The Dissertation Committee for Stephanie Lain certifies that this is the approved version of the following dissertation:

Acoustic correlates of [voice] in two dialects of Venezuelan Spanish

Committee:

Orlando Kelm, Supervisor

David Birdsong, Co-Supervisor

Jean-Pierre Montreuil

Randy Diehl

Frederick Hensey

Acoustic correlates of [voice] in two dialects of Venezuelan Spanish

by

Stephanie Lain, B.A.; M.A.

Dissertation

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2009

Dedication

To my mentor and friend, Professor Jean-Pierre Montreuil.

Acknowledgements

I would like to thank the members of my dissertation committee, especially my supervisors, Professors Orlando Kelm and David Birdsong. Both were instrumental in bringing this project to fruition. In addition, my deep appreciation and thanks go to stats consultant, Michael Mahometa. His help was invaluable. Many people contributed to the successful completion of this project. I would like to acknowledge the speakers from El Tirano and San Rafael de Mucuchíes who participated in the study. Thanks also to Jhonny, Doña Nicolasa, El Capitán, and La Junta Parroquial de Muchuchíes for support in the field. Mil gracias to Sergio Carvajal and his friend Javier for their work as language consultants. I would like to extend thanks to the following friends and fellow graduate students for their helpful comments and/ or participation in the pilot experiment: Melissa Murphy, Jason Brazeal, Tim DiToro, Lori Czerwionka, Jessie Yang, and Mi Jang. Last, but never least, profound thanks to my family for their love and support, especially to Jer and my daughter, Cora, for helping me through the last year.

Acoustic correlates of [voice] in two dialects of Venezuelan Spanish

Publication No. __________

Stephanie Lain, Ph.D
The University of Texas at Austin, 2009

Supervisors: Orlando Kelm, David Birdsong

The present study is an investigation of acoustic correlates corresponding to the category [voice] in two dialects of Venezuelan Spanish. The Andean mountain dialect Mérida (MER) and Caribbean coastal dialect Margarita (MAR) are thought to differ systematically in the phonetic implementation of the Spanish phonological stop series along the lines of lowland and highland divides commonly reported for Latin American Spanish. Specifically, MER has been characterized by a greater percentage of occlusive pronunciations, MAR by more fricative and/or approximant realizations of phonological stops. To test what repercussions these differences in consonant articulation have on the acoustic correlates that encode [voice], a production experiment was run. Informants were 25 adult monolingual speakers of Venezuelan Spanish from the areas of El Tirano (Margarita Island) and San Rafael de Mucuchíes (Mérida state). The materials were 44

CV syllable prompts. Target syllables were analyzed with respect to the following: consonant closure duration, VOT, %VF, RMS, preceding vowel duration, CV ratio, F1 onset frequency, F0 contour, and burst. Statistical analysis using a linear mixed model ANOVA tested for fixed effects of *voicing category*, *dialect* and *condition* (speeded/unspeeded) and interactions of *voicing category* * *dialect* and *dialect* * *condition*. Results showed that the dialects MER and MAR vary significantly in RMS. In addition, the following correlates were significant for the interaction of *voicing category* * *dialect*: consonant duration, VOT, %VF, RMS, CV ratio and burst. Generally, the nature of the differences indicates a greater separation between [± voice] values in MER than in MAR (notably divergent are VOT and RMS). These results imply that while the same acoustic correlates of [voice] are operative in both fortis and lenis dialects of Spanish, [± voice] categories relate differently. Furthermore, with regard to prosody and rate of speech, most significant differences in *condition* occurred in initial position while most significant differences in the interaction of *voicing category* * *dialect* were linked to medial position. The results of this study are relevant to current research on the specifics of dialectal variation in consonant systems. They also have wider implications for the general mapping of phonetics to phonology in speech.

Table of contents

List of Tables

List of Figures

Chapter 1: [Voice] correlates and variability

1.1 Introduction

The feature [voice] is used contrastively in the consonant systems of most of the languages of the world. In the most basic interpretation, [voice] describes the state of the vocal folds during the production of a given segment. [+voice] denotes vibration of the vocal folds throughout the duration of the segment. [-voice] is the absence of such vibration.

In actuality, the phonetic implementation of these categories is rarely absolute (hence frequent mention in phonology of ‘partial voicing’). In perception, many acoustic cues (not only vocal fold vibration) contribute to the listener’s categorization of a consonant as voiced or voiceless (Denes, 1955; Diehl & Rosenberg, 1977; Lisker, 1978b; Haggard, Summerfield, & Roberts, 1991; Whalen, Abramson, Lisker, & Mody, 1993). In production, in addition to vocal fold vibration, there is a long string of acoustic correlates of [voice] in consonants, including (but not limited to): voice onset time (VOT), the presence or absence of a release burst, presence or absence of aspiration, duration of a preceding vowel, consonant vowel duration ratio (CV ratio), F1 onset frequency, F0 contour following closure, relative amplitude. A complete description of the acoustic correlates relevant to Spanish and an explanation of their contrastive importance will be provided in detail in subsequent sections of this chapter, as well as in Chapters 2 and 3. For now, suffice it to say that the phenomenon of voicing is manifold and its relationships complex (Ladefoged & Maddieson, 1996). For example, in one well-known paper, Lisker (1978b) details 16 cues that native English speakers in his study used to disambiguate /b/ and /p/. Each of the cues he measured was found to be robust enough in isolation to accurately allow subjects to distinguish the voicing category, given the absence of other cues.

1.1.1 Articulatory gestures and the contrastive use of [voice]

As mentioned in the introduction, 'voicing' in its strictest sense refers to the vibratory action of the vocal folds. Vocal fold vibration produces a regular energy, or periodicity, in the speech wave. It is usually said that voicing is present during the production of voiced consonants and absent during that of voiceless consonants. The gesture of voicing is controlled by the engagement of muscles in the larynx[1] that either hold the vocal folds in a narrow position (for voicing) or in an open position when voicing is not desirable. This is true in all languages. Acoustic regularities correspond to each of the positions-voiced or voiceless. These will be discussed in more detail in §1.3.1.

1.1.2 Laryngeal gestures also reflect fine adjustments

Despite the near-universality of the voicing contrast in languages and the commonality of the gross gesture that primarily defines consonant voicing in all languages that show a

---

[1] The term 'larynx' refers to the configuration of the thyro-arytenoid ligaments (vocal folds), the muscle-controlled thyroid and arytenoid cartilages, and the glottis (the empty space between the vocal folds), through which phonation occurs.

voicing contrast, the phonetic realization of [voice] does not work exactly like an on/off switch.

In fact, the initiation, maintenance and cessation of phonation all depend on a subtle interplay of articulatory factors. Hoole, Gobl, and Ni Chasaide (1999) assert that in this the phonatory system is fundamentally different from other sub-systems used in speech production. In addition to the muscle-controlled adjustments of the vocal folds, aerodynamic conditions of the glottis (particularly the transglottal pressure), the intrinsic elasticity of the folds and tension held in these by the muscles of the larynx are further contributing factors. Even the slightest change during phonation in any of these factors has the potential of altering the mode of vibration and hence the auditory quality of the sound produced. Under conditions of normal speech, coarticulation, the manner of articulation, prosodic position, the rate of speech and the voicing quality of surrounding segments further impact the quality of consonant voicing.

Many of these qualities will be codified within a particular language system and, along with other subtle nuances in articulation, will influence the quality of voicing, how voiced or voiceless consonants behave in different segmental environments, the nature of the category contrast itself (Ladefoged & Maddieson, 1996).

1.1.3 Language variation in the '2-way voicing contrast'

Ni Chasaide and Gobl (1993), in a review of five languages (French, Italian, Swedish, German, English), discovered considerable differences in the way [voice] contrasts were phonetically enacted, despite the fact that all languages surveyed maintained a '2-way voicing contrast'.[2] In particular, there were appreciable differences in the voice source characteristics for the vowel depending on the voicing category of surrounding consonants, both in terms of directionality (whether it was the preceding or following vowel that was affected) as well as in degree of effect. For example, in Swedish vowels, before the voiceless stop, the authors noted that throughout the duration of the speech signal, the source pulse for these vowels became increasingly weaker, showing more dynamic leakage and an increased symmetry. They remark that this was consistent with a gradually-increasing breathy mode of phonation that meshed with a weakening acoustic signal and loss of energy in the higher frequencies. Given that these effects are absent when the following stop is voiced, one can conclude that in Swedish, considerable information regarding the voicing status of consonants is transferred from the preceding vowel.

By contrast, the French data exhibited virtually none of the effects associated with the voiceless stops in Swedish. Vowels in the study had a constant spectral quality regardless of the upcoming stop. In addition, the duration of the vowel was fairly similar for both contexts and failed to display the differences observed in the Swedish data between [± voice] contexts. Lisker and Abramson (1970b) note:

"In many languages some phoneme categories are distinguished by the timing of glottal adjustments relative to supraglottal articulation, and this timing relation determines not

---

[2] Some languages have contrasts that involve stops with three different laryngeal settings, including Thai and Korean (Ladefoged & Maddieson, 1996). Spanish has a 2-way voicing contrast.

only the voicing state as narrowly defined, but the degree of aspiration and certain features associated with the so-called force of articulation as well" (p. 563).

This variability in timing and tension mechanisms appears to apply not only to the consonant in question, but also to surrounding segments.

In another study, Möbius (2004) undertook the identification and quantification of the major segmental, prosodic, and positional factors influencing the perception of consonant voicing in German. He compared his results for German with those of three other languages: Mandarin Chinese, Hindi, and Mexican Spanish. The voicing profile method he employed was a frame-by-frame report of the voicing status of speech sound realizations in a large corpus. His results show that for [+ voice] stops, the type of left context was the main factor affecting the voicing probability of the entire closure phase.[3] In the case of [- voice] stops, the overall shape of the voicing profiles remained unchanged across left segmental contexts. A vocalic or sonorant consonant context raised the probability of voicing for [p t k] by approximately 10-15% during the 1st half of the closure phase. For voiceless left contexts, the probability of voicing was practically zero. A weak right-context effect for both voiced and voiceless could be seen as well.

---

[3] 'Left context' here refers to the segment (and especially the voicing status) of the segment immediately to the left of the target consonant. The surrounding environment is a major contributor to the properties that can be attributed to a speech sound in any given utterance. Work in concatenative synthesis has shown boundaries between phonemes to be acoustically volatile. Largely for this reason, early attempts to string phonemes together without regard for the dynamics of transition were spectacular failures (Bhaskararao, 1999, [p. 71]).

In the Mexican Spanish data, there was good correspondence between the phonemic specification and correlated phonetic properties (see also Romero, Parrell, and Riera, 2007). The author notes that this differentiation was "perfect" near the end of the closure phase, but much less so at the beginning, where the effect of segmental context was significant. As was observed for the other languages in the study (German, Mandarin Chinese, and Hindi), voiceless stops were likely to show voicing that extended far into the closure.

Möbius found that the languages in his study behaved in a similar way in some aspects. For example, in all the languages he looked at, when the left context was voiced and the consonant phonemically voiceless, a noticeable degree of voicing was present during the beginning of the closure and sometimes farther in (see also Romero [1992] for Spanish). This is one phenomenon that may generalize to a large number of languages. Other aspects of voicing appear to be more language-specific.

Previous work on the voicing mechanism reinforces the notion that, while some aspects of the implementation of [voice] appear to be near-universal, or at least extremely common cross-linguistically (Hirose, Yoshioka, & Niimi, 1979; Löfqvist, 1995; Pickett, 1999), the particulars of maintaining the [voice] contrast are determined on a language-to-language basis (Kingston & Diehl, 1994; Hoole et al., 1999; Ni Chasaide & Gobl, 1993; Möbius, 2004). Furthermore, the realization of voicing is strictly tied to articulatory gestures related to tension and timing, gestures that also come into play in consonant closure.

1.1.4 Implications for dialectal variation

Up to this point, I have commented on research from previous studies that argues that, 1) the phonetic realization of voicing is closely tied to the physical gesture and is perturbed by small adjustments to the tension and timing of articulators, and; 2) the tension and timing of consonant articulation in relation to voicing is language-specific. Given this, what are the implications for dialectal variation in consonant systems? If a language is known to display substantial variation in the way consonants are articulated, and if this variation is predictable by dialect, will not the consonant voicing system be affected as well? This is the main question guiding the current research project.

1.1.5 The case of Spanish: highland and lowland dialects

Spanish is one language where marked differences in consonant articulation are known to exist, depending on the region where the language is spoken. In this case, the difference is of a fortis/ lenis nature. In highland (fortis) varieties of Spanish, occlusive realizations of phonological stops are more prevalent and in lowland (lenis) varieties of Spanish, fricative or approximant realizations prevail, independent of context.

1.1.6 Margarita and Mérida dialects in Venezuelan Spanish

These differences are seen in many different dialects throughout the Spanish-speaking world. I have chosen to focus on two dialects that I am personally familiar with- the Spanish from Margarita Island (situated on the Caribbean coast of Venezuela) and the

Spanish from the region of Mérida in the Venezuelan Andes.[4] These dialects exhibit similar characteristics to other highland and lowland dialects that have been described in the literature (Lipski, 1994). The variability of acoustic correlates of [voice] in these two dialects of Venezuelan Spanish is the focus of the present study. In particular, I consider the correlates of [voice] as they relate to the consonantal gesture. The first dialect, spoken in the Andes Mountains region (Mérida), is known for its strong consonant closure. The other dialect, spoken on Margarita Island, is a coastal variety of the type known for frequent fricative and/or approximant realizations of phonological stops (Lipski, 1994). More details on what is known on the particulars of these dialects will be provided in the literature review in Chapter 2.

1.2 Research questions and expectations

The main research questions to be addressed in the current study are as follows:

- What robust acoustic correlates to [voice] emerge in the two dialects under consideration (Mérida and Margarita- henceforth, MER and MAR)?
- Are the acoustic correlate inventories substantially different in the two dialects?
- Assuming differences are found, what are the implications for [voice] when consonantal parameters of articulation differ?

Secondary research questions include:

- Do observable differences between the two dialects hold across conditions of

---

[4] Mérida is the name of both a city and a state. The data were actually collected in a small town that lies a few hours outside the city by car. More details are given in Chapter 3: Methodology.

prosody and rate of speech?

- Are the observations for dialect consistent across all places of articulation?

The following measurements will be taken: consonant duration, VOT, percent voicing throughout closure (%VF), RMS amplitude (RMS) of the CV window, preceding vowel duration, consonant/ vowel duration ratio (CV ratio), F1 onset frequency, F0 contour following closure (F0 contour), presence/ absence of release burst. The choice of these measures over others that could have been included for analysis was based on three factors: 1) previous demonstrated importance in the literature on voicing in Spanish (consonant duration, VOT, %VF, F1 onset frequency, presence/ absence of release burst); 2) previous demonstrated importance in the general literature on voicing (RMS, preceding vowel duration, CV ratio, F0 contour), and; 3) appropriateness to the study. For example, degree of aspiration, an appropriate and useful measure of [voice] in other languages (including English), is not helpful for Spanish, given that Spanish stops are categorized as either voiced or voiceless unaspirated.

The contexts to be measured in this study are word-initial and word-medial. Speeded and unspeeded conditions are also assessed.

1.2.2 Expectations

1.2.2.1 General expectations

One expectation is that at least some of the target consonant productions for both dialects will not be stop consonants at all, but rather fricatives or approximants, as has been suggested in the literature (Dent, 1976; Trujillo, 1980; Canfield, 1981; Zamora & Guitart, 1982; Romero, 1992; Lipski, 1994; Hualde, 2005). Based on Trujillo's observations for Canary Island Spanish, which supposedly also shares traits with lowland Spanish varieties reported in the Caribbean, and in particular with Venezuela in terms of historical association, it is predicted that MAR will show a higher incidence of fricative and/or approximant values than MER.

Generally speaking, [voice] production in fricatives is similar to that in stops: the vocal folds are held wide apart during the constriction interval for voiceless fricatives and close for voiced fricatives. The vocal folds often vibrate during voiced fricative constrictions, but not always. Many times, the degree and nature of vocal fold vibration in fricatives depends on the position in utterance. One difference is that the amplitude of the devoicing gesture is usually larger for fricatives than for stops (Löfqvist, 1995; Hoole et al., 1999), a phenomenon that correlates in turn with closure duration. Therefore, closure duration may be a more robust correlate of [voice] in fricatives, and less of one in stops (Cole & Cooper, 1975).

In a discussion of cross-linguistic differences in the production of fricatives, Ni Chasaide and Gobl (1993) note that the timing patterns for fricatives can be different from those found in stop consonants. In Swedish, they found that the timing of glottal abduction was very similar in stops and fricatives and very early for both. In French, however, early abduction was observed only in fricative realizations. In Spanish, it is not yet clear what acoustic differences may be observed between the phonation patterns of stops and fricatives.

1.2.2.2 Expectations by measure

1.2.2.2.1 Consonant duration

It is expected that overall consonant durations will be longer in MER than in MAR. [-voice] durations are likely to be longer than [+voice] durations within MER, as is consistent with the literature across languages. However, it is possible that these divisions may not be as sharp in MAR, given that more voicing is expected in both [±voice].

1.2.2.2.2 Voice onset time

In the early studies of stop consonant voicing for Spanish, VOT was established as an important parameter in determining boundaries between voiced and voiceless categories. The majority of these studies were conducted for stops in initial position, in careful speech, spoken by educated speakers. Based on these results, I predict that VOT is likely to be an indicator for [voice] for stops in initial position and that it is more likely to be a defining correlate in MER than in MAR.

1.2.2.2.3 Percent voicing throughout closure

Given that a higher incidence of fricative or approximant realizations for /b d g p t k/ is predicted for MAR, we might expect a higher %VF for MAR than MER, particularly in the [- voice] category. The basis for this assumption emerges from the observation that on an articulatory continuum, as one moves from stop consonants (maximal constriction) on the one end to vowels (minimal constriction) on the other end, sonority will increase. We can expect any sonority increase to be reflected in the relative %VF and RMS measures for the two dialects. Additionally, [+ voice] segments are less likely to be affected than [-voice] segments. Since all target segments occur in a context where they are surrounded by voiced segments, voicing (once started) is unlikely to be interrupted. Therefore, the most reasonable expectation is that %VF will be constant across all [+ voice] segments and will vary only in the [- voice] context.

1.2.2.2.4 RMS amplitude

Amplitude measures reflect relative energy in the signal. Vowels being the loudest of speech segments, followed by approximants > fricatives > stops, we can expect RMS values to be higher in [-voice] for MAR than for MER (no difference in [+voice]).

1.2.2.2.5 Preceding vowel duration

I assume that preceding vowel durations will be longer for [+voice] than for [-voice] in both dialects. Based on the assumption that MAR will show more %VF in [-voice], I expect the category differences to be somewhat blurred with respect to the duration measures.

1.2.2.2.6 Consonant/ vowel duration ratio

Given the predictions for consonant and preceding vowel durations mentioned above, it is anticipated that CV ratios will fluctuate more by [voice] in MER than in MAR.

Furthermore, CV ratio may well be a more reliable measure across speeded and unspeeded conditions than the raw duration measures.

1.2.2.2.7 F1 onset frequency

As was mentioned in the introduction, voicing begins earlier in voiced stops relative to the moment of release. For this reason, we should expect to see lower F1 values in [+ voice] than in [- voice] across dialects. If a greater articulatory tension associated with [-voice] stops (Pickett, 1999) contributes to higher F1 and F0 values for these segments, we may expect the dialect with more stop consonant closure (MER) to also show higher values for F1 and F0. This assumes the existence of a correlation between cavity constriction and frequency. Such a correlation was demonstrated for vowels in a classic paper by House and Fairbanks (1953), where the authors found that duration, fundamental frequency and relative power of vowels were correlated with vowel height (size of cavity opening + tongue height), the voicing attributes of the surrounding consonants, and the manner of articulation.

1.2.2.2.8 F0 contour following closure

One of the common correlates of [- voice] is a higher pitch. As noted in §1.2.2.2.7, this may relate to a higher articulatory tension associated with [- voice] stops (Pickett, 1999), to a higher larynx position (as noted by Westbury, 1983) or be related to the *low frequency property* (Stevens & Blumstein, 1981; Kingston & Diehl, 1994), whereby several acoustic properties combine to boost the [voice] percept. Higher F0 contour values are expected for [- voice] than for [+ voice] in both dialects. If it is the case that greater tension produces the F0 effect associated with higher F0 values for [- voice], we may also expect higher overall F0 values for MER than for MAR (according to the same reasoning given in §1.2.2.2.7 above). If, however, the low frequency property is responsible for an observed difference between MAR and MER, higher values will be seen in [- voice] as compared with [+ voice], but there will be no observable difference between dialects.

1.2.2.2.9 Presence/ absence of release burst

Bursts more often accompany the release of [-voice] consonants. This is related to the greater air pressure build-up that precedes the release of a voiceless stop. A higher proportion of bursts is expected for [- voice] than for [+ voice] across dialects. Since bursts only accompany the release of stop consonants and are not relevant for fricatives and approximants, a higher index of burst activity overall is predicted for MER.

1.2.2.3 According to prosody and rate of speech

Both prosodic position and rate of speech have been found to interact with consonant voicing. For example, Yeou, Honda, Maeda, & Embarki (2007) recently found that speech rate and word boundary had an effect on laryngeal abduction-adduction gestures and on laryngeal-oral coordination in Moroccan Arabic. In their study they used photoglottography to examine laryngeal behavior in different contexts. In particular they found that speech rate and word boundary conditions governed the alignment of the peak glottal opening (whether this occurred during the fricative or the plosive portion of /s/ + glottal stop) as well as the total number of laryngeal gestures (one versus two peaks).

It is expected that acoustic correlates of [voice] will vary within dialect according to prosodic position. These differences may relate to degree of tension held in the vocal folds and in the articulators during and after the pronunciation of the consonant. Speech sounds in initial boundary positions are more emphatic and tend to provide more contrast. For example, Hirose et al. (1979) found that for Japanese voiceless stops in word-initial position, there was more observable PCA activity and a larger glottal width than for stops in medial position.[5] If Spanish stops behave in the same way, we may see the acoustic repercussion of this in measures of duration (longer for [-voice] initial, shorter for [+ voice] or medial) and RMS amplitude (lower for [-voice] initial, reflecting periods of low or no energy in the signal, higher for [+voice] or medial).[6] In addition, the force of the release burst in Spanish has been shown to relate closely to position in utterance (Torreblanca, 1983).

Much work in phonetics has centered around how partitions are elaborated in human speech. Category contrasts are examples of one kind of partition, prosodic boundaries are another. Cho & McQueen (2005) looked at language-specific phonetic

---

[5] PCA stands for posterior cricoarytenoid, the abductor muscle responsible for pulling the vocal folds apart and creating voicelessness during production of a speech sound.

[6] RMS (or root-mean-square) amplitude is a time-domain operation that can be used to provide information on signal loudness. It is calculated by squaring all the values, averaging them, and taking the square root of the average, producing a single value for the signal it is applied to.

enhancement across prosodic environments. The authors compared VOT measures in [t s d z] in Dutch and English. What they found is that VOT as a prosodic boundary measure is manipulated differently in Dutch than in English. Their data show that English stops are generally produced with longer VOTs in prosodic locations corresponding to higher syntactic boundaries. In the Dutch data, however, [t] was produced with shorter VOTs in all such locations. The authors comment:

> “This prosodic shortening leads to a question about the relationship between prosodically conditioned strenghthening and contrast between /t/ and /d/, i.e., whether the hyperarticulation associated with prosodically strong locations may result in an enhancement of paradigmatic contrast...” (p. 148)

What is interesting is that the hyperarticulation resulted in opposite reflexes of VOT in the two languages, Dutch and English, for the segment [t].

Differences in the phonetic implementation of the phoneme inventory across languages or dialects will also vary according to boundary position. For stops in the present study, this could mean a greater release burst and longer VOTs for both voicing categories in higher domain positions. For fricatives, we might expect the contrast to be encoded in larger differences between raw durations, duration ratios, and RMS amplitude measures of each category. What we might expect to see, then, would be a kind of nesting effect where any differences observed between the two dialects would be mirrored at a smaller level by differences according to prosodic position.

1.2.2.4 According to place of articulation

Based on work in stop consonant perception by Pérez (1998), it may be possible to see differences across correlates with respect to the different places of articulation (bilabial, alveolar, velar). In his study, Pérez found that the presence of low-frequency energy had no major influence on the discrimination of /b d g p t k/. Consonant duration did have influence, but the effect was found not to be equal across all places of articulation. Duration had the greatest effect on distinguishing /b/ from /p/ (73%), then /d/ from /t/ (46%). However, it had practically no influence in distinguishing /g/ from /k/. It is not clear how place of articulation would vary across the other measures. I would expect any observed differences to remain consistent across the two dialects.

1.3 Research questions in context

1.3.1 General voicing mechanism

In §1.1.1, voicing was discussed as an engagement of muscles in the larynx that either allows the vocal folds to come together in a narrow position (for voicing) or contracts to enlarge the glottal cavity when voicing is not desirable. Hirose et al. 1979 used electromyographic (EMG) and fiberoptic data to investigate the patterns of adductor and abductor activity during consonant production. They found that the abductor PCA (posterior cricoarytenoid muscle) was suppressed for the voiced portion of consonants and that the adductor IA (interarytenoid muscle) increased in activity during the voiced portions and decreased for the voiceless portions. It was observed that there was a separation of the arytenoids and widening of the glottis for the voiceless portion of the test utterances in their study that included voiceless stops and fricatives, as well as geminates and devoiced vowels. The maximum glottal width was larger when the peak PCA activity was higher. It was also found that the PCA activity was higher and glottal width larger for voiceless stops in word-initial position. In addition, Hirose et al. 1979 found that action of the CT (cricothyroid) could contribute to the increase in vocal fold tension. They concluded that it is plausible to consider the relatively high CT activity in the production of voiceless consonants as one possible contributor to the enhancement of the quality of voicelessness.

Additionally, it is possible for vocal fold vibration to take place without muscular involvement. Kingston and Diehl (1994) note that vocal fold vibration depends on an outward flow of air, which originates from a minimum difference between the subglottal air pressure and the intraoral air pressure. Westbury (1983) points out that, in any given intervocalic segment and assuming the pressures above and below the glottis vary as a function of time, the vocal folds will continue oscillating as long as the pressure drop across them is greater than 2000 dyn/cm^2 (p. 1323). The author notes, “This interval of voicing is due almost entirely to compliance of tissues surrounding the supraglottal cavity, and closely approximates the not uncommon 65-75-ms closure duration for medial /b/ in American English.” (p. 1323)

Furthermore, Westbury (1983), in a study on the articulators involved in consonant closure in English, found that the cavity volume, as determined by positions of the larynx, soft palate and portions of the tongue, was more relevant to voicing maintenance during consonantal closure than were the direction and extent of movements of any single articulator (p. 1331). Voiced stops in the Westbury study were always accompanied by an increase in the volume of the supraglottal cavity. By contrast, Westbury noted that a decrease in cavity volume corresponded to some voiceless stops.

Whether by action of the specific muscles involved in voicing or regulation of the subglottal pressure through manipulation of the supraglottal cavity, voicing tends produce common some acoustic correlates across languages. Pickett (1999) notes that in the case of the voiced consonant, vocal fold vibration continues for some time during the closure and the burst on the release is short and weak. In the case of the voiceless consonant, there is usually no vocal fold vibration during the closure and the burst on the release is strong and of longer duration. F1 energy does not appear until the beginning of the following vowel.[7] The relative difference in position of the folds between voiced and voiceless consonants also tends to produce other acoustic differences. Some differences discussed in Pickett (1999, p. 125) are:

1) Duration of the closure interval is usually slightly longer for voiceless than for voiced stops.
2) The position of the larynx is higher for voiceless than for voiced consonants. This tends to make the mouth pressure higher and to stretch the vocal folds, producing a slightly higher pitch in adjacent vowels.
3) A higher tension may consistently exist in one or more of the articulating factors of voiceless stops and an enlargement of the mouth cavity may occur during the occlusions of voiced stops.

---

[7] An important difference between [±voice] stops in CV syllables emerges from the observation that since voicing begins earlier in voiced stops relative to the moment of release, a greater portion of the transition should be periodic. Consequently, the onset frequency of voiced F1 is predicted to be lower in voiced stops (Harrington & Cassidy, 1999, pp. 91-92).

The next section discusses how the acoustic correlates seen in production may correspond to perception and the categorical encoding of [voice].

1.3.2 Relating acoustic correlates to [voice]

In the previous section I reported on literature intended to show that the articulatory gestures conditioning the contrastive use of [voice] in languages translate into acoustic phenomena that are observable in spectrogram and waveform displays of recorded speech. Many of the measures used to differentiate [± voice] are duration measures, reflecting either a straightforward duration (as in consonant vowel duration, preceding vowel duration) or indirect measures that reflect the relative timing of gestures (for example, VOT). Other acoustic measures target the relative pressure associated with [± voice]- onset frequencies, for example, and contour information.

Of all the acoustic correlates of [voice], perhaps the best-known and most-reviewed is VOT. VOT gained recognition as an important acoustic correlate and cue for [voice] through the work of Lisker and Abramson (Lisker & Abramson, 1970[a,b], 1972; Lisker, 1978[a,b]). Lisker describes VOT as a purely acoustic measure, reflective of but not directly related to the laryngeal and supraglottal events involved in voicing. He notes that VOT is practical for acoustic measurement in that the burst onset is easily located by eye in both the spectrogram and the waveform. In other acoustic measurements, onset and offset transitions can be difficult to discern.

Other acoustic correlates thought to be important in defining a voicing contrast in languages include F0 contour and F1 onset. It has been observed that a falling fundamental frequency usually occurs after a voiceless stop, while a flat or rising F0 usually accompanies voiced stops (Haggard, Ambler, & Callow, 1970; Whalen, Abramson, Lisker, & Mody, 1993). This is known as F0 ‘perturbation’. In perception, perturbation effects have shown that an ambiguous VOT is more likely to be heard as voiceless when the F0 is falling after the onset of voicing than if it is flat or rising. Another important spectral correlate is F1 onset. F1 onset has been established as a more reliable predictor of voicing onset than either F2 or F3. Francis, Ciocca and Yu (2003) found that F1-based measurements of voicing onset were more accurate and less variable than their F2- and F3- based counterparts when compared with time-synced glottal openings in electroglottographic data.

The current project refers to acoustic correlates of [voice] in Spanish, not cues- that is, to the production side of the issue, not the perceptual side. Acoustic correlates are sound repercussions of the physical gestures associated with any particular speech sound, as viewed in a spectrogram + waveform display and measured with the aid of speech analysis software. Observations on how these repercussions tend to group together in language systems can provide information on the way sounds are contrasted in speech.

The perceptual companion to an acoustic correlate is the cue. All acoustic correlates are also potential cues. Any acoustic information that can be perceived may be co-opted to provide information on the sound category. In natural speech, it is thought that cues generally work in tandem. That is, there is some degree of cross-referencing or redundancy in the signal. In natural speech, several cues are present at once. Perceptual studies have shown some acoustic cues to be more auditorily robust than others. For example, VOT has been shown to exist in tandem with thresholds already present in the auditory systems of humans and other animals (Liberman, Harris, Kinney, & Lane, 1961; Kuhl & Miller, 1975; Kuhl, 1981; Sinex, McDonald, & Mott, 1991; Kluender & Lotto, 1994). Other cues work as enhancers, increasing discriminability. For example, Diehl and Kluender (1989) have noted that certain acoustic cues have an optimal perceptual effect only in the presence of other cues. In consonant voicing, the length of the preceding vowel influences the perception of consonant closure duration. A long preceding vowel makes short closure intervals appear shorter. Shorter vowels make them appear longer.

One of the most interesting aspects of language is the trade-off between biology and learned behavior. Despite biological and/or ambient advantages conferred by particular cues, there is nonetheless substantial variation in the gestures speakers use to enact stop consonant categories across languages. With regard to speech perception, Diehl and Kluender (1989) note, “Cross-language differences in the number of categories and in their perceptual extension mean that, although the auditory-phonetic space may be largely given, its functional partitioning is not” (p. 136).

In this dissertation I focus on the production of phonemic stop consonants in two Spanish language systems. Understanding that production and perception phenomena are intricately linked, from time to time it will be necessary to refer not only to the production data, but also to perceptual studies on stop consonant voicing.

1.3.3 Mapping between phonetics and phonology

1.3.3.1 General principles

Given that languages are diverse in the phonetic encoding of what is a near-universal cross-linguistic category ([voice]), what is important to understand about how certain sounds (or combinations of sounds) can be mapped to a symbolic representation? This is a huge question and mostly beyond the scope of the present investigation. Nonetheless, in the present study, I rely on both phonetic and categorical concepts at every level of implementation- from the framing of the research questions, to experimental design, to reporting on and interpreting the results. Therefore, I feel obligated in this section to lay out a basic orientation to the mapping between phonetics and phonology. I will, in addition, provide guidelines as to how each concept (phonetics and phonology) will be discussed throughout the remainder of the dissertation.

There is much that is not yet clear about how speech is learned, how mental representations are stored in the brain, or how phonetic awareness contributes to the establishment of phonological categories (Studdert-Kennedy, 1987; Poeppel & Monahan, 2008; Holt & Lotto, 2008). Nonetheless, by looking at studies across disciplines, it is possible to gain some insight into this complex topic. For the purposes of this paper, the terms *phonetics* and *phonology* are discussed as two aspects of speech processing that are in some ways linked and in other ways diverge. *Phonetics* refers to the handling of physical detail in the production and perception of speech sounds. This detail is directly observable via technologies such as spectrograms, air pressure masks, palatography, glottal illumination, etc. *Phonology*, on the other hand, deals with the perception and classification of behavioral patterns in speech. As a concept, phonology is more abstract, more representative of symbolic thinking. The consequence of this is that the evidence for phonological behavior is less direct and more difficult to access. Phonology refers to categorical behavior. In a discussion of voicing, the phonetics of voicing are the articulatory gestures, correlates and cues of the target and surrounding segments. The phonology of [voice] is what information is taken from these to be meaningful in maintaining contrast between words. Phonology is generally established through perception. Holt and Lotto (2008) demonstrate this link in the following way:

> “Speech sounds are grouped by functional significance within a language; for instance, /l/ and /r/ are distinct in English but not in Japanese. Experience with these regularities tunes perception such that identical acoustic signals may be perceived differently by listeners with different language experience. These changes are thought to reflect functional grouping of speech sounds as categories.” (p. 44)

It is tempting to imagine the mapping between phonetics and phonology as one-to-one, since, after all, the features of phonemes (a concept in phonology) tend to be described in terms of physical or pseudo-physical attributes ([± voice], [± back], [± continuant]). However, the issue is more complicated. Ladefoged (1992) notes that while the interface between phonology and phonetics is,

> “...primarily defined by the physical definitions of the features... [This] is not always the case; sometimes certain sounds are grouped together in ways that cannot be justified by reference to a single physical property, or even a group. It might just be the result of historical circumstances that some sounds, which no longer share any particular phonetic defining characteristics, nevertheless still pattern together.” (p.165)

1.3.3.2 Evidence from neurophysiology

Research in phonology has taken an interesting turn with the advent of brain imaging technologies. Poeppel and Monahan (2008) report on current research in speech involving the use of neurobiological techniques to investigate speech perception and processing. In terms of regional activation during processing, the authors point out that there is no single cortical region that can be argued to be principally responsible for speech perception. The circuitry between areas specialized for language appears to be tightly integrated and at the same time spread throughout whole areas of the brain. This is one key way in which speech perception differs from visual perception. Face-recognition research, for example, has shown that one particular cortical field- the fusiform face area- most likely plays a disproportionately large role in visual perception. No such claim has been made for speech perception.

Ravizza (2001, p. 96), in a review article of impairments to voicing associated with damage to selective areas of the brain, provides a schematic of the differentiation between phonological, phonetic, and auditory/motoric levels of processing and related areas of the brain where each type of processing is thought to take place (p.96).

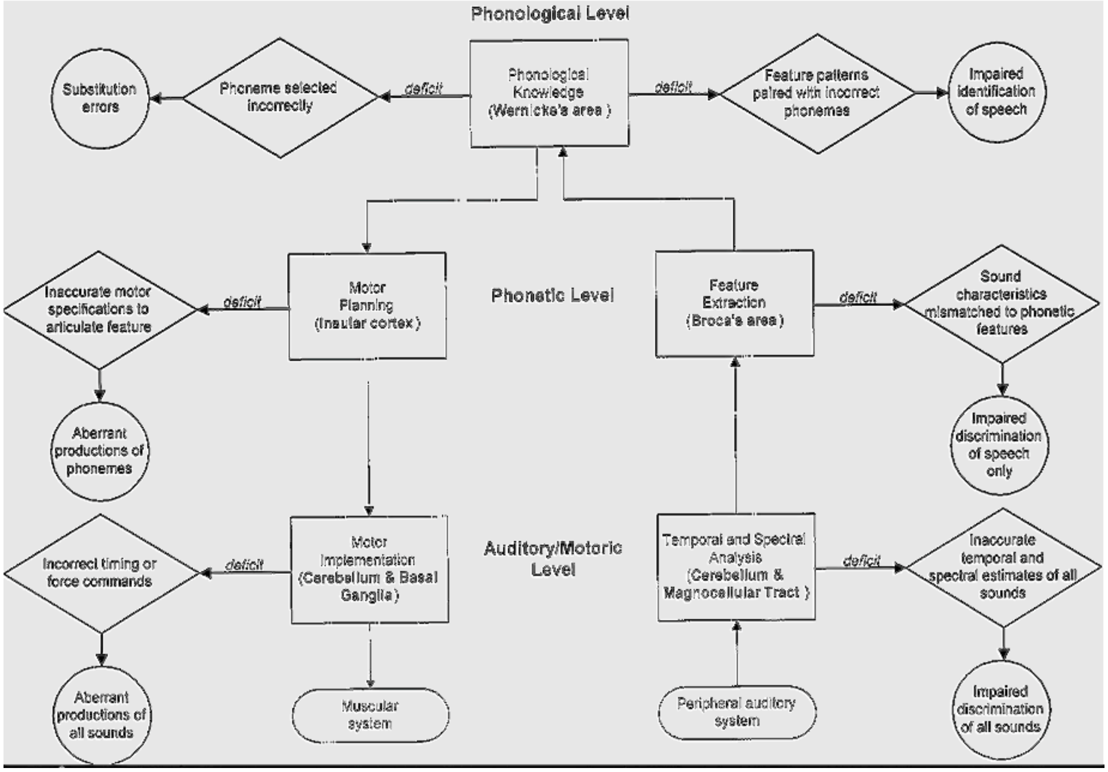


FIG. 1. Neural Areas Affecting Performance at Each Level of Phonemic Processing (taken from Ravizza [2001], p. 96.).

In Figure 1, the information is organized so that the left side of the diagram represents impairments to production and the right side impairments to perception. The bottom level of the model (auditory/motoric) is involved with the production and perception of all sounds and is non-specific to speech. More of interest here are the phonetic and phonological levels. Ravizza notes:

> "The phonetic level is specifically linked to speech and consists of the computations needed to produce or perceive critical linguistic features that comprise phonemes...[T]he ability to match feature bundles to the phonological store or to select the appropriate

phoneme to produce the correct word are claimed to be phonological-level computations.” (p. 96)

Groups impaired at the phonetic level may have deficits in planning speech gestures or extracting distinctive features from the incoming signal. Lesions to the insular cortex are hypothesized to primarily affect the specification of accurate motor commands for speech sounds. Damage to Broca’s area may be involved in the matching of phonetic features to sound characteristics whereas lesions that include Wernicke’s area are thought to result in phonological processing deficits consisting of incorrect phoneme selection and identification.

1.3.3.3 The phonetic/ phonological interface

In this paper, I will have occasion to refer to both phonetic and phonological levels. By *phonetic*, I’ll mean the physical attributes of a sound, ascertained through instrumental analysis. By *phonological*, I’ll mean the category value, or how the sound patterns within the language, without regard to the physical specifications. I assume the category *phonological stops* patterns the same way in each dialect, in the absence of evidence to the contrary. It is most likely that speakers from both dialects in this study share a common phonological store, a set of manipulable segments that combine in a way that is common to the language as a whole. What I hypothesize differs between the two dialects are the feature bundle specifications associated to the phoneme categories for [± voice] stops.

When discussing the feature bundle specifications themselves, and in particular the properties of [± voice] and [± continuant] as relevant features, the ground becomes less sure. There are strong indications that the sub-specifications of these features render their umbrella nomenclature in traditional linguistics misleading or even inaccurate. Historical sound change progressions from stop -> fricative -> approximant are not uncommon cross-linguistically. Ladefoged (1992) remarks that the gradient nature of such changes is not easy to explain in terms of binary features. In the context of this citation, Ladefoged refers specifically to voiced stops in Danish. Matters become potentially more complicated if there is an interaction between feature specifications of the type I hypothesize here- an interaction between manner of articulation and voicing.

Where, then, does the discussion of phonetics end and discussion of phonology begin? Phrased in other terms, how does knowledge of phonetics at a dialectal level infuse the phonology of a language? This is the domain of the interface.

Ladefoged (1992) notes:

"...it is by no means apparent that we can describe the ways in which the sounds of one language are distinguished from another simply in terms of the features that are required for distinguishing lexical items, or for accounting for phonological universals, or for grouping sounds into the natural classes that occur in rules. There are many instances of small but reliable phonetic differences between languages that have not been found to be used for contrasting words within a single language." (p.173)

When Ladefoged is speaking here of 'languages', he might just as easily be speaking of dialects. Perhaps it is the case that these 'small but reliable phonetic differences between languages' (or dialects) remain outside of the patterning behavior that constitutes phonological operations. Or it may be the case that variation in the interface plays a role in defining phonological contrast. This is a question I hope to return to and address in Chapter 6: Implications and directions for further research.

1.4 How the thesis will approach the research questions

In previous sections, stop consonant voicing has been discussed as an extremely common cross-linguistic phenomenon with observed universalities, but also with quite a bit of variability across languages, particularly with regard to language-specific differences in the phonetic implementation of [voice]. Assuming these differences relate directly back to subtle (and sometimes not so subtle) differences in articulation, the question is posed of whether the same cross-linguistic differences in the phonetic implementation of [voice] might not be observed at the level of dialect, when stop consonant articulation in a particular language is known to vary by dialect.

The research questions given in the proposed study ask what acoustic repercussions may be observed in two dialect populations of Spanish, where substantial differences in consonant articulation are thought to exist. In subsequent sections the reader will find a description of the experiment that was designed to address these questions. Specifically, the research questions seek to target what differences may be observed (if any) in 9 known acoustic correlates of [voice] between the two dialects, MAR and MER. Analysis and discussion of the results will include generalizations pertinent to any observed differences, if these are found.

§1.3 provided a contextualization of the research questions in terms of general knowledge on the voicing mechanism in production and how this relates to acoustic cues in perception. A discussion of the link between phonetics and phonology follows, as well as mention of the difficulties that arise in understanding the encoding of a complex phenomenon such as [voice]. In the remainder of §1.4 the basics of the proposed experiment will be re-capped. A detailed description of the methodology may be found in Chapter 3.

1.4.1 Method and measures

This thesis reports on the details and results of an experiment designed to measure 9 acoustic correlates of [voice] in two dialects of Venezuelan Spanish. The measurements were taken from field recordings made in Venezuela, using informants from rural communities in both of the target dialect populations, MAR and MER. The data was then analyzed using speech and statistical analyses software. A detailed description of informants, procedure, and analysis is given in Chapter 3. The measurements taken were: consonant duration, VOT, percent voicing throughout closure (%VF), RMS amplitude (RMS), preceding vowel duration, consonant/ vowel duration ratio (CV ratio), F1 onset frequency, F0 contour following closure (F0 contour), and presence/ absence of release burst. The contexts were word-initial and word-medial. The two conditions were speeded and unspeeded.

1.5 Outline of the dissertation chapters

The remaining sections of this dissertation are organized as follows. Chapter 2 provides a literature review of what is known of the acoustic correlates of and cues to [voice] in Spanish. The literature on Spanish dialectology as it pertains to the proposed topic is also reviewed. Chapter 3 outlines the methodology used in the present study. Chapter 4 presents the results from the present study. Chapter 5 is a discussion of these results. Chapter 6 is a general discussion, with implications and directions for further research.

Chapter 2: Literature review

2.1 Introduction

An overwhelming majority of the studies conducted on voicing production and perception have been carried out with English as the object of study. Nonetheless, since the 70s there has been a growing body of experimental work on voicing distinctions in Spanish. Research on Spanish stops has the potential to be of great interest in that the type of contrast we find in Spanish (plain voiceless/ voiced) is the most commonly-occurring in the consonant systems in the world's languages, present in 72.2% of the languages surveyed by Maddieson (1984) in the UPSID database.

Spanish consonant production is quite different from that of English. A main difference is that Spanish stop consonants are unaspirated. Also, in general, Spanish stops are pronounced with a higher degree of lenition than their English counterparts (Ortega-Llebaria, 2004; Zampini, 1996). Spanish phonological stops have been described in the literature as stop articulations at the beginning of words and after /l n r/ and fricative or approximant realizations in word-medial and especially intervocalic contexts. Important perceptual cues that have thus far been identified for discerning [voice] in Spanish include: VOT (Abramson & Lisker, 1972), the presence/ absence of low frequency energy (or periodicity) (Williams, 1976; Möbius, 2004), relative closure duration (Martínez-Celdrán, 1991a; Pérez, 1998), and articulatory tension (Martínez-Celdrán, 1991a,b).

2.2 Acoustic correlates and potential cues to [voice]

2.2.1 Voice Onset Time

In the area of [voice] perception, several studies investigating Spanish VOT (Lisker & Abramson, 1970a; Hay, 2005; Benkí, 2005) have shown that the phoneme boundary between voiced and voiceless consonants hovers around 0 ms.[8] The result is somewhat at odds with other work in psychology that suggests the underlying basis of the successful perception of VOT in stop consonants may correspond to a basic limitation of the auditory system to respond to differences in temporal order at stimulus onset (Jusczyk, Pisoni, Walley, & Murray, 1980). In a discrimination study with Spanish-hearing infants, Lasky, Syrdal-Lasky, and Klein (1975) found evidence suggesting the presence of three voicing categories. One area of high sensitivity occurred in the region of +20 to +60 ms., the area that corresponds to the English voiced-voiceless distinction. The other area of high sensitivity occurred in the region between roughly -20 and -60 ms. As Jusczyk, Pisoni, Walley, & Murray (1980) note, these discrimination results are interesting because Spanish has only one phoneme boundary separating its voiced and voiceless stops (0 ms.) and that boundary does not coincide with either of the two boundaries that Lasky et al. inferred from their discrimination data.

---

[8] Recall that VOT is calculated as the distance between the release burst and the onset of voicing. Therefore, a VOT of 0 ms. means that the release burst and the onset of voicing occur simultaneously. A 0 ms. boundary indicates a situation where phonologically voiced consonants display a negative VOT (known as ‘prevoicing’) and phonologically voiceless consonants display a positive VOT.

Lisker and Abramson (1970b), in a study of voicing discriminibility in three languages (English, Spanish, Thai), found the data on Spanish to be inconclusive. They report that subjects did show an increase in accuracy of discrimination above chance (33.3%) in the phoneme boundary region, but also showed other peaks along the continuum. According to the authors, "The Spanish subjects often failed to discriminate between variants that they consistently distinguished in the identification tests..." (pp. 18-19). Puzzled by these results, Lisker and Abramson suggest that the Spanish speakers were not well prepared for the task and perhaps had misunderstood the instructions.

In recent work, Hay (2005) found that in a comparison between Spanish- and English-speaking listeners, each group did show increased sensitivity to VOT in the area of values resting on their particular voiced/voiceless category boundary (around -5 ms. for Spanish and +15 to +35 ms. for English). However, her results (unlike those of Lasky et al. for Spanish-hearing infants) did not show evidence in the Spanish-speaking listeners of heightened sensitivity to temporal onset (values of around +20 to +25ms.). In the discussion section, Hay explains that, unlike English-speaking listeners, Spanish-speaking listeners may be more attuned to the presence or absence of low-frequency energy during closure duration than to VOT in making voicing judgments.

2.2.2 Periodicity, relative duration, intensity

Williams (1976) found evidence for the role of low frequency energy throughout closure in maintaining the voicing contrast in Spanish. Additionally, there is a suggested link between the presence of low frequency energy and frication. Dent (1976) maintains that in running Spanish, the occlusive allophones of voiced stop phonemes occur only in absolute initial position and after nasal consonants. In other environments, voiced stop phonemes are phonetically voiced fricatives.

“Therefore the contrast between voiced- and voiceless-stop phoneme categories is maintained not only by the presence or absence of voicing, but also by the presence of frication (voiced phonemes) or its absence (i.e., closure for voiceless phonemes).” p. S41

Despite the early claims, modern researchers in Spanish phonetics have downplayed the role of periodicity and frication in distinguishing voiceless categories from voiced. One complication is that articulatory norms appear to vary by dialect. This suggests the possibility of systematically different values for the acoustic correlates of and potential cues to [voice]. Martínez-Celdrán (2006) remarks that in data from Murcian Spanish, both [± voice] consonants consistently show low-frequency energy throughout the entirety of the closure. This, however, is not something that has been widely reported as characteristic of Spanish as a whole.

In general, Martínez-Celdrán (1991[a, b], 1992) and Pérez (1998) have de-emphasized the role of voicing throughout closure for Spanish stops and have instead insisted on the preeminence of relative closure duration and intensity in the categorization of [voice]. Pérez (1998) found that the presence or absence of low frequency bands had no major influence in the discrimination of /p t k/ /b d g/. His study showed that absolute length (duration) did have influence, though this influence was not equal across all places of articulation.

2.3 Dialectal variation in Spanish consonant production

In Chapter 1, I covered concepts related to phonetics, speech production and perception. My aim was to show that work in different areas of speech science supports the notion that different languages encode stop consonant voicing by means of different acoustic correlates. I then raised the question of if this might not also be true for dialects. This is the main issue behind the research questions guiding the present study. In the next section, my goal is to motivate the choice of dialects by providing a background of work done in Spanish dialectology with regard to the articulation of /b d g p t k/.

2.3.1 Basics of regional variety

Traditional research in Spanish dialectology has reported substantial variability in stop consonant pronunciations, especially between American lowland and highland varieties of the language (Canfield, 1981; Zamora & Guitart, 1982; Lipski, 1993). Interior or mountain dialects in Latin America exhibit 'conservative' language traits (for example, an occlusive pronunciation of /p t k/, /b d g/, especially in initial contexts, full retention of /s/, alveolar pronunciation of /n/) and coastal dialects share 'innovative' features common to Castilian as it was spoken in Southern Spain in the late 15th and early 16th centuries. These features are attributed to an overall lessening of muscular tension in pronunciation, especially noted in sibilants [where aspiration or loss of syllable-final /s/ is common]) (Canfield, 1981; Zamora & Guitart, 1982). The historical circumstances of the colonization of the New World, in combination with the geographical attributes of the region (more versus less isolation) and later development of the language have led to what amounts to substantial diversity in Spanish American dialects. Geographical and an accompanying linguistic diversity is very much in evidence in Venezuela, which features the Andes Mountain range to the west, coastal regions to the north, savannah to the south and the Amazon rainforest to the west.

FIG. 2. Map of Venezuela (Courtesy of lonelyplanet.com [http://www.lonelyplanet.com/maps/southamerica/venezuela/]. Margarita Island [Margarita] is off the northern coast. Mérida lies in the western part of the country, close to the border with Colombia.

Venezuelan dialects have been described as experiencing the same kind of lowland/highland divide that has been observed in other areas of the Spanish-speaking world. Unfortunately, while there are several phonological accounts of language varieties spoken in the country, few phonetic treatises exist. Furthermore, there is no detailed description available for any of the coastal dialects, though these are thought to follow the general trend of Caribbean Spanish. The Caracas standard may be said to be typical of Caribbean varieties in general- exhibiting a softened consonant pronunciation, high incidence of /s/ aspiration, and frequent interchange of /r/ and /l/. Spanish spoken in highland Mérida differs from this standard primarily by virtue of a more emphatic stop pronunciation. In addition, the degree of /s/ aspiration, while present, is less noticeable than in Caracas, or other areas throughout the country.

If few studies have addressed dialectal variation with respect to stop consonant voicing in Spanish, it is certainly the case that even fewer have provided instrumental analysis of the phenomenon. In one good descriptive account, Trujillo (1980) recounts the lax pronunciation of /p t k/ /b d g/ in the Canary Islands, remarking that stop pronunciations there are virtually nonexistent in normal speech, only occurring in certain combinations with other consonants (such as /n r l/). Instead, a fricative or approximant realization of the stop consonant series is the norm in all prosodic positions. Furthermore, /p t k/ are often realized as voiced or partially voiced, especially in intervocalic position, making them confusable with /b d g/. Trujillo notes that this is especially true in rural speech communities, though he documents the phenomenon as widespread throughout the region. In Trujillo's view, most studies involving the weakening or elision of consonants (a process he relates to that of the voicing of phonologically voiceless consonants, especially as they appear in intervocalic position) focus on the realization of /s/ (and sometimes /f/ and /x/).

In fact, in a review of phonological accounts of Venezuelan Spanish, it is the aspiration and elision of /s/ and /f/ that have received most attention, together with the loss of the occlusive element in /x/ ([ks] -> [s] or [h]). With particular regard to the Mérida dialect, researchers from that region point out that this variety shares features with so-called lowland dialects- in particular, the high incidence of aspiration of /s/, approximant pronunciation of /f/ and functional equivalence of /x/ and /s/ (Obediente, 1998; Villamizar, 1988). However, according to these researchers, the weakening of said segments does not extend to another object of their study, the realization of /n/, which tends to be velarized in Caribbean lowland dialects. Instead, in the Mérida dialect, /n/ assimilates to the place of articulation of a following consonant, or is alveolar in intervocalic contexts, just as it would be in a traditional highland, or non-lenited dialect. This indicates that the Mérida dialect may be somewhat of a mixed system, with some aspects characteristic of the lowland Venezuelan dialects and others not. There is no mention of how /p t k/ /b d g/ in the Mérida dialect compare with realizations in other lowland dialects, Venezuelan or otherwise. Nor is this information available for the other dialect in the present investigation, the Margarita dialect, a coastal variety of Venezuelan Spanish. One sociolinguist (Bentivoglio, 1998) notes (at the time of writing) that studies in socio-dialectal variation have only been carried out in four cities in Venezuela: Caracas, Puerto Cabello, Mérida, and two sectors of Maracaibo. The choice of the Margarita dialect for the present study was motivated in part by this absence. Further, the Margarita dialect is potentially of interest because of anecdotal accounts from native speakers of Venezuelan Spanish, who describe this dialect as difficult to understand if one is not a Margarita Islander. The dialect is described as “very fast” and it is said that speakers from that region “don’t produce their consonants.” With this in mind, there is good reason to believe that the Margarita dialect, as representative of a lowland, coastal dialect, provides an excellent contrast for the highland Mérida variety.

2.4 Lewis (2000)- An instrumental study comparing dialects

Lewis (2000) undertook a study of intervocalic voiceless stop consonants in three Spanish dialects- one peninsular (Bilbao), two American (Caracas and Medellín)- in an effort, as he says, “to quantify lenition.” In the study, Lewis evaluated glottal pulses, closure duration, VOT and whether or not a release burst was present for phonologically voiced and voiceless stops in each dialect. Lewis found that for most of the measures, the Bilbao and Caracas dialects tended to pattern together (somewhat surprisingly, from a historical perspective), while the Medellín dialect diverged in certain aspects. Overall, Medellín showed a greater VOT, greater conservation of release bursts, greater mean closure duration, and dramatically fewer instances of voicing during closure. The Lewis study suggests the following points: 1) that there are noticeable dialectal differences in the implementation of [voice] in Spanish, and; 2) these differences correlate directly with articulatory parameters in production; that is, they are actively under speaker control.

In the absence of studies that investigate the physical gestures corresponding to differences in stops across dialects, we rely on evidence from spectrogram and waveform displays to communicate information about subtle differences observed in production.

Supposing that these nuances of articulation determine the acoustic signature of a given sound, it follows that the correlates encoding [voice] will vary according to production norms. The applications of this are far-reaching. The fields of speech perception and psychology are rich with studies showing that experience with a particular language influences the way a listener perceives and categorizes contrasts (e.g. Burki-Cohen, Grosjean, & Miller 1989; Lotto & Holt, 2006). It is known that repeated reliance on a given acoustic cue or set of acoustic cues increases the sensitivity to those cues at the expense of the others. This effect is well-documented in the literature on language acquisition (Flege, Munro, & Fox, 1995), in cross-linguistic studies, and most recently in studies providing electrophysiological support for how native-language linguistic representations constrain auditory processing (Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., & Sinkkonen, J., 1997). To my knowledge, these issues have not been much explored at the level of dialectal variation within language. This is an additional reason the present study may be of interest.

Chapter 3: Methodology

3.1 Introduction

3.2 Informants

The informants were 25 adult monolingual speakers of Spanish between the ages of 20-35, with educational experience ranging between 1st grade and high school. Speakers in Margarita (10 females; 4 males) were recruited from a fishing village, El Tirano, which lies close to Playa el Agua. Speakers in Mérida (7 females; 4 males) were recruited from a town in the Venezuelan highlands, San Rafael de Mucuchíes. An effort was made to choose towns that were small, of roughly the same size, with mostly an indigenous population. All subjects were recruited on-site. No subjects had foreign language experience beyond that required in the public schools. All subjects were paid volunteers. Each was offered two dollars (or the equivalent in bolívares) for their participation in the study. Language background and biographical information were assessed through a questionnaire administered verbally (see Participant Survey, Appendix 3). None of the subjects reported hearing or speech problems.

3.3 Procedure

Materials were 44 CV syllable prompts preceded by the word *son*. *Son* was included for two reasons: 1) to provide a word-initial (but not utterance-initial) context that would permit measurement of VOT, and; 2) to elicit maximum contrast between the word-initial and word-medial positions. Occlusive pronunciations have been noted for both word-initial environments and after /n r l/ (Hualde, 2005).

After the word *son*, each nonsense word that appeared began with a stop consonant and ended with a canonical Spanish vowel (a e i o u). The distribution of prompts throughout the sample was as follows. All stop consonants appeared an equal number of times in the sample (15) and an equal number of times before each vowel (3). All prompts were randomized, both in the regular and in the training blocks. Prompts were viewed in a PowerPoint slideshow administered via laptop computer. The timing of screen changes was controlled by the investigator. During the recording session, subjects were asked to create an alternation, a wordplay whereby Target Syllable 1 (word-initial; appearing on the screen beside *son*) became the first syllable of a nonsense word. The second syllable of the nonsense word would be comprised of /r/ + the vowel in the first syllable. The last syllable (Target Syllable 2 [word-medial]) would be a repetition of Target Syllable 1. Stress would fall on the penultimate syllable (according to the default stress assignment for vowel-final words in Spanish). For example, if the prompt was “Son TO”, subjects said, “Son toróto”, “Son BU”, “Son burúbu”, etc.[9] The choice of nonsense words was motivated by evidence that the frequency of words affects the articulation of all segments within the word. Furthermore, care was taken to devise an experimental method that would require only a minimal level of literacy on the part of

---

[9] A list of prompts and target responses appears in Appendix 1. Appendix 2 shows an example of screen displays from Bloque 3.

subjects. For a study of this nature, it was desirable to work with rural populations, with people who theoretically have had less exposure to print and less contact with speakers from other dialects than might be the case in a cosmopolitan environment.

In ten out of the fifteen times each stop consonant appeared, Target Syllable 1 (henceforth, TS1) was followed by a color (portrayed as a colored square) or a number.

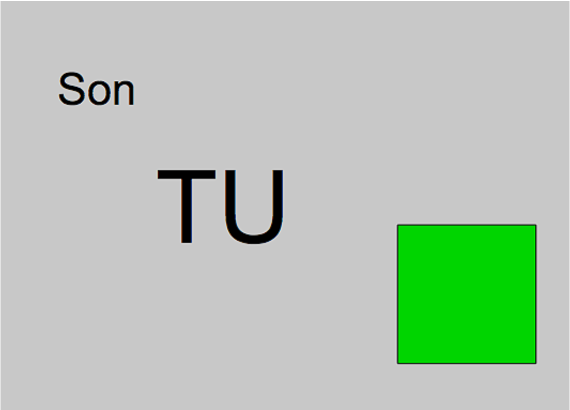


FIG. 3. Example Prompt "Son turútu verde".

The particular colors and numbers used in the sample were chosen for their status as disyllabic trochees ending in vowels, the most frequent word type in Spanish. Half began with either /p/, /t/, or /k/. The other half began with either /b/, /d/, or /g/. On screens where a color or number appeared to the right of the target syllable, subjects were asked to say, "Son toróto verde" or "Son burúbu cinco", etc. The motivation for including an extra word following the target syllable was: 1) to keep subjects engaged in the task and therefore to lessen the possibility of list-like intonation effects; 2) to shift the focus off the target syllable. This measure was thought to increase the likelihood that the nonsense words would be pronounced more like real words in conversational speech. There was an even distribution of beginning /p t k/ and /b d g/ with respect to the target syllables. The prompts were grouped into 4 blocks.

3.3.1 Block 1

Block 1 (23 screens) was a training block consisting of repetitions of *son* plus different target syllables. Subjects received an explanation of what to expect on the screen and what they should say. They then were asked to respond to the prompts. Each subject was given feedback on his or her response in terms of the investigator either repeating the correct response or re-explaining the instructions until these were accurately carried out.

3.3.2 Block 2

Block 2 (24 screens) was a second training block that introduced the prompts with colors and numbers. Feedback was the same as in Block 1.

3.3.3 Block 3

Block 3 (55 screens) was a combination of Blocks 1 and 2, with some prompts containing colors or numbers and others not. The screens in Blocks 1 and 2, as well as the first 10 screens of Block 3 were considered training slides and as such were not included in the measurements.

3.3.4 Block 4

Block 4 (44 screens) was a speeded trial in which subjects were asked to execute the prompts as quickly as possible while maintaining accuracy. As with the inclusion of colors and numbers, the choice to include a speeded trial was prompted by a desire to achieve as natural a speech sample as possible. The hope was that having subjects focus on several tasks at once (accurately interpreting the prompts, making the appropriate alternation, uttering the response as quickly as possible) would disperse attention paid to the nonsense syllable and allow natural speech patterns to emerge.

3.4 Measurements taken

Word-initial and word-medial target syllables from Blocks 3 and 4 were analyzed separately. Word-initial target syllables were analyzed with respect to the following: consonant duration, VOT, %VF, RMS, F1 onset frequency, F0 contour following closure, and presence/ absence of release burst. Word-medial target syllables were analyzed according to the same parameters as the word-initial contexts, with the inclusion of two additional parameters- preceding vowel duration and CV ratio. Measurements were made from a spectrogram display viewed in conjunction with the waveform and pitch track using Wavesurfer speech analysis software (Sjölander & Beskow, 2006). Settings in Wavesurfer were adjusted to view the display in a Hanning window.[10] The bandwidth was set to 250 Hz, a value that is intermediate between the preferred values for male (200 Hz) and female (300 Hz) speakers (Ladefoged, 2003). Figure 4 shows an example of a typical display window.

---

[10] Hanning and Hamming windows provide a smoothing of the edges of a sound wave display, prohibiting zero values. The Hanning window permits a more finely-tuned analysis, allowing values closer to zero.

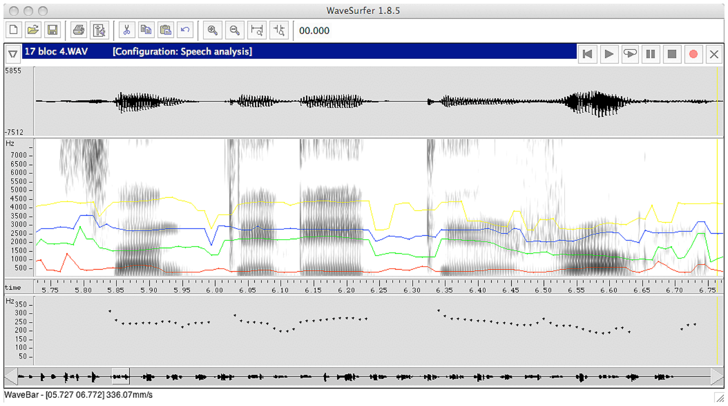


FIG. 4. Example of Wavesurfer Display “Son tiríti rojo” (Speaker 17 [MER]).

3.4.1 Consonant duration

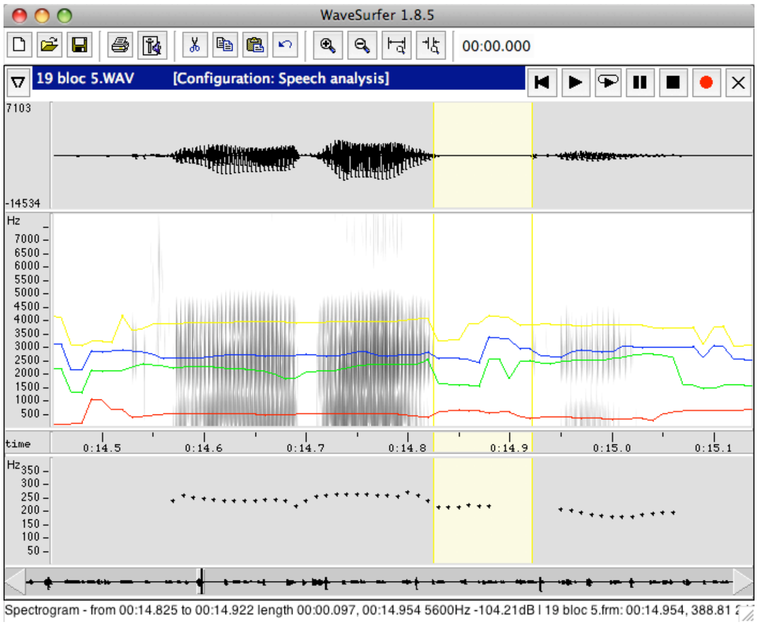
Consonants were measured on the waveform from the point of the last periodic pulsing of the vowel to the burst (where applicable). In cases where there was no burst, the end point was taken to be the place just before periodic energy for the following vowel began. For guidance on segmenting nasal consonants from the oral stops (relevant for TS1), I relied on Ladefoged (2001). Nasals are identified by a low first formant appearing at around 250 Hz accompanied by a large region above containing no energy. End points for nasals, just as for vowels, were judged to take place after the last pulse of periodic energy. In cases where nasals preceded voiced stops, the end point was judged to coincide with the last pulse of F1.

3.4.2 VOT

VOT was measured from the burst to the onset of periodic voicing. In cases of negative VOT, the measurement was judged to begin at the first appearance of F0 energy, continuing throughout the duration of the burst.

3.4.3 % VF

To obtain %VF, I followed a procedure employed by Riede, Mitchell, Tokuda, & Owren (2005). The percentage of voiced frames was quantified by counting the number of glottal pulses throughout the closure gap and dividing this number by the duration in seconds. Figure 5 shows an example of the slice of the spectrogram from which a measurement for %VF was taken.



Spectrogram - from 00:14.825 to 00:14.922 length 00:00.097, 00:14.954 5600Hz -104.21dB I 19 bloc 5.frm: 00:14.954, 388.81

FIG. 5. Example of %VF Measurement Display “Keréke” (Speaker 19 [MER]).

Glottal pulses are shown in the bottom pane. In this example, there are 6 pulses over a time domain of .097 seconds. According to the formula given above, this would yield a %VF of 61.85 for medial /k/.

3.4.4 RMS amplitude

I relied on guidance from Harrington & Cassidy (1999), Gelfand (2001), and Ladefoged (2003) in determining the best way to measure RMS. RMS is a common measure of intensity. It is dependent on the amplitude of the sound wave as measured in decibels (dB). The RMS value is obtained by squaring individual amplitudes in a given time window, averaging these, then taking the square root of the average. As Harrington and Cassidy explain:

> “The purpose of squaring the values is to convert all negative values into positive values, since otherwise the values would tend to cancel each other out (when summed) resulting in an amplitude measure that would be close to zero for most kinds of speech waveforms.” (p.142)

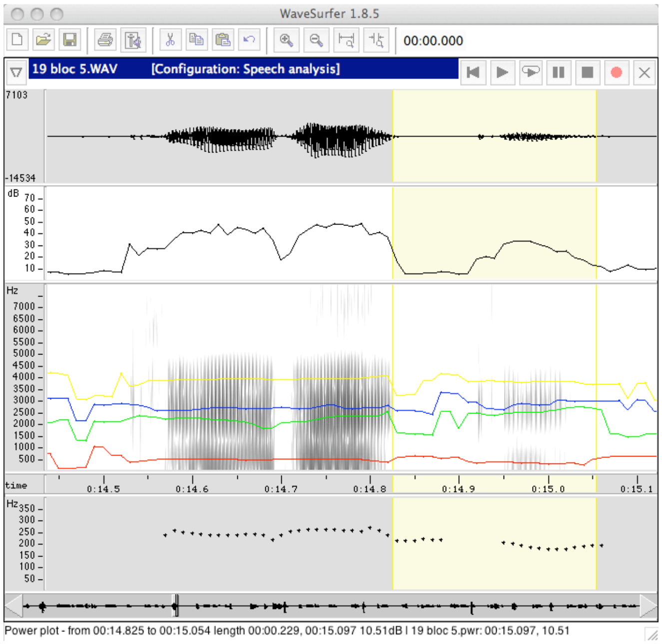
Figure 6 shows an example of how RMS amplitude is calculated.

$$
\begin{aligned}
\text{RMS}\ (x) \quad &= \surd\ (4^2 + (-1)^2 + 0^2 + 8^2)\ /\ 4 \\
&= \surd\ (16 + 1 + 0 + 64)\ /\ 4 \\
&= \surd\ 20.25 \\
&= 4.25
\end{aligned}
$$

FIG. 6. Sample RMS for Signal x = [4, -1, 0, 8] (Harrington & Cassidy, 1999; p. 142).

By this procedure, the intensity of a sound relative to a given reference sound is calculated not by comparing the relative amplitudes but instead by comparing the relative powers of the two sounds (the power of a sound = the square of its amplitude).

The Wavesurfer program provides a power display where relative amplitudes for a given time window are plotted on a scale of 10 to 70 dB. Most values in this sample are expected to fall between 20 and 60 dB. For the present study I chose to examine the unstressed CV window for both TS1 and TS2. Three points were measured: 1) initial trough signifying amplitude low following closure; 2) peak representing maximum aperture of the vocalic gesture, and; 3) final trough signifying closure of the gesture and transition to the following segment. The choice of window was carefully considered. Choosing only peaks could give a false picture of the amplitude. An initial high would reflect energy coming off the preceding segment (vowel or nasal) prior to the initiation of the consonantal gesture. Choosing only troughs, on the other hand, would force amplitudes down into the lower values and would not provide an accurate gauge of the range. I believe a trough-peak-trough analysis is the one that most appropriately captures the overall amplitude profile throughout the CV window. Figure 7 shows an example of a CV window used to calculate RMS on a spectrogram/ waveform/ power display.



```
Power plot - from 00:14.825 to 00:15.054 length 00:00.229, 00:15.097 10.51dB I 19 bloc 5.pwr: 00:15.097, 10.51
```

FIG. 7. Example of CV Window in a Spectrogram/ Waveform/ Power Display “Keréke” (Speaker 19 [MER]).

The power display appears in the second pane from the top, below the waveform and above the spectrogram display. The CV window is highlighted. In this example, three measurement points are taken: the initial trough during the consonant closure (5.36 dB), the peak during the following vowel (33.53 dB), and the final trough at the end of the vowel (13.83 dB).[11] These points are then averaged to provide an RMS amplitude value of 17.57 dB for this particular TS2.

3.4.5 Preceding vowel duration

Vowel duration measures were initiated from the onset of regular periodic energy and concluded at the end of the last cycle of regular pulsing. Preceding vowel duration measures are only applicable to TS2 measures, as TS1 tokens were separated from the vowel by an intervening nasal consonant.

3.4.6 CV ratio

CV ratio was measured by dividing the consonant duration by the preceding vowel duration. This measure is only applicable to TS2 for the same reason given in the previous description, namely that there was no preceding vowel in TS1.

3.4.7 F1 onset frequency

F1 onset frequency was measured at the point where F1 stabilizes following the release of the target consonant.

3.4.8 F0 contour

F0 contour is the average pitch taken over the duration of the vowel following the target consonant.

---

[11] Actual values not shown on this display.

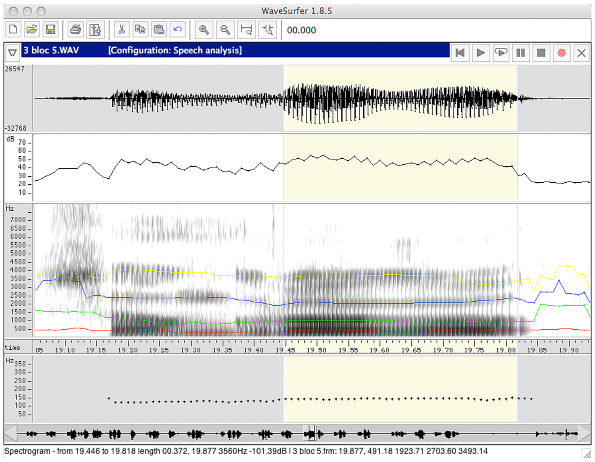
3.4.9 Presence/ absence of release burst

Yes or no depending on whether a release burst (sudden spike in frequency) was visible on the spectrogram prior to the onset of regular voicing.

3.5 How measurements were made

3.5.1 Segmentation issues

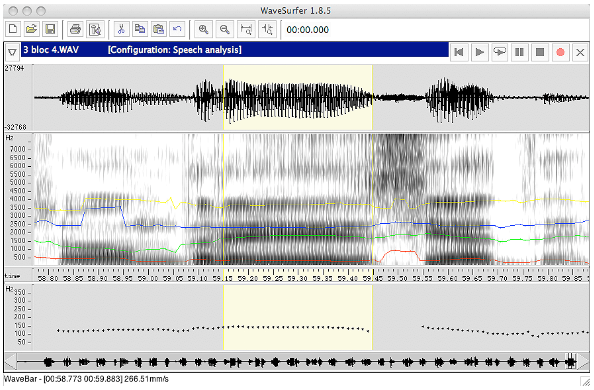
As will be discussed throughout the remainder of this dissertation, approximant realizations of phonological stops were anticipated, particularly in [+ voice] and in MAR speakers more than in MER speakers. In choosing whether or not to classify a segment as a stop consonant (and hence, measureable under the parameters I have stated here) or an approximant (where approximant status is noted, but measures are not possible), I relied heavily on the profile of the waveform. For example, in responses such as the one in Figure 8, the response was included on the basis of evidence of a consonant-like gesture (dip in the waveform, similar to what would be seen in a traditional voiced stop).



Spectrogram - from 19.446 to 19.818 length 00.372, 19.877 3560Hz -101.39dB I 3 bloc 5.frm: 19.877, 491.18 1923.71 2703.60 3493.14

FIG. 8. Speaker 3 (MAR) /ogo/ of “Son gorógo (siete)”.

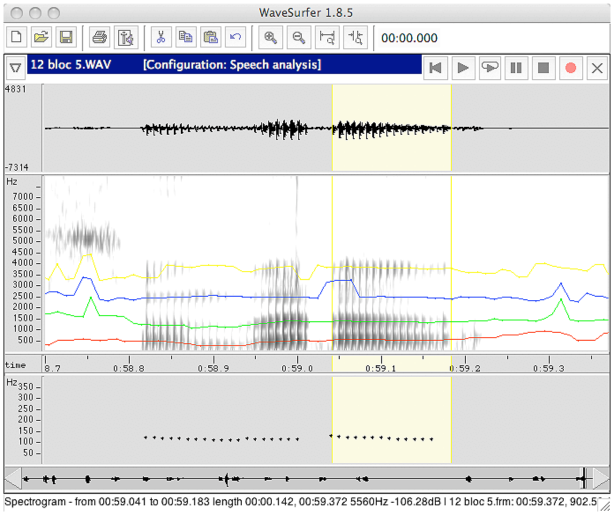
Unlike as in a traditional voiced stop, notice that this spectrogram shows evidence of formant structure throughout the duration of the /VCV/ sequence. Normally this presence and type of formant structure is used as a main classifier for sonorants, especially vowels. Therefore, the sounds I considered here to be consonants may not be thought of as such in a traditional sense. Their display is nonetheless more consonant-like than the other kind of realization I chose to label as ‘approximant’. In this second type of response, there may be dips in the waveform energy or not, but these dips do not correspond to any consistent pattern. Figures 9 and 10 are examples of responses that were labeled as approximants.



WaveBar - [00:58.773 00:59.883] 266.51mm/s

FIG. 9. Speaker 3 (MAR) /idi/ in "Son dirídi siete".

Figure 9 shows some variation in energy, of a narrowing consistent with a closing gesture. There is no clear corresponding opening, however. From the waveform, it is evident that the periodic energy does build and fall again before the frication of the /s/ in *siete*, but it is difficult to pinpoint where this action begins.



Spectrogram - from 00:59.041 to 00:59.183 length 00:00.142, 00:59.372 5560Hz -106.28dB l 12 bloc 5.frm: 00:59.372, 902.5

FIG. 10. Speaker 12 (MAR) /ada/ Sequence of "Son daráda".

Finally, the pronunciation of /ada/ in Figure 10 is consistent with what some researchers have referred to as consonant deletion. The energy on the waveform begins high and tapers off, similar to what might be seen in a spectrogram of a final vowel.

3.6 Statistics

Statistical analysis was performed with SPSS software, using a linear mixed model ANOVA and nesting tokens within subject. A linear mixed model handles data where observations are not independent. This is the case, for example, when tokens are presented repeatedly to subjects. Tokens are considered not to be independent of one another, by virtue of being produced by the same subject. In the current study, subjects produce the same target consonant (/b d g p t k/) multiple times.

One advantage to a linear mixed model is that it correctly models correlated errors, whereas procedures in the general linear model family usually do not (Garson, 2008). The current model postulates the existence of random effects, where the set of values of a categorical predictor variable are seen not as the complete set but rather as a random sample of all values (for example, in this study, the variable "speaker" has values representing only 23 of an unknown number of possible subjects). Through random effects models in linear mixed modeling, the researcher can make inferences over a wider population.

The model for the present study tested 3 independent variables: voicing category ([± voice]), dialect (MAR, MER), and condition (unspeeded, speeded). Additionally, this model tested for interactions between dialect and voicing category and dialect and condition. Alpha levels for all analyses were set at < .05.

3.7 How the data are presented

In this chapter I have described the design of the experiment, including information on the informants, the procedure, measurements and statistics. In the next chapter, I report the results for each of these sections in turn.

Chapter 4: Results

4.1 Informants

This section reports on the data gathered from informants, including notes on task performance and dialectal and individual speaker profiles. With regard to the task- two participants (Speakers 1 [MAR] and 20 [MER]) were unable to successfully complete the training blocks. Therefore no data from these speakers was included in the statistical analysis. Another participant (Speaker 21 [MER]) was able to carry out only 3 usable responses (out of a potential 90). Therefore, the data from this subject was excluded as well. Furthermore, technical difficulties with the recording equipment (batteries or memory cards running out) resulted in a loss of data for Speaker 6 (MAR) and incomplete data for Speaker 12 (MAR).

For all speakers, certain response types were excluded. Responses where the /r/ + vowel syllable became transposed with one of the other syllables (e.g. *son da**dára***) were removed from further analysis. Likewise responses where the primary stress was perceived by the researcher to reside anywhere other than in the penultimate syllable (e.g. *son dara**dá***). There were some responses where vowels in the nonsense word did not match the one given in the prompt (e.g. *son darád**o***) and cases where an extra word was added in the same breath group (e.g. *son kiríki, **¿no?***). These responses were excluded. Lastly, in a very few cases, V2 exceeded 500 ms. In these instances the speaker appeared to be processing the prompt at the same time they were giving the response, resulting in an irregular delay. These responses were omitted from the sample.

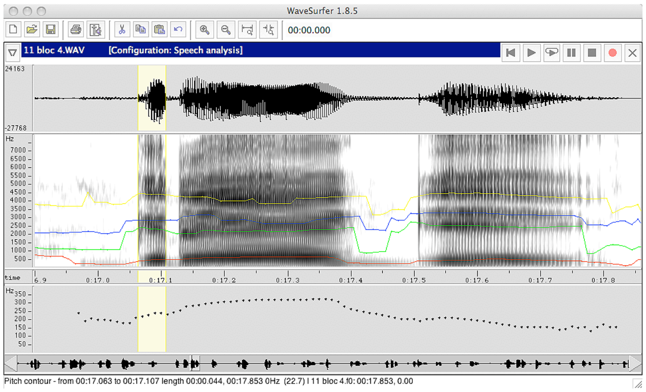
In TS1 environments I excluded responses where there was no preceding nasal (e.g. *son d**a** daráda*) or when *son* and the target response occurred in different breath groups. Some speakers pronounced *son* right away, then took a second or more to pronounce the target response, presumably for reasons of processing. TS1 responses were not measured, but TS2 responses were measured in both of the scenarios mentioned above.

In terms of general observations on the nature of the language data recorded for this study, it was noted that there was more variability in the speech of female participants than in that of the male participants. This was particularly true of MAR speakers. However, it was also the case that there were more speakers overall from MAR than from MER and that most of these were female.[12] It was observed that males in both dialects tended to pattern more like MAR speakers than MER speakers.

In the MAR population of speakers, the incidence of approximant realizations is very high, particularly in male speakers. V1 in the MAR population is often unusually short, giving the auditory impression of a consonant (C + /r/) cluster, rather than a CV sequence (see Figure 11).

---

[12] For descriptive statistics on all speakers, see Appendix 4.



Pitch contour - from 00:17.063 to 00:17.107 length 00:00.044, 00:17.853 0Hz (22.7) l 11 bloc 4.f0: 00:17.853, 0.00

FIG. 11. Speaker 11 (MAR) Short V1 Typical of MAR Speakers in "Guerégue".

In MAR, the /r/ is pronounced as a sharp tap, showing a clear, almost stop-like break between vowels, as opposed to the /r/ of MER realizations, where evidence of formant structure is sometimes seen. To my knowledge, this characteristic has not been reported before.

In MAR medial position, there is a great deal of variation with respect to manner of articulation for the phonological VCV sequence. Sometimes a medial consonant is recognizable on the spectrogram, but more often than not (and as noted in §3.5.1) the "consonant" is nothing more than a dip in energy, indistinguishable from a long vowel, or vowel-like utterance. Sometimes there is frication on the end, as is shown in Figure 12.

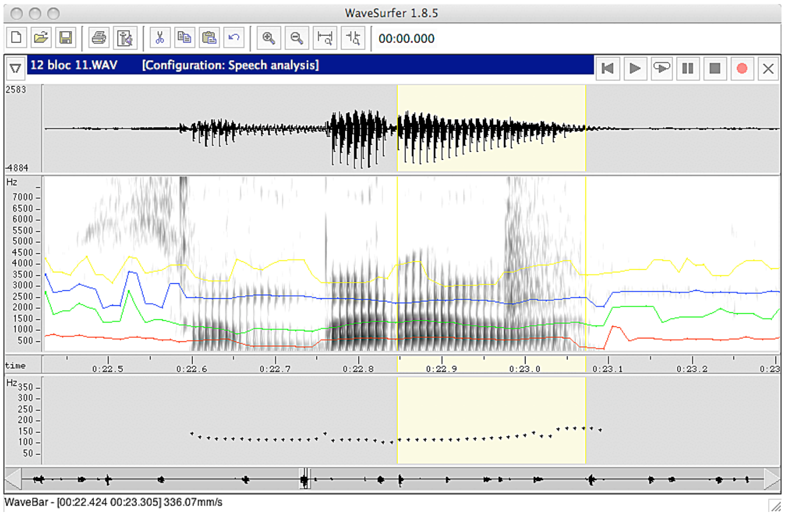
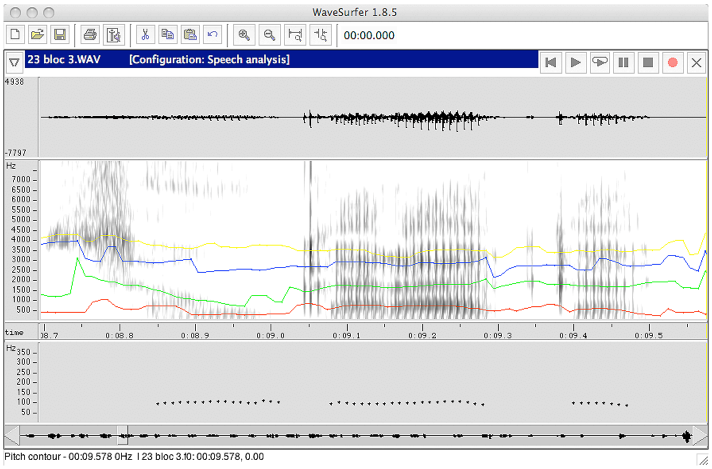


FIG. 12. Speaker 12 (MAR) Frication on Final Vowel. Shown here: /aba/ of "Son barába".

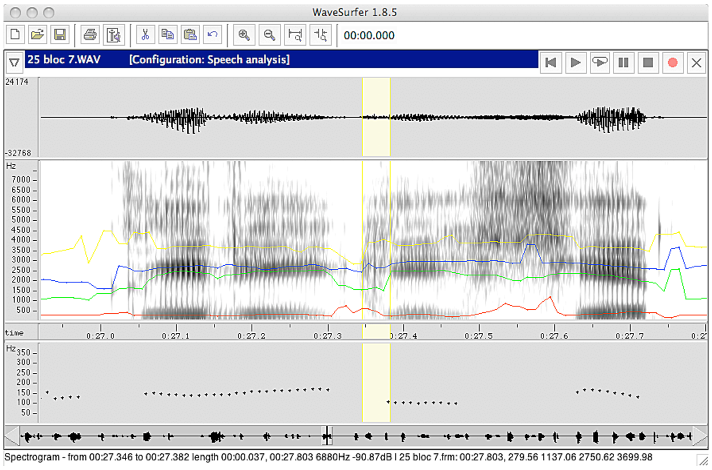
The MER dialect pronunciation shows less inter-speaker variation than MAR, with the exception, as mentioned before, of the pronunciation of male speakers. Some interesting features of this dialect include frequent incidence of creaky voice (as shown in Figure 13) and aspiration (in Figure 14).



Pitch contour - 00:09.578 0Hz I 23 bloc 3.f0: 00:09.578, 0.00

FIG. 13. Speaker 23 (MER) Creaky Voice.

Creaky voice is indicated by the separation of pulsing periods on the spectrogram and the consequent striated appearance.



Spectrogram - from 00:27.346 to 00:27.382 length 00:00.037, 00:27.803 6880Hz -90.87dB I 25 bloc 7.frm: 00:27.803, 279.56 1137.06 2750.62 3699.98

FIG. 14. Speaker 25 (MER) Aspiration Following Medial /g/ of “Guirígui siete”.

There is not much evidence of prevoicing in either dialect, as has been reported in other studies on Spanish (Williams, 1976), though there is evidence of partial frication. With regard to prevoicing, the absence is likely a result of limitations on the experimental design. In this study, target consonants always appear preceded by sonorants, either a nasal consonant, as in TS1, or a vowel. This is not unlike the usual situation in natural speech. Any prevoicing would be masked by the formant structure of preceding segments. In cases where a pause preceded the target response (cases that were omitted from the data set), there is occasional evidence of prevoicing, especially among MAR speakers, as in Figure 15.

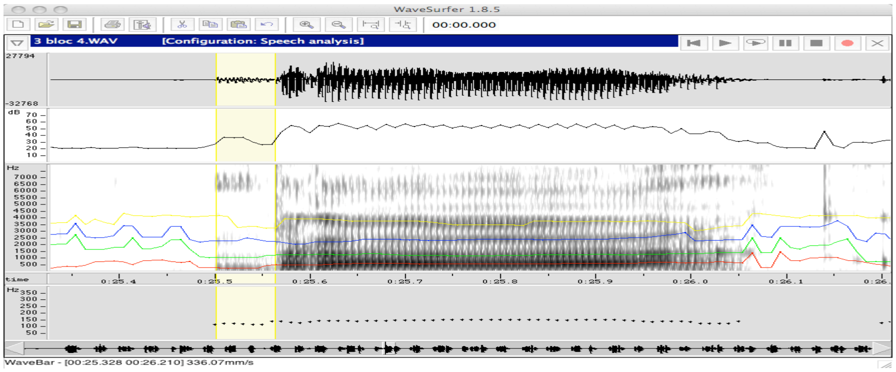


FIG. 15. Speaker 3 (MAR) Prevoicing of /b/ in “Barába (cuatro)”.

In addition to variation by dialect, there was individual variation in the sample. It was observed that speakers often pronounce TS2s as velar (/g/ or /k/) in cases where TS1 is coronal or bilabial (ex/ “tiríki”, “birígui”). One speaker (Speaker 7 [MAR]) consistently pronounces TS2 as a voiceless velar (i.e., “doróko”, “piríki”). Otherwise, individuals differ in speech rate and in how carefully or casually they respond, in where they pause as they process the prompts.

4.2 Procedure

Subjects were recorded with a Shure head-mounted dynamic microphone adjusted to the left corner of the mouth, at approximately ½ inch from the lips. Responses were recorded onto a compact flash card using a Marantz PMD 660 steady-state recorder. They were later transferred via a Macintosh G4 PowerBook onto an external hard drive. The sound files were recorded as WAV files at 48 kHz. All recordings took place in the field, in places that the subject and researcher agreed upon as being both amenable and relatively quiet. Most often, this was on the sidewalk outside the subject's residence or place of work. The recordings were generally of good quality, though some contained sporadic background noise. In cases where noise was excessive and interfered with the measurements, the responses were discarded.

Measurements were made from a spectrogram display viewed in conjunction with the waveform and pitch track using Wavesurfer speech analysis software (Sjölander & Beskow, 2006). Settings in Wavesurfer were adjusted to view the display in a Hanning window[13]. For duration measurements the bandwidth was set at 250 Hz, a value that is intermediate between the preferred values for male (200 Hz) and female (300 Hz)

---

[13] Hanning and Hamming windows provide a smoothing of the edges of a sound wave display, prohibiting zero values. The Hanning window permits a more finely-tuned analysis, allowing values closer to zero.

speakers (Ladefoged, 2003). Figure 16 shows an example of a typical speech analysis display consisting of waveform and spectrogram.

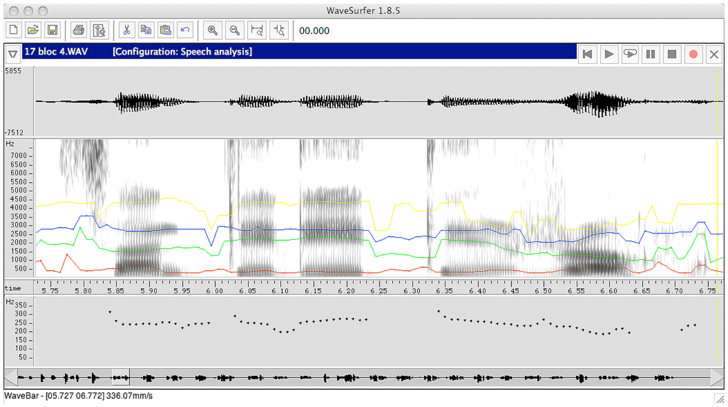


FIG. 16. Example of Wavesurfer Display "Son tiríti rojo" (Speaker 17 [MER]).

There were many cases where extensive coarticulation between nC or CV syllables prevented accurate segmentation in the waveform/ spectrogram display. A total of 14.92% of the initial sample data was affected. In cases where it was not possible to separate segments, the measurements associated with that target consonant and surrounding vowel (if applicable) were omitted.

Preliminary analysis showed a slight positive skew for the consonant duration data in word-initial position. This might be attributed to a general propensity for elongation during the processing of stimuli (i.e., lengthening is one strategy for buying more time to think). Consonant duration measurements were converted to a standardized score (Z-score). Any Z-scores that fell more than three standard deviations from the mean were removed from the analysis. When the descriptive analysis was run again, the data fell within a normal distribution.

A large positive skew was also found for *VOT* in initial position. Three exceptionally large outliers were responsible for the skew. One large outlier was also found for *VOT* in medial position. Most likely these anomalies arose out of data entry error. Removing values with Z-scores > 3 corrected the problem.

Statistical analysis was performed with SPSS software, using a linear mixed model ANOVA and nesting tokens within subject. The model tested 3 independent variables: *voicing category* ([± voice]), *dialect* (MAR, MER), and *condition* (unspeeded, speeded). In addition, the model tested for interactions between *voicing category* * *dialect* and *dialect* * *condition*. The analysis was based on a total of 3780 observations (2 contexts (initial and medial) X 21 speakers [12 MAR, 9 MER] X 90 tokens each).

4.3 Measurements taken

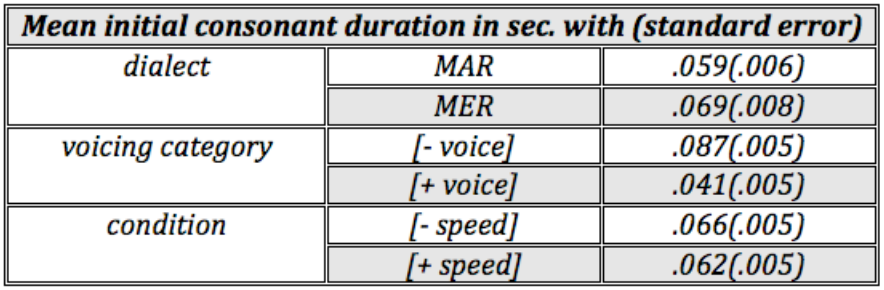
4.3.1 Consonant duration

Consonant duration was the only correlate to return significant values for three different variables (*voicing category*, *condition*, and *voicing category * dialect*) in both initial and medial contexts.[14] [-Voice] measurements were significantly longer than [+voice] measurements in both initial and medial contexts. MER consonant durations were longer than those found in MAR, though the difference was nonsignificant and the difference in

---

[14] Significance levels for all measures may be found in Tables 17 and 18 in §4.4.

means in medial position was slight. Unspeeded was significantly longer than speeded. The level of significance was greater in initial than in medial position. Table 1 shows the mean consonant durations in initial position, broken down by category and accompanied by the standard error.



Mean initial consonant duration in sec. with (standard error)		
dialect	*MAR*	*.059(.006)*
	MER	*.069(.008)*
voicing category	*[- voice]*	*.087(.005)*
	[+ voice]	*.041(.005)*
condition	*[- speed]*	*.066(.005)*
	[+ speed]	*.062(.005)*

TABLE 1. Initial Consonant Duration.

Figure 17 shows the interaction of *voicing category* * *dialect* in initial position. [+ Voice] values for MAR and MER were roughly equivalent (with MER being slightly longer). [-Voice] values, however, were substantially longer for MER, causing a wider separation between [± voice] in this dialect than in MAR.

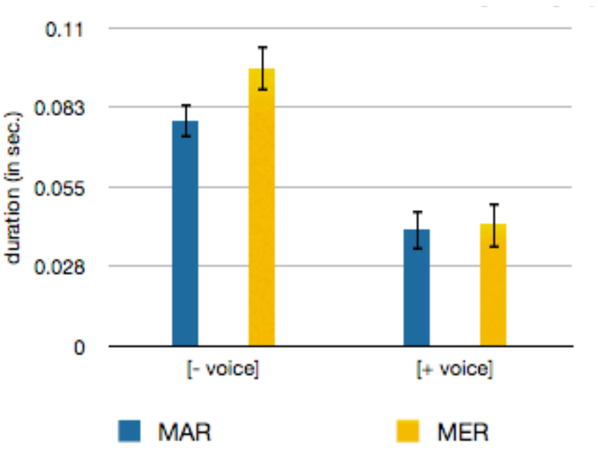


FIG. 17. Initial Consonant Duration Interaction Voicing Category * Dialect.

Figure 18 below shows the interaction of *voicing category* * *dialect* in medial position. Consonant durations were longer for medial consonants than for consonants in initial position. Durations were roughly equal between dialects for consonants in medial position. The differences observed were nonsignificant. MAR was slightly longer than MER in the [+ voice] context. In the [- voice] context, MER was longer than MAR, though the difference was less than in initial position. The separation between [± voice] was slightly greater in initial than in medial position.

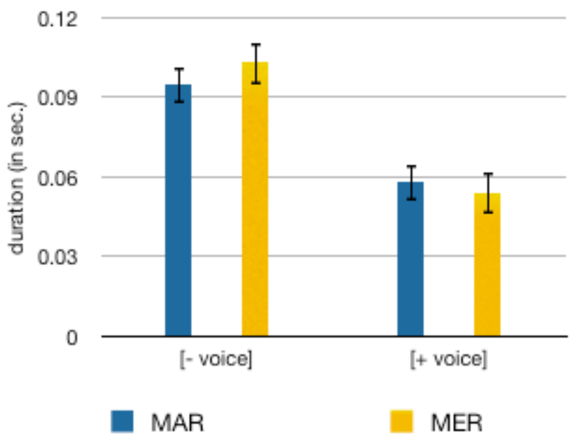
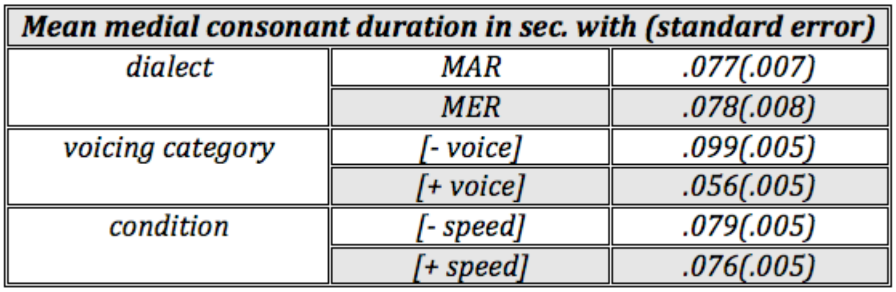


FIG. 18. Medial Consonant Duration Interaction Voicing Category * Dialect.

Table 2 indicates the consonant duration in medial position.

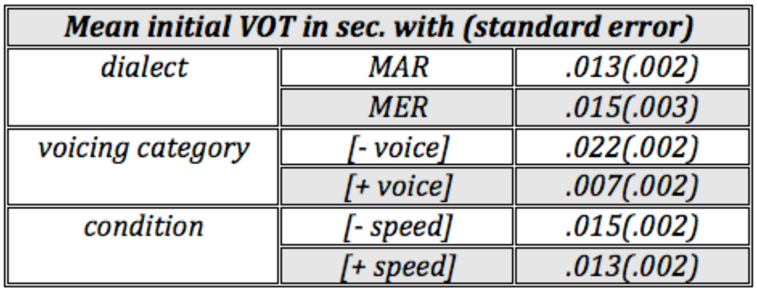


Mean medial consonant duration in sec. with (standard error)		
dialect	*MAR*	*.077(.007)*
	MER	*.078(.008)*
voicing category	*[- voice]*	*.099(.005)*
	[+ voice]	*.056(.005)*
condition	*[- speed]*	*.079(.005)*
	[+ speed]	*.076(.005)*

TABLE 2. Medial Consonant Duration.

4.3.2 Voice onset time

Initial values for VOT were significant for the following variables: *voicing category*, *condition*, and the interaction of *voicing category * dialect*. In medial position, VOT was significant for *condition* only. *Voicing category* approached significance at .092. As was the case with *consonant duration*, medial VOT values were slightly longer than those in initial position. However, the [± voice] means, with *dialect* and *condition* collapsed, were shorter than those found in initial position. MER and MAR were roughly the same length in initial position. In medial position, MAR was slightly longer, though the difference was nonsignificant. The separation between [± voice] categories in initial position was substantially larger for MAR than for MER. Table 3 presents means for VOT in initial position.



Mean initial VOT in sec. with (standard error)		
dialect	*MAR*	*.013(.002)*
	MER	*.015(.003)*
voicing category	*[- voice]*	*.022(.002)*
	[+ voice]	*.007(.002)*
condition	*[- speed]*	*.015(.002)*
	[+ speed]	*.013(.002)*

TABLE 3. Initial VOT

Figure 19 shows the interaction of *voicing category* * *dialect* in initial position. For VOT there is a wider separation of [± voice] values in MAR than in MER. This interaction was significant. [-Voice] values are slightly higher in MAR, but [+ voice] values are lower. In medial position, the difference between dialects is the same, but the relationship between MAR and MER is inverted, with the mean in MAR being slightly higher.

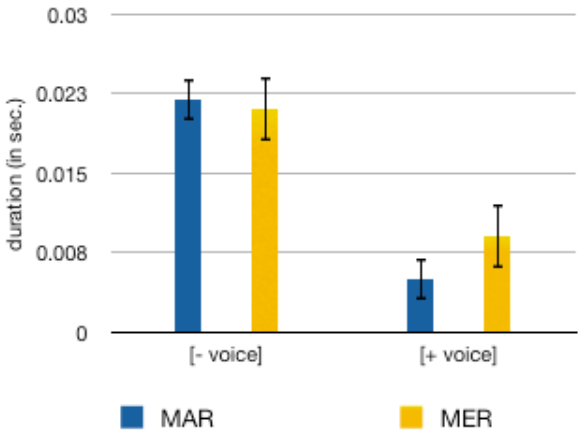
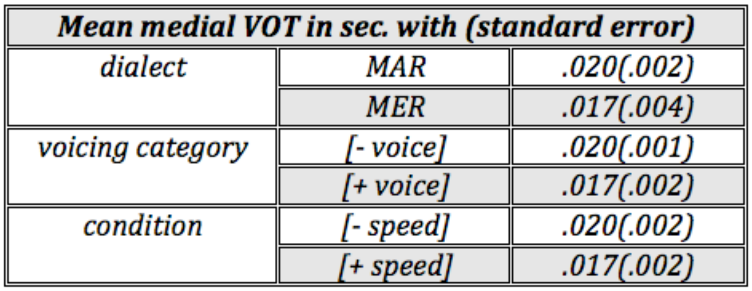


FIG. 19. Initial VOT Interaction of Voicing Category * Dialect.

Table 4 shows mean values for VOT in medial position.



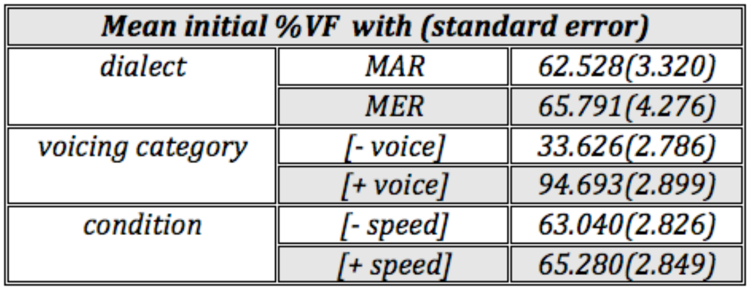
Mean medial VOT in sec. with (standard error)		
dialect	*MAR*	*.020(.002)*
	MER	*.017(.004)*
voicing category	*[- voice]*	*.020(.001)*
	[+ voice]	*.017(.002)*
condition	*[- speed]*	*.020(.002)*
	[+ speed]	*.017(.002)*

TABLE 4. Medial VOT.

4.3.3 Percent voicing throughout closure

%VF was closely tied to *voicing category*, with significant differences in both initial and medial positions, as well as significant interactions for *voicing category * dialect* in medial position. [+ Voice] was significantly higher than [- voice]. Speeded was higher than unspeeded, though the difference was nonsignificant. The discrepancy between speeded and unspeeded values was greater in initial than in medial position. It may be noted that the reflex for %VF in response to *voicing category* and *condition* is the opposite of what is seen in consonant duration. That is, while consonant durations decrease under the speeded condition, the percent of voicing throughout closure increases.

In initial position, MER values were greater than MAR values, but in medial position this trend was reversed. Differences in both initial and medial contexts were nonsignificant. As in VOT, the standard error for %VF was higher in medial position than in initial position, indicating greater variability. Table 5 shows mean %VF in initial position.



Mean initial %VF with (standard error)		
dialect	*MAR*	*62.528(3.320)*
	MER	*65.791(4.276)*
voicing category	*[- voice]*	*33.626(2.786)*
	[+ voice]	*94.693(2.899)*
condition	*[- speed]*	*63.040(2.826)*
	[+ speed]	*65.280(2.849)*

TABLE 5. Initial %VF.

Figure 20 shows the interaction of *voicing category * dialect* in medial position. As was the case with previous measures, [+ voice] values for both dialects are roughly equivalent. [- Voice] values are different, however, with MAR values being higher and MER lower. The interaction of *voicing category* * *dialect* in medial position was significant.

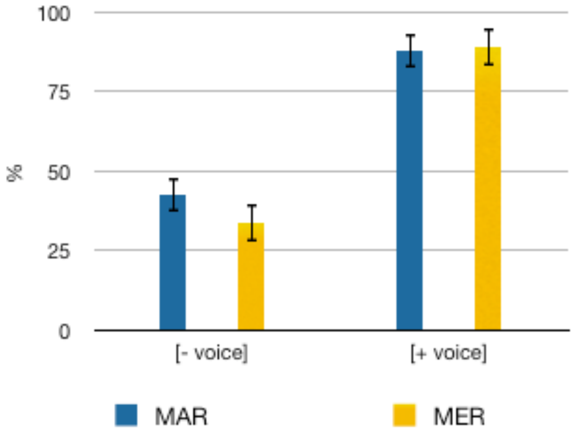
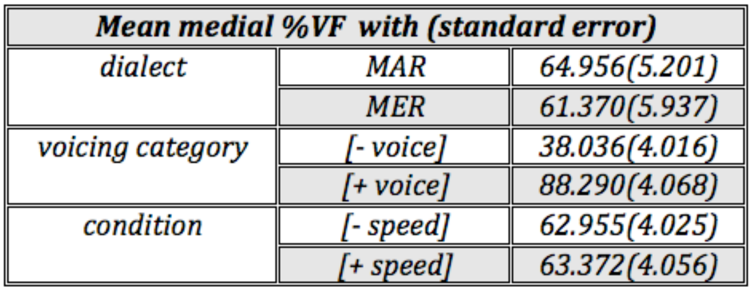


FIG. 20. Medial %VF Interaction Voicing Category * Dialect.

Table 6 displays the means for %VF in medial position.



Mean medial %VF with (standard error)		
dialect	*MAR*	*64.956(5.201)*
	MER	*61.370(5.937)*
voicing category	*[- voice]*	*38.036(4.016)*
	[+ voice]	*88.290(4.068)*
condition	*[- speed]*	*62.955(4.025)*
	[+ speed]	*63.372(4.056)*

TABLE 6. Medial %VF.

4.3.4 RMS amplitude

RMS was the one acoustic correlate of those surveyed in this sample that reliably distinguished between MAR and MER. Figure 21 illustrates these differences, elaborated on a spectrogram display (axes time and frequency).

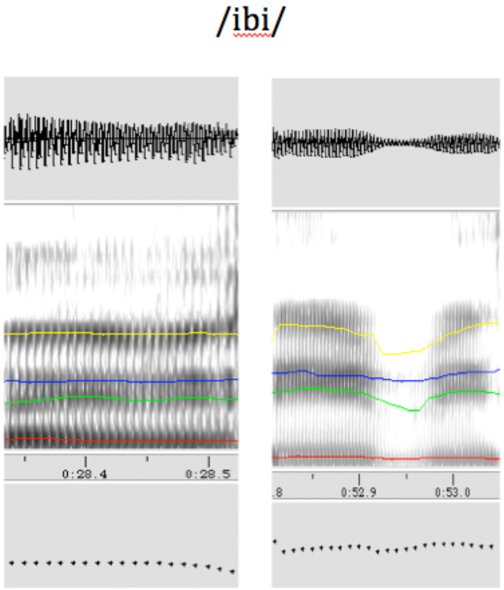
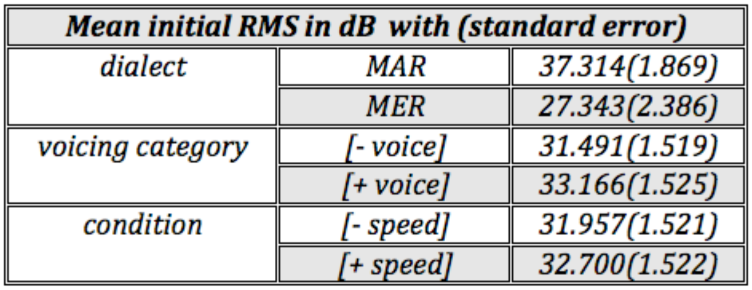

/ibi/

FIG. 21. Approximant and Consonant Realizations of /ibi/ Sequence. Speakers 3 (left, MAR) and 17 (right, MER) pronounce "Son biríbi cinco".

The spectrogram slice on the right represents a clear VCV sequence with a dramatic dip in spectral energy consistent with consonant closure. The spectrogram slice on the left shows no such break between segments, only a faint lessening of spectral energy and for a fraction of the time.

A main difference between the RMS profiles in the MAR and the MER populations (apart from the averages themselves) is that the initial measurement points for MER tended to begin at a much lower starting point than for MAR. The MAR starting points often resemble MER peaks. The arc is generally the same in shape (low, gradual rise, mid), but the overall climb is much steeper for the MER population (in other words, there is less excursion) and the upper limit values generally do not reach those found in MAR.

In addition to distinguishing between dialects, RMS was significant for *voicing category* in both initial and medial contexts. For *condition* it was significant in initial context. For the interaction of *voicing category * dialect*, RMS was significant in medial context and approached significance in initial context (.075). Table 7 shows mean RMS values in initial position.



Mean initial RMS in dB with (standard error)		
dialect	*MAR*	*37.314(1.869)*
	MER	*27.343(2.386)*
voicing category	*[- voice]*	*31.491(1.519)*
	[+ voice]	*33.166(1.525)*
condition	*[- speed]*	*31.957(1.521)*
	[+ speed]	*32.700(1.522)*

TABLE 7. Initial RMS.

Figure 22 shows the interaction of *voicing category * dialect* in initial position. In initial position, [± voice] values are more or less equivalent for MER. For MAR, [+ voice] values are higher.

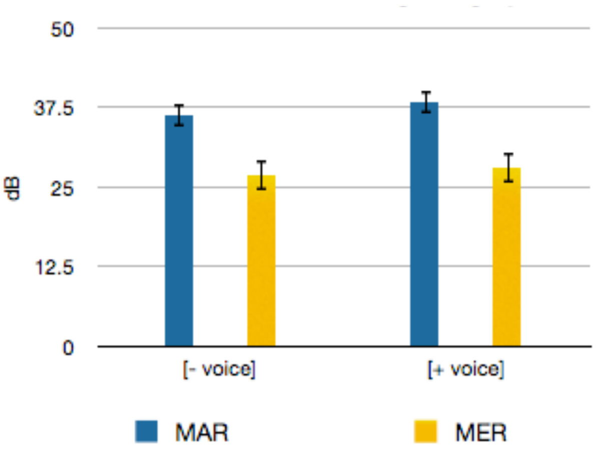
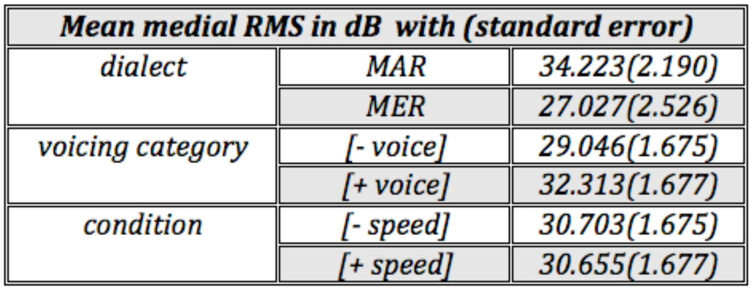


FIG. 22. Initial RMS Interaction Voicing Category * Dialect.

Table 8 show mean RMS values in medial position.



Mean medial RMS in dB with (standard error)		
dialect	*MAR*	*34.223(2.190)*
	MER	*27.027(2.526)*
voicing category	*[- voice]*	*29.046(1.675)*
	[+ voice]	*32.313(1.677)*
condition	*[- speed]*	*30.703(1.675)*
	[+ speed]	*30.655(1.677)*

TABLE 8. Medial RMS.

RMS values in medial position are slightly lower than those in initial position. MAR values in medial position are lower than those in initial position. MER values remain mostly unchanged across position. [+ Voice] RMS values are significantly higher than [-voice] values. This difference is greater in medial position. Speeded values are significantly larger than unspeeded values in initial position. In medial position there is no difference between speeded and unspeeded. Figure 23 shows the interaction of *voicing category* * *dialect* in medial position.

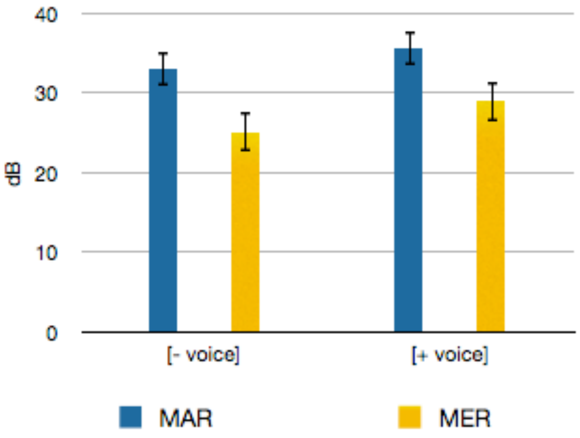
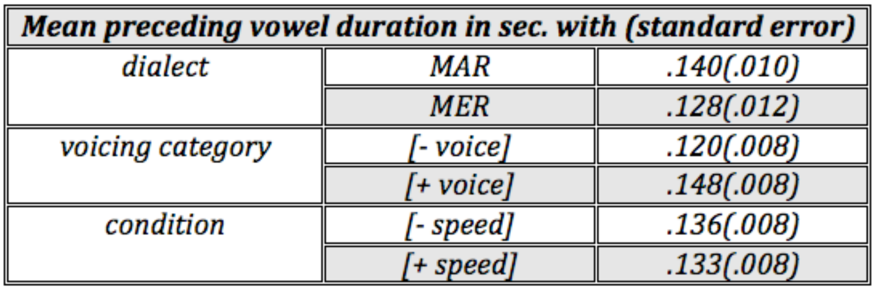


FIG. 23. Medial RMS Interaction Voicing Category * Dialect.

The main difference between the interaction of *voicing category* and *dialect* in initial and medial positions is seen in MER. In initial position, [± voice] are differentiated only slightly. In medial position, there is a greater separation between the category boundaries. [+ Voice] is substantially higher. In MAR, the relationship between [± voice] remains the same across initial and medial contexts ([+ voice] higher). The interaction of *voicing category * dialect* was significant for RMS in medial position.

4.3.5 Preceding vowel duration

Given the limitations on the design of this study, preceding vowel duration measures were valid for medial context only. In general, values were longer for MAR than for MER, though the difference did not achieve significance. Table 9 presents the means for preceding vowel duration.



Mean preceding vowel duration in sec. with (standard error)		
dialect	*MAR*	*.140(.010)*
	MER	*.128(.012)*
voicing category	*[- voice]*	*.120(.008)*
	[+ voice]	*.148(.008)*
condition	*[- speed]*	*.136(.008)*
	[+ speed]	*.133(.008)*

TABLE 9. Preceding Vowel Duration.

A main effect seen here was the significant interaction between *dialect * condition*, shown in Figure 24. The results for *preceding vowel duration* differ from those seen previously in this sample in that there is a qualitative difference in how the values for *condition* are aligned in each dialect. In MAR, values are larger in the unspeeded than in the speeded condition. In MER this trend is reversed. Additionally, the wide separation in values usually seen in MER is curiously absent here.

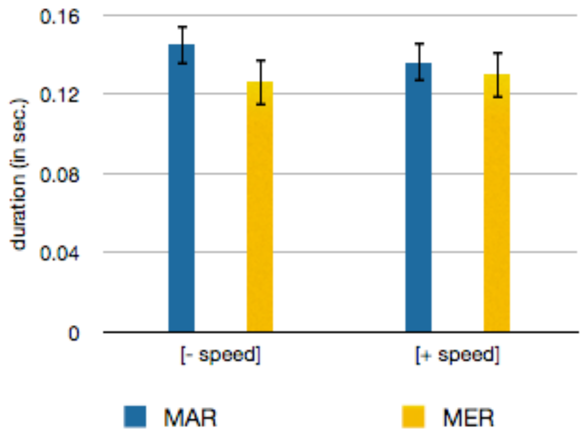
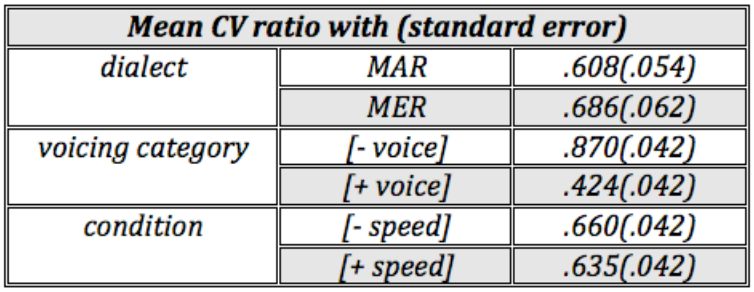


FIG. 24. Preceding Vowel Duration Interaction Condition * Dialect.

4.3.6 Consonant/ vowel duration ratio

Values for *CV ratio* were significant for *voicing category* and the interaction of *voicing category* * *dialect*. Values were higher overall for MER than for MAR (i.e., consonants longer and vowels shorter). Slightly higher CV ratios were seen in unspeeded than in speeded contexts. The difference between unspeeded and speeded was nonsignificant at .096. Table 10 presents the means for CV ratio.



Mean CV ratio with (standard error)		
dialect	*MAR*	*.608(.054)*
	MER	*.686(.062)*
voicing category	*[- voice]*	*.870(.042)*
	[+ voice]	*.424(.042)*
condition	*[- speed]*	*.660(.042)*
	[+ speed]	*.635(.042)*

TABLE 10. CV Ratio.

There was a significant interaction between *voicing category* * *dialect*, as shown in Figure 25. Values for [+ voice] converge, but the [- voice] values differ substantially. Note the wider separation between [± voice] values in MER than in MAR, as has been observed in other correlates.

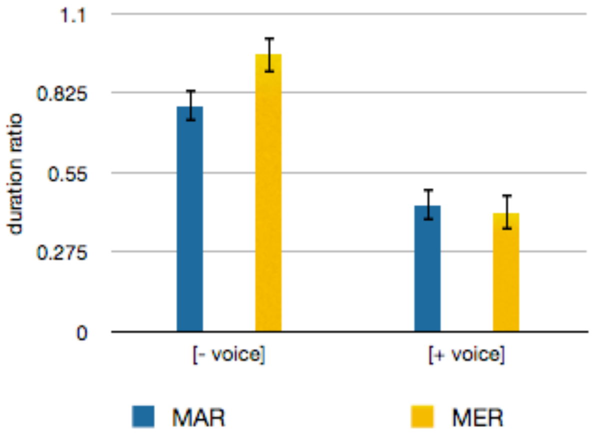
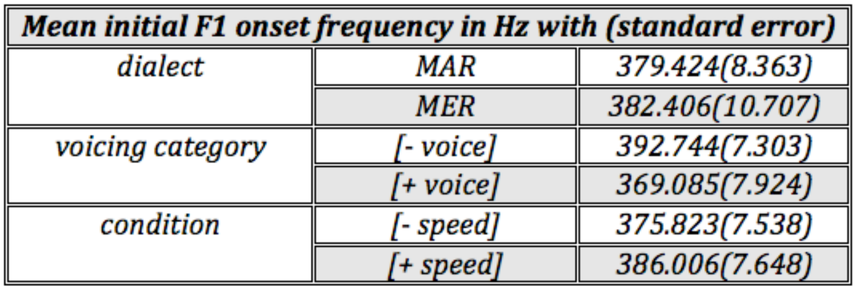


FIG. 25. CV Ratio Interaction Voicing Category * Dialect

4.3.7 F1 onset frequency

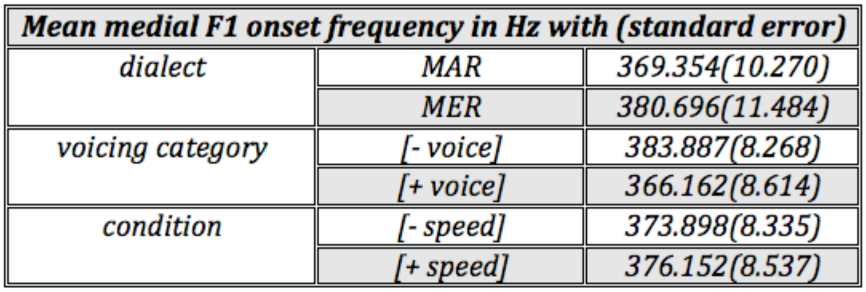
F1 onset values were significant for *voicing category* in both initial and medial contexts. In initial context, [- voice] was significantly greater than [+ voice], but this trend was reversed in medial context. In both contexts speeded F1 onset values were higher than unspeeded, though these differences were nonsignificant. MER values were greater than those seen in MAR. The difference was nonsignificant in both initial and medial contexts. Table 11 shows the initial means for F1 onset.



Mean initial F1 onset frequency in Hz with (standard error)		
dialect	MAR	379.424(8.363)
	MER	382.406(10.707)
voicing category	[- voice]	392.744(7.303)
	[+ voice]	369.085(7.924)
condition	[- speed]	375.823(7.538)
	[+ speed]	386.006(7.648)

TABLE 11. Initial F1 Onset Frequency.

Overall, initial values of F1 onset frequency were higher than those in medial context. This was sustained across all categories. There was a bigger difference in F1 onset frequency values by context for MAR than for MER. There was a greater difference between speeded and unspeeded in initial position than in medial position. Table 12 shows the means for F1 onset values in medial position.

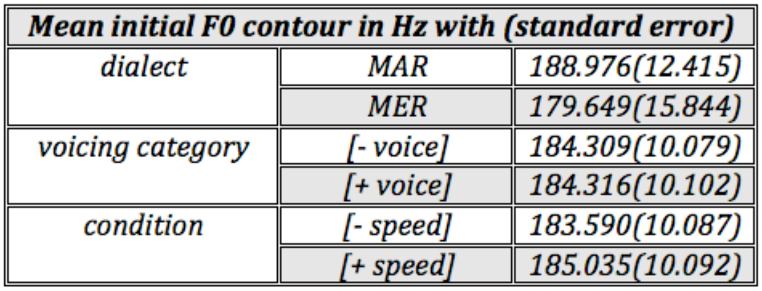


Mean medial F1 onset frequency in Hz with (standard error)		
dialect	MAR	369.354(10.270)
	MER	380.696(11.484)
voicing category	[- voice]	383.887(8.268)
	[+ voice]	366.162(8.614)
condition	[- speed]	373.898(8.335)
	[+ speed]	376.152(8.537)

TABLE 12. Medial F1 Onset Frequency.

4.3.8 F0 contour following closure

F0 contour values were discovered not to differ significantly for any of the independent variables or interactions. Values were roughly equal for the different subcategories of voicing category and condition in initial position. In medial position, [+ voice] and speeded values were slightly higher. Table 13 shows F0 contour means in the initial context.



Mean initial F0 contour in Hz with (standard error)		
dialect	*MAR*	*188.976(12.415)*
	MER	*179.649(15.844)*
voicing category	*[- voice]*	*184.309(10.079)*
	[+ voice]	*184.316(10.102)*
condition	*[- speed]*	*183.590(10.087)*
	[+ speed]	*185.035(10.092)*

TABLE 13. Initial F0 Contour.

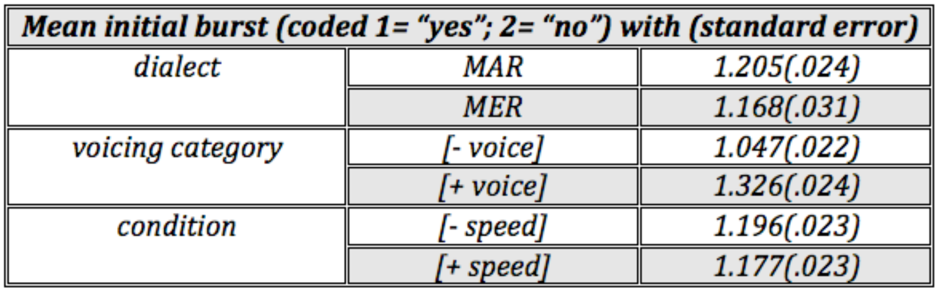
MAR values for F0 contour were found to be higher than MER values in initial position. However, this trend was reversed in medial position. The difference between the two dialects was not large enough to be statistically significant. In general, values across all categories were higher in medial than in initial position. Table 14 shows F0 contour means in medial position.

Mean medial F0 contour in Hz with (standard error)		
dialect	MAR	197.205(14.666)
	MER	217.159(16.887)
voicing category	[- voice]	206.470(11.231)
	[+ voice]	207.894(11.263)
condition	[- speed]	206.038(11.237)
	[+ speed]	208.326(11.257)

TABLE 14. Medial F0 Contour.

4.3.9 Presence/ absence of release burst

Burst was significant for voicing category in both initial and medial contexts. In addition, there was a significant interaction between *voicing category * dialect* in medial context. The means for *burst* in initial position are lower than those in medial position across all categories. According to the way bursts were coded in this data set, this means there were fewer bursts in medial than in initial position. In initial position, there were more bursts in MER than in MAR, but in medial position, this trend was reversed. There were significantly more bursts in [- voice] than in [+ voice] in both contexts. There were more bursts in the speeded condition than in the unspeeded condition, though the difference was nonsignificant. In medial context the difference between speeded and unspeeded was slight. Table 15 presents the means for bursts in initial position.



Mean initial burst (coded 1= "yes"; 2= "no") with (standard error)		
dialect	*MAR*	*1.205(.024)*
	MER	*1.168(.031)*
voicing category	*[- voice]*	*1.047(.022)*
	[+ voice]	*1.326(.024)*
condition	*[- speed]*	*1.196(.023)*
	[+ speed]	*1.177(.023)*

TABLE 15. Initial Burst.

Figure 26 shows the interaction of *voicing category* * *dialect* in medial position. Bursts were coded as "2= no burst" and "1= burst". There were more bursts in MER [- voice] than in MAR and a greater separation between [± voice] categories with regard to burst. MAR showed a higher incidence of bursts in [+ voice].

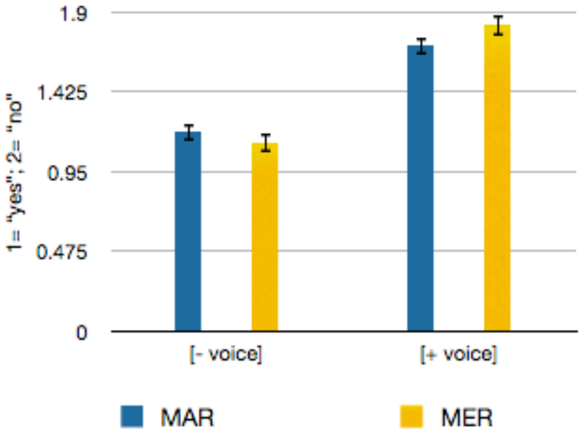
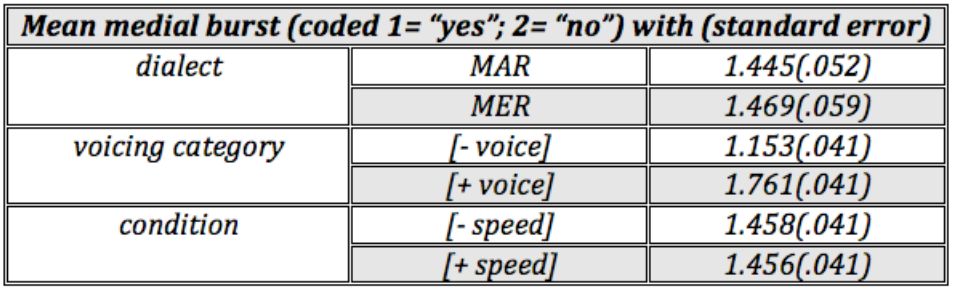


FIG. 26. Medial Burst Interaction Voicing Category * Dialect.

Table 16 shows the means for bursts in medial position.



Mean medial burst (coded 1= "yes"; 2= "no") with (standard error)		
dialect	*MAR*	*1.445(.052)*
	MER	*1.469(.059)*
voicing category	*[- voice]*	*1.153(.041)*
	[+ voice]	*1.761(.041)*
condition	*[- speed]*	*1.458(.041)*
	[+ speed]	*1.456(.041)*

TABLE 16. Medial Burst.

4.4 Statistics

The substantial number of approximant realizations in the sample has already been mentioned. 7.25% of overall responses in initial position were catalogued as approximants. Of these, 95% of these fell into the [+ voice] category, as compared with 4% [- voice]. MAR responses accounted for 82.48% of responses labeled as approximants. 17.5% came from MER. The responses were almost evenly split along *condition*, with 47.44% unspeeded and 52.55% speeded.

Fewer medial approximants were identified in the sample than initial approximants- 6.03% compared to 7%. There was a much higher number of [- voice] approximants in medial than in initial contexts- 14.91% as opposed to 4%. 85.08% of medial approximants landed in the [+ voice] category. More MER responses were labeled as approximants in the medial context than in the initial- 37.71% (62.28% for MAR). Slightly more unspeeded approximants were recorded in medial context than were seen in the initial context (49.12% unspeeded, 50.87% speeded, compared to 47.44% and 52.55%, respectively). Results for condition may not be entirely reflective of the behavior of approximants under different rate conditions. Reasons for this are: 1) the number of unspeeded to speeded prompts was unequal (46 unspeeded to 104 speeded). This may give an advantage to the speeded condition. Another factor to consider is that, 2) there were many more speaker errors in the speeded condition than in the unspeeded condition. A number of the speeded responses had to be discarded.

The concepts of place of articulation, speaker variation and variation by gender were not primary objects of analysis in this investigation and thus were not included in the main model. In this section, however, I will make some general remarks concerning their distribution. This may be of use to future researchers.

In the current study, a multivariate ANOVA test indicated that there were significant differences between place of articulation values for most of the correlates surveyed, including: consonant duration, %VF, VOT, F1 onset, preceding vowel duration and CV ratio. As was the case with place of articulation, a multivariate ANOVA analysis of the effect of speaker on the dependent variables included in this study showed results that were highly significant, even when the data was sorted by dialect. All acoustic correlates were affected, with the exception of burst and VOT in medial position. There were gender differences as well- related to consonant duration, VOT, F0, F1 onset, preceding vowel duration and RMS. In this section I describe the results of the statistical analysis, beginning with fixed effects and interactions, followed by tables of the *p* values.

The main statistical model used in this study tested fixed effects of *voicing category*, *dialect* and *condition*. Interactions between *voicing category * dialect* and *dialect * condition* were tested as well. In initial and medial positions, the only dependent variable to return a statistical difference for dialect was *RMS* (*F[1, 877.338] = 36.159, p< .05 initial; F[1, 1152.618] = 168.551, p< .05 medial*).

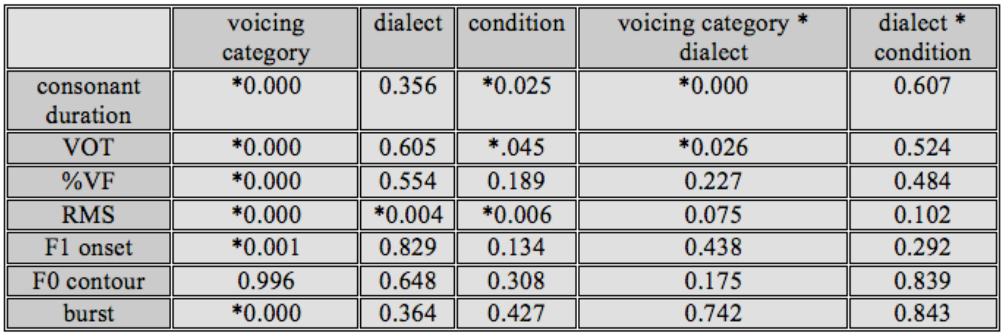
The following measures showed significant differences between [± voice] categories when the measures occurred in initial position: *consonant duration* (*F[1,*

$860.954] = 581.420, p < .05$), *VOT* ($F[1, 733.716] = 147.771, p < .05$), *%VF* ($F[1, 886.540] = 1234.792, p < .05$), *RMS* ($F[1, 877.338] = 36.159, p < .05$), *F1 onset* ($F[1, 892.229] = 11.741, p < .05$), and *burst* ($F[1, 893.853] = 138.102, p < .05$). In medial position, the following dependent variables were significant for *voicing category*: *consonant duration* ($F[1, 1178.892] = 597.501, p < .05$), *%VF* ($F[1, 1181.959] = 825.554, p < .05$), *RMS* ($F[1, 1152.618] = 168.551, p < .05$), *preceding vowel duration* ($F[1, 1184.492] = 149.473, p < .05$), *CV ratio* ($F[1, 1179.300] = 873.641, p < .05$), *F1 onset* ($F[1, 1165.069] = 6.578, p < .05$), and *burst* ($F[1, 1184.423] = 819.863, p < .05$).

For the fixed variable *condition*, *consonant duration* and *VOT* showed significant differences between speeded and unspeeded for both initial and medial contexts: ($F[1, 860.954] = 581.420, p < .05$ *initial consonant duration*; $F[1, 1178.892] = 597.501, p < .05$ *medial consonant duration*; $F[1, 733.716] = 147.771, p < .05$ *initial VOT*; $F[1, 671.473] = 9.522, p < .05$ *medial VOT*). *RMS* was significant for *condition* in initial position only ($F[1, 877.338] = 36.159, p < .05$).

For the interaction of *voicing category * dialect*, *consonant duration* and *VOT* were significant in initial context ($F[1, 860.954] = 16.442, p < .05$; $F[1, 733.716] = 4.988, p < .026$). In medial context, *voicing category * dialect* was significant for *consonant duration* ($F[1, 1178.892] = 11.783, p < .05$), *%VF* ($F[1, 1181.959] = 7.557, p < .05$), *RMS* ($F[1, 1152.618] = 5.334, p < .05$), *CV ratio* ($F[1, 1179.300] = 45.428, p < .05$), and *burst* ($F[1, 1184.423] = 18.723, p < .05$). *Dialect * condition* was found to be significant for *preceding vowel duration* only ($F[1, 1184.125] = 7.861, p < .05$).

Table 17 shows *p* values for all variables in initial position.

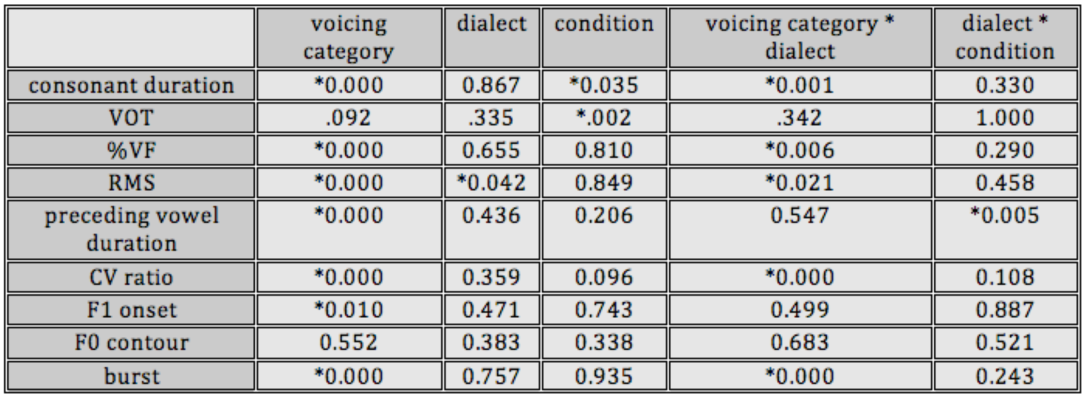


	voicing category	dialect	condition	voicing category * dialect	dialect * condition
consonant duration	*0.000	0.356	*0.025	*0.000	0.607
VOT	*0.000	0.605	*.045	*0.026	0.524
%VF	*0.000	0.554	0.189	0.227	0.484
RMS	*0.000	*0.004	*0.006	0.075	0.102
F1 onset	*0.001	0.829	0.134	0.438	0.292
F0 contour	0.996	0.648	0.308	0.175	0.839
burst	*0.000	0.364	0.427	0.742	0.843

*Represents a significance level ≤ 0.05.

TABLE 17. *p* Values: Initial Position.

Table 18 shows *p* values for all variables in medial position.



	voicing category	dialect	condition	voicing category * dialect	dialect * condition
consonant duration	*0.000	0.867	*0.035	*0.001	0.330
VOT	.092	.335	*.002	.342	1.000
%VF	*0.000	0.655	0.810	*0.006	0.290
RMS	*0.000	*0.042	0.849	*0.021	0.458
preceding vowel duration	*0.000	0.436	0.206	0.547	*0.005
CV ratio	*0.000	0.359	0.096	*0.000	0.108
F1 onset	*0.010	0.471	0.743	0.499	0.887
F0 contour	0.552	0.383	0.338	0.683	0.521
burst	*0.000	0.757	0.935	*0.000	0.243

*Represents a significance level ≤ 0.05.

TABLE 18. *p* Values: Medial Position.

Table 19 shows how the results compare with predictions made in Chapter 1.

	prediction	**results**
consonant duration	longer in MER than in MAR; [- voice] longer than [+ voice] within MER; less separation between [voice] in MAR	*longer in MER than in MAR; [- voice] longer than [+ voice] within MER; less separation between [voice] in MAR*
VOT	important indicator of [voice] in initial position; more important indicator of [voice] in MER than in MAR	*important indicator of [voice] in initial position; more important indicator of [voice] in MAR than in MER*
%VF	higher for [- voice] in MAR than in MER	*higher for [- voice] in MAR than in MER in medial position only*
RMS	higher for [- voice] in MAR than in MER; no difference in [+ voice]	*MAR higher for both [± voice]*
preceding vowel duration	longer for [+ voice] than for [- voice] in both dialects; less separation between [voice] in MAR	*longer for [+ voice] than for [- voice] in both dialects; more separation between [voice] in MAR*
CV ratio	more variability by [voice] in MER than in MAR; more reliable across speeded and unspeeded conditions	*more variability by [voice] in MER than in MAR; no statistical difference between speeded and unspeeded conditions*
F1 onset	lower in [+ voice] than in [- voice] across dialects; higher in MER [- voice] than in MAR	*lower in [+ voice] than in [- voice] across dialects; no statistical difference between MER and MAR [- voice]*
F0 contour	higher in [- voice] than in [+ voice] in both dialects; higher in MER [- voice] than in MAR	*no statistical differences observed*
burst	more bursts for [- voice] than for [+ voice] across dialects; higher proportion of bursts overall for MER	*more bursts for [- voice] than for [+ voice] across dialects; higher proportion of bursts overall for MER*

Indicates a divergence between results and predictions.

TABLE 19. Comparison of Predictions with Results.

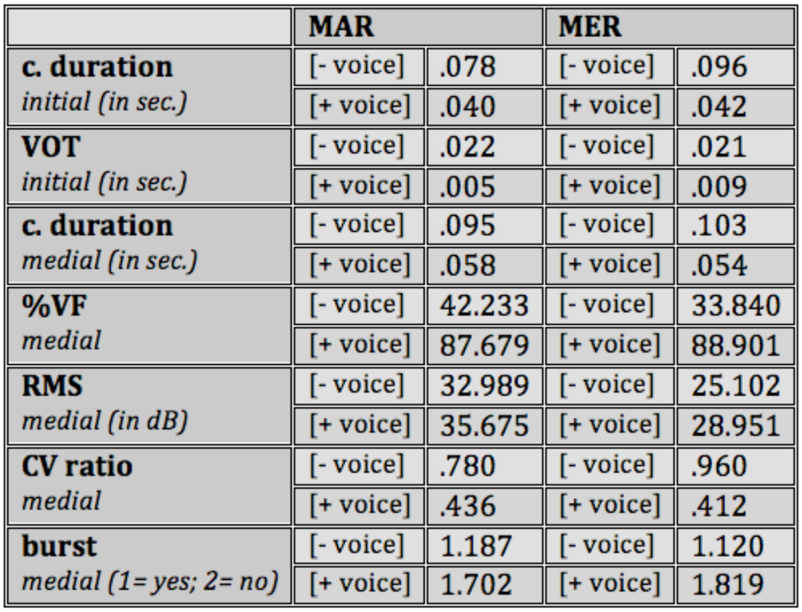
Chapter 5: Discussion of results

5.1 [± voice] findings

Eight of the nine proposed acoustic correlates of [voice] evaluated in this study showed significant differences between values for [± voice] categories. The only acoustic correlate to not show a significant difference between [± voice] was F0 contour. Of the correlates that showed significant differences between [± voice], almost all upheld these differences in both initial and medial contexts. The exception was VOT, which showed a significant difference in initial context only. Medial VOT values did, however, approach significance.

5.2 Mérida vs. Margarita findings

In this study it was found that the main differentiator between MAR and MER dialects was RMS. RMS values were higher in MAR than in MER, in both initial and medial contexts, in both [± voice]. Six out of the nine acoustic correlates surveyed in this study showed significant interactions between *voicing category* * *dialect*. In other words, six out of nine of the correlates show substantially different behavior between [± voice] categories when taking into account the factor of *dialect*. Table 20 shows [± voice] values for MAR and MER for the measures that showed a significant interaction between *voicing category* * *dialect*.



	MAR		MER	
c. duration *initial (in sec.)*	[- voice]	.078	[- voice]	.096
	[+ voice]	.040	[+ voice]	.042
VOT *initial (in sec.)*	[- voice]	.022	[- voice]	.021
	[+ voice]	.005	[+ voice]	.009
c. duration *medial (in sec.)*	[- voice]	.095	[- voice]	.103
	[+ voice]	.058	[+ voice]	.054
%VF *medial*	[- voice]	42.233	[- voice]	33.840
	[+ voice]	87.679	[+ voice]	88.901
RMS *medial (in dB)*	[- voice]	32.989	[- voice]	25.102
	[+ voice]	35.675	[+ voice]	28.951
CV ratio *medial*	[- voice]	.780	[- voice]	.960
	[+ voice]	.436	[+ voice]	.412
burst *medial (1= yes; 2= no)*	[- voice]	1.187	[- voice]	1.120
	[+ voice]	1.702	[+ voice]	1.819

TABLE 20. Comparison of MAR and MER in 6 Key Measures.

It is of note that most significant interactions occurred in medial position. With the exception of RMS and %VF, MER values were generally higher and the differences between [± voice] greater. The higher RMS values for MAR have already been noted. With respect to %VF, initial MER values were found to be higher than those in MAR. In medial position, however, MAR values were higher than MER in [- voice] and slightly lower in [+ voice], in effect lessening the gap between [± voice]. In fact, that is the main generalization to be made in this study about the way [voice] is encoded in the production of two consonant subsystems: highland and lowland. In MER, the highland dialect, the gap between categories is lessened. This is true in consonant duration, %VF, CV ratio and burst. It is not, however, true in VOT, where the gap is slightly wider between MAR

[± voice], or in RMS, where the difference between the two dialects lies in overall values, not in the way the categories relate to one another. In terms of which [voice] values tend to differ from one another, it is sometimes the [- voice] value (as in consonant duration, %VF, and CV ratio). Other times it is the [+ voice] value (as in VOT). The burst measure shows inconsistencies. There is a wide separation between [± voice] values in MER, but neither the [- voice] nor the [+ voice] values are close to those of MAR. In general, MER showed a wider separation between values for [± voice]. More of these differences seemed to converge on the [- voice] rather than the [+ voice] category values, confirming results by Lewis (2000). Finally, in the interaction of *dialect * condition* in preceding vowel duration, the unspeeded values for MAR were greater than those in the speeded condition. In MER, however, this trend was reversed.

General findings not related to the research questions were discussed in §4.1. These include durational and spectral patterns particular to each dialect, for example the short V1 typical of MAR speakers and the aspiration and creaky voice associated with MER speakers. Such observations fall outside of the umbrella of analysis for the present study, but may merit further investigation (see §6.2). It was noted in §4.1 that there was variability within the dialect populations and that more overall variability occurred in the speech of female participants than in that of male participants. In addition, approximant realizations were more prevalent among male speakers than among female speakers. With regard to the question of variability, this result may simply arise from the fact that there were more MAR participants than MER participants and more female speakers than male speakers overall. It is possible that the similarities observed between the groups of male speakers in the increased number of approximant realizations occur as a matter of chance. To determine whether or not this is the case, a more thorough examination is necessary. A new investigation would benefit from a more balanced subject pool and would necessarily exercise stricter control over potentially confounding factors such as socioeconomic status, job history and education.

With regard to the greater variability found in the female MAR population as compared to the female MER population, it may be that one or more of the above-mentioned socioeconomic factors influenced the results. As part of the selection process, participants in the current study were interviewed about their occupations and level of education, in addition to other demographic information gathered- such as age, place of residence, and the amount of time residing in the community. This information was used to ensure that the participants in the study would form a reasonably homogenous group. Overall, the goal was met, as evidenced by the results reviewed in the beginning of this section. MAR and MER speakers are differentiated from one another in this study by RMS and by the interaction of *voicing category * dialect* in consonant duration, VOT, %VF, CV ratio and burst. Nonetheless, a closer examination of the language patterns that are seen within and across the two dialects would be informative.

5.3 Blocks findings

In the present study, some of the acoustic correlates found to be significant for *voicing category* were also significant for *condition*. These include: consonant duration, VOT, and RMS. RMS was also the main indicator of dialect in this study. One question these results raise is how reliable acoustic correlates may be as voicing cues in perception if they are subject to substantial variability in rate.

Only one correlate (preceding vowel duration) returned a significant difference for the interaction of *dialect * condition*. In MAR, preceding vowel durations were longer in the unspeeded condition than in the speeded condition. In MER, the reverse was true. One other difference to note is that the separation between [± voice] boundaries was greater for MAR than for MER. This is a divergence from the behavior of most of the other acoustic correlates under survey here. It may be the case that since the gap between boundaries tends to be greater in MER, the durational relationships between consonant and preceding vowel in that dialect are of less importance. Indeed, there is only a slight difference in MER with regard to [± voice] values in the preceding vowel duration measure.

Studies in clear speech have shown that acoustic correlate values can shift dramatically under different speaking conditions (Krause & Braida, 2002, 2003; Lewis, 2001 for Spanish). In a speeded situation, where intelligibility of speech may be compromised, duration and other properties of the speech signal have been identified as contributing to speech clarity. Speech rate modifications tend to affect vocalic segments more than consonantal segments of speech (Werner & Keller, 1994). Within stop consonants, VOT and other durations linked to the transitional portions of VC or CV boundaries in stop consonants vary less than the durations of fricatives or nasals. The different results for preceding vowel duration make sense in this light. A more vowel-like pronunciation (such as that seen in MAR) will show greater fluctuation in relative duration measures under duress. Furthermore, consonant duration measures in the current study varied significantly in both initial and medial contexts according to *condition* (speeded/ unspeeded). VOT and RMS varied according to *condition*- RMS in initial context only. There were no differences between MAR and MER with regard to how consonant duration, VOT or RMS behaved in speeded or unspeeded conditions.

5.4 Findings for 9 acoustic correlates

5.4.1 Consonant duration

Consonant duration was an important indicator of *voicing category*, *condition* and *dialect*. [- Voice] values were longer than [+ voice] values, confirming results from previous studies. MER consonant durations were longer than those in MAR, as was predicted in §1.2.2.2.1. In particular, the MER [- voice] values were substantially higher than those for MAR. The difference was greater in initial than in medial context.

5.4.2 Voice onset time

VOT has been identified as an important acoustic cue for distinguishing stop consonant boundaries in a variety of languages with different laryngeal settings (Lisker & Abramson, 1970a,b; Ladefoged & Maddieson, 1996). In the present study, there were significant differences in values across several categories: *voicing category*, *condition*, and the interaction of *voicing category * dialect*. Significant differences showed up in both initial and medial positions. As occurred with consonant duration, medial values for VOT tended to be longer than those in initial position. The exception was [- voice] initial, which was longer than its medial counterpart. One twist is that MAR [± voice] values show a larger category separation than is seen in the MER values. In initial position, MAR [- voice] is longer and [+ voice] shorter than in MER. Furthermore, in medial position, MAR VOT durations are slightly longer than those found in MER.

VOT mean values in this study were higher than has been previously reported for VOT values in Spanish. In initial position, overall VOT means were 7 ms. (voiced) and 22 ms. (voiceless). In medial position, the overall means were 17 ms. (voiced) and 20 ms. (voiceless). The range of values was somewhat extreme- from -74 ms. to 38 ms. (in initial position) and from -56 ms. to 108 ms. (in medial position). Despite this, extreme values were not common. Most minimum values hovered around 0 ms. Maximums were more variable- fluctuating from around 10 ms. to 40 ms. The presence of just a few extreme values, especially on the positive end of the scale, seemed to pull the means higher than what is usual for Spanish. Normally, the [± voice] VOT boundary for Spanish hovers around 0 ms (Lisker & Abramson, 1970b; Lasky et al., 1975; Hay, 2005; Benkí, 2005).

In the present study, high VOT values were observed for both MAR and MER participants. It may be that something about the structure of the design of the experiment prompted the unusually high values. Recall that in §4.1, irregularities dealing with speaker response to prompts were discussed. Many times these irregularities were of a durational nature- excessively long pauses or vowel durations. Duration measures seem especially sensitive to conditions that slow down or speed up speech. Part of the task for participants in this study was to create a nonsense word following a specific pattern. This required speakers to craft the appropriate response either before they spoke or as they were speaking. Many speakers seemed to process the stimuli “online”, sounding the response out. The idea behind the experimental design for this study was to create a situation where potential effects of literacy and/or word-frequency would be neutralized, allowing spontaneous speech patterns to emerge. The relative cognitive load of the task, however, may have tilted the results in other, unexpected ways.

5.4.3 Percent voicing throughout closure

In the present study, %VF was significant for voicing category in both initial and medial positions. Values in initial position were higher. Normally, in an utterance-initial position, it is the case that overall voicing is less when compared with that of segments in other contrasts, because it takes some time for vocal fold vibration to engage. For this reason, there is almost always an initial period of voicelessness. One reason the results from the present study may diverge from those of previous studies is that word-initial segments in this study were not utterance-initial. They were preceded by /n/. Extensive coarticulation was seen between the nasal and TS1. Therefore, voicing was already engaged when the gesture for the oral stop closure began.

One thing to note is that while consonant durations tended to decrease under the speeded condition, the percent of voicing throughout closure increased. Cho and Keating (2001) reported a similar result for Korean. Additionally, they found that the interval of voicing during stop closure was shorter in higher domain-initial positions and longer in lower domain-initial positions. They attribute this result to a timing difference between domain positions with regard to glottal abduction.[15] Aerodynamic constraints may be a factor as well. The shorter the segment, the easier it is to maintain voicing throughout the duration of the closure. Longer durations are subject to energy leakage. Periodicity often lessons or is lost towards the end of a closure.

In initial position, MER values were greater than MAR values, but in medial position, this trend was reversed. The interaction in medial position between voicing category and dialect indicates that the differences between the two dialects with respect to %VF lies in the [- voice] category. [+ Voice] values in MER and MAR are functionally equivalent, but the [- voice] values for MAR are higher than those for MER. It is of interest to note that there is more overall separation between voicing categories and speeded and unspeeded conditions in the initial position, but less separation between dialects. As in many of the other acoustic correlates under review here, dialectal differences tend to surface in medial position.

5.4.4 RMS amplitude

RMS was shown to uphold the difference between the two dialects, MAR and MER in both initial and medial contexts, though the level of significance in the medial context was lower. MAR means for both [± voice] were higher than in MER. Additionally, it was

---

[15] An earlier glottal abduction gesture in the higher positions will contribute to a longer voiceless interval, providing more contrast with the shorter voiced interval.

observed that the initial measurement points for MER RMS occurred at much lower values than those for MAR. This is consistent with an account of MER as a strong-closure dialect. If there is complete closure before the beginning of the opening gesture, the beginning amplitude measure will be much lower than if the closure is incomplete at the time of the initial trough measurement (recall that in this study, the RMS amplitude window was taken as a trough-peak-trough measure). RMS, then, is a way of getting at sonority (or intensity) through amplitude.

RMS has not been discussed much as an acoustic correlate of [voice] in Spanish. This is one reason the results of the present study may be of interest. RMS was found to be significant for *voicing category* in both initial and medial contexts. It was also significant for the interaction of *voicing category* and *dialect* in medial context and approached significance in initial context. It was found that the relationship between the categories remained the same in the two dialects ([+ voice] higher than [- voice]), but that the MER values were shifted down from what was seen in MAR. The difference was greater in medial position than in initial position.

In a study on the acoustic properties of clear speech, Krause and Braida (2003) have shown that higher RMS levels can provide an intelligibility advantage in both clear/normal and clear/slow speech situations. This may account for the strong RMS showing in the current results across *voicing category*, *dialect* and *condition* (RMS significant for *condition* in initial position). As occurred with %VF, RMS tended to be higher for initial over medial contexts and for speeded versus unspeeded conditions.

5.4.5 Preceding vowel duration

The most important effect for preceding vowel duration was the interaction between *dialect* and *condition*. One unusual aspect is that this is the only correlate that shows an inverse shift in the way [± voice] relate to each other in the two dialects. MAR preceding vowel durations were longer in the unspeeded than in the speeded condition. In MER the reverse was true. Unspeeded values were shorter and speeded values were longer. It must be noted, however, that the difference between [± voice] is slight.

5.4.6 Consonant/ vowel duration ratio

CV ratio values were significant for *voicing category* and the interaction of *voicing category* and *dialect*. MER values were higher than those for MAR. In other words, the consonants tended to be longer in that dialect and the vowels shorter. Higher CV ratios were also found in unspeeded contexts than in speeded contexts. In the interaction between *voicing category* and *dialect*, [+ voice] values appear to converge, but the [-voice] values diverge in that the MER means are much higher than those found for MAR. This is consistent with the consonant duration data and with the preceding vowel duration data (MAR values longer). Preceding vowel duration, however, did not show significance for the interaction of *voicing category * dialect* and one key difference between the consonant duration measure and the CV ratio measure is in the size of the effect. The effect of the duration difference between MER and MAR is larger when taken as a ratio with the preceding vowel than when the consonant duration stands alone. This result is not inconsistent with the view that the frequently occurring feature combinations in the languages of the world come about because those combinations maximize perceptual distinctiveness through the mechanism of feature enhancement (Stevens & Keyser, 1989; Diehl & Kluender, 1989; Kingston & Diehl, 1995; Diehl & Lindblom, 2004).

5.4.7 F1 onset frequency

F1 onset values were significant for voicing category in both initial and medial contexts. In initial context, [- voice] was greater than [+ voice], but the opposite was true in medial context. [+ Voice] was greater than [- voice]. In both contexts, speeded F1 onset values were higher than unspeeded. There was a greater difference between speeded and unspeeded in initial position than in medial position. MER values were greater than MAR values.

F1 onset values were higher in initial than in medial contexts, as expected. Higher F1s have been observed for clear/normal speech than for conversational/normal speech (Krause & Braida, 2003). In the present study, F1 onsets were higher in the speeded condition and lower in the unspeeded condition (though in medial context the difference was marginal). Krause and Braida (2003) attribute their results to clear/normal speech being produced at higher intensities than conv/normal speech since louder speech is typically obtained with a larger jaw opening, resulting in decreased tongue height and increased F1. In the current study, the higher F1 onset values found in the speeded condition may correspond to greater articulatory tension as speakers endeavor to respond to the prompt quickly. The larger F1 onset values for MER were predicted based on the impression of more consonant closure for this dialect. A gesture of complete stop closure versus partial closure would entail higher F1 onset values upon the pressure of release.

5.4.8 F0 contour following closure

F0 contour values were found not to differ significantly for any of the fixed variables or interactions. Values were roughly equal for the different subcategories of voicing category and condition in initial position. In medial position, [+ voice] and speeded values were slightly higher. MAR values for F0 contour were higher than MER values in intial position. The trend was reversed in medial position. Overall, the difference between the two dialects was not large enough to be statistically significant. F0 contour values tended to be higher across all categories in medial than in initial position.

F0 contour might have been expected to be an indicator of [voice] based on research that F0 contributes to a phenomenon known in speech perception as the low-frequency property (Stevens & Blumstein, 1981; Kingston & Diehl, 1994). Under the low-frequency property, vocal fold vibration during the consonant constriction, a low F1 and a low F0 all contribute to the presence of low-frequency energy in or near the constriction. The positive correlation of two or more of these properties has been shown to influence voicing judgments in perception (Diehl, Castleman & Kingston, 1995). F0, however, besides being a possible acoustic correlate of [voice] is also the main carrier of intonation in languages. The absence of an F0 contour effect in the present study is inconclusive. It may mean that F0 contour is not an important acoustic correlate of [voice] in Spanish, or it may simply indicate a flat intonation in the way informants delivered the prompts. In other words, the null effect of F0 contour on [± voice] may have come about as an artifact of the task.[16]

5.4.9 Presence/ absence of release burst

Burst was significant for *voicing category* in both initial and medial contexts. In addition, there was a significant interaction between *voicing category * dialect* in medial context. The means for burst in initial position are lower than those in medial position across all categories; there were fewer bursts in medial than in initial position. In initial position, there were more bursts in MER than in MAR, but in medial position, this trend was reversed. There were more bursts in [- voice] than in [+ voice] in both contexts. There were more bursts in the speeded condition than in the unspeeded condition, though in medial context the difference was slight. There were more bursts in MER [- voice] than in MAR and a greater separation between [± voice] categories with regard to burst. MAR showed a higher incidence of bursts in [+ voice].

5.5 Answers to research questions

Recall that the research questions were as follows:

- What robust acoustic correlates to [voice] emerge in the two dialects under consideration (Mérida and Margarita- henceforth, MER and MAR)?
- Are the acoustic correlate inventories substantially different in the two dialects?
- Assuming differences are found, what are the implications for [voice] when

---

[16] Thanks to Randy Diehl (personal correspondance) for help in understanding this issue.

consonantal parameters of articulation differ?

Secondary research questions included:

- Do observable differences between the two dialects hold across conditions of prosody and rate of speech?
- Are the observations for dialect consistent across all places of articulation?

The results of the present study show that the following acoustic correlates display statistical differences between [± voice] at a level above chance (when $\alpha = .05$): consonant duration, VOT, %VF, RMS, preceding vowel duration, CV ratio, F1 onset, and burst. The only acoustic correlate that was significantly different between the two dialects was RMS. However, it was shown that the way [± voice] relate to one another differs according to dialect along the following acoustic correlates: consonant duration, VOT, %VF, RMS, CV ratio and burst. In these cases, it most often occurred that there was a greater separation between [± voice] values. Usually the MER values were more widely separated than the MAR values. This was the case for consonant duration, %VF, CV ratio and burst. In all correlates but burst, the difference between dialects was more salient in [- voice]. VOT had a different reflex from these other acoustic correlates. Separation between [± voice] was greater in MAR than in MER and the difference was more striking in the [+ voice] category. In RMS, the relationship between [± voice] was the same in both dialects but there were substantial differences in the means for each dialect (with MAR RMS being much higher). Finally, with respect to preceding vowel duration, it was shown that the relationship between [± voice] was reversed in each dialect. In MAR, durations were longer in the speeded condition. In MER, there was very little difference between unspeeded and speeded. Speeded was slightly longer.

The differences observed between dialects in this study indicate that while the same acoustic correlates of [voice] are operative (at least in production) in both fortis and lenis dialects of Spanish, [± voice] relate to each other differently. Category differences tend to be maximized in MER. [+ Voice] values from both dialects tend to converge (as has been reported previously in Lewis, 2000). Notably divergent are the acoustic correlates VOT, preceding vowel duration and RMS. Here, the way [± voice] category values intersect with dialect is either reversed (VOT, preceding vowel duration) or the means are very much separated between one dialect and the next (RMS).

The observable differences between dialects in this study generally held across conditions of rate of speech and (to a lesser extent) prosody. Three of the nine acoustic correlates measured here varied significantly depending on position (consonant duration, VOT and RMS). This occurred most often in initial position, though consonant duration did also vary significantly in medial position. For the duration measurements, the unspeeded condition was longer than the speeded condition. For the amplitude measurement, speeded was higher than unspeeded. It is interesting that, of the correlates evaluated in the present study there are a few that encode a lot of information. Consonant duration, VOT and RMS returned significant differences for *voicing category*, *condition*, *dialect* (in the case of RMS) and the interaction of *voicing category * dialect*. One question for further research is whether or not the responsiveness of these correlates to so much change makes them more or less effective as perceptual cues.

In the present study, it was found that significant returns for *voicing category* held equally across both initial and medial contexts (the exception was VOT, which was significant in initial context only). RMS also held levels of significance for *dialect* in both initial and medial position. *Condition* was more variable. Most significant differences occurred in initial position (though consonant duration was significant in both initial and medial position). In dialectal differences between the way that categories related to one another (interaction of *voicing category * dialect*), most levels of significance occurred in medial position. The exceptions were VOT, which was significant in initial position only and consonant duration, which was significant in both initial and medial positions. Therefore, it seems that most dialectal information with regard to how [± voice] relate to one another is given in medial position. In initial position, the behavior of dialects with regard to [± voice] is relatively homogenous. Bigger differences are seen with changes in rate.

The main statistical model did not test for changes according to place of articulation. However, a follow-up multivariate ANOVA test indicated there were significant differences between place of articulation values for many of the correlates surveyed, including: consonant duration, %VF, VOT, F1 onset, preceding vowel duration and CV ratio. This information may be of use to future researchers.

Chapter 6: Implications and directions for further research

6.1 Summary and implications

The current study was designed to investigate how presumed physical differences in articulation (in particular, a highland/ lowland contrast in two dialects of Spanish) impact acoustic correlates related to the phonetic implementation of [voice]. It was predicted that MER and MAR would show different values for the nine different correlates depending on how each correlate responded to greater or lesser consonant closure. Some of the predictions made in §1.2.2 turned out to be accurate (greater separation of [± voice] means in MER than in MAR; duration predictions for [± voice] confirmed in results). Others conflicted with the actual results (no statistical differences observed in F0 contour, VOT better indicator of [voice] in MAR than in MER, more separation in [± voice] in MAR preceding vowel duration than in MER). Additionally, some of the phenomena observed in the results were not predicted in §1.2. For example, RMS turned out to be the key defining correlate that separated the two dialects, returning much higher values for MAR in both [± voice] categories. The different behavior of acoustic correlates in word-initial vs. word-medial position was not explicitly predicted in most cases. It was found that a few of the acoustic correlates examined here (consonant duration, VOT and RMS) varied significantly along the context of more than one independent variable and in the interaction of *voicing category * dialect*. In these cases, VOT and RMS varied significantly according to *condition* in initial position only. On the other hand, there were significant interactions between *voicing category * dialect* that emerged only in medial position (RMS, %VF and burst).[17]

The main problem under investigation in this study (how [voice] is implemented under consonant systems in Spanish that vary in degree of closure) was described in Chapter 1 in traditional linguistic terms, calling on feature theory and the concepts of fortis and lenis dialects. While these terms are convenient labels, they underspecify the range of results found in the present study. One problem with the fortis/ lenis categorization, noted by Ladefoged and Maddieson (1996), is the lack of consensus in the literature as to how the terms should be applied. In a general sense, *fortis* often refers to either increased respiratory or articulatory energy, while *lenis* denotes less overall energy, without regard to type. In §2.3.1, I referred to previous work in dialectology in defining fortis and lenis types for Spanish. A fortis dialect such as may be found in interior or highland Latin American dialects is characterized by occlusive pronunciations of /p t k/ and /b d g/, especially in initial contexts. The lenis, mostly coastal, dialects are described as displaying less overall muscular tension in the pronunciation of phonological stops. In the current study, the fortis/ lenis distinction between representative highland and lowland dialects MER and MAR was captured by the acoustic correlate RMS amplitude. RMS was lower in the fortis dialect (MER) and higher in the lenis dialect (MAR). In the context of the present study, then, the fortis/ lenis debate takes an interesting twist. An

---

[17] CV ratio also reached significance for *voicing category * dialect*, but since it was only measured in word-medial context (not word-initial), I do not include it in the same group.

elevated RMS amplitude is reflective of high, sustained energy in the signal (in other words, sonority). If energy in the signal were commensurate with articulatory strength, MAR would be the fortis dialect and MER the lenis. Counting back from acoustic to articulatory properties, it would probably be most accurate to say that MER is fortis based on shorter, sharper transitions (indicative of stop consonants and perhaps articulatory energy) and MAR lenis based on the pervasiveness of speech sounds that share some characteristics with vowels but are not vowels. The margins are far from clear and it is not at all certain whether the type of fortis/ lenis that applies in these dialects also applies cross-linguistically (although the hope is that the results here can be generalized to at least some other situations between dialects).

To further complicate matters, Ladefoged and Maddieson (1996) remark that in Germanic languages, many writers have preferred to use the terms *fortis* and *lenis* rather than [± voice] when describing opposing series of stops. This choice has evolved in part out of the revelation that vocal fold vibrations meant to define the [+ voice] series are more often absent than not. In the present study, vocal fold vibration did play a role in distinguishing [± voice], but so did seven other acoustic correlates. Ladefoged and Maddieson conclude:

> "As little is known about the articulatory dynamics of most languages, we would caution against making the assumption that phonological voicing differences are associated with articulatory strength differences in any particular case." (p. 96)

In Spanish, the status of [voice] continues to be under debate. As is the case with the Germanic languages, some researchers have disputed the role of vocal fold vibration in its ability to uphold the voicing contrast, especially in certain dialects and under conditions of running speech (Martínez-Celdrán, 2006, 1992, 1991[a, b]; Pérez, 1998). Lately, discussion has arisen as to whether the voicing contrast in Spanish has instead come to resemble a manner contrast ([-voice] = occlusives, [+voice] = continuants) (Hualde, 2005; however, see Romero, 2007 for a contrary view). In the current study, despite the proliferation of continuants (roughly 20% of the sample consisted of approximants that were inseparable on a waveform/ spectrogram display from the surrounding vowel segments), the boundary between [± voice] means was statistically upheld by %VF. The implication from this result is that even in a dialect where continuants are frequent (as is the case of MAR), a statistical difference in vocal fold vibration can be maintained between [± voice] categories. They may not, however, be the absolute (0 or 100%) that the terms [- voice] or [+ voice] imply. The means for the lowland dialect MAR in the present study, for example, were 42.233% [- voice] versus 87.679 [+ voice] in medial position.

Likewise, the results of the present study indicate several problems for feature theory. The most severe and overreaching of these is that feature theory does not address speech as a dynamic system with complex, interactive parts. It does not address the many-to-one relationship of acoustic correlates to phonological category reviewed in §1.1.2, §1.1.3, and §1.3.1 of the present study. Nor does it comment on how acoustic properties may work together to promote redundancy (and hence, robustness) in the speech signal (as has been suggested by Stevens & Keyser, 1989; Diehl & Kluender, 1989; Kingston & Diehl, 1995; Diehl & Lindblom, 2004, and others). Lastly, feature theory provides no insight into how languages or dialects can differ in the phonetic encoding of phonological contrasts or for the existence of different phonetic behavior in lower versus higher prosodic domains. These points are of key interest in the present study.

Prosodic boundaries are known to be particularly influential on the acoustic quality of the speech sound (Lisker & Abramson, 1972; Beckman & Edwards, 1990; Keating, Linker, & Huffman, 1983; Lindblom, 1990; Fougeron & Keating, 1997). Phonemic contrasts are often reflected differently depending on the prosodic domain in which they appear. In the present study, it was shown there were usually larger differences between [± voice] values across correlates in word-initial position than in word-medial position. This result is consistent with word recognition research which shows that segments at beginnings of words are more important for word recognition than are segments that occur later in the word. Another difference is that in the present study, acoustic correlates appearing word-initially were more sensitive to changes in *condition* (speeded/ unspeeded) than were their word-medial counterparts. Miller and Dexter (1988) suggest a link between these two. That is, recognition of early segments is especially important in faster speech. On a side note, it is interesting to see that %VF and RMS in the present study grouped together with respect to certain values. Both of them had higher values for initial than for medial position. This goes against a widespread view that consonants become more lenited through contact with surrounding vowels. If that were true, we could expect medial segments to have higher values for both %VF and RMS. Instead, the opposite is true. It appears in the current study that differences in [± voice] values are maximized in initial position and minimized in medial position.

On the other hand, most of the contrasts separating MAR from MER (significant interactions of *voicing category * dialect*) with regard to [voice] took place in medial position, suggesting some sort of tradeoff between phonemic and dialectal contrast. In his H & H theory, Lindblom (1990) suggested that speech production functions as an adaptive response to a variable task. Listeners normally do not use the speech signal to convey great chunks of information, but rather to augment and fill in communicative gaps. This is a common thread throughout many types of human behavior. For example, in the book *The Design of Everyday Things*, the cognitive scientist Donald Norman discusses the way humans approach the tools in their environment:

> "In everyday situations, behavior is determined by the combination of internal knowledge and external information and constraints... People can deliberately organize the environment to support their behavior... There is a tradeoff between the amount of mental knowledge and the amount of external knowledge required in performing tasks. People are free to operate variously in allowing for this tradeoff." (Norman, 1988: 55-56)

If we apply Norman's notion of tools to speech, it would make sense to expect variation in speech production. We don't have to try to account for such variation. What we account for is: 1) the patterns we observe, and; 2) to what we can attribute them.

6.2 Limitations

There are several limitations to the present study. One is that the study- from foundation to design to results- is based only on acoustic evidence. Ideally, a comprehensive study would involve different kinds of data: articulatory, acoustic, perceptual and/ or neurophysiological. This point will be addressed in more detail towards the end of the section.

Other limitations of the study emerge from the design. One goal of this study was to elicit language data from informants who were likely to have a strong regional accent. Therefore, participants with a high literacy level, or who had lived in different places or traveled abroad or who had contact with speakers of other languages were excluded from consideration in the study. The design had to be geared towards a low-literacy population. In fact, many of the participants in the current study had no more than a first- or second-grade level of education. Therefore, for the task, a kind of word game was invented with easily readable prompts. The game was intended to draw on the participants' pattern generation skills rather than relying on literacy. This would also serve as a distractor, to take the emphasis off the fact that informants were being recorded. Unfortunately, the task had a somewhat steep learning curve. Some informants were unable to complete the task. Others sometimes switched patterns in the middle of the task. Still others compensated for the cognitive load by inserting pauses or drawing out durations. This was seen especially in the consonant duration and VOT measures.

A last problem in the design emerged from the fact that word-initial and word-medial segments had to be analyzed separately. This arose out of the difficulty in setting up an environment where all conditions could be controlled for. It was desireable to have a word-initial (but not utterance-initial) measure. The nonsense word had to begin with a stop consonant (as that was the target segment) and ideally be preceded by a vowel (as was the case word-medially). It was very difficult to find a preceding word that was vowel final. *La* would have been one possibility, but because of gender agreement constraints in Spanish, it would have sounded strange to have a final vowel other than /a/ in the nonsense word. The decision was made to use *son* as the preceding word and to use an equal distribution of canonical vowels in the nonsense syllables. This worked fairly well, but for the extensive overlap between the nasal segments and oral stops in the sample. The overlap made segmentation difficult in initial position. At the design stage, it seemed desireable to create an initial context that would encourage occlusive pronunciations (thus highlighting the contrast between word-initial and word-medial target segments). A preceding nasal would provide just such a context. However, subsequent difficulties in segmentation and in having to analyze the two groups (word-initial and word-medial) separately most likely negated the potential benefits of using /n/. In hindsight, perhaps an adjective/ noun prompt such as "Su boróbo" would have been a better choice.

One limitation that was mentioned in §5.2 is the difficulty in generalizing about the variability observed in this study. Because of the way the subject pool was balanced (more female speakers than male speakers), the type of demographic information collected (general) and how it was used (for screening purposes only), it was not possible to make detailed comments on the nature of the variability observed in the study. There is some indication that subpatterns exist along gender lines (with regard to approximant realizations) and within dialect (for various measures). It is unclear what forces are driving the variability observed in this study- whether it arises from chance or from other demographic factors not controlled for in the initial screening process. Though the results for *dialect* and the interaction of *voicing category * dialect* in this study were robust, it may be of interest for future researchers to pursue the question of variability within dialect and/ or across gender lines.

Lastly, a production study is limited in what it is able to add to an understanding of speech processing. The present study adds to the body of acoustic data on Spanish consonant systems. This in turn contributes to knowledge of dialectal differentiation and to the behavior of consonant systems as a whole, at least to those that share a 2-way voiced/ voiceless unaspirated contrast. In terms of defining the [voice] contrast in Spanish, a speech production study can only give indications as to what listeners should or should not be able to tell apart in perception. The present study provides information on what categories are kept reliably apart (statistically-speaking) and by which acoustic correlates. It also indicates trends. This is helpful in determining which acoustic phenomena merit further study.

In order to investigate how category contrasts are actually encoded phonetically, different methods are used. In §1.3.3.1 and §1.3.3.2, I indicated that questions regarding phonology must be taken up by research in speech perception and/ or neuroimaging. In the remainder of this section I indicate areas where further research is warranted.

The results from the present study hold a few important implications for phonology. First, it was found that despite substantial differences in consonant articulation (confirmed here through the acoustic RMS measure), the sound systems in MAR and MER have somehow evolved to maintain a core set of boundary contrasts (active in eight of nine acoustic correlates surveyed) that separates [+ voice] from [-voice] in Spanish stop consonants. Stop consonant articulation in these dialects must have shifted and diverged over time, but only in ways that do not critically affect the phonemic contrast. The differences observed here correspond to another kind of contrast, one of dialectal identity. The results from the present study suggest that the dialect contrast may be housed in a different location (word-medially) and encoded differently (by means of different durational properties: for example different distributions of means, how speakers speed up or slow down word-medially, in V1 vowel contrasts). How listeners use sounds in speech, how they actually tell sounds apart and use them in a contrastive framework is the domain of studies in speech perception and categorization. Follow-up studies in speech perception are needed to investigate the reliability of the acoustic correlates examined in cueing a voice percept.

Speech perception relies on the successful integration of spectral and temporal cues in the signal. Researchers in Spanish phonetics have claimed that consonant duration, intensity, and vocal fold vibration are all important cues to voicing in Spanish. The present study suggests that, in addition to these correlates, VOT (in initial position), RMS, preceding vowel duration, CV ratio, F1 onset and burst are all likely contenders in helping to sustain the voicing contrast. The present study does not make claims as to which of these acoustic correlates (as cues) are likely to provoke a stronger [± voice] response than others. It may be of interest to know this information. A comprehensive study of Spanish speech perception would need to factor in details of prosodic domain and rate of speech as well.

Similarly, studies in neuroimaging are providing new and relevant information on where and how speech contrasts exist in the brain. Neuroimaging data is needed to understand how listeners partition and are generally able to make sense of incoming speech stimuli. It is thought that learners discover sound categories in the language input by identifying statistical peaks in the distribution of sounds in acoustic space. Studies performed with human infants as subjects show them to be extremely sensitive to the statistical properties of incoming speech stimuli (Saffran, Aslin, & Newport, 1996; Kuhl, 2000). Furthermore, current research on brain imaging indicates that experiential learning modifies which natural boundaries are important in a particular language (Ravizza, 2005; see also Kluender et al., 1998). Despite the progress made in understanding how the distributional properties of speech encourage learning, the difficulty remains in understanding exactly *how* sounds are linked to phonemes. The problem goes back to one of many-to-one mapping. As Kazanina, Phillips, and Isdardi (2006) in a study on mismatch negativity note:

> “There are different possible mappings between phonemes and speech sounds, and therefore sets of sound categories with similar acoustic distributions may map onto different sets of phonemes across languages.” (p. 11381)

Technologies such as magnetoencephalographic (MEG) brain recordings can be useful in sorting out the intricacies of perception and processing. MEG measures a response in brain activity when listeners are presented with certain auditory stimuli. In most studies using this technology, the presence or latency of a mismatch negativity (MMN) response is set up to reflect the speaker’s ability to discriminate an acoustic-phonetic contrast between standard and deviant tokens (of the type in-category or out). In Kazanina et al. (2006), the authors used a magnetic MMN (MMNm) to sort different [d] and [t] tokens into phoneme or allophone categories. Participants in the study were native speakers of Russian (where the [d]/ [t] contrast is phonemic) and Korean (where the contrast is allophonic and context-dependent). The aim of the study was to test whether the early stages of speech sound processing are governed by purely acoustic properties of the sound or whether they are affected by the functional role of the sound in word representations- in other words, by the phonemic status. Results from this test show that a speaker’s perceptual space (as reflected in early auditory brain responses) is shaped not only by a bottom-up analysis of the distribution of native language sounds, but also by an abstract analysis of the functional significance of those sounds.

The Kazanina et al. (2006) study deals with the question of the relation between the code used for storing words (i.e., the phonemes of the language), the statistical distribution of the sounds of the language in acoustic space, and the preattentive perceptual abilities of the speaker (as measured by MEG).

6.3 Suggestions for further research

How best to sort out information about what is phonemic, dialectal, individual, or accidental? What is needed is a way to visualize data in a kind of overlay, where it is possible to separate (or at the least see connections to) the different factors that influence speech production. How far, for example, and in what directions can a sound system be stretched and still maintain contrast? How is it possible to maintain a statistical contrast between [± voice] with different articulatory systems? A necessary interrelatedness of sounds and acoustic properties is a condition of speech sounds having evolved in the context of a blended environment. In fact, context-dependent speech perception seems to be a general property of the auditory system, non-specific to speech contexts or to human listeners. Holt and Lotto (2008) note:

> "The auditory system appears to represent acoustic signals not in terms of absolute values, but relative to sounds that precede (and follow) them (Wade & Holt, 2005). This context sensitivity is a consequence of the general operating characteristics of human (and bird) auditory systems." (pps. 43-44)

Ultimately, researchers want to ask how speech sounds function and interact with one another within the confines of a system. As has been seen, there are many different details that somehow come together to allow humans to communicate with one another through speech. The questions remains: how is this information organized?

Lindblom (1992), in an article on the traditional divide between phonetics and phonology, advocates an approach to speech research that addresses fundamental questions of how speech inventories evolve and, in particular, what makes a good contrastive network. He poses the question, “If phonological systems were seen as adaptations to universal performance constraints on speaking, listening and learning to speak, what would they be like?” (p.181)

Lindblom and Sundberg (1971) and Lindblom and Maddieson (1988) argued for models that address how cross-linguistic language propensities for certain sounds or combinations of sounds have evolved in the light of constraints on production and perception. A question that continuously begs asking is what pull certain combinations of sounds have on one another? How does language as a system evolve and self-maintain? Lindblom attempted to address that question by investigating an a priori range of physical sounds universally available for the selection of vowel contrasts in languages. Together with Bladon, he developed the idea of a *dispersion principle* and described this as providing an idiosyncratic shaping of the vowel space, encouraging room for more open-close than for front-back and rounding gestures. He defines the aim of this body of work and that of the subsequent *size principle* for consonant systems[18]:

> "Our position is that phonetic systems are the way they are, not because of implicational laws or markedness conventions (which are data-driven and therefore in principle non-explanatory), but because the values of phonetic segments evolved in response to universal, non-linguistic input/output constraints." (p.188)

For Lindblom and Maddieson, consonant inventories are adaptations to perceptual constraints as well as to production factors. In the case of both consonant and vowel systems, a single principle, such as that of maximizing inter-vocalic perceptual contrast, does not suffice to account for the most-favored patterns. In a speech system, there are many such patterns that are interconnected and allow the system to self-maintain.

One emerging body of work, of interest to researchers across vastly different scientific disciplines, seeks to explore the nature of systems, especially complex systems. *Network theory*, originating in the fields of applied mathematics and physics, has in recent years spread to all areas of modern scientific inquiry. Concepts from network theory have been used to describe behavioral phenomena from many different fields, including: social networking, traffic patterns, neural networks, the spread of infectious disease, and others (Watts and Strogatz, 1988). The breadth of application of such studies indicates that networks appear to be a common feature of all complex systems (Barabási and Fowler, 2009). Barábasi explains,

---

[18] The *size principle* states that consonant systems can be classified as Basic, Elaborated, or Complex. The number of contrasts shifts accordingly, forming an ascending series along a continuum of articulatory complexity.

> “One of the fundamental surprises, which certainly excites the physics community, is that we keep finding similar organizing principles across widely different systems. That is, if for a moment you forget that one node is a metabolite, the other is a gene, and the third is a person, the networks behind metabolism, genetics, and social systems are very much alike. And this has allowed people- social scientists like you *(directed towards James Fowler)*, physicists like me, as well as biologists and economists- to talk together on equal terms.” (p. 92)

An important feature of networks is the emphasis on relationships and on how information is integrated across nodes. In a literature review of papers concerned with the structure and function of complex networks, Newman (2003) describes commonalities in the current body of theory. According to the author, the current theory has three aims: 1) to find and highlight statistical properties, such as path lengths and degree distributions, that characterize the structure and behavior of networked systems, and to suggest appropriate ways to measure these properties; 2) it aims to create models of networks that can help us to understand the meaning of these properties- how they came to be as they are and how they interact with one another; 3) it aims to predict what the behavior of networked systems will be on the basis of measured structural properties and the local rules governing individual vertices.

What application does this have to speech? If, as Lindblom and Bladon say, ‘values of phonetic segments evolved in response to universal, non-linguistic input/output constraints’, the task of phoneticians remains to identify and describe the non-linguistic principles that guide speech output. Lindblom and others have been instrumental in laying the groundwork for this kind of research in phonetics. Unfortunately, most of the work that has been conducted up to date has been internal to the field, without regard to research in other scientific disciplines. One area that has remained largely unrecognized and unaddressed in phonetics is how a given element (without regard to content) operates in a network of contrast and the degree to which the organizational principles of such networks can be considered independent of content. This emerges as a promising area for future research in speech.

In the current study, I began by posing the question of how control of the voicing mechanism, which is rooted in physical gestures common to all languages, yet sensitive to subtle manipulations on a language-to-language basis, responds to dialectal differences in consonant articulation. It was shown that the two dialects under investigation vary significantly along the acoustic parameter RMS. Furthermore, it was shown that voicing category interactions for the two dialects varied significantly along the following parameters: consonant duration, VOT, %VF, RMS, CV ratio, and burst. Some of the changes involve a shifting up or down of values. Others entail a qualitative difference in the way the categories relate to one another, as in the case of preceding vowel duration under speeded and unspeeded conditions. Further studies are needed in perception and/ or brain imaging to determine what repercussions such dialectal differences may have on the way category information is encoded for each of the different speech populations. This study is of potential interest to researchers in that the kind of 2-way voicing contrast represented here is extremely common throughout the world’s languages.

What does phonetics in the 21st century look like? Some of the same questions continue to be of interest: What makes a language a language? How do sound inventories work in speech? What makes for a good system of contrast? In summary, and as an extension towards further work, one would like to be able comment on how the structures found within any given sound inventory impact the behavior of the system as a whole. Certainly it is no coincidence that clusters of properties correlated with [voice] reoccur across languages. A modern account of phonetics must be prepared to incorporate patterns observed in speech and reconcile these with not only previous language research, but also with what is known of the mechanisms that impact speech behavior- biology, systems, connections and constraints.

Appendix 1: Testing Materials Experiment 1

Prompt (presented visually)	Target nonsense word (spoken by subject)
‘son pa’	/son pa'rapa/
‘son pe’	/son pe'repe/
‘son pi’	/son pi'ripi/
‘son po’	/son po'ropo/
‘son pu’	/son pu'rupu/
‘son ba’	/son ba'raba/
‘son be’	/son be'rebe/
‘son bi’	/son bi'ribi/
‘son bo’	/son bo'robo/
‘son bu’	/son bu'rubu/
‘son ta’	/son ta'rata/
‘son te’	/son te'rete/
‘son ti’	/son ti'riti/
‘son to’	/son to'roto/
‘son tu’	/son tu'rutu/
‘son da’	/son da'rada/
‘son de’	/son de'rede/
‘son di’	/son di'ridi/
‘son do’	/son do'rodo/
‘son du’	/son du'rudu/
‘son ka’	/son ka'raka/
‘son ke’	/son ke'reke/
‘son ki’	/son ki'riki/
‘son ko’	/son ko'roko/
‘son ku’	/son ku'ruku/
‘son ga’	/son ga'raga/
‘son gue’	/son ge'rege/
‘son gui’	/son gi'rigui/
‘son go’	/son go'rogo/
‘son gu’	/son gu'rugu/

Colors and numbers appearing in picture prompts (second part of task)

nueve
doce
negro
verde
rojo
blanco

cero
cuatro
cinco
siete
trece
quince

Appendix 2: Example of screens shown to participants

Bloque 3 50	Son PO 51	Son DI 7 52	Son GA 🟥 53	Son PE 54	Son BI 55	Son KO 56
Son TA 13 57	Son DO ⬜ 58	Son KA 59	Son TA 15 60	Son BE 9 61	Son DO 62	Son BE 13 63
Son GA 🟩 64	Son KA 12 65	Son TO 66	Son TO 🟥 67	Son DO 🟩 68	Son PI 69	Son GA 0 70
Son KI 71	Son TA 5 72	Son GO 9 73	Son PO 15 74	Son PE 5 75	Son KU ⬛ 76	Son BU 13 77

Appendix 3: Participant Survey

¿Cómo se llama Usted?
(*What's your name?*)

¿Cuántos años tiene?
(*How old are you?*)

¿Hasta qué año llegó en la escuela?
(*What was the last grade you completed in school?*)

¿De dónde viene Usted?
(*Where are you from?*)

¿Dónde vive ahora?
(*Where do you live now?*)

¿Cuánto tiempo tiene viviendo en el pueblo?
(*How long have you been living in this town?*)

¿Ha tenido Usted alguna vez problemas con el oído o con el habla?
(*Have you ever had any problems with your hearing or speech?*)

¿Habla Usted algun idioma que no sea el castellano?
(*Do you speak any foreign languages?*)

¿Tiene Usted familiares o amigos cercanos que sean hablantes nativos de un idioma que no sea el castellano?
(*Do you have any family members or close friends that are native speakers of a language other than Spanish?*)

¿Cuál es el trabajo de Usted?
(*What's your occupation?*)

¿Puede contar algo de las responsabilidades que desempeña en su trabajo?
(*Can you tell me something about the work that you do?*)

¿Puede decir algo sobre esta comunidad donde Usted vive? ¿Cómo es? ¿Quiénes son ustedes que viven aquí?
(*Can you say something about the community where you live? What's it like? What are the people who live here like?*)

Appendix 4a: Descriptive Statistics- Word-initial

Descriptive Statistics

spkr		N	Minimum	Maximum	Mean	Std. Deviation
2	cdur	22	.016	.180	.09155	.047356
	vf	23	0	100	61.17	33.001
	burst	23	1	2	1.09	.288
	vot	20	.000	.067	.03290	.018595
	f0	22	73.1	154.1	135.445	19.8505
	f1	23	200	520	375.65	89.230
	rms	22	32.88	48.22	41.2305	3.57049
	Valid N (listwise)	19				
3	cdur	63	.008	.190	.04732	.032895
	vf	63	0	100	65.86	29.808
	burst	63	1	2	1.13	.336
	vot	55	.000	.059	.02476	.011030
	f0	63	122.5	146.2	135.511	4.9913
	f1	63	240	480	363.17	64.379
	rms	63	31.43	46.81	40.2366	3.15567
	Valid N (listwise)	55				
4	cdur	32	.019	.175	.08134	.037363
	vf	33	0	100	38.82	43.283
	burst	33	1	2	1.03	.174
	vot	31	.000	.069	.01774	.014578
	f0	33	31.5	240.4	190.538	49.0960
	f1	33	280	640	413.33	71.880
	rms	33	25.50	40.12	35.0517	3.21706
	Valid N (listwise)	31				

5	cdur	51	.006	.206	.08694	.049917
	vf	54	0	100	40.46	44.028
	burst	54	1	2	1.11	.317
	vot	48	-.074	.060	.01254	.022205
	f0	54	180.6	245.6	215.213	13.4623
	f1	54	160	640	358.52	108.774
	rms	54	32.64	51.00	42.4300	3.40465
	Valid N (listwise)	45				
7	cdur	59	.016	.110	.05458	.021034
	vf	59	0	100	46.81	45.468
	burst	59	1	2	1.07	.254
	vot	55	-.006	.056	.02558	.015165
	f0	59	106.4	192.6	168.037	14.0137
	f1	59	280	680	412.88	94.960
	rms	59	33.28	48.94	41.7680	3.87437
	Valid N (listwise)	55				
8	cdur	47	.011	.094	.04134	.017580
	vf	47	0	100	43.70	45.262
	burst	47	1	2	1.21	.414
	vot	37	-.047	.180	.01838	.032026
	f0	47	158.0	193.9	180.713	7.0236
	f1	47	200	640	400.85	107.942
	rms	47	31.08	48.96	41.0484	4.33083
	Valid N (listwise)	37				
9	cdur	64	.008	.147	.04469	.029205
	vf	64	0	100	71.64	33.051
	burst	64	1	2	1.23	.427
	vot	49	-.029	.029	.00314	.014746

	f0	64	189.3	247.7	213.552	13.3666
	f1	64	200	640	388.75	106.778
	rms	64	33.20	51.97	44.1330	4.54167
	Valid N (listwise)	49				
10	cdur	8	.013	.076	.02612	.020691
	vf	8	100	100	100.00	.000
	burst	8	1	2	1.38	.518
	vot	5	.000	.008	.00220	.003493
	f0	8	198.1	226.1	209.688	9.7690
	f1	8	320	480	385.00	56.315
	rms	8	41.86	49.41	45.2383	2.99069
	Valid N (listwise)	5				
11	cdur	32	.032	.172	.10253	.042370
	vf	39	0	100	54.28	46.406
	burst	39	1	2	1.31	.468
	vot	27	-.022	.056	.01241	.014178
	f0	39	158.7	292.3	238.372	21.0606
	f1	39	240	600	396.92	96.849
	rms	39	31.92	47.51	39.9837	4.07882
	Valid N (listwise)	23				
12	cdur	47	.016	.073	.04166	.015378
	vf	47	0	100	52.87	44.491
	burst	47	1	2	1.32	.471
	vot	32	.000	.044	.01694	.011022
	f0	47	107.6	272.9	127.674	22.7769
	f1	47	240	520	330.21	71.975
	rms	47	17.04	31.39	23.6265	3.22869
	Valid N (listwise)	32				

13	cdur	10	.012	.101	.06290	.024995
	vf	10	14	100	46.40	28.316
	burst	10	1	2	1.20	.422
	vot	8	.005	.024	.01437	.007130
	f0	10	200.1	248.5	222.550	14.7544
	f1	10	280	600	424.00	108.648
	rms	10	23.68	38.36	30.5087	4.58904
	Valid N (listwise)	8				
14	cdur	45	.013	.160	.07138	.039756
	vf	47	0	100	52.02	40.554
	burst	47	1	2	1.17	.380
	vot	39	-.008	.150	.01210	.023454
	f0	47	112.7	225.9	209.170	17.0418
	f1	47	280	520	385.53	62.094
	rms	47	19.81	41.39	28.5835	5.11786
	Valid N (listwise)	37				
15	cdur	39	.009	.143	.07128	.040447
	vf	39	0	100	54.05	38.629
	burst	39	1	2	1.10	.307
	vot	35	.000	.062	.01986	.014167
	f0	39	109.2	231.7	212.900	18.1446
	f1	39	200	560	363.08	82.275
	rms	39	20.92	39.01	27.8856	4.89094
	Valid N (listwise)	35				
16	cdur	13	.037	.194	.12615	.053690
	vf	13	0	55	13.77	18.869
	burst	13	1	1	1.00	.000
	vot	13	.012	.083	.03815	.025023

	f0	13	91.4	201.7	184.415	29.3084
	f1	13	200	640	335.38	139.081
	rms	14	18.54	55.00	27.3505	8.70480
	Valid N (listwise)	13				
17	cdur	83	.012	.188	.05811	.034638
	vf	85	0	100	73.15	31.772
	burst	85	1	2	1.20	.402
	vot	67	.000	.088	.01381	.011105
	f0	85	113.0	274.0	236.293	24.9211
	f1	85	200	720	416.94	118.302
	rms	85	17.28	35.51	25.3661	3.68875
	Valid N (listwise)	66				
18	cdur	49	.013	.075	.03784	.014203
	vf	49	0	100	83.80	32.470
	burst	49	1	2	1.29	.456
	vot	35	.000	.029	.01237	.006495
	f0	49	100.4	183.2	121.380	12.3613
	f1	49	280	520	404.90	75.114
	rms	49	11.27	27.43	17.1199	3.53379
	Valid N (listwise)	35				
19	cdur	31	.025	.149	.08319	.035206
	vf	31	0	100	68.97	41.181
	burst	31	1	2	1.16	.374
	vot	26	.000	.037	.01242	.009949
	f0	31	235.3	274.4	252.681	9.1891
	f1	31	240	680	423.23	139.245
	rms	31	16.12	34.99	28.6391	3.98435
	Valid N (listwise)	26				

22	cdur	47	.009	.122	.06181	.032087
	vf	47	0	100	59.13	39.284
	burst	47	1	2	1.09	.282
	vot	43	.000	.026	.00805	.006102
	f0	47	180.1	223.5	200.270	9.1485
	f1	47	160	520	358.30	92.861
	rms	47	25.16	46.00	34.8902	4.27731
	Valid N (listwise)	43				
23	cdur	65	.013	.112	.06108	.028737
	vf	67	0	100	53.00	41.951
	burst	67	1	2	1.10	.308
	vot	59	.000	.088	.02707	.019801
	f0	67	.0	119.4	91.943	36.8787
	f1	67	200	720	403.58	123.922
	rms	67	11.68	29.42	21.1589	3.71349
	Valid N (listwise)	57				
24	cdur	70	.016	.205	.08326	.051176
	vf	71	17	100	63.99	35.565
	burst	71	1	2	1.14	.350
	vot	61	.000	.058	.01431	.010734
	f0	71	139.8	228.5	201.476	10.0333
	f1	71	200	640	345.35	100.340
	rms	71	20.88	37.89	27.2353	3.68352
	Valid N (listwise)	60				
25	cdur	4	.042	.143	.10225	.042937
	vf	5	0	100	42.00	43.480
	burst	5	1	1	1.00	.000
	vot	5	.009	.022	.01560	.005225

f0	5	131.3	152.8	145.380	8.4325
f1	5	240	440	328.00	86.718
rms	5	29.94	39.87	36.1620	4.13602
Valid N (listwise)	4				

Appendix 4b: Descriptive Statistics- Word-medial

Descriptive Statistics

spkr		N	Minimum	Maximum	Mean	Std. Deviation
2	vdur	37	.059	.162	.10208	.017870
	cdur	36	.028	.178	.07514	.027562
	vf	36	36	100	81.97	21.907
	burst	36	1	2	1.25	.439
	vot	22	.000	.082	.02795	.018758
	cvratio	36	.260	1.500	.74694	.292973
	f0	35	0	173	125.40	39.932
	f1	33	240	520	345.45	69.870
	rms	32	30	45	36.96	4.528
	Valid N (listwise)	22				
3	vdur	70	.071	.199	.12816	.025507
	cdur	70	.019	.104	.05511	.019755
	vf	70	18	100	87.39	21.825
	burst	70	1	2	1.47	.503
	vot	37	.006	.095	.02062	.014413
	cvratio	70	.120	.920	.46286	.209569
	f0	70	130	154	141.89	5.304
	f1	70	200	920	314.29	95.653
	rms	70	27	51	39.08	4.694
	Valid N (listwise)	37				
4	vdur	67	.068	.396	.19761	.056500
	cdur	67	.026	.226	.09394	.045927
	vf	67	0	100	58.15	48.348
	burst	67	1	2	1.57	.499

	vot	29	.009	.076	.02490	.015653
	cvratio	67	.110	1.300	.52313	.301959
	f0	67	37	302	217.67	60.282
	f1	67	240	760	416.12	115.483
	rms	67	29	45	35.70	3.700
	Valid N (listwise)	29				
5	vdur	81	.119	.275	.16925	.031030
	cdur	81	.023	.200	.08864	.032253
	vf	81	0	100	60.68	38.219
	burst	81	1	2	1.35	.479
	vot	53	-.045	.045	.01725	.013863
	cvratio	81	.130	.930	.54123	.205690
	f0	81	0	307	224.50	43.830
	f1	81	160	680	365.43	118.280
	rms	81	32	52	41.01	4.271
	Valid N (listwise)	53				
7	vdur	80	.067	.165	.11346	.018659
	cdur	80	.044	.138	.09673	.020599
	vf	80	0	100	22.61	32.312
	burst	80	1	2	1.06	.244
	vot	72	.005	.070	.03057	.014749
	cvratio	80	.370	1.600	.87737	.248747
	f0	77	74	259	191.11	43.547
	f1	77	240	640	402.60	96.973
	rms	77	28	50	39.10	4.134
	Valid N (listwise)	72				
8	vdur	60	.056	.256	.12458	.050139
	cdur	56	.017	.126	.05523	.024465

	vf	57	0	100	74.02	37.931
	burst	58	1	2	1.62	.489
	vot	18	-.055	.040	.01167	.019036
	cvratio	56	.110	1.160	.50643	.245789
	f0	53	138	246	194.49	20.747
	f1	53	200	640	375.85	95.544
	rms	53	31	54	41.80	5.685
	Valid N (listwise)	18				
9	vdur	71	.059	.285	.11794	.033893
	cdur	71	.015	.156	.06239	.031821
	vf	71	0	100	62.46	36.454
	burst	71	1	2	1.41	.495
	vot	42	-.025	.079	.01093	.014757
	cvratio	71	.080	1.290	.57169	.318936
	f0	70	0	286	237.65	36.423
	f1	70	280	720	402.86	100.623
	rms	70	34	55	42.10	4.606
	Valid N (listwise)	41				
10	vdur	9	.098	.172	.12967	.023749
	cdur	9	.016	.096	.04222	.030388
	vf	9	42	100	89.11	22.071
	burst	9	1	2	1.78	.441
	vot	2	.020	.028	.02400	.005657
	cvratio	9	.130	.780	.32889	.244307
	f0	9	150	227	190.66	30.733
	f1	9	280	560	408.89	86.667
	rms	9	33	50	43.30	5.490
	Valid N (listwise)	2				

11	vdur	70	.101	.482	.24507	.083619
	cdur	70	.034	.333	.11139	.063840
	vf	70	0	100	62.34	40.251
	burst	70	1	2	1.47	.503
	vot	37	.000	.051	.01662	.013783
	cvratio	70	.110	1.370	.49029	.275996
	f0	70	146	315	236.79	43.431
	f1	70	160	760	392.57	151.253
	rms	70	27	48	35.86	4.143
	Valid N (listwise)	37				
12	vdur	42	.055	.201	.08676	.022721
	cdur	42	.011	.077	.04890	.014593
	vf	42	0	100	76.00	38.266
	burst	42	1	2	1.69	.468
	vot	13	.004	.048	.01738	.012939
	cvratio	42	.065	1.350	.60274	.240061
	f0	42	0	193	142.72	29.293
	f1	42	160	480	297.14	80.556
	rms	42	14	28	18.48	3.060
	Valid N (listwise)	13				
13	vdur	19	.077	.167	.11400	.024051
	cdur	19	.014	.132	.07705	.036581
	vf	19	0	100	58.26	34.931
	burst	19	1	2	1.21	.419
	vot	15	-.005	.048	.01947	.013495
	cvratio	19	.110	1.250	.70632	.353572
	f0	19	96	317	228.20	68.443
	f1	19	200	720	395.79	168.071

	rms	19	18	43	27.74	5.993
	Valid N (listwise)	15				
14	vdur	55	.074	.278	.12636	.044171
	cdur	55	.027	.189	.11093	.036984
	vf	55	0	100	42.15	43.442
	burst	55	1	2	1.27	.449
	vot	40	.003	.108	.01478	.016395
	cvratio	55	.190	2.000	1.01145	.496165
	f0	55	110	252	222.05	28.635
	f1	55	120	600	381.82	106.616
	rms	55	16	42	26.43	5.022
	Valid N (listwise)	40				
15	vdur	54	.084	.213	.13430	.027677
	cdur	54	.019	.193	.08750	.039259
	vf	54	0	100	43.04	43.205
	burst	54	1	2	1.30	.461
	vot	35	.012	.055	.03166	.012723
	cvratio	54	.140	1.500	.69333	.349344
	f0	51	129	323	249.78	54.147
	f1	51	200	800	367.84	147.517
	rms	51	16	37	25.31	4.287
	Valid N (listwise)	35				
16	vdur	26	.065	.303	.12727	.068640
	cdur	26	.061	.364	.14342	.064686
	vf	26	0	100	21.46	33.500
	burst	26	1	1	1.00	.000
	vot	24	-.039	.069	.02246	.021439
	cvratio	26	.430	3.500	1.30731	.662731

	f0	22	180	271	229.13	31.031
	f1	22	160	800	354.55	173.829
	rms	22	18	30	23.31	3.344
	Valid N (listwise)	22				
17	vdur	85	.049	.180	.11789	.028607
	cdur	85	.022	.168	.08301	.036031
	vf	85	0	100	74.41	33.378
	burst	85	1	2	1.48	.503
	vot	45	.000	.055	.01831	.011281
	cvratio	85	.180	1.600	.77765	.429650
	f0	82	126	320	254.15	45.216
	f1	82	160	720	402.44	119.315
	rms	82	14	34	22.13	4.145
	Valid N (listwise)	42				
18	vdur	51	.037	.239	.10269	.041044
	cdur	51	.021	.098	.05614	.021329
	vf	51	0	100	71.37	36.777
	burst	51	1	2	1.51	.505
	vot	25	.000	.054	.01484	.010554
	cvratio	51	.140	1.300	.60922	.266900
	f0	51	100	164	126.96	16.444
	f1	51	200	520	396.86	68.746
	rms	51	10	27	17.39	4.322
	Valid N (listwise)	25				
19	vdur	68	.091	.349	.15872	.053686
	cdur	68	.019	.139	.07594	.026757
	vf	68	0	100	53.03	40.584
	burst	68	1	2	1.43	.498

	vot	39	.003	.043	.01821	.009825
	cvratio	68	.110	.980	.52912	.246490
	f0	68	115	341	268.44	65.412
	f1	68	200	680	424.71	133.299
	rms	68	14	43	23.62	5.643
	Valid N (listwise)	39				
22	vdur	55	.069	.166	.10442	.022436
	cdur	55	.016	.145	.08189	.033896
	vf	55	0	100	40.27	40.753
	burst	55	1	2	1.29	.458
	vot	38	-.056	.033	.01176	.014414
	cvratio	55	.130	1.760	.85691	.427542
	f0	53	159	305	260.75	33.523
	f1	53	200	880	374.34	127.965
	rms	53	21	45	28.91	4.694
	Valid N (listwise)	37				
23	vdur	71	.075	.210	.12546	.025662
	cdur	70	.023	.177	.06250	.029302
	vf	70	0	100	68.79	44.525
	burst	70	1	2	1.73	.448
	vot	19	.007	.048	.02463	.009341
	cvratio	70	.160	1.450	.53914	.322026
	f0	70	13	136	118.13	16.152
	f1	70	200	720	369.14	117.396
	rms	70	14	27	20.63	3.054
	Valid N (listwise)	19				
24	vdur	77	.103	.332	.15839	.040083
	cdur	77	.026	.306	.09826	.052412

	vf	77	0	100	56.43	40.725
	burst	77	1	2	1.44	.500
	vot	43	.009	.067	.02235	.012906
	cvratio	77	.150	1.400	.64727	.336505
	f0	77	192	340	281.16	28.998
	f1	77	200	760	372.99	125.897
	rms	77	21	41	28.42	4.239
	Valid N (listwise)	43				
25	vdur	59	.076	.183	.11353	.022063
	cdur	59	.020	.210	.08669	.044394
	vf	59	0	100	55.54	44.358
	burst	59	1	2	1.34	.477
	vot	39	.000	.045	.01741	.009869
	cvratio	59	.130	1.740	.83424	.492325
	f0	58	79	213	174.79	32.776
	f1	58	160	600	327.59	91.906
	rms	58	23	48	33.69	5.492
	Valid N (listwise)	38				

References

Abramson, A. & Lisker, L. (1972). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics, 1*, 1-8.

Aslin, R., Jusczyk, P., & Pisoni, D. (1998). Speech and auditory processing during infancy: Constraints on and precursors to language. In D. Kuhn & R. Siegler (Eds.), *Handbook of Child Psychology, 5th edition, vol.2: Cognition, Perception and Language* (pp. 147-198). New York: Wiley.

Barabási, A.-L., & Fowler, J. (2009). Interview in the Seed Salon. *Seed, 20*, 91-96.

Beckman, M., & Edwards, J. (1990). Lenthenings and shortenings and the nature of prosody. In J. Kingston and M. Beckman (Eds.), *Between the Grammar and Physics of Speech (Papers in Laboratory Phonology I)* (pp. 152-178). Cambridge: Cambridge University Press.

Beddor, P. (2007). Perceiving coarticulated speech. Paper presented at *La coarticulation: Indices, direction et représentation*. Montpellier, France, December 7th 2007.

Benkí, J. (2005). Perception of VOT and first formant onset by Spanish and English speakers. In J. Cohen, K. McAlister, K. Rolstad, J. MacSwain (Ed.s), ISB4: Proceedings of the 4th International Symposium on Bilingualism (pp. 240-248), Somerville, MA: Cascadilla Press.

Bentivoglio, P. (1998). La variación sociofonológica. *Español Actual, 69*, 29-42.

Best, C., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception in Psychophysics, 29*, 191 211.

Bhaskararao, P. (1994). Subphonemic segment inventories for concatenative speech synthesis. In E. Keller (Ed.), *Fundamentals of Speech Synthesis and Speech Recognition: Basic concepts, state of the art and future challenges* (pp. 69-85). Chichester: John Wiley & Sons.

Burki-Cohen, J., Grosjean, F. & Miller, J. (1989). Base-language effects on word identification in bilingual speech: Evidence from categorical perception experiments. *Language and Speech, 32*, 355-371.

Cancho, R., & Solé, R. (2001). The Small world of human language. *Proc. R. Soc. Lond. B, 268*, 2261-2265.

Canfield, D. (1981). *Spanish Pronunciation in the Americas*. Chicago: University of Chicago Press.

Cho, T., & Keating, P. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics, 29*, 155-190.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27*, 207-229.

Cho, T., & McQueen, J. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics, 33(2)*, 121 157.

Cole, R., & Cooper, W. (1975). Perception of voicing in English affricates and fricatives. *J. Acoust. Soc. Am., 58*, 1280-128.

Denes, P. (1955). Effect of duration on the perception of voicing. *J. Acoust. Soc. Am., 27*, 761-764.

Dent, L. (1976). Voice Onset Time of spontaneously spoken Spanish voiceless stops. *J. Acoust. Soc. Am., 59*: S41.

Diehl, R. (1981). Feature detectors for speech: A critical reappraisal. *Psychological Bulletin, 89*, 1-18.

Diehl, R., Castleman, W., & Kingston, J. (1995). On the internal perceptual structure of phonological features: The [voice] distinction. *J. Acoust. Soc. Am., 97*, 3333 33334 (abstract).

Diehl, R. & Kluender, K. (1989). On the objects of speech perception. *Ecological Psychology, 1(2)*, 121-144.

Diehl, R. & Lindblom, B. (2004). Explaining the structure of feature and phoneme inventories: The role of auditory distinctiveness. In S. Greenberg, W. Ainsworth, A. Popper, & R. Fay (Eds.), *Speech Processing in the Auditory System* (pp. 101 162). New York: Springer-Verlag.

Diehl, R. & Rosenberg, D. (1977). Acoustic feature analysis in the perception of voicing contrasts. *Perception and Psychophysics, 21(5)*, 418-422.

Eimas, P. & Corbit, J. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4*, 99-109.

Flege, J., Munro, M., & Fox, R. (1994). Auditory and categorical effects on cross language vowel perception. *J. Acoust. Soc. Am., 95*, 3623-3641.

Fougeron, C., & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *J. Acoust. Soc. Am., 106*, 3728-3740.

Francis, A., Ciocca, V., & Yu, J.-M.-C. (2003). Accuracy and variability of acoustic measures of voicing onset. *J. Acoust. Soc. Am., 113(2)*, 1025-1032.

Franco, F. (1998). Reseña histórica de Mérida. In C. Domínguez & E. Mora (Eds.), *El habla de Mérida* (pp. 331-340). Mérida, Venezuela: Universidad de los Andes.

Garson, D. (2008). Linear Mixed Models: Random Effects, Hierarchical Linear, Multilevel, Random Coefficients, and Repeated Measures Models. http://faculty.chass.ncsu.edu/garson/PA765/multilevel.htm

Gelfand, S. (2001). *Essentials of Audiology 2nd ed*. New York: Thieme.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *J. Acoust. Soc. Amer., 47(2)*, 613-617.

Haggard, M., Summerfield, Q., & Roberts, M. (1991). Psychoacoustical and cultural detriments of phoneme boundaries: Evidence from trading F0 cues in the voiced voiceless distinction. *J. Phonet., 9*, 49-62.

Harrington, J & Cassidy, S. (1999). *Techniques in Speech Acoustics*. Dordrecht: Kluwer Academic Publishers.

Hay, J. (2005). How auditory discontinuities and linguistic experience affect the perception of speech and non-speech in English- and Spanish-Speaking listeners. PhD diss., University of Texas at Austin.

Hirose, H., Yoshioka, H. & Niimi, S. (1979). A cross-language study of laryngeal adjustment in consonant production. In H. Hollien & P. Hollien (Eds.), *Current Issues in the Phonetic Sciences*. Amsterdam: John Benjamins.

Hoole, P., Gobl, C., & Chasaide, A. (1999). Laryngeal coarticulation. In W. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, Data and Techniques* (pp. 105-143). Cambridge: Cambridge University Press.

Holt, L. & Lotto, A. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science, 17(1)*, 42-46.

Holt, L., Lotto, A., & Diehl, R. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *J. Acoust. Soc. Am., 116*, 1763-1773.

House, A. & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *J. Acoust. Soc. Amer., 25*, 105 113.

Hualde, J. (2005). *The Sounds of Spanish*. Cambridge: Cambridge University Press.

Jongman, A., Sereno, J., Raaijmakers, M., & Lahiri, A. (1992) The phonological representation of [voice] in speech perception. *Language and Speech, 35*, 137-152.

Jusczyk, P., Pisoni, D., Walley, A., & Murray, J. (1980). Discrimination of relative onset time of two-component tones by infants. *J. Acoust. Soc. Am., 67(1)*, 262-270.

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The Influence of meaning on the perception of speech sounds. *PNAS, 103(30)*, 11381-11386.

Keating, P., Linker, W., & Huffman, M. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics, 11*, 277-290.

Kingston, J., & Diehl, R. (1994). Phonetic Knowledge. *Language, 70(3)*, 419-454.

---------(1995). Intermediate properties in the perception of distinctive feature values. In B. Connell & A. Arvaniti (Eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV* (pp.7-27). Cambridge: Cambridge Univ. Press.

Kingston, J., Diehl, R., Kluender, K., & Parker, E. (1990). Resonance versus source characteristics in perceiving spectral continuity between vowels and consonants. *J. Acoust. Soc. Am., 88*:S54-S55 (Abstract).

Kluender, K. & Lotto, A. (1994). Effects of first formant onset frequency on [voice] judgments result from general auditory processes not specific to humans. *J. Acoust. Soc. Am., 95*, 1044-1052.

Kluender, K., Lotto, A., Holt, L., & Bloedel, S. (1998). Role of experience in language specific functional mappings for vowel sounds as inferred from human, nonhuman, and computational models. *J. Acoust. Soc. Am., 104*, 3568-3582.

Krause, J. & Braida, L. (2002) Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *J. Acoust. Soc. Am., 112(5),* 2165-2172.

---------(2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *J. Acoust. Soc. Am., 115(1)*, 362-378.

Kuhl, P. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America 97*, 11850-11857.

Kuhl, P. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *J. Acoust. Soc. Am., 70*, 340-349.

Kuhl, P. & Miller, J. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190*, 69-72.

Ladefoged, P. (1983). The linguistic use of different phonation types. In D. Bless and J. Abbs (Eds.), *Vocal Fold Physiology* (pp. 351-360). San Diego: College Hill Press.

---------(1992). The many interfaces between phonetics and phonology. In W. Dressler, H. Luschützky, O. Pfeiffer & J. Rennison (Eds.) *Phonologica 1988: Proceedings of the 6th International Phonology Meeting* (pps. 165-180). Cambridge: Cambridge University Press,.

---------(2001). *A Course in Phonetics*. Boston, MA: Heinle & Heinle.

---------(2003). *Phonetic Data Analysis*. Malden, MA: Blackwell Publishing.

Ladefoged, P. & Maddieson, I. (1996). *The Sounds of the World's Languages*. Malden, MA: Blackwell Publishing.

Lasky, R., Syrdal-Lasky, A., & Klein, R. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology, 20*, 215-225.

Lewis, A. (2001). Weakening of intervocalic /p, t, k/ in two Spanish dialects: Toward the quantification of lenition processes. Unpublished doctoral dissertation. University of Illinois at Urbana-Champaign.

---------(2000). Acoustic variability of intervocalic voiceless stop consonants in three Spanish dialects. In H. Campos, E. Herberger, A. Morales-Front, T. Walsh (Eds.), *Hispanic Linguistics at the Turn of the Millennium: Papers from the 3rd Hispanic Linguistics Symposium (pp. 101-114)*. Sommerville: Cascadilla Press.

Liberman, A., Harris, K., Kinney, J. & Lane, H. (1961). The discrimination of relative onset time of the components of certain speech and non-speech patterns. *Journal of Experimental Psychology, 61*, 379.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle and A. Marchal (Eds.), *Speech Production and Speech Modeling* (pp. 403-439). Dodrecht: Kluwer.

----------(1992). Phonetic content in phonology. In W. Dressler, H. Luschützky, O. Pfeiffer & J. Rennison (Eds.), *Phonologica 1988: Proceedings of the 6th International Phonology Meeting* (pp. 181-196). Cambridge: Cambridge University Press.

Lindblom, B. & Maddieson, I. (1988). Phonetic universals in consonant systems. In L. Hyman & C. Li (Eds.), *Language, Speech and Mind* (pp. 62-78). London: Routledge

Lipski, J. (1994). *Latin American Spanish*. New York: Longman.

Lisker, L. (1978a). In qualified defense of VOT. *Language and Speech, 21(4)*, 375-383.

----------(1978b). 'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Haskins Laboratories Status Report on Speech Research*, 86-87.

Lisker. L., & Abramson, A. (1970a). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 563-567). Prague: Academia.

----------(1970b). Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the 6th International Congress of Phonetic Sciences*. Prague: Academia.

----------(1972). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics, 1*, 1-8.

Löfqvist, A. (1995). Laryngeal mechanisms and interarticulator timing in voiceless consonant production. In F. Bell-Berti and L.J. Raphael (Eds.), *Producing Speech: Contemporary issues*. Woodbury, NY: Contemporary Press.

Lotto, A. & Holt, L. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *J. Acoust. Soc. Am., 119(5)*, 3059-3071.

MacNeilage, P. (2008). *The Origin of Speech*. Oxford: Oxford University Press.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Martínez Celdrán, E. (2006). Some chimeras of Spanish traditional phonetics. Paper presented at the *3rd Conference on Laboratory Approaches to Spanish Phonology*. Toronto, Canada.

---------(1992). Experimentos sobre la duración de las oclusivas y aproximantes del español. *Revista Española de Lingüística, 22(1)*, 211-12.

---------(1991a). Relevancia de los elementos de las oclusivas sordas del castellano según la discriminación auditiva. In E. Martínez-Celdrán (Ed.), *Fonética experimental: Teoría y práctica* (pp. 115-130). Madrid: Síntesis (Lingüística, 6).

---------(1991b). Tensión frente a sonoridad en las consonante mates del castellano. In E. Martínez Celdrán (Ed.), *Fonética experimental: Teoría y práctica* (pp. 131-141). Madrid: Síntesis (Lingüística, 6).

Miller, J., & Dexter, E. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human perception and performance, 14(3)*, 369-378.

Möbius, B. (2004). Corpus-based investigations on the phonetics of consonant voicing. *Folia Linguistica, 38(1-2)*, 5-26.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., & Sinkkonen, J. (1997). Language specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432-434.

Newman, M. (2003). The Structure and function of complex networks. *SIAM Review, 45(2)*, 167-256.

Ni Chasaide, A. & Gobl, C. (1993). Contextual variation of the vowel voice source as a function of adjacent consonants. *Language and Speech, 36(2-3)*, 303-330.

Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *J. Acoust. Soc. Am., 115*, 1777-1790.

Norman, D. (1988). *The Design of Everyday Things*. Basic Books: New York.

Obendiente Sosa, E. (1998). Fonetismo segmental. *Español Actual, 69*, 11-18.

Ortega-Llebaria, M. (2004). Interplay between phonetic and inventory constraints in the degree of spirantization of voiced stops: Comparing intervocalic /b/ and intervocalic /g/ in Spanish and English. In T. Face (Ed.), *Laboratory Approaches to Spanish Phonology* (pp. 237-253). New York: Mouton de Gruyter.

Pérez, H. (1998). Influence of two acoustic features in the perception of /p t k/ vs. /b d g/ correlation. *ICSLP 98 Proceedings of the 5th International Conference on Spoken Language Processing*, 30th November- 4th December 1998. Sydney, Australia.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., & Roberts, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *J. Cognit. Neurosci., 12*, 1038-1055.

Pickett, J. (1999). *The Acoustics of speech communication: Fundamentals, speech perception theory, and technology*. Boston: Allyn and Bacon.

Pisoni, D. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *J. Acoust. Soc. Am., 61*, 1352-1361.

Pisoni, D., Aslin, R., Perey, A. & Hennessy, B. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human perception and performance, 8*, 297-314.

Pisoni, D. & Luce, P. (1987). Trading relations, acoustic cue integration, and context effects in speech perception. In M. Schouten (Ed.), *The Psychophysics of Speech Perception* (pp. 155-172). Dordrecht: Martinus Nijhoff.

Poeppel, D., & Monahan, P. (2008). Speech perception: Cognitive foundations and cortical implementation. *Current Directions in Psychological Science, 17(2)*, 80 85.

Ravizza, S. (2001). Relating selective brain damage to impairments with voicing contrasts. *Brain and Language, 77*, pps. 95-118.

Repp, Bruno H. & Lin, H.-B. (1991). Effects of preceding consonant on the Voice-Onset Time category boundary. *Journal of Experimental Psychology: Human perception and performance, 17(1)*, 289-302.

Riede, T., Mitchell, B., Tokuda, I., & Owren, M. (2005). Characterizing noise in nonhuman vocalizations: Acoustic analysis and human perception of barks by coyotes and dogs. *J. Acoust. Soc. Am., 118(1)*, 514-522.

Romero, J. (1992). An experimental analysis of spirantization in Spanish. *J. Acoust. Soc. Am., 92(4)*, 2pSP14.

Romero, J., Parrell, B., & Riera, M. (2007). What distinguishes /p, t, k/ from /b, d, g/ in Spanish? Poster presented at *PaPI 2007*. Braga, Portugal. Sept. 6th, 2007.

Saffran, J., Aslin, R., & E. Newport. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Sinex, D., McDonald, L. & Mott, J. (1991). Neural correlates of nonmonotonic temporal acuity for Voice Onset Time. *J. Acoust. Soc. Am., 90*, 2441-2449.

Sjölander, K. & Beskow, J. Wavesurfer (2006). An open source speech tool, Centre for Speech Technology, KTH, Drottning Kristinas väg 31, SE-100 44, Sweden.

Stevens, K. & Blumstein, S. (1981). The search for invariant acoustic correlates of phonetic features. In P. Eimas & J. Miller (Ed.s), *Perspectives on the Study of Speech* (pp. 1-38). Hillsdale, NJ: Erlbaum.

Stevens, K. & Keyser, S. (1989). Primary features and their enhancement in consonants. *Language, 65*, 81-106.

Studdert-Kennedy, M. (1987). Phoneme as a perceptuomotor structure. In A. Allport, D. Mackay, W. Prinz & E. Scheerer (Ed.s), *Language Perception and Production: Relationships between Listening, Speaking, Reading and Writing* (pp. 67-84). Orlando, Florida: Academic Press.

Torreblanca, M. (1983). La sonoridad de las consonantes oclusivas españolas. *Hispania, 66*, 391-395.

Trujillo, R. (1980). Sonorización de sordas en Canarias. *Anuario de Letras de México, 18*, 247-254.

Villamizar, T. (1998). Fonetismo. In E. Obediente Sosa (Ed.), *El habla rural de la Cordillera de Mérida* (pp. 27-89). Mérida, Venezuela: Editorial Venezolana.

Watts, D. & Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature, 393(4)*,440-442.

Werner, S. & Keller, E. (1994). Prosodic aspects of speech. In E. Keller (Ed.), *Fundamentals of Speech Synthesis and Speech Recognition* (pp. 23-39). West Sussex, England: John Wiley & Sons Ltd.

Whalen, D., Abramson, A., Lisker, L., Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *J. Acoust. Soc. Amer., 4(1)*, 2152 2159.

Williams, L. (1976). Prevoicing as a perceptual cue for voicing in Spanish. J. *Acoust. Soc. Am., 59*, S41.

Yeou, M., Honda, K., Maeda, S. & Embarki, M. (2007). Laryngeal activity in the production of consonant clusters and geminates in Moroccan Arabic. Presented at *La co-articulation: Indices, direction et représentation*. Montpellier, France, December 7th 2007.

Zamora Munné, J. & Guitart, J. (1982). *Dialectología hispanoamericana*. Salamanca: Ediciones Almar, S.A.

Zampini, M. (1996). Voiced stop spirantization in the ESL speech of native speakers of Spanish. *Applied Psycholinguistics, 17*, 335-354.

Vita

Stephanie Lain was born in Houston, Texas on September 11, 1969. After graduating from Westlake High School, Austin, Texas in 1988, she enrolled in the University of Texas at Austin. She received a Bachelor of Arts degree from the University of Texas in 1993. During the years 1993-1996, Stephanie lived in Mérida, Venezuela, where she taught English as a foreign language. In September, 2000, she entered the Graduate School at the University of Texas at Austin. She received a Master of Arts degree in Spanish in 2003. During her time at the University of Texas, Stephanie has been employed by the Department of Spanish and Portuguese as a Teaching Assistant, an Assistant Instructor and a Graduate Assistant. She has over twelve years of foreign language teaching experience. She has given conference talks at international conferences in phonetics and phonology and was the recipient of Professional Development Awards from the office of Graduate Studies in 2007 and 2008.

Permanent address: 2503-C Bridle Path, Austin, TX 78703

This dissertation was typed by Stephanie Lain.