voicing patterns exhibited in HES and Peninsular varieties, which they consider to be innovations. For HES, all word-final /s/ input to the postlexical level is assumed as /S/. However, this is mapped postlexically to [z] in the output when followed by a vowel to preserve the contrast between word-final and word-initial /s/, for example in minimal pairs such as has ido ‘you have gone’ [a.zi.ðo] and ha sido ‘it has been’ [a.si.ðo]. This is based on Robinson’s (1979) claim in his original work on HES that Cuencano speakers can perceive the difference between [s] and [z] in word-final prevocalic contexts. Bradley and Delforge interpret this observation as a re-ranking of their systemic contrast preservation constraint to a higher position post-lexically, rather than lexically. In order to explain the difference among subdialects of HES, Bradley (2005) takes this analysis one step further and claims that prefixes are incorporated at the word level in Quiteño Spanish and postlexically in Cuencano Spanish. Finally, for the Peninsular voicing dialects, where intervocalic /s/ can be realized as [z] in all word positions, Bradley &
Delforge (2006) predict [S] to be preferred across the board in intervocalic contexts because the *[αvoice] constraint is very highly ranked in these dialects.

The strong point of this analysis is that it accounts for diachronic changes, gradient voicing assimilation before voiced consonants (where they predict [S]), and modern intervocalic voicing phenomena. The re-ranking of constraints offers a clear visualization of how sound change can occur. Nevertheless, even without looking at data from more recent studies, there are some issues with this proposal. First, according to the principles of OT, all of the faithfulness and markedness constraints used should be phonetically or perceptually motivated. In both articles, the authors maintain that a sibilant specified for voicing presumably requires more effort, which justifies the *[αvoice] constraint. This assumes that sibilant voicing will be more “effortful” in general and does not take into account how context could affect degree of effort. Clearly, voicing post-pausal /s/ is not the same as voicing intervocalic /s/ (cf. Westbury & Keating 1986). Also, the constraint that specifies syllable-initial sibilants as voiceless is based on the fact that voiceless sibilants are preferred aerodynamically in utterance-initial position. The OT constraint references syllable-initial position while the phonetic reasoning is related to utterance-initial position. Finally, the whole theory hinges on the preservation of contrasts such as *ha sido* ‘it has been’ versus *has ido* ‘you have gone’, which are few and far between. It seems inefficient to preserve contrast in a context where there are so few minimal pairs. Even Robinson agrees in later work that [s] and [z] do not truly contrast, but that they are merely allophonic of the same phoneme /s/ (Robinson 2012: 133).

Another formal analysis of intervocalic /s/ voicing to consider is Colina (2009).
Like previous analyses, Colina assumes underspecification but gives a different reasoning in support of underspecification. She states that coda position cannot license a laryngeal node, making coda /s/ underspecified [S]. Under this account, any /s/ in the coda must obtain its voicing specification from neighboring segments through a shared laryngeal node. The author further motivates a ternary voicing distinction, despite it being more complex than a system with only [+voice] and [-voice], by showing that an OT analysis without [0voice] as an option cannot account for voicing in HES. This is in line with previous analyses’ reliance on three voicing categories, although Colina uses somewhat different OT constraints. The main innovation in her analysis is that she claims that all word-final sibilants are underspecified in HES, and therefore does not posit a later step of postlexical mapping to [z] as Bradley (2005) and Bradley & Delforge (2006) do.

According to Colina’s constraint hierarchy, underspecified [S] is always selected as the optimal candidate for HES. Contrastively, in other dialects of Spanish [s] is the optimal candidate for prepausal, word-final prevocalic, and before a voiceless consonant, and [z] is predicted before a voiced consonant.

Colina concludes that her analysis is more explanatorily adequate and more elegant than previous accounts because it does not rely on two levels of representation (lexical and postlexical). One issue with Colina’s analysis is that it allows for gradient voicing of /s/ before voiced consonants only in HES, when this has been found for other dialects of Spanish. Additionally, the output-output voicing identity constraint she employs states: “the output voice specification is identical to that of its correspondent in isolation” (p. 13), which seems to imply two levels of representations even if they are not explicitly stipulated.
One final analysis to examine is that of Strcharczuk et al. (2013). As mentioned previously, these authors find that some Quiteño speakers may voice categorically, while others voice gradiently. More specifically, three of their speakers show an effect of speech rate manipulation, while the other three speakers exhibit an “optional phonological rule” (p. 36), and therefore do not show any effect of speech rate. The authors utilize this transitional stage of Quiteño Spanish, in which the community exhibits more than one pattern, to propose the stages of this linguistic change. They claim that coda delaryngealization for /s/ started in pre-consonantal position at the phrase level, and then extended to word level (Stage 2), which would allow for the voicing assimilation found in sequences such as gas noble ‘noble gas’ [gaS.no.βle]. From there, Strycharczuk et al. propose an analogical change in which prevocalic word-final /s/ suffers delaryngealization due to the parallel between preconsonantal and prevocalic word-final /s/ (Stage 3). In both cases, what starts as a gradient process stabilizes over time as listeners reinterpret partial voicing as full voicing, that is, as [z] (Stage 4). The authors conclude by saying that Quiteño Spanish is at the moment in between Stages 3 and 4, coda delaryngealization and categorical reinterpretation, which is why they find two different patterns among their participants. Furthermore, the fact that these two patterns exist in the community is evidence of a sound change in progress, whereby the speakers that voice categorically, but optionally are on the forefront of the categorical reinterpretation.

The strength of this analysis is that it is able to capture why intervocalic /s/ voicing only occurs word-finally in Quiteño Spanish, and not medially, while also highlighting the similarity between intervocalic and preconsonantal voicing. It is also
able to propose a plausible trajectory for /s/ voicing in Quiteño Spanish with motivations for each stage. One issue, however, is that the assumptions of the analysis are based on seven speakers of this dialect, one of which does not follow either pattern, and data from a more representative sample would be necessary to strongly support the conclusion that two distinct patterns exist in this variety. Additionally, while the authors account for why voicing only occurs in word-final prevocalic /s/, they do not sufficiently explain why this is an optional phonological rule.

The previous analyses reviewed here are unable to account for the patterns I have found in my production data. The analyses presented in Lipski (1989), Bradley (2005), Bradley & Delforge (2006), and Strycharczuk et al. (2013) cannot account for variable intervocalic /s/ voicing in Lojano Spanish, because they would predict [z] exclusively in word-final position, which is not what we have seen to be true. At first glance, it seems that a modified version of Colina’s (2009) analysis or the Peninsular analysis of Bradley & Delforge (2006) could explain the patterns seen in Lojano Spanish. If we assume that intervocalic /s/ is always underspecified for voicing, this would allow /s/ to voice gradiently in initial, medial and final positions. In fact, this would predict that voicing of /s/ is dependent on “phonetic factors such as sibilant duration and intensity, stress, adjacency to major prosodic boundaries, speech register, and speaking rate” (Bradley & Delforge 2006: 9). There are two potential issues with this revised proposal, however. First, as these authors admit, requiring an underspecified [S] may add a level of complexity to the system that is not necessary. Second, if /s/ is underspecified and realized just through the interpolation between the surrounding vowels or by sharing a laryngeal node with these vowels, it may be difficult to explain any completely voiceless
productions. As we saw in Figure 29, there are indeed quite a few tokens that are produced with 0% voicing. If /s/ is not specified as voiceless, we might expect it to show even more influence of the surrounding vowels than it does, i.e. a higher degree of voicing than what we observe.

The former analyses suffer from a more fundamental flaw: they assume that phonetics and phonology must be kept separate, not allowing any interaction between the two. For Lipski (1989), Bradley (2005), and Bradley & Delforge (2006), /s/ voicing in HES is an exclusively phonological process. In the view of Colina (2009), it is a purely phonetic phenomenon. Strycharczuk (2012) sums up this issue nicely saying, “Spontaneous voicing operating on phonetically underspecified targets is expected to be variable and gradient. A phonological voicing process, on the other hand, is expected to be categorical” (p. 29). I believe a more profitable approach should incorporate both phonological and phonetic factors. I will develop the model to explain intervocalic /s/ voicing in the next section.

5.1.4.2 Proposed analysis

The fact that intervocalic /s/ voices more in higher speech rates, between unstressed vowels, and before non-high vowels suggests that this might be a reduction process. This idea is best understood under the framework of Articulatory Phonology, as described in Browman & Goldstein (1989). I will first describe the basic principles of this framework and then explain how it can be applied to intervocalic /s/ voicing. The central hypothesis of this seminal work is that gestures are the basic phonological units of contrast instead of segments. In this framework, gestures are defined as “an abstract
characterization of coordinated task-directed movements of articulators within the vocal tract” (p. 206). Gestures are coordinated both spatially and temporally in a “larger coordinated structure, or constellation” (p. 206), and these constellations are represented in a gestural score. As such, two languages might have the same gestures for a given sound but differ in the relative timing of the gestures. Crucially, the specified spatiotemporal coordination of gestures may change in the implementation due to the relative size and overlap of the gestures. Thus, this model can easily account for a speaker’s need to increase speed and fluency, through reducing the duration and increasing the overlap of gestures, particularly in casual speech. Since this first proposal by Browman & Goldstein, increase in the overlap of gestures and/or decrease in gestural magnitude, which may result in gestural hiding or gestural blending, have been invoked to explain a wide variety of phenomena, such as coarticulatory effects, allophonic variation, deletion, insertion, and assimilation.

Following Campos-Astorkiza (2014), I propose that a gestural approach can provide a better representation of the gradient and variable nature of Lojano voicing. In her study of pre-consonantal /s/, Campos-Astorkiza (2014) concludes that for /s/ before voiced consonants, “increased gestural overlap between two adjacent and contradictory glottal gestures results in gestural blending and, consequently, gradient surface assimilation” (p. 19). Extending this to intervocalic /s/, if we assume /s/ is underlyingly voiceless, its voiceless glottal gesture is flanked on either side by the voiced glottal gestures of the surrounding vowels. A particular coordination of these glottal gestures is specified, but as speech rate increases, there will be more gestural overlap between /s/ and the surrounding vowels, and in addition the gestural magnitude of the glottal gestures
will be reduced. The voiced glottal gestures of the vowels overlap with the voiceless glottal gesture of the /s/, which results in the gradient voicing of intervocalic /s/. If /s/ is influenced by the voiced glottal gestures on either side, we might expect voicing to be evident on the transitional edges between vowel-/s/ and /s/-vowel. This is generally the pattern I have observed in tokens that are not fully voiced, as is the case with the example token in Figure 5 in section 3.2.3.

This representation seems to suggest that intervocalic /s/ voicing in Lojano Spanish is a reduction process, which goes along with the results for speech rate, stress, and following vowel. This interpretation gains some support from two previous studies of intervocalic /s/ that mention articulation. Torreblanca (1986) affirms that most cases of modern intervocalic /s/ voicing are due to articulatory relaxation, and Widdison (1997) asserts that a “greatly abbreviated syllable-final [s]” (p. 262) may spontaneously voice even when it is not followed by a voiced consonant.

This analysis is also validated by many scholars’ accounts of voicing as a type of weakening. In his review of the fortis/lenis feature in many different languages, Kohler (1984) discusses how voicing goes against the air-stream requirements of obstruents, and yet voicing is characteristic (in some languages) of lenis obstruents. Ultimately, he decides that we must consider the sound as a whole when deciding whether it belongs to the lenis or fortis category. As such, the other characteristics of voiced obstruents (such as duration) might sway them towards being classified as lenis. While most of the examples that Kohler gives concern stops, this point is important to consider for fricatives as well since voiced fricatives are generally shorter than voiceless ones.
Westbury & Keating (1986) challenge conventional views on the markedness and naturalness of stop voicing. In their view, “natural” should be defined in terms of articulatory ease, with “easier” meaning fewer changes in state from segment to segment. This emphasizes the importance of context in determining whether something is more “natural” or “easy” – since individual sounds are very infrequently produced in isolation. Using the equivalent network model and breath-stream dynamics model, Westbury & Keating calculate differential equations of the changes in the system over time in order to see what these models would predict for utterance-initial, intervocalic, and utterance-final stops. According to their calculations, 60 milliseconds of voicing is ideal for a vowel+stop+vowel sequence, and so any more voicing or voicelessness would require an extra “cost”. Based on these results, the authors conclude “there is a well-defined sense in which it is more difficult to terminate than to allow (or maintain) voicing during a stop” (p. 158). The potential extension to intervocalic fricatives is clear: in order to produce a voiceless fricative between vowels, one would have to in a sense turn off voicing, which could imply an extra effort. Another important point made by Westbury & Keating is that there are always multiple and complex constraints to consider. Where the predictions of their model fail, they cite that likely communicative efficiency, acoustic invariability, and perceptual requirements are coming into play.

Building on these remarks, Recasens (2002) reviews a myriad of Romance processes and decides whether the terms “weakening” or “strengthening” should be applied in each case. Again, he highlights that we must look at the context where a process occurs before concluding it is “weakening” or “strengthening”. For example, in the case of liquids, it is hard to say if [l] > [ɾ] or [ɾ] > [l] are cases of strengthening or
weakening since these sounds often happen in identical environments, and so Recasens believes instead that liquid processes can be explained through acoustic motivation.

Ultimately, the author decides that the term weakening “should be applied exclusively to segmental replacements resulting from a decrease in duration in constriction degree, and from a change in glottal activity yielding voicing” (p. 355). Although he does not mention intervocalic /s/ voicing, we can conclude from this definition that he would consider this process to be weakening since both voicing and a decrease in duration are involved.

Summarizing, the view of intervocalic /s/ voicing as a reduction phenomenon espoused here is supported by arguments from several studies. Seen in this light, it is possible to draw parallels with other intervocalic processes in Spanish and more generally in Pan-Romance. Hualde et al. (2011) and Hualde (2013) both examine the voicing of intervocalic /p, t, k/. Hualde et al. (2011) suggest that examining incipient voicing in voiceless stops in modern-day Spanish may allow us to understand the historical case of lenition that happened in the evolution of Latin to Romance. Their data comes from a corpus of twenty speakers from Majorca, with both unscripted and scripted speech. The authors coded tokens of intervocalic /p, t, k/ and /b, d, g/ as voiced (uninterrupted F0 tracker), partially voiced (some interruption but mostly present), or voiceless (less than half voiced). While they find that /p, t, k/ are at times realized as partially or fully voiced, the other acoustic measures taken show that there is a significant difference in constriction and duration between /p, t, k/ realized as voiceless, /p, t, k/ realized as voiced, and underlying /b, d, g/. Hualde et al. (2011) take this to mean that intervocalic /p, t, k/ and /b, d, g/ have not yet merged because they are differentiated by features other
than voicing. Lastly, they claim that sound change starts at first as an online effect of
gestural overlap, which is then conventionalized: intervocalic /p, t, k/ is in the process of
conventionalization since voicing was the only type of reduction (and not for instance
fricativization) seen in their data. Although these authors do not provide much detail
about how the gestural overlap affects intervocalic /p, t, k/, we might assume that their
model is similar to the one presented here.

Hualde (2013) provides evidence that voicing of intervocalic /p, t, k/ also occurs
in Rome Italian and Goizueta Basque. Similar to the effect of speech rate that I have
found, Hualde (2013) also finds in this preliminary study that rates of /p, t, k/ voicing are
higher in fast as compared to normal speech. In addition, Blecua & Rost (2013) show that
intervocalic /f/ in Peninsular Spanish is realized on a gradual scale of reduction from the
least relaxed [f] to the most relaxed [β], with 37% of realizations being partially or fully
voiced. Finally, though his analysis did not pan out for Lojano Spanish, Lipski (1989)
was right in making a connection between intervocalic /s/ voicing and /s/ aspiration.
Beyond the fact that these two phenomena sometimes occur in the same linguistic
context, Recasens (2002: 349) cites that [h] may voice before being deleted all together in
dialects where aspiration happens in intervocalic contexts.20 Indeed, File-Muriel &
Brown (2011) regard voicing as indicative of weakening in their continuous analysis of
/s/ weakening in the Spanish of Cali. All of these studies serve to demonstrate that
intervocalic /s/ voicing is actually part of a much broader panorama of intervocalic
reduction in Spanish.

20 In an unrelated project (Walker et al. 2014), my colleagues and I have noted voiced
productions of intervocalic [h] in Puerto Rican Spanish.
Having concluded that intervocalic /s/ voicing in Lojano Spanish can and should be considered a reduction process, I will now return to the proposed gestural model to explain how the results for each conditioning factor fit into this model. As mentioned previously, the predetermined gestural alignment for adjacent sounds can be effected in the implementation by factors such as a speaker’s need to increase speed and fluency. More precisely, as speech rate increases, there will be more overlap between adjacent gestures (cf. Harcastle & Hewlett 1999). Again, more overlap in the case of intervocalic /s/ voicing implies gestural blending between the voiced glottal gesture of the surrounding vowels and the voiceless glottal gesture of the /s/. Thus, a gestural account would predict more voicing as speech rate increases, which is precisely what is seen in the results here.

In unstressed positions, the magnitude of gestures is less than in stressed positions, which allows for more reduction in the former, weaker positions. For instance, Cooper (1991) finds that in English the glottal gesture is affected by stress in two ways, that is, that the length and the magnitude of the glottal gesture decrease in unstressed positions. In the case of intervocalic /s/, this predicts more voicing in unstressed positions because there will be less time allotted to reach the voiceless glottal target and because the voiceless gesture is reduced in magnitude. Cole et al. (1999) find more weakening of stops in unstressed positions in Spanish, supporting the fact that unstressed contexts are articulatorily weak in this language. The results seen here for intervocalic /s/ voicing in Lojano Spanish follow from this articulatory analysis. The greatest amount of reduction (voicing) is found in unstressed-unstressed tokens, followed by stressed-unstressed tokens, and much less so in unstressed-stressed and stressed-stressed tokens. When /s/ is
both preceded and followed by an unstressed vowel, the reducing tendency of unstressed position is doubled, allowing for the greatest amount of voicing to occur. On the other hand, in unstressed-stressed and stressed-stressed tokens, the magnitude and length of the gestures is increased as a consequence of being in or next to a stressed position, which produces less voicing because the voiceless glottal gesture has more time and greater magnitude to be more fully realized and suffers less gestural overlap with the surrounding vowels’ gestures.

The relationship between vowel height and voicing of /s/ can also be understood in reference to the articulation of these sounds. Given the closer constriction of high vowels as compared to non-high vowels, voicing is harder to maintain before a high vowel (Koenig et al. 2011). That is, when an intervocalic /s/ is followed by [i] or [u], the articulators are preparing for the close articulation of the upcoming vowels, which makes voicing of the /s/ dispreferred aerodynamically. For the interview data, this is the effect of following vowel that we find. There is more voicing before non-high vowels than before high vowels.

While the first three conditioning factors are easily accommodated into a gestural approach, the results for word position are a bit more complicated. It seems that if voicing in this dialect were purely due to articulatory reduction, there would be no effect of word position, or if anything word-medial position should exhibit the most voicing. Within the word, the gestures between adjacent sounds might be more coarticulated, which would allow for more overlap between /s/ and the surrounding vowels, leading to more voicing of that /s/. Nevertheless, given that speech is continuous and that resyllabification occurs in Spanish, there may not be a substantial difference between the
coordination of gestures within a word as opposed to between words. However, this still leaves us with the problem of the different behavior between medial, final, and initial positions.

A few previous studies have invoked incipient phonologization to explain the higher rates of /s/ voicing in certain word positions (cf. Torreira & Ernestus 2012). These scholars have claimed that as intervocalic /s/ voicing becomes a phonological process, as opposed to just a phonetic reduction process, it will affect a given /s/ differently based on where it falls in the word. If this feature is indeed entering into Lojano Spanish, as will be discussed later in this section, the effect of position might be even stronger in the future as intervocalic /s/ voicing becomes limited to certain positions. The idea of phonologization does not explain or predict which positions will exhibit more voicing. It is possible that word-final voicing is preferred in Lojano Spanish because this dialect is in contact with Quiteño and Cuencano Spanish, where voicing is found almost exclusively in word-final position. Furthermore, word-final position is known to be a weak position in Spanish, which Hualde & Prieto (2014) explain as being the result of timing differences between /VCV/, /V#CV/, and /VC#V/ sequences. While these authors consider several potential reasons why word-final intervocalic /s/ might voice more than word-initial or medial, they ultimately decide that the most plausible hypothesis is that word-final /s/ is more susceptible to voicing given its gestural coordination, even if resyllabification, which they claim may be incomplete, eventually occurs. Subsequently, I believe that word-initial voicing may be caused by analogy with word-final position. Since cases of word-final /s/ occur before a vowel, they would resyllabify into the next syllable, which makes them appear on the surface like instances of word-initial /s/, which
could fuel the symmetry between final and initial position. That is, word-initial and final positions are interpreted as symmetrical and reanalyzed as word edges. While this result is not necessarily predicted by a gestural approach, it can be accommodated within it by assuming that the gestural coordination specified for word-final and initial positions (between words) is different than that of word-medial position (within words). This pattern of voicing in Lojano Spanish mirrors what has been found for /s/ aspiration in a few dialects of Spanish, such as New Mexican Spanish and some rural dialects of Highland Colombian Spanish (cf. Brown & Torres Cacoullos 2003). In these varieties, /s/ aspiration began as a word-final, pre-consonantal phenomenon, extended to include word-final prevocalic contexts, and eventually began to affect even word-initial /s/, typified by the classic example of señor ‘sir’ pronounced as [heɲoɾ]. This parallel with /s/ aspiration gives more weight to the idea that word-initial voicing may be the result of analogy as it shows another case in Spanish where this has occurred.

It has also been shown that social factors influence intervocalic /s/ voicing in Lojano Spanish, namely gender and age. There are many potential factors that could cause male Lojanos to produce more voiced tokens than their female counterparts. Since the effect of speech rate is so important, a first thought is that perhaps the men in this sample have higher speech rates, which could cause them to voice intervocalic /s/ more. This hypothesis is supported by the average speech rates: female participants average 5.6 syllables per second in the read data and 6.5 syllables per second in the interview data, while male participants average 6.5 and 7.0 syllables per second in the read and interview data respectively. That is, the average speech rate of women in the interview data is
comparable to the men’s average speech rate when reading. Nonetheless, the effects of
gender and speech rate are actually independent. The first piece of evidence for this is
that the addition of gender to the statistical models significantly improves the overall fit
of the model to the data even when speech rate is already present as a factor, suggesting
that the effect of participant gender goes above and beyond the effect of speech rate.
Additionally, testing for collinearity shows that speech rate and gender are not collinear,
which provides further evidence that their effects are independent. Clearly speech rate
alone cannot provide a justification for why male Lojanos voice more than female
Lojanos.

For another potential explanation, we must turn to physiology. File-Muriel et al.
(2012) state that men are more likely to voice intervocalic /s/ because on average their
vocal cords are thicker than women’s vocal cords, which would make it more difficult to
turn “off” voicing for the /s/ and then turn voicing back “on” for the subsequent vowel in
a vowel-/s/-vowel sequence. Similarly, Jessen (2009) finds gender differences in the
production of voicing in stops in German that he claims are due to different sizes of oral
cavity. In the case of small oral cavity volume, typical of women, there is an aerodynamic
preference for short voicing duration because the air pressure at the closure or channeling
of air increases quite rapidly, triggering the cessation of voicing. This is the same reason
that voiced fricatives are aerodynamically dispreferred overall, discussed earlier in this
section, however, this is magnified in the case of a smaller oral cavity. On the other hand,
apart from being physiologically motivated, voicing could be a gendered practice that
male Lojanos learn and emulate, and in fact Johnson (2006) concludes that anatomical
characteristics cannot fully account for gender differences in speech. Foulkes & Docherty
(2006) alert that it is often very difficult to separate the effect of non-learned, non-arbitrary factors, such as vocal tract characteristics, from the effect of arbitrary, learned behaviors. Regardless of whether voicing is physiological or learned, the fact that male Lojanos voice more creates the situation where voicing could be socially associated with male speech.

The inverse relationship between age and voicing suggests that this feature may be coming in to Lojano Spanish from neighboring dialects. Usually when new linguistic features are adopted by a community, it is the youngest generations who will be the first group to show high rates of adoption, and from there the feature diffuses throughout the community as time progresses. This is the pattern that we see with intervocalic /s/ voicing in Lojano Spanish; however, this conclusion remains tentative given the number of speakers in the dataset. Bailey (2002) warns that any conclusions made based on comparing age groups (apparent-time construct, see section 2.4.5) are only warranted for adult participants in large, representative samples. Thus, the effect of age should be corroborated in the future by a larger sample of Lojanos, although the pattern seen here is already quite suggestive and is also supported by social changes in Loja. While it was once very difficult to traverse the roads between Loja and other neighboring cities, and indeed Loja was not connected by highway at all to other highland cities until the mid-twentieth century (Toscano Mateus 1953: 15), that is no longer the case due to the rapid improvement of highways in this region. Lojanos can now travel much more freely and easily than in the past, which has increased overall communication between Lojanos and their neighbors in Cuenca and Quito, priming their dialect for linguistic change. This idea is also supported by the fact that there are high rates of voicing in word-final position,
which is the environment where voicing is present in Cuencano and Quiteño Spanish, and by the fact that there are other linguistic changes (e.g. assibilation of /r/) that are transpiring in Lojano Spanish.

The intersection of age and gender can give us additional information about this change. As mentioned in section 3.3.8, the direction of the effect of age in the interview data is similar for males and females, however the age cutoff for males and females is quite different. The female participants are divided into younger and older than 20 years old, while the breakoff point for males is younger or older than 60 years old (see Figure 13). This illustrates that young and middle aged men show similar rates of voicing in comparison to older men, while the youngest women are markedly different than the women in middle and oldest age groups. Nevertheless, the effect of gender is so great that even the youngest female participants do not reach the same rates of voicing as the oldest male participants. Thus, we could say that the leaders of this change are the male Lojanos under the age of 60. Linguistic change in which the men of the community are leading the change is not completely unheard of (cf. Michnowicz 2011), but it is certainly less common in sociolinguistic studies than changes lead by women. Nonetheless, given the complexity of gender in this particular case, it is not safe to make any firm assumptions about men being the leaders of this change in Loja. As mentioned previously, it is difficult, if not impossible to disentangle physiological and learned gender differences, and in the case of Loja we would need evidence of the social evaluation of voicing in order to make any strong claims about who is leading this change. At this point in Loja, it is unlikely that there is any prestige or stigma associated with intervocalic /s/ voicing.
since it is not a feature that is explicitly commented on. As the sound change progresses, however, it is possible that this feature could take on an association with social class.

In conclusion, the model proposed here to explain intervocalic /s/ voicing in Lojano Spanish combines the gestural approach of Articulatory Phonology and the principles of sound change. It is evident that this process cannot be understood without considering the interaction of phonetic, phonological, and social factors. It is only at the nexus of these factors that we can arrive at a complete analysis of this feature. Intervocalic /s/ voicing in Lojano Spanish likely started as the effect of gestural overlap, but then was advanced even further by social changes in the community. As Lojanos begin to interact more with other communities in Highland Ecuador where this feature is present, intervocalic /s/ voicing is even more present in Lojano speech patterns. In this way, Lojano Spanish is not as far along in the process of sound change as Quiteño and Cuencano Spanish, or perhaps Lojano Spanish is on an entirely different trajectory than other HES varieties. For this reason, the effect of word position is different in Lojano Spanish than in Quiteño and Cuencano Spanish.

5.2 Perception of intervocalic /s/ voicing

In Chapter 4, the results of an online perception experiment were presented. This experiment consisted of six experimental blocks, each of which included six to eight training items. Blocks 1-3 are similarity rating tasks in which the participants heard two audio files and had to respond on a scale of 1 ‘very similar’ to 6 ‘very different’. Blocks 4-6 are discrimination tasks in which the participants heard two audio files and had to decide if they sounded ‘same’ or ‘different’. The different blocks correspond to stimuli in
different word positions. Blocks 1 and 4 contain stimuli with /s/ in word-medial position (e.g. [așa]), Blocks 2 and 5 consist of stimuli with /s/ in word-initial position (e.g. [la _saka]), and Blocks 3 and 6 are made up of stimuli with /s/ in word-final position (e.g. [laș _ata]). Twenty-four native Lojanos completed Blocks 1-3, and twenty-one of these same participants also completed Blocks 4-6. As a reminder, pairs that included two audios of the same production ([asa]-[asa]) are referred to as “identity pairs”, while pairs that included two audios of different productions ([asa]-[aza]) are referred to as “difference pairs”. For Blocks 1-3, Welch Two Sample t-tests and permutation tests were performed on the responses to target stimuli to determine whether there was a significant difference between mean response for identity and difference pairs, and a linear mixed effects model was fit to the data to examine the fixed effect of block and the random effect of participant. For Blocks 4-6, a two-tailed Pearson’s chi-squared test with Yates’s continuity correction was performed to test whether participants responded ‘different’ significantly more to difference pairs than to identity pairs, and a logistic mixed effects model was fit to the data to determine the effects of block and participant.

5.2.1 Summary of perception results

The results of the training items indicated that the participants understood the task at hand, and in the case of Blocks 1-3 understood appropriately how to utilize the 1-6 scale. In the case of the similarity rating tasks, Blocks 1-3, participants responded significantly higher on this scale to difference pairs than to identity pairs, exhibiting that they hear difference pairs as more different. This is most robust in word-final position (Block 3), somewhat less so in word-initial position (Block 2), and much less so in word-
medial position (Block 1). The linear mixed effects model for Blocks 1-3 shows a significant difference between responses for word-medial and word-initial position, and a significant random effect of participant. Nevertheless, the random effect of participant is likely to be due to the fact that participants had six options to choose from in this task, and therefore it is not likely that this effect tells us anything noteworthy about the results. For the discrimination tasks, Blocks 4-6, participants responded “different” significantly more to difference pairs than to identity pairs, confirming that they hear difference pairs as different more frequently than identity pairs. This is most robust for word-initial position (Block 5), followed by word-final position (Block 6), and least robust in word-medial position (Block 4). The logistic mixed effects model for Blocks 4-6 shows a significant effect of block (word position): the responses to word-medial position are significantly different than word-final and initial position, and the responses to word-final positions are significantly different than word-initial position. The random effect of participant in Blocks 4-6 is not significant.

5.2.2 Effect of word position

Comparing the experimental blocks, there is an effect of word position such that the difference between [s] and [z] is heard more readily in some blocks as compared to others. Combining the results from the similarity rating and discrimination tasks, it seems that the participants hear this difference most readily in word-final and word-initial position. Given that it is precisely in these two word positions that Lojanos voice intervocalic /s/ most frequently, we can see there is a correlation between production and perception. The participants who completed the online experiment would be hearing
word-final and initial /s/ being voiced more often in their everyday interactions with other Lojanos, which may allow them to hear [s]-[z] more readily in these word positions. On the other hand, since word-medial /s/ is voiced less frequently, the participants would be hearing less voicing in that context and thus they do not hear the difference between [s] and [z] in medial position as readily.

Another reason that the perception may show an effect of word position is that the Lojanos who took the online experiment are likely also in contact with speakers of Cuencano and Quiteño Spanish. Since these two subdialects voice intervocalic /s/ more often in word-final position, Lojanos who are exposed to these dialects may more readily hear [s]-[z] in word-final position because they have heard voicing in this environment when interacting with Cuencanos and Quiteños.

It is interesting that perception of the difference between [s] and [z] for Lojanos is different based on what word position the /s/ is in. While it is possible that this effect is an artifact of these particular stimuli, and thus the results should be corroborated with further testing, the reader will recall that there were two distinct stimuli for each [s]-[z] pair, and thus the results are not dependent solely on one set of stimuli. These results serve as preliminary evidence that in the case of allophonic variation, listeners are in some way attuned to the context in which they hear one variant over another. The effect of linguistic environment more broadly has also been explored in previous studies looking at sociolinguistic perception (Callier 2014; García et al., forthcoming). In his work on the social meaning of creaky voice in Mandarin Chinese, Callier (2014) finds that social evaluation of this feature is dependent on location within the intonational phrase, showing that listeners access contextual information when perceiving
sociolinguistic variation. Preliminary evidence from Puerto Rican Spanish also shows that listeners socially evaluate /s/ aspiration differently when it is heard in preconsonantal, prevocalic, and prepausal contexts (García et al., forthcoming).

5.2.3 Comparison to previous studies

Beyond the effect of word position, it is useful to consider how the results seen here for the perception of [s]-[z] compare to other studies looking at phonemic and allophonic perception in Spanish and other languages.\(^{21}\) First, I will consider how the present results are different than what we would expect for languages where /s/ and /z/ represent a phonemic contrast, or for other sounds that contrast phonemically in Spanish. In the first regard, previous studies looking at the perception of /s/ and /z/ in English report extremely high accuracy rates across participants (Cole & Cooper 1975, Jongman 1989, Stevens et al. 1991). For the complete details of these studies, see section 2.5.2. As a reminder, these studies all look at the effect of manipulation of acoustic characteristics on the perception of fricative voicing, and for this reason they are not fully comparable to the present study. Using an identification task, Jongman (1989) shows that participants correctly identify [s] 92% of the time even when the fricative is manipulated to have a duration of only 20 ms. Similarly, Stevens et al. (1991) find that participants judge [asa] as [s] and [aza] as [z] over 90% of the time regardless of fricative duration. While the tasks used in these studies are different than the present study, it is clear that the perception of [s]-[z] is markedly different in English and Spanish. For instance, in the

\(^{21}\) Due to their inconclusive nature, the results of Widdison (1995, 1996, 1997) will not be included in this discussion. See section 2.5.2 for details on these studies.
discrimination tasks (Blocks 4-6), participants’ accuracy (responding ‘different’ to difference pairs) is highest at 47.6% in Block 5. Thus, while native Lojanos hear [s]-[z] as different more frequently than [s]-[s], their perception of these sounds is much more variable than if the sounds were phonemic, unsurprisingly.

Further evidence for this point comes from studies looking at phonemic perception in Spanish. Martínez-Celdrán (1993) uses a categorical perception experiment to examine the effect of closure duration on the perception of /b/ and /p/ in Spanish. In the discrimination task, participants’ accuracy at the extreme values, the shortest and longest closure duration, reaches 100%. That is, there is no question that listeners clearly distinguish /p/ and /b/ in Spanish. Again, the perception of intervocalic [s] and [z] in Lojano Spanish is markedly different.

How then does the perception of [s] and [z] compare to other cases of allophonic perception? To answer this question we must reconsider the results of Boomershine et al. (2008), who also use similarity rating and discrimination tasks. Boomershine et al. (2008) examine the perception of [d, ð, r] by Spanish and English participants, although only the Spanish results for [d-ð] are immediately pertinent here. Even though the tasks used are the same as the present study, the results are not entirely comparable since these authors report normalized reaction times for the discrimination tasks and normalized similarity ratings. In Experiments 1 and 3, the similarity rating tasks, the mean normalized rating for [d-ð] pairs is roughly 0.5, where the value 0 represents neither similar nor different (the middle of the 1-5 scale). Accordingly, [d] and [ð] are perceived as different, but not

Martínez-Celdrán (1993) also looks at the perception of /pp/, however, those results are not relevant for this discussion.
as different as the pairs containing a phonemic contrast (e.g. [ð]-[ɾ]). Experiments 2 and 4, the discrimination tasks, exhibit similar results. While it is impossible to say whether [d-ð] or [s-z] is heard as more different in Spanish given the methodological differences between Boomershine et al. (2008) and the present study, the results of both of these studies, as well as those of Johnson & Babel (2010), confirm that subphonemic differences are heard to some degree.

In conclusion, similar to the perception of other allophonic pairs, the Lojano participants in this study perceive [s]-[z] as more different than [s]-[s], however, this perception is much more subtle and variable than in the case of phonemic perception. Furthermore, the fact that no participant explicitly commented on [s] or [z] in the open-ended survey questions suggests that the perception of intervocalic /s/ voicing in this dialect is not as explicit as other features. Still, the evidence provided here that Lojanos do hear to some degree a difference between intervocalic [s] and [z] opens up the possibility that this variable may carry social associations, which is a point future studies should consider. Finally, the effect of word position in the perception of this variation suggests that listeners to some degree pay attention to context when processing allophonic variation.

5.3 Conclusions

5.3.1 Status of intervocalic /s/ in Lojano Spanish

Considered together, the production and perception results presented here paint a broader picture of the status of this feature in Lojano Spanish. On the production side, we
have seen that intervocalic /s/ voicing in this dialect is a gradient and variable phenomenon, which can be explained through a gestural approach that also takes into account the principles of sound change. This variation in production cannot be understood unless we consider the complex interaction of phonetic, phonological, and social factors. For the perception of this feature, we can conclude that participants hear [s]-[z] pairs as more different than [s]-[s], and this perception is dependent on context, that is, word position. It is evident that there is a connection between production and perception, since [s]-[z] is most readily heard in word-final and initial positions, which is where we find more voicing in the production results. At the present moment, as intervocalic /s/ voicing is entering into the community of Loja, it is still somewhat under the radar of most Lojanos. They do not explicitly comment on this particular feature, and instead look to other sounds when describing what “the Spanish of Loja” sounds like. This does not mean, however, that intervocalic /s/ voicing is not associated with particular social groups or qualities, since language attitudes can be implicit and not easily put into words (Campbell-Kibler 2012). Future studies then should consider whether this feature is associated in Loja with certain social characteristics, such as sounding masculine.

In the panorama of the changes currently taking place in the society and speech of Loja, intervocalic /s/ voicing is only one of the many sounds that is on the move. In my fieldwork experience, I have also noted that the younger generations’ production of /r/ and /ʎ/ is changing. As mentioned in section 2.2, Lojano Spanish is traditionally known for having a normative trill /r/ and for being a lleista dialect, that is, preserving the
distinction between the phonemes /ʎ/ and /ɟ/ (cf. Lipski 1994). However, the youngest generations of Lojanos can be heard assibilating /r/ and merging /ʎ/ with /ɟ/. These linguistic changes are accompanied by social changes. As mentioned before, communication between Loja and neighboring highland towns has increased with the advent of transportation improvements, and now young Lojanos tend to orient more towards outside of the city than to the city itself. Indeed most of the university students that I interviewed expressed an interest in leaving Loja at some point in the near future to pursue further education or employment. For this reason, now is the ideal time to take this snapshot of Loja, as it is in the transitional period between the isolation of the past and the interconnectedness of the future. As these connections with other places are expanded even further, there is no doubt that Lojano Spanish will continue to change in important ways. Therefore, intervocalic /s/ voicing is not only one of a few distinctive features of Lojano Spanish, but it is one of the features that will be important to continue to monitor in the future as an index of how Lojano Spanish is changing.

5.3.2 Contributions

The results and analysis presented here speak to many different fields within linguistics. At the most basic level, this study fills a gap in the literature by providing the first systematic look at the production and perception of intervocalic /s/ voicing in Highland Ecuadorian Spanish. By combining diverse methodologies, I am able to provide a more complete analysis of this feature. Moreover, this study provides important gains to the field of dialectology since no previous study to my knowledge has systematically
examined *any* feature in Lojano Spanish. The assumptions of modern sociolinguistics and variationist studies have more often than not been based on communities in large cities, and thus these assumptions need to be confirmed with the study of small-scale communities and speakers of languages other than English. The fact that Loja is a smaller, conservative community allows for a different perspective to the question of how variation disperses throughout a community during a change in progress.

As mentioned in Chapter 2, one of the goals of this work is to show that the dividing line between phonetics and phonology is not as clear-cut as was assumed in the past, and that both phonetic and phonological factors may be important in understanding a particular process. In direct contrast to previous studies claiming intervocalic /s/ voicing is a categorical phonological process, the acoustic analysis presented here proves that at least in the case of Lojano Spanish, intervocalic /s/ voicing is gradient and variable, the result of gestural overlap and incipient sound change. Thus, this work adds to the growing number of studies in the field of laboratory phonology that aim to unify experimental results with linguistic theory.

Along these lines, this study contributes to the field of speech perception, by providing further evidence as to the differences between allophonic and phonemic perception. While production studies of allophonic variation are now commonplace, we still do not have a clear understanding of how and if native speakers perceive subphonemic differences. Following Boomershine et al. (2008) and Johnson & Babel (2010), the experimental perception results presented here confirm that listeners do hear subphonemic differences, but that this perception is dependent on the distribution of the sounds in the language and dialect. Furthermore, the present study adds to previous work
by showing that allophonic perception can be sensitive to phonological context. Finally, an important methodological contribution of this perception experiment is that despite the challenges of testing for allophonic perception (i.e. not being able to rely on orthography), it is possible to develop a sound methodology that is successful at capturing this.

Another goal of this dissertation is to push the boundaries of both sociolinguistics and phonetics. As many scholars working in the rapidly growing field of sociophonetics have warned (cf. Foulkes et al. 2012), this field must straddle the methodological aims of both sociolinguistics and phonetics, without doing disservice to either one. Unfortunately, all too often sociophonetic studies fall more towards one side of the fence or the other, giving preference to phonetics in some cases and sociolinguistics in other cases. While laboratory data was not used, the fine-grained acoustic analysis conducted here adheres to the high standards of phonetic studies on Spanish /s/, such as Schmidt & Willis (2011), File-Muriel & Brown (2011), and Campos-Astorkiza (2014). This is not at the expense of sound sociolinguistic methodology, however, as the data were collected in as naturalistic an environment as possible, and with social factors in mind.

The most substantial area in which this study pushes these boundaries is the statistical analysis. Previous studies in sociophonetics looking at diverse sounds have frequently limited their analyses to the statistical tools that are commonplace in modern phonetics or sociolinguistics, without questioning which tools are actually most appropriate for the data. Many studies on Spanish /s/ have only considered two variants (whether that be [s] and [z] for voicing, or aspirated and unaspirated for weakening) for the sole purpose of using traditional sociolinguistic tools (such as Varbrul) that only
allow for a binary response variable. On the other hand, even those more recent studies that utilize a continuous dependent measure (duration or percent voicing) do so at the expense of sound statistical analysis since many times the assumptions of linear regression (such as normality) are violated by the distribution of phonetic data. Here I adopt innovative statistical tools that are better suited to sociophonetic variation. For the continuous analyses, I utilize inflated beta regression, which is specifically designed to work with percentages (Stasinopoulos et al. 2015). In the categorical analyses, I employ ordinal logistic regression since it allows for more than two categories and assumes there is a hierarchical order to these categories, which is ideally suited to voicing category (unvoiced < partially voiced < fully voiced). Although these are the best tests for the dataset given its composition, the downside of using these tests is that the results are not directly comparable with previous studies that use other methods. However, utilizing sound statistical modeling moves the field forward and for this reason, it is worth the trade-off. As the field of sociophonetics expands, it should do so by challenging the traditions of both phonetics and sociolinguistics.

5.4 Future research

Future research building off of this study can be separated into two areas: work on Lojano Spanish, and the cross-dialectal study of intervocalic /s/ voicing. In the first regard, there is much work to be done in documenting the features of Lojano Spanish and how they are tied to social factors. As mentioned in section 5.3.1, other phonetic features of interest in Lojano Spanish are the assimilation of /t/, the loss of /ʎ/, and the velarization
of syllable-final /n/. Since Loja finds itself at a turning point, the detailed examination of these features will no doubt be incredibly fruitful not only in mapping how language and society interact in Loja, but also in shedding light about how sound changes occur in smaller, less metropolitan communities. For intervocalic /s/ voicing in Loja, the conclusions made here would be further supported by a larger sample of Lojanos from different social strata of the city. I was not able to look at the effects of factors such as education level because of the homogeneity of the participants in terms of social class. It would be interesting to see whether the same patterns seen here hold for other sectors of the population in Loja. Also, since the perception experiment here is limited to allophonic perception, future research should consider the social perception of intervocalic /s/ voicing in Lojano Spanish. Here we find that male Lojanos voice significantly more than female Lojanos, which could cause intervocalic /s/ voicing to be associated with male’s speech and masculinity. This feature may also index with regional identity within the highlands since each subdialect of HES realizes intervocalic /s/ in a slightly different way. This investigation would be particularly merited given that most sociolinguistic perception studies have focused on features that community members are readily aware of (Barnes 2013, Walker et al. 2014, among others), while studies of features that are more under the radar (like intervocalic /s/ voicing) are much less common (although see Campbell-Kibler 2012 on /t/ release). Apart from Loja, additional acoustic analyses of intervocalic /s/ voicing in Cuencano and Quiteño Spanish are needed in order to evaluate the claims of earlier impressionistic studies.

Given that previous studies of intervocalic /s/ voicing have been limited to Spain, Ecuador, and Colombia, more cross-dialectal comparisons are needed. The model
proposed here makes certain predictions in this regard. As detailed in section 5.1.4.2, my analysis assumes that intervocalic /s/ voicing started in Lojano Spanish as the result of overlapping glottal gestures, and this initial voicing was then amplified due to contact with other voicing dialects. Assuming the coordination of gestures does not differ between varieties of Spanish, this predicts that all modern dialects of Spanish should exhibit some degree of voicing of intervocalic /s/, which will be dependent on factors such as stress and speech rate. This has already been partially corroborated by the Colombian and Argentine data presented in A. García (2013) and Rohena-Madrazo (2011) since these authors find minimal voicing of intervocalic /s/ in these two varieties respectively. This hypothesis needs to be further tested by carrying out acoustic analyses of the production of intervocalic /s/ in dialects that are not known to have this feature, such as Mexican or Argentine Spanish. In addition, the model makes predictions about other factors that may condition the realization of intervocalic /s/. If this is indeed a reduction process, a usage-based model predicts that it should be manifested more in high frequency words than in low frequency words, since “those words that are used more often have more opportunity to be affected by phonetic processes” (Bybee 2001: 11). Future research then should take into account the effect of frequency.

A cross-dialectal comparison is also warranted in the area of perception. The assumption made here is that Lojanos hear [s]-[z] as more different than [s]-[s] because they hear intervocalic /s/ voicing on a daily basis. This predicts that listeners in dialects without this feature should not hear the difference between [s] and [z] as readily as Lojanos do. The effect of word position in the perception results also carries implications for other dialects. Here we find a correlation between production and perception in that
Lojanos produce intervocalic /s/ as voiced significantly more in word-final and initial position, and they most readily hear the difference between [s] and [z] in these same positions. To test this hypothesis further, it would be useful to replicate this perception experiment with listeners from Cuenca or Quito, where we would expect the most robust difference to be heard only in word-final position. This would be quite easy to do given that the perception experiment is hosted online. More broadly, further research in allophonic perception in all languages is needed to see whether this effect of word position, or more generally linguistic environment, is seen with other sounds. The investigation of allophonic perception also has important implications for work on social perception. Therefore, as mentioned before, a study of the social evaluation of intervocalic /s/ voicing is needed to confirm whether the effects of age and gender seen here are also seen in the perception of this feature.
REFERENCES


Bailey, Guy, Tom Wikle, Jan Tillery, and Lori Sand. 1991. The apparent time construct. Language Variation and Change 3.03.241-64.


Bradley, Travis. 2005. Sibilant voicing in highland Ecuadorian Spanish. Lingua(gem) 2.2.9-42.


Brown, Esther L., and Rena Torres Cacoullos. 2003. Spanish /s/: A different story from beginning (initial) to end (final). A Romance Perspective in Language Knowledge and Use. Selected Papers from the 31st Linguistic Symposium on Romance Languages (LSRL), Chicago, 19-22 April 2000.


Chappell, Whitney, and Christina Garcia. 2014. Intervocalic /s/ voicing in Costa Rican Spanish. The University of Texas San Antonio, MS.


Eckert, Penelope. 2012. Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. Annual Review of Anthropology 41.87-100.


García, Christina, Abby Walker, and Kathryn Campbell-Kibler. 2014. Effects of linguistic environment on the social meaning of /s/ aspiration. The Ohio State University, MS.


Hualde, José Ignacio, Miguel Simonet, and Marianna Nadeu. 2011. Consonant lenition and phonological recategorization. Laboratory Phonology 2.2.301-29.


Instituto Nacional de Estadística y Censos (INEC). 2006. La población indígena en el Ecuador.


Jessen, Michael. 2004. Instability in the production and perception of intervocalic closure voicing as a cue to bdg vs. ptk in German. Folia linguistica 38.1/2.27-42.


APPENDIX A: QUESTIONS USED IN THE INTERVIEWS

A. Información General
   • ¿Dónde nació? ¿En qué año nació? ¿Hasta cuándo vivió allí?
   • ¿Qué es su lengua materna? ¿Cuáles otras lenguas habla?
   • Cuénteme de los lugares donde ha vivido.

B. Niñez/Juventud
   • ¿Cómo era de niño? ¿Sus padres eran muy estrictos?
   • ¿Fue a la escuela pública o privada? ¿Era buen estudiante en el colegio?
   • ¿Dónde hizo su pre-grado? ¿Maestría?

C. Gustos
   • ¿Le gusta viajar? ¿Dentro de Ecuador? ¿Otros países?
   • ¿Qué es su equipo de fútbol favorito?
   • ¿Qué le gusta hacer en su tiempo libre?

D. Empleo
   • ¿Cuál fue su primer trabajo? ¿En qué trabaja ahora?
   • ¿En qué trabajan sus padres?
   • ¿Qué son los trabajos más comunes en Loja?

E. Pueblo/Vecindario
   • ¿Cuáles feriados se celebran en Loja? ¿Cómo se celebran?
   • ¿Por qué es conocida Loja?
   • ¿Cómo es la ciudad de Loja? ¿Cómo es su gente? ¿Cómo ha cambiado Loja a lo largo de su vida? ¿Qué ha sido el impacto de la nueva carretera?

F. Lenguaje
   • ¿Nota diferencias en cómo hablan los lojanos y cómo hablan los ecuatorianos de otras partes?
   • ¿Habla usted diferente que sus hijos/sobrinos? ¿que sus padres? ¿que sus amigos?
APPENDIX B: LIST OF STIMULI IN READING TASK

1. Dice que la yemita está allí.
2. Lo halló dos veces.
3. Dice que ya pasó el momento.
4. Dice que lo haya hecho.
5. Dice que la hienita está allí.
6. Cuando compre una de estas casas inolvidables, estaré feliz.
7. Repite abollar por favor.
8. La valla está allí.
9. Dice que es el único salmón que tienen.
10. Siempre cuando quiero que no hayas armado bronca, ya lo has hecho.
11. Lo ralló dos veces.
12. Siempre ha sido una buena muchacha.
13. La maya está allí.
14. Lo hallo dos veces.
15. Dice que está llenita siempre.
16. Siempre has ido cuando llego.
17. Dice que habrás armado bronca.
18. Siempre tiene el análisis adecuado.
19. A: ¿Dicen que es uno de ellos? B: Dicen que puede ser uno de ellos.
20. Dice que está llena siempre.
21. Dice que el chico sabe todo.
22. Llenita está la iglesia.
23. Yoga es su favorito.
25. Yemita repite dos veces.
26. Lo rayó dos veces.
27. Por favor no te olvides de los casos especiales, esto es importante.
29. Yogur es el mejor.
30. Cuando ya hayas ido, voy yo también.
31. Cuando mi hija no vuelve a casa, estoy preocupada.
32. Cuando no ha sido, yo lo sabía.
33. Por favor muéstrame los fotos de los manatís, quiero acordarme del viaje.
34. Ella lo halla dos veces.
35. Por favor no puede ser, es imposible.
36. La malla está allí.
37. Hierro no hay.
38. Dice que puede ser la cable que está mal.
39. Yacidio dijo dos veces.
40. Dice uno sospecha lo que ha pasado.
41. Por favor no dejes este desorden, arréglalo.
42. Hienita dijo dos veces.
43. Cuando está deshecho, me voy.
44. Llena está la iglesia.
45. Siempre a las siete y media tomo el desayuno.
46. Cuando ella salió, yo estaba trabajando.
47. La hiena está allí.
48. Siempre quiere que pase yo primero.
49. A: ¿Dice que parece a uno de sus hijos? B: No, dice que es uno de sus hijos.
50. Cuando termine el proceso, voy de vacaciones.
51. Repite aboyer por favor.
52. Cuando no desayuno, me siento mal.
53. Siempre termina en desastre.
54. Dice que tu cuarto está hecho un desorden.
55. A: ¿Es el trámite más complicado? B: No, es el proceso más complicado.
56. Llamo dos veces.
57. La yemita está allí.
58. Siempre me quedo con la memoria de estas casa inolvidables.
59. Siempre viene a mi casa para cenar.
60. Cuando no quieres que pase, dime nomás.
61. Lo ralló dos veces.
62. Yema dijo dos veces.
63. A: ¿Dijiste que son unas casas inolvidables. B: No, dije que son unos casos inolvidables.
64. Se cayó dos veces.
65. No está llenita todavía.
66. Cuando no lo cosí, ella lo hizo.
67. A: ¿Dijiste que tu cuarto está hecho un desastre? B: No, dije que toda la casa está hecha un desastre.
68. Siempre está deshecho.
69. Dice que se calló.
70. Lloró dos veces.
71. Cuando el chico no sabe, lo admite.
72. Cuando esté listo el otro saco, te llamo.
73. Siempre hay que pasar los autobuses.
74. Rayó el disco dos veces.
75. Siempre el proceso dura seis meses.
76. Dice que es un proceso muy largo.
77. Dice que es desigual que el otro.
78. Dice que estos son casos especiales.
79. Por favor no lo dejes desigual, así no me gusta.
80. Dice que el piso está mojado.
81. Cuando vemos un desastre así, queremos ayudar.
82. Siempre mi bisabuelo viene para Navidad.
83. Cuando tenga el análisis adecuado, te aviso.
84. Dice que los vasos son desechables.
85. A: ¿Dijo que el almuerzo es bueno? B: No, dijo que el desayuno es bueno.
86. Dice que los mejores somos nosotros.
87. Cuando puedes pasar, te aviso.
88. Por favor no pongas tus zapatos en el piso, están sucios.
89. Siempre tenía la sonrisa más brillante.
90. Callo dos veces.
91. Dice que está allá a la vuelta.
92. Llamo dos veces.
93. Cuando mi bisabuelo está aquí, yo estoy feliz.
94. Lluvia repite dos veces.
95. Por favor explicame el proceso, no lo entiendo.
96. Dice que la yema está allí.
97. Dice que vaya por favor.
98. Por favor no hables a este muchacho, no es la mejor idea.
99. Dice que el acuario tiene unos manatís únicos.
100. Lloro dos veces.
101. Por favor si tiene un problema con nosotros, avísanos.
102. Dice que hayas armado bronca.
103. Hiena dijo dos veces.
104. Dice que ella salió ya.
105. Por favor no me digas que pasó, no quiero saber.
106. Por favor no botas los vasos desechables, algunos vamos a usar otra vez.
107. A: ¿Dijiste que ella llegó? B: No, dije que ella salió.
108. Llamó dos veces.
109. Por favor no dejes que uno sospecha, la gente es chismosa.
110. Cuando caiga el único salmón, sabemos que estamos en problemas.
111. Dice que es un muchacho muy bueno.
112. La yema está allí.
113. Siempre cosí mi propia ropa.
114. Siempre me pongo otro saco con estos pantalones.
115. Llamado dijo dos veces.
116. Cuando habrás armado la mesa, vamos a almorzar.
117. Cuando veas la sonrisa, vas a saber.
APPENDIX C: STIMULI USED IN THE PERCEPTION EXPERIMENT

BLOCK 1/4

1. aɾa2-ada1
2. asa2-aza2
3. ara1-aða2
4. aða1-ada2
5. asa2-aza1
6. ara1-ada1
7. ada2-aða1
8. aza1-asa1
9. aða1-ara1
10. ada1-ara1
11. asa1-asa2
12. ara1-aða1
13. aða2-ara2
14. aza1-aza2
15. ada1-aða1
16. ada2-ara1
17. asa1-aza2
18. aða1-aða2
19. asa2-asa1
20. aða2-ara1
21. aða1-ara2
22. aza2-aza1
23. ara2-aða2
24. aza1-asa2
25. ada2-ara2
26. ada2-ada1
27. aza2-asa1
28. aða1-ada1
29. ara2-aða1
30. aza2-asa2
31. ada1-aða2
32. aōa2-ada2
33. asa1-aza1
34. ara2-ara1

**BLOCK 2/4**

1. la ōaβa2-la daβa2
2. la zaka2-la saka2
3. la daβa1-la ōaβa1
4. la daβa2-la daβa2
5. la saka2-la zaka1
6. la ōaβa1-la daβa2
7. la ōaβa1-la ōaβa2
8. la zaka2-la zaka1
9. la daβa2-la daβa2
10. la ōaβa1-la daβa1
11. la saka2-la saka1
12. la daβa1-la daβa2
13. la daβa2-la daβa1
14. la zaka1-la saka2
15. la daβa2-la daβa1
16. la daβa2-la daβa1
17. la saka1-la zaka2
18. la daβa1-la ōaβa1
19. la zaka2-la saka1
20. la daβa1-la daβa2
21. la daβa2-la daβa1
22. la saka2-la zaka2
23. la daβa1-la daβa1
24. la daβa1-la daβa2
25. la saka1-la saka2
26. la daβa2-la daβa1
27. la saka1-la zaka1
28. la daβa2-la daβa1
29. la daβa1-la daβa2
30. la zaka1-la saka1
31. la daβa2-la daβa1
32. la zaka1-la zaka2
BLOCK 3/6

1. tañ alto2-tan alto2
2. las ata2-laz ata1
3. coñ oro2-coñ oro1
4. con oro2-coñ oro1
5. las ata1-laz ata2
6. tan alto1-tañ alto1
7. las ata2-laz ata2
8. tan alto2-tañ alto1
9. coñ oro1-con oro2
10. laz ata2-laz ata1
11. tañ alto1-tan alto2
12. laz ata2-las ata2
13. tañ alto2-tan alto1
14. coñ oro2-con oro1
15. las ata2-las ata1
16. con oro1-con oro2
17. laz ata1-laz ata2
18. tañ alto1-tan alto1
19. con oro1-coñ oro2
20. las ata1-las ata2
21. coñ oro1-coñ oro2
22. las ata1-laz ata1
23. coñ oro2-con oro2
24. tan alto2-tan alto1
25. laz ata1-las ata1
26. tañ alto2-tañ alto1
27. tan alto2-tañ alto2
28. laz ata1-las ata2
29. tan alto1-tan alto2
30. con oro2-coñ oro2
31. laz ata2-las ata1
32. con oro1-coñ oro1