Effects of contrast and articulatory precision on the realization of sibilants

Q-Exam

13 May 2019

Katherine Blake

Abstract

This study investigates the effects of two hypothesized competing pressures on the phonetic realization sibilants: inventory size and articulatory precision, via a production experiment. Voiceless sibilant fricatives in three languages of differing sibilant inventory sizes are examined: Spanish /s ʧ/, Catalan /s z ʃ ʒ ʦ ʤ ʧ ʤ/ and English /s z ʃ ʒ ʧ ʤ/. Results indicate a clustering effect on the within-category variation of the /s/ measured via center of gravity in Spanish, compared to that in Catalan or English, suggesting an effect of inventory size rather than an underlying strict articulatory precision requirement for /s/.

The constraint of inventory size on production is the hypothesis put forth by Dispersion Theory (Liljencrants and Lindblom (1972)). The notion that more phonemic contrasts result in an expansion effect on the distance between categories, or a clustering effect on the tokens within them so as to avoid perceptual overlap is an intuitive one. However, it lacks significant empirical support (e.g., Manuel (1990); Bradlow (1995); Evers et al. (1998)). Alternatively, Keating (1983) claims that sibilants require a relatively high level of articulatory precision, which is specified underlyingly. We see conflicting results of variability and precision of sibilants in various studies (e.g., Keating (1983), low variation in jaw height, and Tabain (2001) low variation in center of gravity; but Iskarous et al. (2011), variable jaw height).

Given these competing theoretical claims and empirical findings, a production experiment collected both acoustic (center of gravity; CoG) and articulatory data (lip shape/aperture) to investigate the effects of inventory size (H1) and articulatory precision (H2) hypothesized in the literature. Eight Spanish speakers (4F; 4M), three Catalan (2F; 1M), and eight English (4F; 4M) participated. Standard deviations were calculated over speaker-normalized CoG (nCoG) values by-speaker for /s/. A greater standard deviation expresses a wider distribution, and greater variance. Results show nCoG of Spanish /s/ has a mean standard deviation of 799.63 across speakers, Catalan /s/ 496.15, and English /s/ 429.15. Results of an F-test conducted on the nCoG data to test if the variances in CoG are significantly different between languages are below.

Language Pair	F statistic	Degrees of Freedom	p value	Mean Standard Deviation Difference
Catalan - Spanish	0.72	df1 = 88 df2 = 221	p = 0.08	-303.48
English - Spanish	0.68	df1 = 235 df2 = 221	p < 0.01	-370.48
English - Catalan	0.95	df1 = 235 df2 = 88	p = .76	-67.00
Catalan/English - Spanish	0.70	df1 = 324 df2 = 221	p < 0.01	-352.21

F-test results indicate that nCoG of Spanish /s/ has significantly more variance than English /s/. Pooling Catalan and English together to test against Spanish, results of the F-test show that nCoG of an /s/ that forms part of a larger inventory has a statistically lower variance (smaller standard deviation) than the /s/ in a smaller inventory language. This result indicates that the articulation of /s/ may be quite flexible depending on the language or inventory size, contra strict articulatory precision claims. This result lends support to the DT hypotheses (H1): larger inventories show a clustering effect on within-category variation.

1 Introduction

A long-standing issue in phonology is what governs the organization of inventories of speech sounds. This paper investigates how the phonetic production of sounds may guide the inventory. The relationship between inventory size and phonetic production is an intuitive one: if there are many categories to contrast, they should be sufficiently far apart in the acoustic space to be perceptually distinct, and their tokens are expected to be more tightly produced in order to avoid overlap in the signal with other categories.

For example, vowels in a ten-vowel system, compared to the same vowels in a five-vowel system, would show an expansion effect on their distance from other vowels, and the tokens of these vowels would show a clustering effect in their phonetic realization due to crowding of the phonological system in order for there to be sufficient perceptual contrast. This is the key idea behind Dispersion Theory (Liljencrants and Lindblom (1972); Lindblom (1986)).

Despite this widely held intuitive assumption, empirical results of the proposed effect of inventory size on phonetic production lack compelling support. Several studies show no expansion effect on the distances between categories of larger inventories, and no clustering effect on the within-category variation of those categories (e.g., Manuel (1990); Bradlow (1995); Lindau and Wood (1977); Disner (1983)).

The aforementioned work, and the original conception of Dispersion Theory, has attempted to find this effect in vowel systems, but others have extended DT to consonant inventories as well (e.g., Lisker and Abramson (1964); Lahiri et al. (1984); Jongman et al. (1985); Utman and Blumstein (1994)). Vowels and consonants are fundamentally different, both phonologically (in how they pattern and what processes they undergo) and phonetically (in their articulation and acoustic correlates). We may therefore expect effects of inventory size on production also to be different between vowel and consonant inventories. For example, the boundaries between different vowel qualities are often hard to clearly delineate and the vowel space has much more room to expand and contract. The articulation and acoustics of consonants are more constrained.

This constraint is particularly evident for the sibilant consonant class, which requires a relatively greater level of articulatory and acoustic precision (Keating (1983); Bjorndahl (2018)). Because of this reported precision, if variation is then found in the data, this may indicate a clearer result in support of the hypothesized relationship between inventory size and production. Sibilants may then better help us understand this issue of the interaction between inventory size and production. This issue of inventory size and production variance is investigated here in turn, this time using sibilants due to their reported relative precision.

To test this effect, a production study was conducted to probe the within-category variation of the sibilant fricatives in Spanish, Catalan, and English. These three languages have differently-sized sibilant inventories: Spanish /s ʧ/, Catalan /s z ʃ ʒ ʦ ʤ ʧ ʤ/ and English /s z ʃ ʒ ʧ ʤ/. Only the voiceless sibilant fricatives of these languages are the focus of study. DT predicts a greater degree of within-category variation for the Spanish /s/ than its Catalan and English counterparts, due to the relatively smaller siblant inventory.

The paper is organized as follows. Section 2 lays out the relevant background literature. The methods of the experiment are provided in Section 3, the results of which are presented in Section 4. Section 5 offers a discussion. Section 6 concludes.

2 Background

To lay the groundwork for this study, we need to know how sibilants are generally organized within a given system, what previous phonetics studies have found on the languages investigated here and cross-linguistically, what the theories about the organization of sound systems predict, and what measures should be taken to test these predictions in the current study.

In this section, then, the typological and phonological situation of sibilants in the broader scheme of attested inventories is discussed in 2.1. Section 2.2 highlights previous phonetic work on sibilants, cross-linguistically and language-specific to this study. Then, in order to understand how phonological inventories may be governed, two competing theories are pre-

sented, Dispersion Theory (2.3) and Quantal Theory (2.4), with a note on Feature Economy (2.5). General phonetic properties of sibilants are laid out in 2.6, leading to the hypotheses and predictions of the current study in 2.7.

2.1 Phonology of Sibilants

Strident fricatives, or sibilants, are quite different from non-strident fricatives. Their opposition has long been noted in the phonological literature. Jakobson et al. (1951) had strident/mellow as one pair of binary acoustically-defined features, strident referring to sibilant fricatives and mellow referring to non-sibilant fricatives. Sibilant and nonsibilant fricatives are also articulated quite differently, and Chomsky and Halle (1968) maintained strident as a binary *articulatory*-defined feature, which persists in more recent frameworks (e.g., Clements (1990); Hume (1994); Clements and Hume (1995)). This opposition is generally grounded in the difference in the type of frication produced (i.e., the relative amount of noisiness, Bjorndahl (2018)); these phonetic characteristics that underlie the phonological classification are examined in more detail in section 2.6.

Typologically, the presence of a fricative in a language's inventory is extremely common: inventories with at least one fricative make up 93.4% of those in Maddieson (1984)'s survey. Of particular interest to this study, that one fricative is usually an /s/ sound (88.5% of languages), followed by /f/ or /ʃ/ (Bjorndahl (2018); Maddieson (1984)). The sibilants investigated in this study, /s/ and /ʃ/, number among the top 20 most common speech sounds in the same survey, with one of the two occurring in 87% of languages surveyed, and both of them in 47%.

There is something interesting to note between the nature of fricative versus stop organizations within inventories: stops usually occur in a series, contrasting at multiple places of articulation and with multiple phonation specifications; however, fricative inventories of the world seem to fall into two categories. (1) they display a series organization like that of stops and nasals, or, more commonly, (2) they are one-offs or occur in pairs (though not

necessarily a voiceless/voiced pair). Rarely does a language contrast fricatives at more than the three major places of articulation (labial, coronal, velar) like stops (Bjorndahl (2018)). This organizational difference may be an effect of the articulatory precision required for fricatives, and sibilants more specifically.

2.1.1 A note about voiced sibilants

Catalan and English voiceless sibilant fricatives and affricates have complete corresponding voiced series; however, these are not considered here. There are several reasons for this. For one, Spanish does not have contrastive /z/, it is only allophonic, so cross-language comparison of voiced sibilants is not possible in these three languages. And, unlike non-strident fricatives, voiced sibilants are generally well-established counterparts to the opposing voiceless sounds (Bjorndahl (2018)). Given these two facts, the voiced members of the Catalan and English sibilant inventories are not the focus here, but the reader is referred to Bjorndahl (2018) for a compelling discussion of the nature of voiced fricatives.

Previous phonetic work on various languages with differing sibilant inventories are now examined in turn, cross-linguistically (2.1.1) and specific to this study (2.1.2 - 2.1.4).

2.2 Previous Phonetic Work on Sibilants

2.2.1 Cross-linguistic

Much of what we know about the acoustics and production of sibilants comes from English (e.g., Hughes and Halle (1956); Shadle (1985, 1990, 1991); Jongman et al. (2000)). This body of work has largely found that correlates of sibilant place of articulation can be reliably measured in the noise of the sibilant itself, using spectral information like center of gravity, kurtosis and spectral tilt. Measures that are not taken from the noise of the sibilant, such as amplitude and duration, are less reliable.

While the majority of previous work on fricatives remains heavily biased towards English, there has been some notable work on a more diverse set of languages. Ladefoged and Mad-

dieson (1996) present a summary of several cross-linguistic studies on sibilants describing their various articulatory possibilities in the world's languages, reproduced below in Table 1.

		"Place of articulation"	Exemplifying languages
1	s̪	apical dental	Chinese, Diegueño, Polish
2	s	apical or laminal alveolar	English, Ubykh
3	s̻	laminal alveolar	Toda
4	s̺	laminal flat post-alveolar	Chinese, Polish, Ubykh
5	s̼	apical post-alveolar	Diegueño, Toda
6	ʃ	apical or laminal domed post-alveolar (or palato-alveolar)	English
7	ʃ̻	laminal domed post-alveolar	Toda
8	ɕ	laminal palatalized post-alveolar (alveolo-palatal)	Chinese, Polish, Ubykh
9	ŝ	laminal closed post-alveolar ('hissing-hushing')	Ubykh
10	ʂ	sub-apical palatal (sub-apical retroflex)	Toda

Table 1: Types of Sibilants from Ladefoged and Maddieson (1996), p. 164

Gordon et al. (2002) conducted a cross-linguistic acoustic study of seven mostly unrelated languages: Aleut (Western dialect), Apache (Western dialect), Chickasaw, Scottish Gaelic, Hupa, Montana Salish, and Toda[1]. Fricative inventories for these languages range between four (Western Aleut, Chickasaw) and nine (Toda) members; however, sibilants number between two and four members of these inventories. Sibilants contrast at alveolar and postalveolar places of articulation /s ʃ/ (Western Apache, Chickasaw, Hupa, and Montana Salish); alveolar and palatal /s ç/ (Western Aleut); alveolar, postalveolar, and palatal /s ʃ ç/ (Scottish Gaelic); and lamino-dental alveolar, apical alveolar, laminal postalveolar, and sub-apical retroflex /s̪ s ʃ s/ (Toda). Acoustic analyses carried out on fieldwork-sourced recordings of these languages (generally controlled for vowel environment) included duration, center of gravity, and spectral shape[2]. Results from this study are summarized in Table 2.

In general, duration was found to poorly differentiate between different sibilant places

[1]The qualifier "mostly" is used because Hupa and Western Apache are both Athabaskan (Na Dene phylum) languages, but belonging to different branches (Pacific coast and southern, respectively).

[2]Formant transitions were also analyzed in languages with velar and uvular fricatives, as this has been shown to help distinguish between these two places of articulation. None of these back fricatives are examined in the present study, so this measure is not discussed in detail here.

Language	Sibilant Inventory	Reliable Measures
Aleut (Western)	/s ɕ/	center of gravity, spectral shape
Apache (Western)	/s ʃ/	center of gravity, spectral shape
Chickasaw	/s ʃ/	center of gravity, spectral shape
Scottish Gaelic	/s ʃ ɕ/	center of gravity, spectral shape
Hupa	/s ʃ/	center of gravity, spectral shape
Montana Salish	/s ʃ/	center of gravity, spectral shape
Toda	/s̻ s ʃ s̺/	center of gravity, spectral shape

Table 2: Results summary from Gordon et al. (2002)

of articulation; however, center of gravity was a reliable differentiator. /s/ generally had the highest center of gravity in all the examined languages, except for Toda, where /s̻/ was highest. /ʃ/ also had a significantly higher center of gravity than /ɕ/ in Western Aleut. The authors also compared spectral shapes of fricatives, which were calculated by averaging the spectral slice at the center of a fricative across tokens for a speaker or across speakers of the same gender. Across languages, support for predictions made by the vocal tract model was found: generally, the height of the most pronounced peak correlated with the anteriority of the place of articulation of the fricative. Speakers and languages showed considerable uniformity in spectral shapes, and thus it was generally a good differentiator in place of articulation. There was, however, notable inter-speaker variation found in spectral shape of /s/ within languages. This is attributed to variation in place of articulation, and length, shape, and position of the tongue. Different languages and different sounds may therefore employ different places of articulation along which to contrast members of a category, as shown by the cross-linguistic data presented by Gordon et al. (2002). This must be kept in mind when examining fricative data that is not English.

There are further distinctions possibly relevant for this study, apical and laminal. In her electropalatographic and acoustic study on English and French sibilant fricatives, Dart (1991) found differences in the spectrum between apical and laminal articulations. Within dental sibilants, "the laminal articulation has a steeper slope and rises higher in the high frequencies" compared to apico-dental sibilants. Within alveolars, "it is the apicals which

have the greater amount of high frequency energy, the laminals having an essentially flat spectrum" (Dart (1991), p. 83).

Dart (1991) and Ladefoged and Maddieson (1996) both found that the English speakers in their studies were split roughly 50-50 in whether they produced an apical or laminal /s/. The /s/ in Catalan is described as apical, while the /s/ in Latin American Spanish is described as laminal (Wheeler (2005); Hualde (2005)). All three are generally described as alveolar. An apical-laminal distinction between fricatives within a language is rare, but the distinction is made in O'odham (Dart (1993)) and existed in Ubykh (now extinct; Ladefoged and Maddieson (1996)). Acoustically, the distinction between the apical and laminal alveolar sibilant fricatives can be seen in different spectral shapes in Dart (1993), and the apical post-alveolar sibilant fricative in Ubykh has a higher center of gravity (around 2500Hz) than the laminal post-alveolar (around 2000Hz; Ladefoged and Maddieson (1996)). Center of gravity results reported in section 4 do not show the same patterning as Ubykh: the apical /s/ in Catalan has a lower center of gravity than laminal /s/ in Spanish, though this difference could be driven by other factors (e.g., constriction location or lip shape).

2.2.2 Spanish

Spanish has only the sibilants /s ʧ/, and there is no corresponding contrastive voiced series; however, in many dialects [z] is an allophone of /s/. Perhaps relevant in the speech of the participants in this study, in some parts of Mexico (central highlands), /s/ may alternate with [h] ("/s/ aspiration"), but this alternation is subject to variation, occurring more often in casual speech and in syllable-final position, especially pre-consonantally (not the phonological environment examined in this study; Hualde (2005)). For these reasons, it is not a concern here. Another few facts worth mentioning are: the *ceceo* phenomenon – a distinction between /s/ and /θ/ made in many parts of Spain – is not present anywhere in Latin America, the broad dialect region of all the speakers in the present study; and, the fricative /x/ is also described as "less strident" in Latin American Spanish, so it does not form part of this

study either (Hualde (2005)). The reader may also be curious about the nature of /ʝ/. The phonemic status of this highly variable sound is debated in the literature. In some regions of Mexico it may be variably affricated, but in northern Mexico, the southwestern United States, and parts of Central America, it is realized as a glide (Hualde (2005)). Given its marginal status at best, it is also not considered in this study. The full inventory of contrastive consonants is below[3]. The full sibilant inventory is in the dashed box, but the sibilant analyzed in the current study is circled.

	Bilabial	Labio-dental	Dental	Alveolar	Palatal	Velar
Plosive	p b		t d			k g
Affricate					ʧ	
Fricative		f		s	(ʝ)	x
Nasal	m			n	ɲ	
Tap				ɾ		
Trill				r		
Lateral				l		

Figure 1: Phonemes of Spanish

Much of the previous work on Spanish /s/ characterizes the sounds as extremely variable, in both its phonological patterning – Mason (1987) cites at least three allophones at various places of articulation in different dialects – and its phonetic realization. Phonetically, center of gravity was found to be quite variable between speakers and dialects, ranging averages of 2700 - 3400Hz between New York City Spanish speakers, and those who had recently immigrated there. Both groups had an equal distribution of Caribbean and mainland Latin American origins (Erker (2012)). Center of gravity measurements in the present data are expected to be similar to this range.

[3]There are, of course, many Spanishes in the world. The above inventory is a basic one. Most of the participants in the present study spoke Chicano/Mexican Spanish, or other dialects of Spanish with the same two sibilants.

2.2.3 Catalan

Catalan has four voiceless sibilants /s ʃ ʦ ʧ/, and a corresponding voiced series /z ʒ ʣ ʤ/. The full inventory of contrastive consonants in Catalan is below (Wheeler (2005))[4]. The complete sibilant inventory is in the dashed box, but the sibilants analyzed in the current study are circled.

	Bilabial	Labio-dental	Lamino-dental	Apico-alveolar	Lamino-alveopalatal	Dorso-palatal	(Labio-) Velar
Plosive	p b		t d				k g
Affricate				ʦ ʣ	ʧ ʤ		
Fricative		f		s z	ʃ ʒ		
Nasal	m			n	ɲ		ŋ
Glide						j	w
Tap				ɾ			
Trill				r			
Lateral				l	ʎ		

Figure 2: Phonemes of Catalan

Previous work on fricatives in Eastern Catalan (as spoken in and around Barcelona, and the dialect examined here) shows that the /s/ has a more retracted pronunciation and is more /ʃ/-like, but still distinct from the lamino-alveopalatal (Recasens and Mira (2013)). This anterior place of articulation makes the front cavity larger and thus results in a lower spectral peak around 3500 - 4500Hz (Recasens and Espinosa (2007)). In their combined electropalatographic (EPG) and acoustic study, Recasens and Mira (2013) found that Catalan speakers made clear articulatory and acoustic distinctions between the two sibilants. Acoustic data showed a significant different in center of gravity between the two fricatives, but not duration. This is similar to English and other languages of the world.

[4]Some varieties of Catalan (Balearic and Valencian) retain /v/, but this contrast is merged with /b/ and [β] in other dialects Wheeler (2005). Speakers in this study are all from Barcelona, so /v/ is not included in the above inventory.

2.2.4 English

English has three sibilants /s ʃ ʧ/, and corresponding voiced /z ʒ ʤ/; the entire contrastive inventory of consonants is shown below (Quirk et al. (1972)). The full sibilant inventory is in the dashed box, but the sibilants analyzed in the current study are circled.

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palatal	(Labio-) Velar	Glottal
Plosive	p b			t d			k g	
Affricate					ʧ ʤ			
Fricative		f v	θ ð	s z	ʃ ʒ			h
Nasal	m			n			ŋ	
Glide						j	w	
Rhotic				ɹ				
Lateral				l				

Figure 3: Phonemes of English

There are many well-established measures in the noise of the sibilant fricatives that can significantly differ based on place of articulation. Jongman et al. (2000) in their thorough experimental study on 20 native speakers of English, found that English sibilants have well-defined and distinct spectral shapes, with the alveolar having a relatively higher primary spectral peak around 7000Hz and the post-alveolar having a lower spectral peak around 4000Hz. Spectral tilt, or skewness, was also significant in distinguishing POA between the sibilants in English, with the post-alveolar having a lower spectral tilt. Conversely, the alveolar has a significantly higher kurtosis value than the post-alveolar.

In addition to the four spectral moments and spectral peak location, other acoustic cues may also help determine POA. Jongman et al. (2000) also found that second formant transitions from sibilant to vowel may be significant in determing POA, with a higher F2 transition value for the posterior fricative. The relative amplitude of the alveolar was also found to be significantly higher than for the post-alveolar, but other studies show mixed results on this (e.g., Behrens and Blumstein (1988). Duration, however, was not significant in distinguishing between /s/ and /ʃ/ (Jongman et al. (2000)).

In their perception study, Heinz and Stevens (1961) tested English speakers on synthesized tokens of [f], [θ], [s], and [ʃ] in isolation and prevocalic position. They found that speakers differentiated between the sibilants [s] and [ʃ] primarily based on cues in the frication noise alone, rather than in formant transitions. Other work corroborates these findings (e.g., Hughes and Halle (1956); Bladon and Seitz (1986)).

2.2.5 Summary

The previous production work on sibilants indicates several reliable spectral measurements that discriminate place of articulation of sibilants across languages including spectral peak location, spectral shape, and perhaps most consistently, center of gravity. Spectral information has also been affirmed in work on perception to be a primary cue for place of articulation distinction in sibilants. Given these two pieces – production and perception – the results presented in section 4 will focus on spectral data, emphasizing center of gravity.

I now move on to review two prominent theories in the literature about how the organization of phoneme inventories is governed: *Dispersion Theory* and *Quantal Theory*.

2.3 Dispersion Theory

The theoretical relationship between inventory size and variance was first investigated in the early work of Liljencrants and Lindblom (1972), formalized as Dispersion Theory (DT). The central claim in DT is that the relative distance between categories is such that there is maximal, or sufficient, perceptual contrast. DT was initially proposed to model the structure of vowel inventories given a certain number of vowels (ranging from three to twelve) contrastive in the inventory, claiming that even in small vowel inventories, the vowels are maximally distant from each other (rather than maintaining some set distance) in order for there to be maximal perceptual contrast. This idea is schematized in Figure 4. DT was intended to account for the typological cross-linguistic trends based on this theory of maximal, or later sufficient, perceptual contrast.

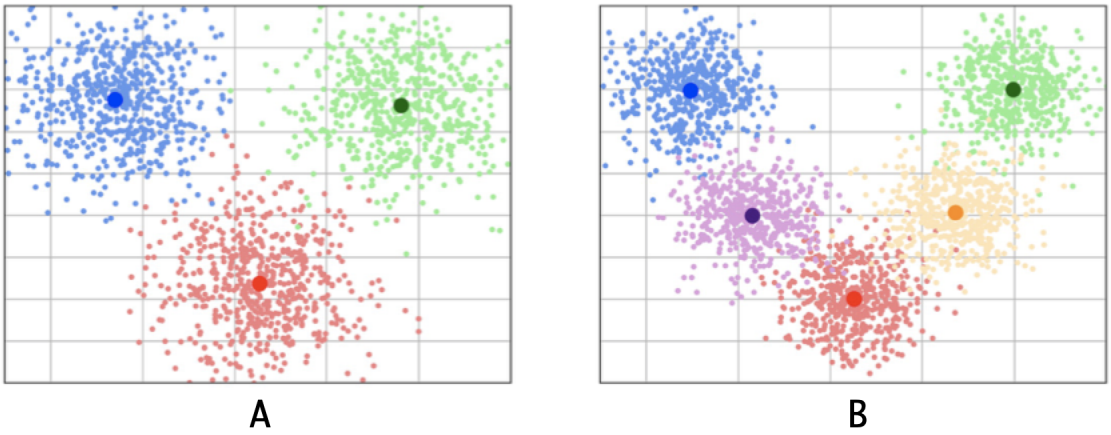

Figure 4: Schematic of original DT predictions. Large dots mark the mean of their distributions, which are represented by smaller dots of the same color. Figure 4A schematizes a smaller language inventory (three categories), than Figure 4b (five categories).

Figure 4a schematizes the dispersion of a smaller language inventory with only three contrasting categories, and Figure 4b the dispersion of a larger inventory, which has the same three contrasting categories of A, but with an additional two. This figure shows larger mean-to-mean distances between the same three categories in Language B (red, blue, and green) compared to Language A, as well as less within-category variation. In essence, DT predicts that larger inventories will tend to show an expansion effect on the distances between their categories (Figure 4a) and/or less variation (i.e., tightening or clustering effect; Figure 4b) within them. However, various works show that these predictions do not hold even for vowels.

2.3.1 Dispersion of Vowels

Oft-cited in the literature for experimental support of DT hypotheses, I argue that Manuel (1990) needs to be reexamined. Her experimental study looked at vowels in three Southern Bantu languages with different vowel inventories. Shona and Ndebele have /i e ɑ o u/ and Sotho /i e ɛ ɑ ɔ o u/. Shona and Ndebele are in opposition with Sotho, which has a more crowded inventory, in the low and mid-vowel space, shown in Figure 5.

Three male speakers of each language read nonse words of the skeleton pVpVpV, where

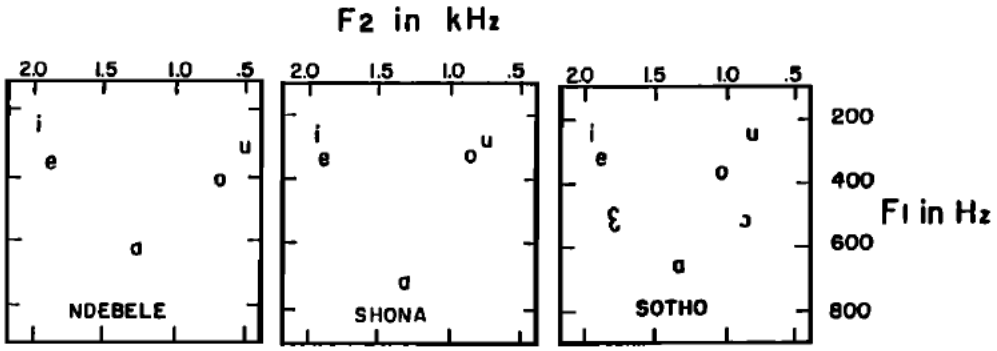

Figure 5: "Examples of phonemic vowels of Ndebele, Shona, and Sotho. Data are from one speaker of each of the languages" (Manuel (1990), p.1287; original caption).

V could be any of the vowels in the system, within a carrier phrase. However, there was an important issue with the experimental design:

> "[S]ubsequent listening to the audio recordings revealed that, for two of the Sotho speakers, there was a very large variability in the production of orthographic *e*, ranging from [i] - [e], and one Sotho speaker produced a lower mid vowel, more like English /ɛ/. This observation indicates that, for the vowel /e/, we cannot be certain whether subjects were consistently producing a single phonemic vowel, and, if not, which one in a given case. Though this results in some noise in the data, it does not crucially affect the ability to test the hypotheses under consideration" (Manuel (1990), p. 1288).

I argue that this does, in fact, crucially affect the testability of the hypotheses. The study is specifically looking at coarticulatory effects on these vowels, predicting that Shota and Ndebele vowels /e/ and /ɑ/ are coarticulated more than their counterparts in Sotho; but how can this be examined when the *e* vowel for Sotho speakers was ambiguously /i/ ∼ /e/ ∼ /ɛ/? It is critical that this vowel be /e/, given that /ɛ/ is not contrastive in the other two languages, and /i/ is a high vowel and therefore not in the crucially "more crowded" mid/low vowel space. This noise in the data was also, seemingly, not removed from the analysis, obscuring the results which showed greater anticipatory coarticulation for /ɑ/ in Shona and Ndebele (as expected), but no difference in the midvowels /e/ and /o/ (where the same result as /ɑ/ was expected).

Moreover, the more basic question asked by DT is how inventory size affects the distance

between vowels, rather than their degree of coarticulation. But even here, there appears to be no effect (Figure 6). There is no noticeable difference in the distance between /e/ - /ɑ/ - /o/ between the three languages, when the predicted result is an expansion effect on the vowels in the more crowded system, Sotho. Then again, it is hard to interpret these results given the above methodological error. In any case, this work does not provide support to DT hypotheses *per se*.

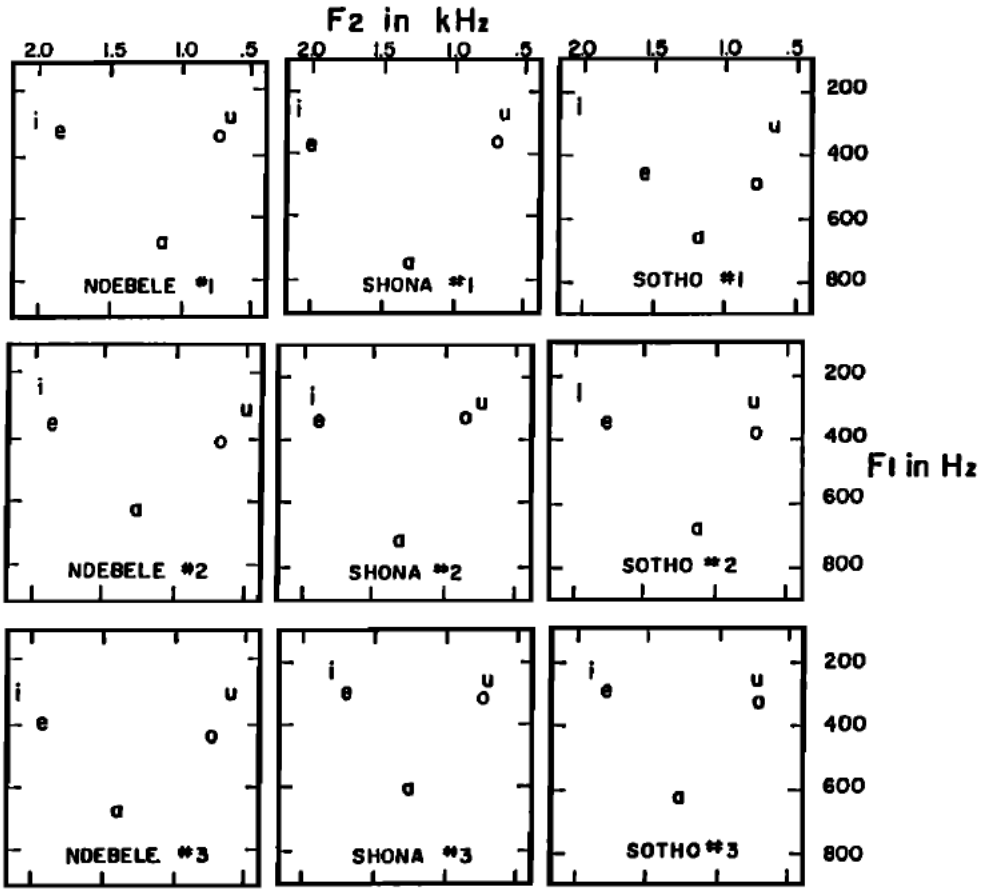

Figure 6: "Average F1 and F2 values for the middle of target vowels /a/and /e/for the nine subjects. Values for the vowels /i/,/o/, and /u/ are from the medial vowel in /pipipi/, /popopo/, and /pupupu/, respectively, and are based on from two to five tokens each. Values for /a/ and /e/ are based on more contexts, as indicated in the text" (Manuel (1990), p.1289; original caption).

In another work, Bradlow (1995) focuses on within-category variation (rather than distance) as an effect of inventory size on the production of vowels between Spanish, which has

a relatively common five-vowel system, and English, which has a relatively rare eleven-vowel system (excluding diphthongs). She found that "the data showed no difference between tightness of within-category clustering." Her results also indicate no consistent effect of dispersion on the relative distances between vowels in each system. Thus, these works on the dispersion of vowel systems provide no empirical support for DT, along either the distance or within-category variation dimensions, and neither do other cross-linguistic acoustic studies (e.g., Lindau and Wood (1977); Disner (1983)).

Though work on the bearing of DT hypotheses on vowel inventories lacks empirical support, some have extended DT to consonant inventories. This work is reviewed in the following subsection.

2.3.2 Dispersion of Consonants

While DT was originally conceived to account for the structure of vowel inventories, some work has applied the framework to consonants. Given that consonants and vowels have different phonological behaviors and phonetic properties, we may expect DT predictions to be different for consonant systems than vowel systems. These predictions have been tested in studies like Lisker and Abramson (1964), who compared voicing distinctions in 11 languages; Lahiri et al. (1984), stop consonants in three languages; Jongman et al. (1985), coronal stops in three languages; and Utman and Blumstein (1994), labiodental fricatives in two languages; and others.

In her forthcoming dissertation, Hauser (2019) tests for effects of inventory size on phonation and place of articulation with a production experiment on speakers of Hindi, which has a four-way contrast at four different places of articulation, and English, which has a two-way contrast at three places of articulation. DT predicts that larger inventories will show less variation in the phonetic realization of their categories than smaller inventories due to pressure to maintain distinctive perceptual categories via less overlap. Therefore, the expected results on Hindi and English are that, among the parallel sets of stops between the two

languages, Hindi will show less variation in phonation and place of articulation because its stop inventory is much larger (16 versus 6). In terms of looking at phonation, Hauser (2019) does not find this when looking at voiceless lag time: Hindi speaker data show just as much variation as the English speaker data. However, when examining variation in the prevoicing dimension, Hindi does exhibit less variation than English. This is important because prevoicing has been shown in the perception literature to be the primary cue for the phonological voicing contrast in Hindi (and a secondary cue in English), and motivations behind DT come from the perception domain. Taking into account what acoustic cues have been shown to matter for perceptual distinction of contrast in a particular language may reveal different results related to DT. However, it may also be the case that phonation and place of articulation behave quite differently in this respect. While breathy/aspirated and creaky phonation exist, most languages contrast only two phonation types, voiced or modal, and voiceless (Berkson (2013)). Compared to the much larger set of attested place of articulation contrasts, it is possible that the pressures that guide the organization of phoneme inventories affect the realization of a multi-modal set and an often bi-modal set differently. Namely, phonation may not be subject to the same organizational pressures as place of articulation.

More specifically relevant to the current study, Evers et al. (1998) compared [s] and [ʃ] in three languages: English, Dutch, and Bengali. The idea was to look for acoustic differences in these two phones, where they are contrastive in English, and allophonic (/s/ $\sim$ [ʃ]) in Dutch and (/ʃ/ $\sim$ [s]) in Bengali. Using spectral steepness, which quantifies gross spectral shape, the authors found that there was not a systematic difference between phonemic sibilant language English and allophonic sibilant languages Dutch and Bengali. All three equally varied on where they placed the spectral "boundary" between these two sounds. However, it appears that English speaker data showed greater acoustic distance between the two sounds and less overlap than Dutch and Bengali, which generally showed closer acoustic distance and more overlap, as quantified by the spectral steepness metric (see Figure 3, p.356).

2.3.3 Dispersion Metrics

Studies in the previous two subsections have touched both on results of the hypothesized expansion effect (i.e., measuring distance) and tightening effect (i.e., measuring within-category variation). This study focuses on the latter.

DT was eventually updated to include this within-category variation as a factor as in addition to mean-to-mean distances (Lindblom (1986)). Distributional information of a category is important. Hauser (2017) compares traditional mean-to-mean measures to a newly proposed metric for measuring the dispersion of consonant inventories which takes within-category variation into account, and finds that it changes the resultant ranking of most dispersed inventories (thought the results still do not support DT hypotheses).

Still, within-category variation information is relevant. In its original conception, DT was perceptually motivated: dispersed categories facilitate perceptual distinction of these categories (Liljencrants and Lindblom (1972)). Studies on speech perception have shown that perception is affected by a category's distribution (Clayards et al. (2008); McMurray et al. (2002); Pisoni and Tash (1974); Clarke and Luce (2005)). On these grounds, Hauser (2017) proposes that our metric for measuring dispersion should include within-category distribution information (i.e., variation). The results presented in this paper therefore focus on within-category (within-speaker) variation.

Previous studies on both vowels and consonants taken as a whole highlight the empirical issue of DT: the actual phonetic effects it predicts are not clearly observed in experimental data. A competing theory on what drives the shape of inventories, Quantal Theory, is now presented.

2.4 Quantal Theories of Speech

Quantal Theory (QT), as laid out in the work of Stevens and Keyser, is a principle by which distinctive features arise from interactions in the acoustic-articulatory domain and the acoustic-auditory domain (Stevens (1989); Stevens and Keyser (2010)). Take the acoustic-

articulatory domain: here, articulation is gradient, and there are regions where these small changes in articulation do not disturb the acoustic signal. These plateaus are referred to as *stable*. The boundaries between these stable regions arise from a small change in articulation that results in a drastic change in the acoustic signal. This relationship is schematized in Figure 7.

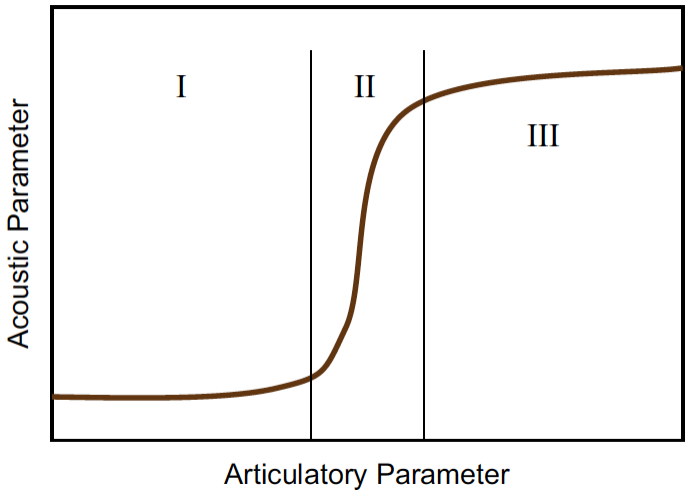

Figure 7: Schematic relationship between the articulatory parameter (x-axis) and the acoustic parameter (y-axis) as proposed by Stevens (1989). Figure from Bjorndahl (2018), p. 299.

The sigmoidal curve in this schematic figure represents that for a gradual change in the articulatory parameter, such as tongue height, there are regions where the acoustic result, such as turbulence, changes relatively little, but at some point there is a rapid change. These "regions of insensitivity" of the parameter on the y-axis (acoustic) are marked as regions I and III, where the sigmoidal curve shows a plateau. The rapid change in the acoustic parameter (y-axis) as a result of a change in values in the aritculatory parameter (x-axis) occurs in region II. The difference between regions I and III is proposed to be drastic, and in fact each stable region is hypothesized to correspond to a distinctive feature:

> "We suggest that this tendency for quantal relations between articulatory and acoustic parameters or between acoustic and auditory parameters is a principle factor shaping the inventory of articulatory states or gestures and their acoustic consequences that are used to signal distinctions in a language. The articulatory and acoustic attributes

> that occur within the plateau-like regions of the relations are, in effect, the correlates of the distinctive features" (Stevens (1989), p. 5).

The supposed effect of inventory size on production variance may be imagined in a QT-framework in the following way: the sigmoidal curve between two stable regions is not as steep, meaning that plateaus blend together or the difference between them is not so drastic so as to not fit the requirements needed to correlate with a distinctive feature.

We may find support for this scenario in the other domain proposed by QT: acoustic-auditory. Speakers with smaller inventories may not perceive the two different dimensions, such as /s/ and /ʃ/, because they are not contrastive. This may result in a less-steep S curve in this domain so that the two distinctive categories that are anatomically and acoustically possible, perceptually blend together for speakers of a certain language. The effect on this domain could then spread to the articulatory-acoustic domain in terms of greater variation within those parameters, given that perception and production are closely related.

However, a fundamental issue with QT is just how powerful it is:

> "We hypothesize that a quantal acoustic/articulatory relation underlies each distinctive feature, and consequently each feature can be said to be based on a defining articulatory range and a defining acoustic attribute. [...] These defining attributes are properties of the human speech production system and are expected to be universal in language. It is hypothesized that the human speech production system is structured in such a way that the sounds that it can generate and the articulatory attributes that produce these sounds define a set of quantal states" (Stevens and Keyser, 2010, pg. 15).

Given the weak predictive power of QT, Bjorndahl (2018) proposes an additional alternative, combining QT with Emergent Feature Theory (Mielke (2008)). The quantal regions as schematized in Figure 7 simply constitute the *basis* for possible distinctive features, but are not features in and of themselves. Therefore, languages employ quantal regions to distinguish sets of sounds differently and learners must induce these relationships. This may be the best way that QT could model the proposed relationship between inventory size and production variance. Spanish employs the quantal distinction between, say, stops and sibilants, and but does not employ the quantal distinction between finer places of articulation within sibilants, like [+/- anterior]. We may then expect to see freer oscillation, or more variation, between

these regions in Spanish than Catalan or English since it does not distinguish /s/ from /ʃ/.

2.4.1 Articulatory Precision

Relevant to sibilants, another claim of QT is that typologically common sounds require less articulatory precision (Stevens and Keyser (2010)). QT predicts /s/ and /ʃ/ to have relatively variable articulations because they are cross-linguistically common speech sounds (Maddieson (1984); see section 2.1). Contrastively, other theories predict their relative articulatory precision. Articulatory variation may be constrained by contrast, as theorized by DT; and Keating (1983) suggests that segments may have specified articulatory targets for jaw position, the one for sibilants being quite fixed. Keating (1983) reports that in Fijian, which has only one sibilant, the jaw height of /s/ across coarticulation contexts varied "hardly at all" (Condax (1980), Fijian data; Keating (1983), quote); and in her own experiment on English, she notes that "/s/ places strong demands on jaw position, and other segments accommodate it," meaning that jaw height of /s/ varies little across coarticulation contexts in this language also. Experimental work by Tabain (2001) also supports the relative articulatory precision of sibilants. She compared electropalatographic and acoustic data between the six coronal fricatives of Australian English /θ ð s z ʃ ʒ/. The typologically more common sibilants showed very low variability compared to the typologically rarer nonsibilants: "it is suggested that sibilant fricatives do not lend themselves to the articulatory imprecision which, according to [the Hyper- and Hypo- and Quantal theories of speech], characterizes perceptually salient, and typologically common, speech sounds" (Tabain (2001), p.57). In their comprehensive book, Ladefoged and Maddieson (1996) also contrast the relative articulatory precision of fricatives, and sibilants especially, with that of stops and nasals (p. 137). With these conflicting claims in mind, the current study also tests the articulatory variability of sibilant fricatives, as constrained by inventory size.

2.5 Feature Economy

The principle of FEATURE ECONOMY (FE), proposed by Clements (2003), is also worth mentioning here, as it too is a principle that is proposed to guide the organization of sound systems. FE argues that sound systems take advantage of the features they already exploit: they tend to "maximize" the number of phonemes that combine already utilized features. This is relevant for the present study when considering the phonological feature [+/- strident], whose opposition is generally grounded in terms of their articulatory turbulence and acoustic noisiness (Bjorndahl (2018)). The Spanish /s/ presents an interesting case in terms of FE. While the full inventory can be found in Figure 1 in section 2.5.2, Spanish has dental obstruents /t d/ and palatals /ʧ ʝ ɲ/ in terms of other [coronal] phonemes, and fricatives /f ʝ x/ in terms of other [+continuant] obstruents. It would seem, then, that there is a gap in [+ strident] phonemes at other coronal places of articulation that already exist such as [+ anterior, + distributed] (dental) and [- anterior, - distributed] (palatal). Spanish also has contrastive [+/- voice] feature, and yet there is no [+ voice, + strident] phoneme (Hualde (2005)). By FE, the system may be working to fill these gaps and for this reason too, we may see more variation in Spanish /s/ than Catalan and English /s/, whose inventories do not show the same gaps.

Sections 2.3 and 2.4 presented two competing theories about how inventories may be governed, Dispersion Theory and Quantal Theory, with a note on Feature Economy in section 2.5. This study aims to test if predictions made by these theories are borne out in the actual phonetic data; therefore, an overview of sibilant phonetics is provided in the next subsection.

2.6 Phonetics of Sibilants

2.6.1 An Overview

The general production of a fricative involves a narrow constriction, through which there is rapid airflow. This creates acoustic turbulence, and the random fluctuations in velocity in the

airflow act as a source of sound; the amount of randomness serves to acoustically distinguish sibilant fricatives from nonsibilant fricatives, which often display formant structure (Jongman et al. (2000); Kim et al. (2015)). Sibilants also have a much higher amount of turbulence when compared to nonsibilant fricatives, which can be attributed to their articulation: the presence of the teeth as an obstacle to airflow in sibilants versus the absence of this obstacle in nonsibilants (Shadle (1985); Kim et al. (2015)).

2.6.2 Acoustics

Measurements suited for reliably interpreting place of articulation of sibilants have been shown to come from the sound spectrum: center of gravity, kurtosis, and spectral tilt (Shadle (1985, 1990, 1991); Jongman et al. (2000); Koenig et al. (2013), among many others).

Center of gravity (CoG) is where the center of energy is concentrated in the spectrum (Grey and Gordon (1978); Van Son and Pols (1996)). Conceptually, it can be thought of as weights balanced on a fulcrum: if the signal is chunked into windows, and each window is weighted based on the amount of energy it holds, the center of gravity is where the fulcrum would have to be placed in order for the weights of the windows to be balanced on either side. A schematic of CoG using this metaphor is shown in Figure 8. The windows are represented by the blue bars, which correspond to the intensity in the spectrum of an English /s/ in black. The red dashed line represents the placement of the fulcrum: here the weights on either side would be balanced.

More scientifically, CoG is calculated by using a Fourrier transform to take the weighted mean of the frequencies in the signal, with the weights being the magnitudes of the frequencies (Peeters (2004)). The farther back in the mouth the constriction is made, the lower the center of gravity: anterior sibilant /s/ has a higher center of gravity than posterior /ʃ/. Lip shape can also affect center of gravity: rounding has been shown to lower CoG (Ladefoged and Maddieson (1996)).

Given that CoG has been shown to be a correlate of place of articulation in sibilants in

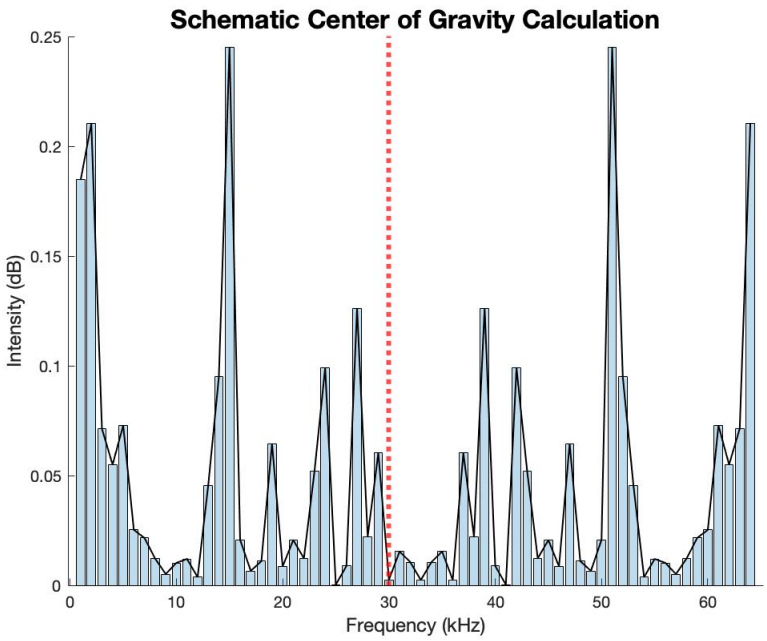

Figure 8: A schematic of how center of gravity is calculated. The spectrum of an English /s/ is plotted with a black line. A bar plot showing its intensity in decibels (y-axis) at various frequency points (kHz; x-axis) is in blue. The red dashed line marks the center of gravity.

several unrelated languages (Gordon et al. (2002); Recasens and Espinosa (2007); Jongman et al. (2000)), this is the spectral measure reported in the results in section 4.

2.6.3 Articulation

In order to better understand the relationship between the articulatory precision of sibilants and pressures on production variance, the present experiment also measures lip rounding and retraction, and jaw height. Jaw height may provide interesting results on the basis that it has been claimed to have a high level of articulatory precision for sibilants. As already laid out in section 2.4.1, Keating (1983) suggests that sibilants may have precisely specified articulatory target for jaw position. This articulatory precision requirement is supported by work on Fijian (Condax (1980)), American English (Keating (1983)), and Australian English (Tabain (2001)).

However, in their X-ray Microbeam (XRMB; Wisconsin database) study of English sibilants, Iskarous et al. (2011) found that jaw height varied more in low vowel contexts, and

showed a arc-shaped trajectory over the course of the sibilant, while the trajectory of the jaw for /ʃ/ remained steady. This finding is important because it shows that sibilants may not be so articulatorily precise as was previously claimed by Keating (1983). While it may be the case that jaw height varies little for sibilants when compared to other sounds in a language as Keating (1983) has reported, variance in production in terms achieving a jaw-height articulatory target may not be so constrained as to escape the hypothesized effect of inventory size on variance. Results on jaw movement are also reported in section 4.

Lip retraction and lip rounding were also recorded in the present experiment, due to the increased prominence of these gestures in hyper-articulated speech (Green et al. (2010)). It is a secondary hypothesis of this work, then, that lip retraction of /s/, and lip rounding of /ʃ/, for English and Catalan, will be more pronounced and/or occur more often at slower speaking rates.

2.7 Hypotheses and Predictions

This study reports on the variance in production of sibilant fricatives in three languages: Spanish, Catalan, and English. As hypothesized by DT, variance in Catalan and English may be constrained by contrast (as they have larger sibilant inventories than Spanish), and/or by articulatory precision requirements as hypothesized by Keating (1983). Since variance is the focus, it is induced in the three languages in two ways: (1) by having three repetitions of each token, and (2) by varying speaking rate.

Two competing theories of sibilant production are of interest here. DT hypothesizes that there is relationship between inventory size and phonetic production, such that a larger inventory is predicted to have less within-category variation in production of its categories than a smaller inventory (Lindblom (1986)). A schematic outcome of the results predicted by DT is shown in Figure 9. English and Catalan are in blue, with /s/ represented by the solid line and /ʃ/ by the dashed. These two distributions are narrower and the mu of the blue /s/ curve is farther away compared to the distribution and mu of the Spanish /s/ in

red. These results are predicted to be borne out in both acoustic (e.g., center of gravity) and articulatory (e.g., jaw displacement, lip shape) domains, according to DT.

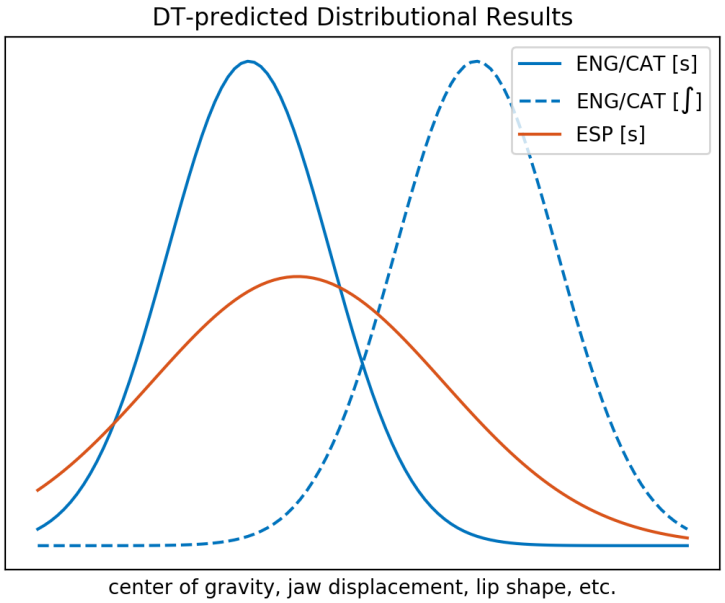

Figure 9: DT-predicted distributional results

However, this hypothesized effect has not found much in the way of empirical support in the literature (e.g., Manuel (1990); Bradlow (1995); Evers et al. (1998)). Alternatively, it has been proposed that sibilants require relatively high articulatory precision (Keating (1983)). This claim finds empirical support in some studies (e.g., Keating (1983); Tabain (2001)), but not others (e.g., Iskarous et al. (2011)).

This work therefore aims to investigate the effects that (1) inventory size, and (2) articulatory precision, may have on the phonetic realization of sibilants in terms of within-category variation. The primary hypotheses are enumerated below:

1. Sibilant variation is constrained due to inventory size effects on the phonetic realization of phonemic categories (DT-based Hypothesis).

2. Sibilant variation is constrained due to their requirement of a relatively high level of articulatory precision (Articulatory Precision-based Hypothesis).

Predictions following from these hypotheses are two-fold, acoustic and articulatory, but acoustic predictions are the focus here:

1. Intra-speaker within-category variation of center of gravity will be more constrained for a sibilant in a larger inventory (DT-based prediction).

2. Intra-speaker within-category variation of center of gravity will not significantly differ as an effect of inventory size, due to strict articulatory precision requirements (Articulatory Precision-based prediction).

We may, however, find that articulation and acoustic realizations do not work the same way. If center of gravity is shown to be a reliable measure of place of articulation, then they should closely correlate, but still we may see relatively more within-category acoustic variation than articulatory variation given that articulation may not exactly map to acoustic output (as is predicted by QT).

These predictions are grounded in the previous findings presented in the antecedent sections. They are tested via the experiment outlined in the next section.

3 Methodology

In order to investigate the interaction between inventory size and variation and the effects of articulatory precision, data need to be both articulatory and acoustic, the languages of study need to have different inventory structures, and both inter- and intra-speaker variation needs to be compared. The following subsections outline a production experiment with participants of three languages with different sibilant inventories: Spanish, Catalan, and English. Articulatory data on jaw movement and lip shape were collected, as well as acoustic data. In addition to different phoneme inventory structures, speaking rate was manipulated as an independent variable in order to put added pressure on the system to vary.

3.1 Experimental Procedure

Data to test the above hypotheses was collected in a production experiment, conducted in a sound-attenuated booth located in the Cornell Phonetics Laboratory. In addition to collecting audio via a headset-mounted AKG C520 condenser microphone, visual data was collected via a Logitech Pro 1080p Webcam, attached to the top of a computer monitor the participants used for the experiment. The webcam was used to track the movement of 3mm round black stickers placed on the face: three stickers on the forehead in a triangle, one on the nasion, three each on the upper and lower lips just inside the vermilion, and two on the chin. All stickers were centered on the midsagittal line, and a triangle cutout was used to ensure that the forehead stickers were stuck in the same place relative to each other on each participant. A schematic of sticker placement in shown in Figure 10.

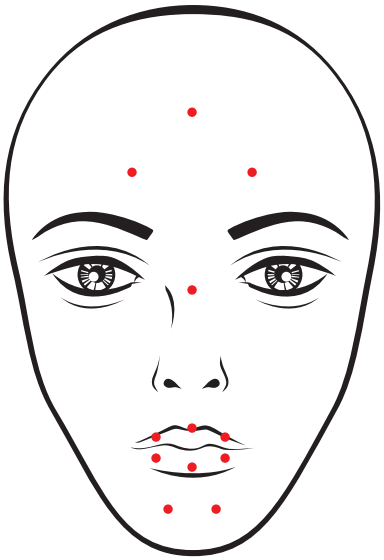

Figure 10: Schematic of sticker placement for articulatory data collection. Stickers are shown in red for visual salience, but were black in the experiment.

This sticker methodology is similar to that used by Parker and Mielke (submitted) in production experiments on high vowels in Bora (Bora-Witoto, South America). Tape was

placed on the carpet to ensure that the chair was the same distance from the computer monitor for each participant.

Before the full experiment, participants completed a training session on a set of three stimuli not included in the experiment, presented in eight trails (so two stimuli were repeated three times, and the third only twice). The full experiment required a range of speaking rates, but all trials during the training were the same speaking rate, the fastest one, so that participants would be prepared for the speed of the cue. Many of the participants completed the training session twice before becoming comfortable enough with the cueing to move on.

The visual cue was used to get participants to vary their speaking rate. In each trial, the cue was displayed before the participant was prompted to respond. The visual cue consisted of a box moving across the screen at ten different speeds, and participants were instructed before they began the experiment to approximate their speech to be as fast or as slow as the box moves. The training session was designed for participants to get used to the visual cue, and to learn the carrier phrase. The carrier phrase was given during the pre-trial instructions only, and only the target word appeared simultaneously with the visual cue for the participant to insert into the carrier phrase during each trail.

3.1.1 Stimuli

Stimuli are controlled for preceding vowel, word position, word length, and frequency. All stimuli have the target sound in initial position, as the onset of a stressed syllable with an /a/ nucleus, in a disyllabic word[5]. Only the voiceless series of sibilant fricatives and affricates from each language was recorded.

The stimuli in Table 3 were presented to the participants on the computer monitor in three blocks of 80 trials, with optional breaks between each block. Three repetitions of each word was recorded per speaking rate, at 10 different speaking rates. Stimuli were randomized within each block. Six fillers: *taxi* 'taxi' /ˈtak.si/, *taza* 'cup' /ˈta.sa/, *pase* 'pass' /ˈpa.se/,

[5]Except for Spanish *sabor* 'flavor', which has final stress. Spanish does not have such extreme nonstressed syllable reduction as seen in English, so this effect is hoped to be minor.

papel ‘paper’ /ˈpa.pel/, *tablas* ‘tie/draw’ /ˈta.blas/, *pacto* ‘pact’ /ˈpak.to/ were included for Spanish. Four fillers: *tasques* ‘duties’ /ˈtas.kəs/, *tardor* ‘autumn’ /ˈtar.dor/, *passat* ‘past’ /pa.ˈsat/, *parets* ‘walls’ /pa.ˈrets/, were included for Catalan. There were five fillers for English: *toddlers* /ˈtɔd.lɛɹz/, *topics* /ˈtɔ.pɪks/, *popcorn* /ˈpɔp.kɔɹn/, *pockets* /ˈpɔ.kɪts/, and *toxins* /ˈtɔk.sɪnz/.

Each language had eight stimuli (targets plus fillers). 8 stimuli x 3 repetitions x 10 speaking rates = 240 tokens, per speaker. There were 19 total participants, yielding 4560 total tokens collected. Data collection took between 45 and 60 minutes per speaker, including training and paperwork.

Language	Vowel	Sibilant	Word	Transcription
Spanish	/a/	/s/	*sabor* ‘flavor’	/**s**a.ˈbor/
	/a/	/tʃ/	*Acha* name	/ˈa.**ʧ**a/
Catalan	/a/	/s/	*santa* ‘saint’	/ˈ**s**an.ta/
	/a/	/ʃ/	*xarxa* ‘net’	/ˈ**ʃ**ar.ʃa/
	/a/	/ts/	*atzar* ‘chance’	/ˈa.**t͡s**ar/
	/a/	/tʃ/	*atxa* ‘torch’	/ˈa.**ʧ**a/
English	/a/	/s/	*sockets*	/ˈ**s**ɔ.kɪts/
	/a/	/ʃ/	*shamans*	/ˈ**ʃ**ɔ.mɪnz/
	/a/	/tʃ/	*matcha*	/ˈmɔ.**ʧ**ə/

Table 3: Stimuli

The carrier phrase for Spanish was: *Canta __ otra vez* /kan.ta __ o.tra ves/, and Catalan: *Canta __ un altre cop* /kan.ta __ un al.tre kop/. Both mean ‘Sing __ again,’ where the form of ‘sing’ is the informal imperative. The English carrier phrase was: *I will draw __ again* /aɪ wl̩ dɹɔ __ ə.gɛn/.

Stimuli were also controlled for frequency. Corpora were used to make sure that the chosen stimuli were not among the most frequent words or the least frequent. The cutoff depended on the corpus. Those words not found in the Catalan corpus were verified with a native speaker. The Spanish words were verified using The Corpus del Español: NOW, created by Dr. Mark Davies in the Linguistics department at Brigham Young University. The corpus contains over 5.5 billion words from 21 different Spanish-speaking countries,

collected from online texts and updated monthly. The corpus is openly available online at: corpusdelespanol.org (Davis (2001)). The Catalan words were verified using the Catalan WikiCorpus (v. 10), created by many collaborators in the Computer Science department at the Universitat Politècnica de Catalunya (Reese et al. (2010)). The corpus contains over 750 million words from Wikipedia articles. A Python script was used to read, clean, and search the corpus. English words were verified using the COCA corpus (Davies (2008)), which at the point of access contained over 560 million words. The corpus maintains a balanced word count between spoken, fiction, popular magazine, newspaper, and academic journal sources, from 1990 to 2017.

3.1.2 Participants

For Spanish and English there were eight participants each, four male and four female. Unfortunately it was difficult to find local Catalan speakers: three participated in the experiment, two female and one male.

Most Spanish speakers spoke Mexican Spanish (five out of eight); others among them spoke Spanish from Venezuela, Bolivia, and Columbia. Almost all Spanish participants were current undergraduates at Cornell and were between the ages of 18 and 30. All three Catalan speakers were from Barcelona, Spain. Two were visiting graduate students, one male and one female both aged between 21 and 30. The third Catalan speaker was a lecturer at Cornell and over 61 years old. English speakers were also mostly current undergraduates at Cornell, aged between 18 and 30. Most were from the New England/Mid Atlantic region (five out of eight), with few exceptions (Oregon, Utah, Michigan). Speakers for each language all had that language as their L1, with no current speaking or hearing issues[6]. A summary of participant data is shown in Table 4. Where a degree is listed, that indicates it was completed, otherwise the level of education in progress is indicated. NBC stands for 'Natural

[6]Two English speakers (one male, one female) reported undergoing speech therapy when they were young. They had no apparent speaking issues at the time of recording, and did not receive therapy for the target sounds.

Speaker	Language (Dialect)	Gender	Age	Origin	Education	Time in U.S.
AB_ESP	Spanish (Mexican)	Male	21-30	Mexico	Graduate	23 years
AV_ESP	Spanish (Mexican)	Female	21-30	California	Undergraduate	NBC
JR_ESP	Spanish (Venezuelan)	Male	< 21	Florida	Undergraduate	18 years
JT_ESP	Spanish (Colombian)	Male	21-30	Colombia	Undergraduate	NBC
LA_ESP	Spanish (Mexican)	Female	21-30	Texas	B.A.	NBC
LT_ESP	Spanish (Mexican)	Male	< 21	Mexico	Undergraduate	14 years
LV_ESP	Spanish (Bolivian)	Female	< 21	Bolivia	Undergraduate	10 years
MM_ESP	Spanish (Mexican)	Female	< 21	Illinois	Undergraduate	NBC
CA_CAT	Catalan (Eastern)	Male	21-30	Barcelona	Graduate	< 1 year
LC_CAT	Catalan (Eastern)	Female	61+	Barcelona	Ph.D	25 years
MM_CAT	Catalan (Eastern)	Female	21-30	Barcelona	Graduate	1 year
CC_ENG	English	Female	< 21	New York	Undergraduate	NBC
AM_ENG	English	Male	31-50	Utah	B.A.	NBC
FL_ENG	English	Female	< 21	New Jersey	Undergraduate	NBC
IM_ENG	English	Male	< 21	Oregon	Undergraduate	NBC
KL_ENG	English	Female	21-30	New Jersey	Undergraduate	NBC
LH_ENG	English	Female	< 21	New Jersey	Undergraduate	NBC
VB_ENG	English	Male	21-30	Michigan	Undergraduate	NBC

Table 4: Summary of Participant Data

born citizen,' meaning they were born in the United States.

3.2 Data Processing

Acoustic processing and analyses were conducted in MATLAB. The signal was subjected to various filters. A fourth-order high-pass Butterworth filter with a cutoff of 70Hz was applied to the normalized signal as a first pass in order to remove any electronic noise. A bandpass filter was then applied to the signal with cutoffs at 7000Hz and 10000Hz for female speakers, and 6000Hz and 9000Hz for male speakers; this was done to create a sibilant envelope, which was normalized. A vocalic envelope was also created using a finite impulse response bandpass filter with cutoffs at 300 and 1000Hz; this envelope was also normalized. The sibilant and vocalic envelopes were then separated out from each other by subtracting the vocalic envelope from the sibilant envelope; this was also normalized. All of the above filtering was done in order to facilitate the detection of the response and the sibilant.

The envelopes generated by this filtering were then used to automatically identify the sibilant peaks of interest and the onset and offset of the response, since speaking rate was an independent variable. This was generally done by identifying peaks over set thresholds. To delineate the response itself, boundaries were simply drawn the first and last time a peak in the vocalic envelope hit the set threshold. Thresholds were also used to identify and delineate the target sibilant peaks in the sibilant envelope. A set of the top number of peaks, depending on carrier phrase (i.e., the bursts in 'draw' in English and 'canta' in some Spanish speakers; and the /s/ in 'otra vez' in Spanish) and target word (i.e., two sibilants in 'sockets' and 'shamans' in English and 'xarxa' in Catalan) were generated and then labelled in an order depending on these factors. A sample output of these boundaries and envelopes plotted on the audio signal is in Figure 11 for a male English speaker.

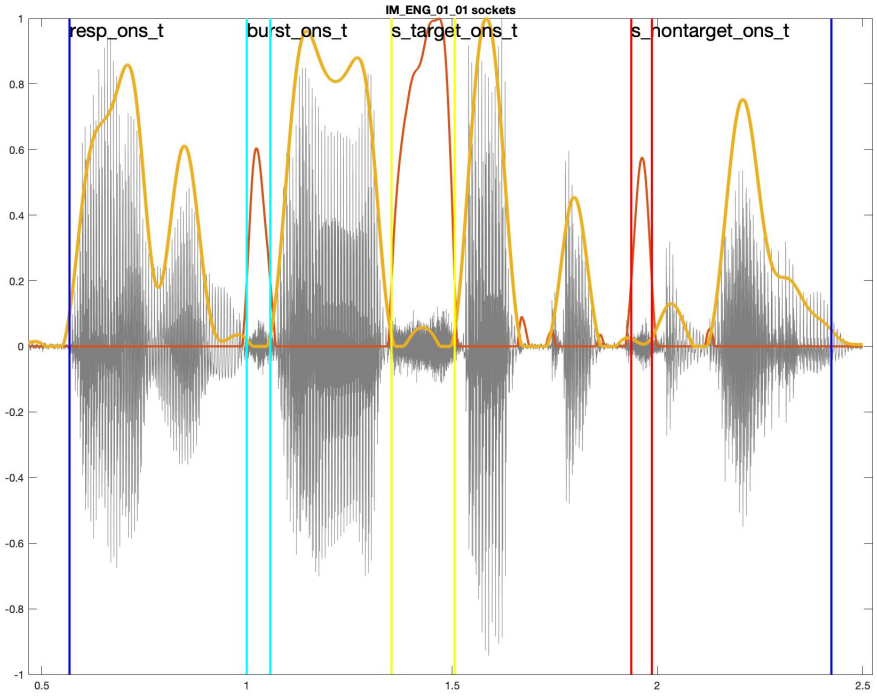

Figure 11: A single-trial example plot with sibilant envelope (red), vocalic envelope (orange), response boundaries (blue), burst boundaries (cyan), target sibilant boundaries (yellow), and non-target sibilant boundaries (red) for the English phrase 'I will draw sockets again.'

Boundaries identified using the thresholds described above were then fed into a script that generated textgrids in Praat (Boersma and Weenink (2009)) for corresponding audio files (one trial equals one audio file). Plots like the example shown in Figure 11 were used to visually inspect the accuracy of the script. Problematic boundary positions were then hand-adjusted in Praat.

Boundary positions were used to collect duration data, and the audio signal they delineated was used to conduct spectral measurements. Measures were taken over a 40ms window centered at the middle of the fricative. This window length was chosen based on Jongman (1989)'s findings that the first 40ms of fricative noise was sufficient for listeners to discriminate between [s] and [ʃ] (71% accuracy for [s] identification and 89% for [ʃ] identification). The first *70*ms raised accuracy to above 80% for [s] and to almost 100% for [ʃ]. The entire duration of the fricative was necessary for 100% accuracy. Therefore, a 40ms window in the center seems sufficient for perceptual differentiation, and this windowing has been used in at least one other study on [s] versus [ʃ] (Evers et al. (1998)).

3.3 Statistical Measures

This subsection describes the outlier-removal processes used, and the statistical tests the data were subjected to. Outliers were determined using z-score, and those with an absolute z-score value over 2.5 were eliminated. 14 trials were also hand-eliminated on the basis of speech errors, electronic noise, or too much cutoff[7] of the target word; these trials constituted less than 1% of the data. Those trials with cutoffs of the carrier phrase were included, as long as the target word was produced in full. Outliers totaled 685 trials (out of 4560 total), making up 15% of the data.

The results section makes reference to nCoG values, which are the residuals of the raw CoG data (outliers removed), when the effect of subject is factored out. These values are useful because speakers are expected to have different CoG values based on unique vocal

[7]Target words where any part of the sibilant was cutoff, as well as those with a full syllable or more cutoff, were not included.

tract lengths, so while the nCoG values factor out this effect, they importantly still preserve intra-speaker variance. These residuals are plotted in Figure (12).

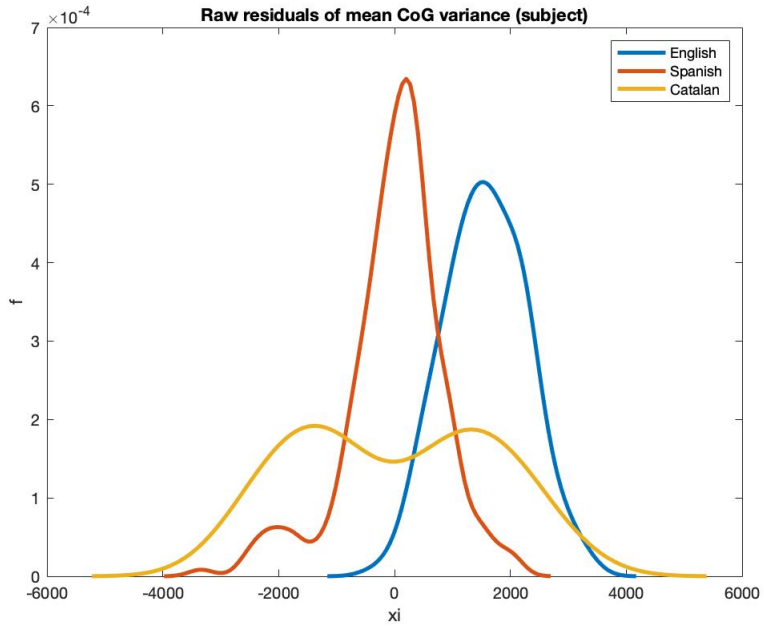

Figure 12: Raw residuals of mean center of gravity with the effect of subject filtered out by language. Spanish in red; Catalan in yellow; English in blue.

In order to test the predictions of the hypotheses, a two-sample F-test was conducted on the nCoG data grouped by language to see if the variances in CoG were significantly different between language pairs. Following DT, the variances between Spanish and Catalan, and between Spanish and English should be significantly different, with Spanish having significantly higher variance (here, this is quantified in terms of greater standard deviation). Following the articulatory precision hypothesis, the variances are predicted not to be significantly different between the languages: all should be relatively low.

3.4 Dispersion Metrics

Given the information outlined in section 2.3.3, if there is hope of finding an effect of inventory size on phonetic realization of its categories, within-category variation must be considered; therefore, the results presented in section 4 crucially compare within-category variation of

the sibilant fricatives, given the importance of distributional information outlined by Hauser (2017), supported by claims from Liljencrants and Lindblom (1972), and various speech perception studies.

4 Results

4.1 Speaking Rate

This subsection presents results showing the success of the visual cue used in the experiment to get participants to vary their speaking rate. In order to induce variance in the speech of the participants, speaking rate was varied using a visual cue of a box moving across the screen at 10 different rates. In addition to variation in duration, the visual cue may also induce clearer speech (or hyperarticulation) at slower rates and hypoarticulation at faster rates (Scarborough and Zellou (2013)).

Linear regression models were run on the participants of each language to see if (1) cue duration and response duration were positively correlated, then (2) response duration and sibilant duration were positively correlated. Plots of the models in (1) are shown in Figures (13-15), and (2) in Figures (16-18). (Adjusted) R-squared values for the models are provided in Table 5.

Table 5 shows that cue duration and response duration were significantly positively correlated for all three languages, meaning that participants produced a shorter response when they saw a shorter visual cue, and vice versa. This response duration was also significantly positively correlated with sibilant duration in Catalan and English; a shorter response correlated with a shorter sibilant, and a longer response with a longer sibilant. However, all R-squared values are relatively close to zero, indicating that not much of the variance in response duration was accounted for by cue duration or in sibilant duration by response duration.

Figure (13) shows the significantly positive correlation between cue duration and response

Language	Parameters	R-squared	p-value
Spanish	cue duration by response duration	0.33	$p < 0.001$
Spanish	response duration by sibilant duration	0.27	$p < 0.001$
Catalan	cue duration by response duration	0.19	$p < 0.001$
Catalan	response duration by sibilant duration	0.05	$p < 0.01$
English	cue duration by response duration	0.35	$p < 0.001$
English	response duration by sibilant duration	0.19	$p < 0.001$

Table 5: Summary of linear regression model results for speaking rate

duration in Spanish, Figures (14) and (15) show the same results in Catalan and English.

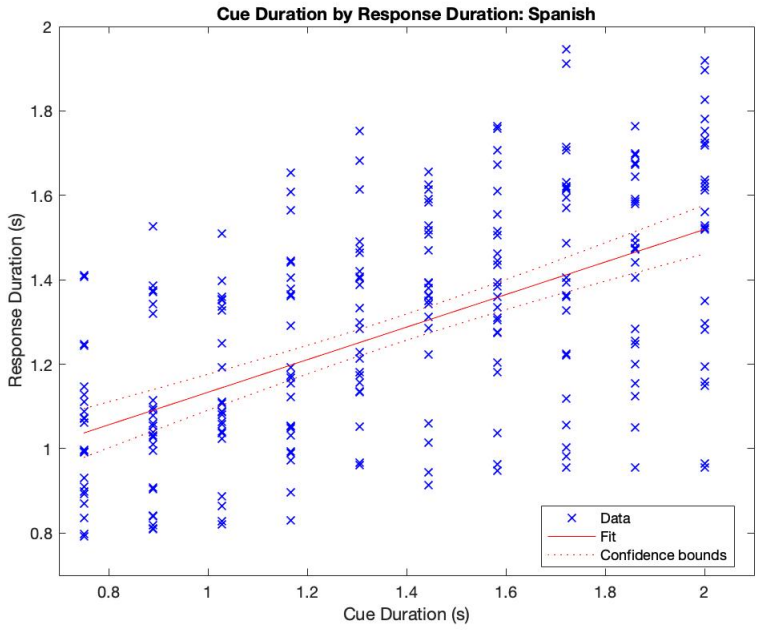

Figure 13: Scatter plot of response (x-axis) and cue (y-axis) durations with a best fit line (solid red) determined by linear regression model: Spanish

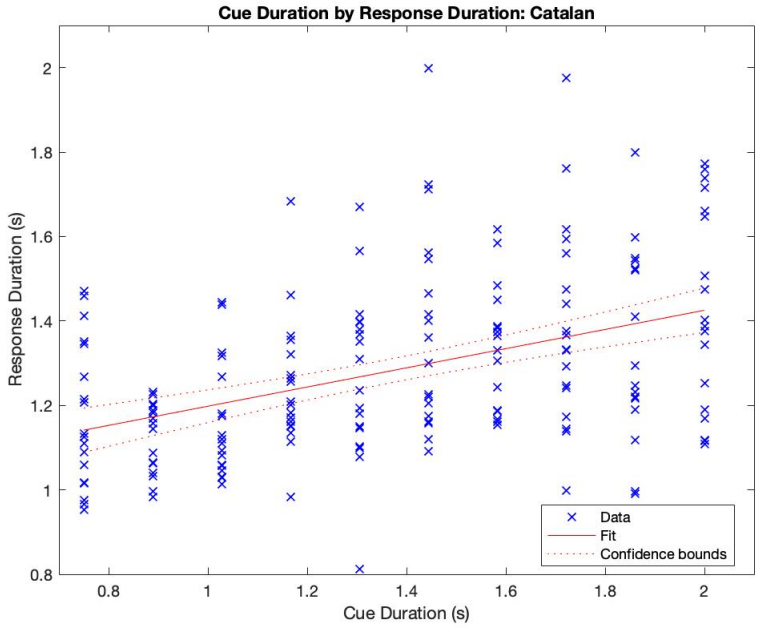

Figure 14: Scatter plot of response (x-axis) and cue (y-axis) durations with a best fit line (solid red) determined by linear regression model: Catalan

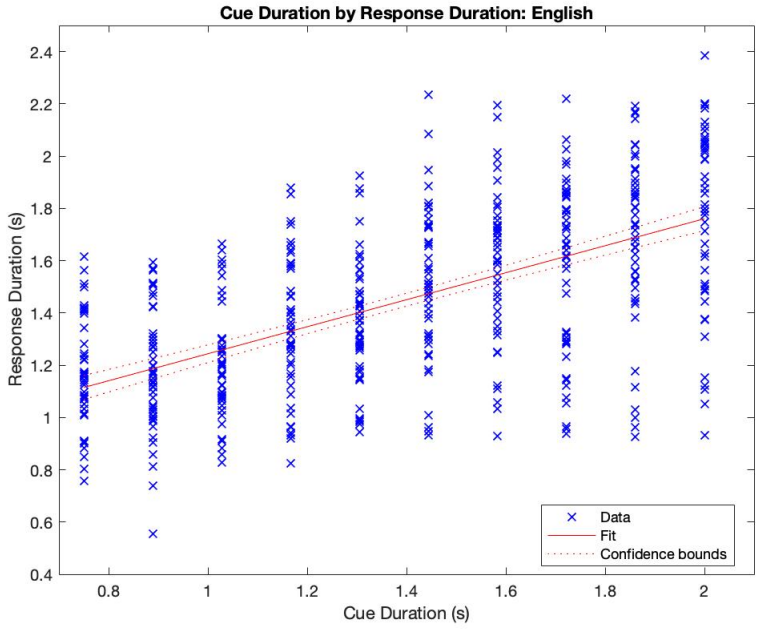

Figure 15: Scatter plot of response (x-axis) and cue (y-axis) durations with a best fit line (solid red) determined by linear regression model: English

Figure 16 shows a significant positive correlation between response duration and sibilant duration in Spanish. Figures (17) and (18) show the same results for Catalan and English.

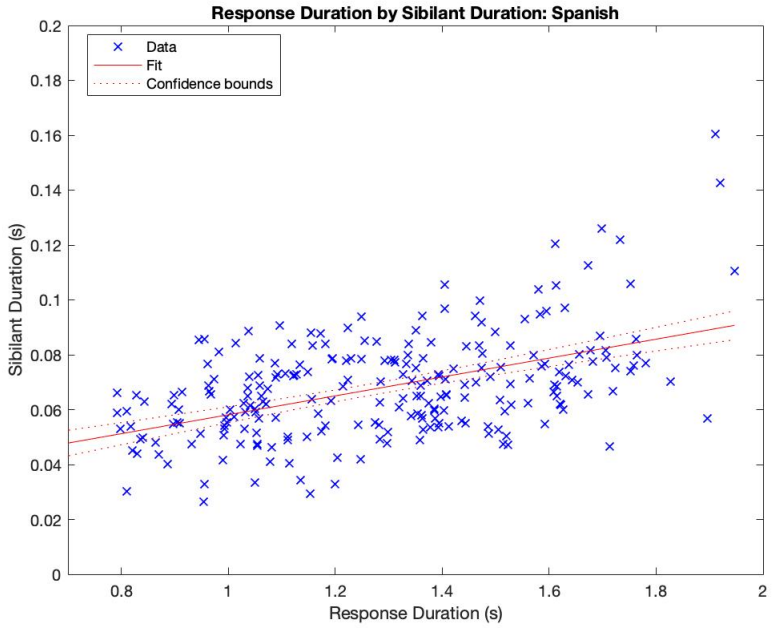

Figure 16: Scatter plot of response (x-axis) and sibilant (y-axis) durations with a best fit line (solid red) determined by linear regression model: Spanish

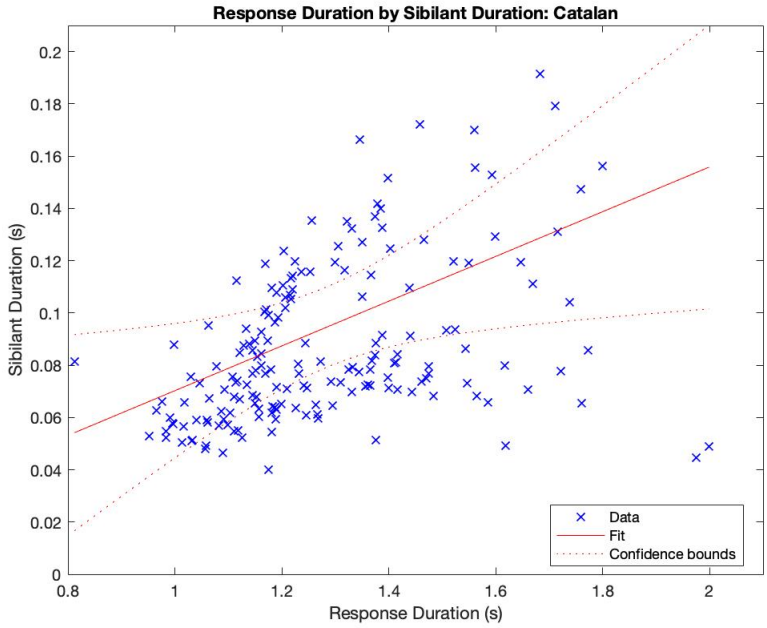

Figure 17: Scatter plot of response (x-axis) and sibilant (y-axis) durations with a best fit line (solid red) determined by linear regression model: Catalan

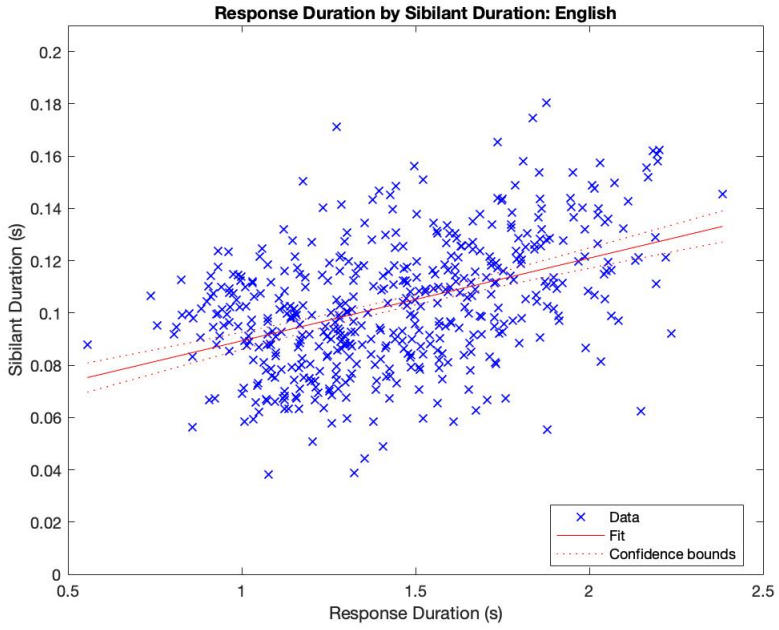

Figure 18: Scatter plot of response (x-axis) and sibilant (y-axis) durations with a best fit line (solid red) determined by linear regression model: English

The results in Table 5, visually represented by Figures (13 - 15), show that the visual cue was successful in getting participants to respond at various speaking rates, positively correlated with the cue (i.e., faster cues equaled shorter response durations) in all three languages. Response duration also significantly positively correlated with sibilant duration (i.e., shorter response duration equaled shorter sibilant duration) in all three languages (Figures (16 - 18)).

In order to see if speaking rate contributed to sibilant place of articulation variation, a linear regression model was run on sibilant duration and nCoG (collapsed for language and sibilant). Figure (19) plots this linear regression best fit line over a scatterplot of the duration and nCoG data. There was no significant correlation between sibilant duration and nCoG found (R-squared value < -0.001; p = 0.63).

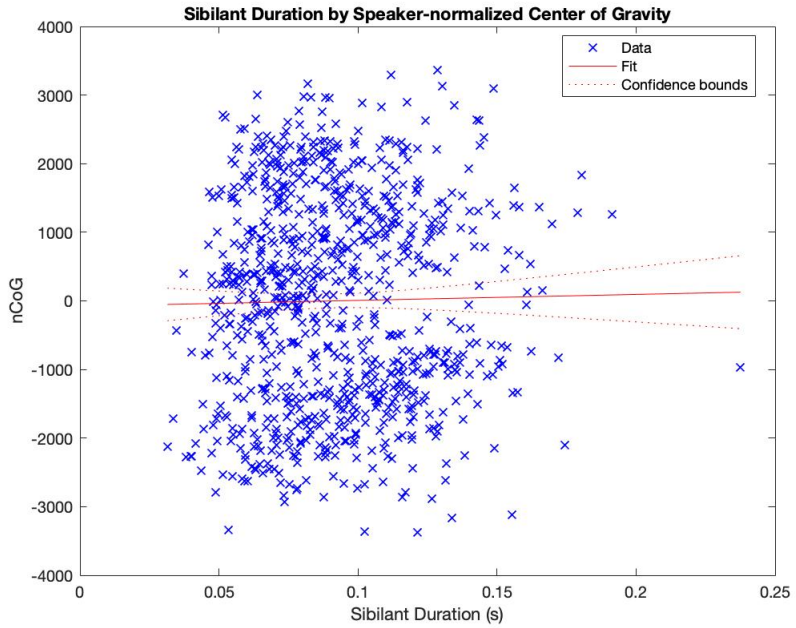

Figure 19: Scatter plot of sibilant duration (x-axis) and nCoG (y-axis) with a best fit line (solid red) determined by linear regression model.

While the visual cue was successful in achieving correlated variance in duration of speakers' responses and sibilants, this did not reliably result in variance in center of gravity; variance occurred anyway.

4.2 Acoustics Results

This section presents results of center of gravity measurements. Both the mean and maximum CoGs were calculated, and these values are reported first by speaker, to show intra-speaker variation, then by language. Only the mean CoG values are presented graphically; results from maximum CoGs reflect the same patterns. Boxplots of CoG reveal speaker-specific, and language-specific productions of the sibilants (no two speaker or languages are exactly alike). CoG of Spanish /s/ at roughly 7000Hz is significantly lower than English at roughly 7300Hz; Catalan /s/ has the lowest CoG of the three at around 6700Hz, averaged across speakers. This likely correlates with anteriority of constriction location, moving front to back: Spanish /s/, English /s/, Catalan /s/; but, CoG can also be influenced by lip shape, so the addition of articulatory data is needed to confirm this. CoG of /ʃ/ was significantly lower than /s/ in both Catalan (4100Hz) and English (3500Hz).

4.2.1 Center of Gravity

Figure (20) schematically shows the distribution of the mean CoG of /s/ via boxplots for each individual speaker in all three languages. Figure (21) shows the mean CoG values of /ʃ/ by speaker in Catalan and English. Subjects are grouped by language (Catalan, English, Spanish), then by gender (Female, Male). Red lines in the center represent the median of the data, and the top and bottom blue lines of each box represent the 25 and 75 percentiles, respectively. If notches do not overlap, the difference between the medians is statistically significant, with 95% confidence (i.e., the difference would be significant in a two-sample t-test, but without correcting for multiple comparisons).

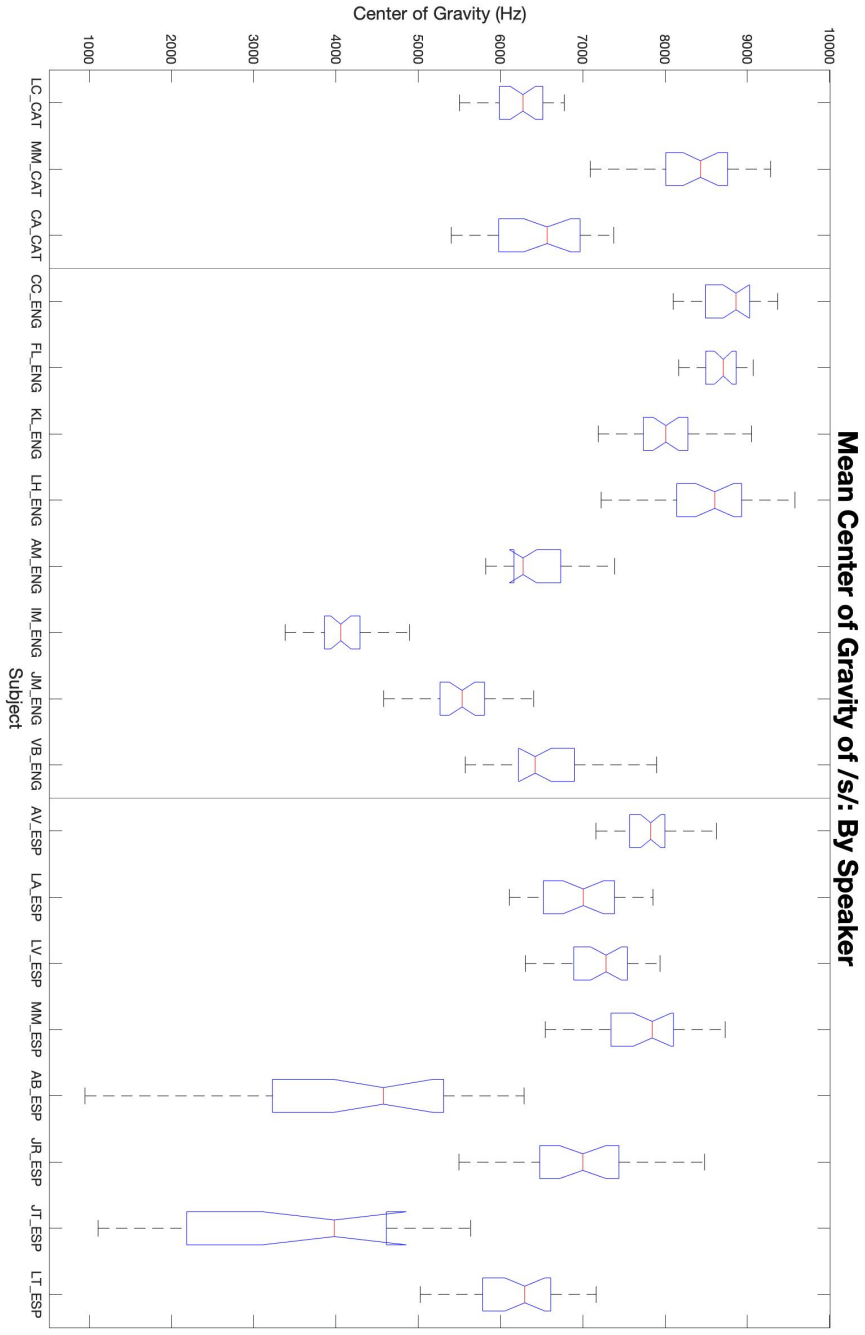

Figure 20: Boxplot of mean center of gravity of /s/ for each speaker. Speakers are grouped by language: Catalan, English, Spanish; then by gender: Female, Male. Recall there are two female Catalan speakers and one male; English and Spanish are gender-balanced.

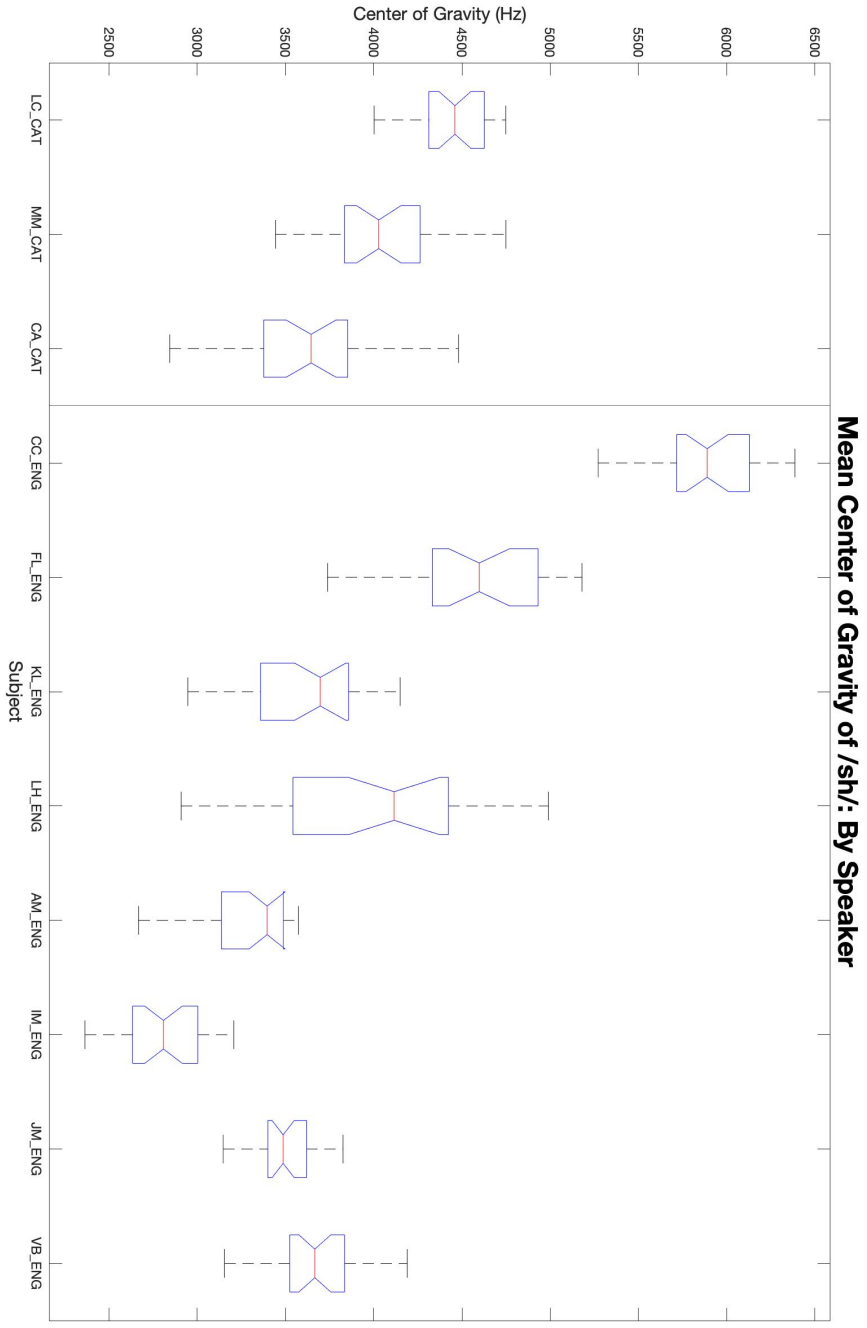

Figure 21: Boxplot of mean center of gravity of /ʃ/ for each speaker. Speakers are grouped by language: Catalan, English; then by gender: Female, Male. Recall there are two female Catalan speakers and one male; English is gender-balanced.

Figure (22) shows boxplots of the mean center of gravity for /s/ then /ʃ/ across languages.

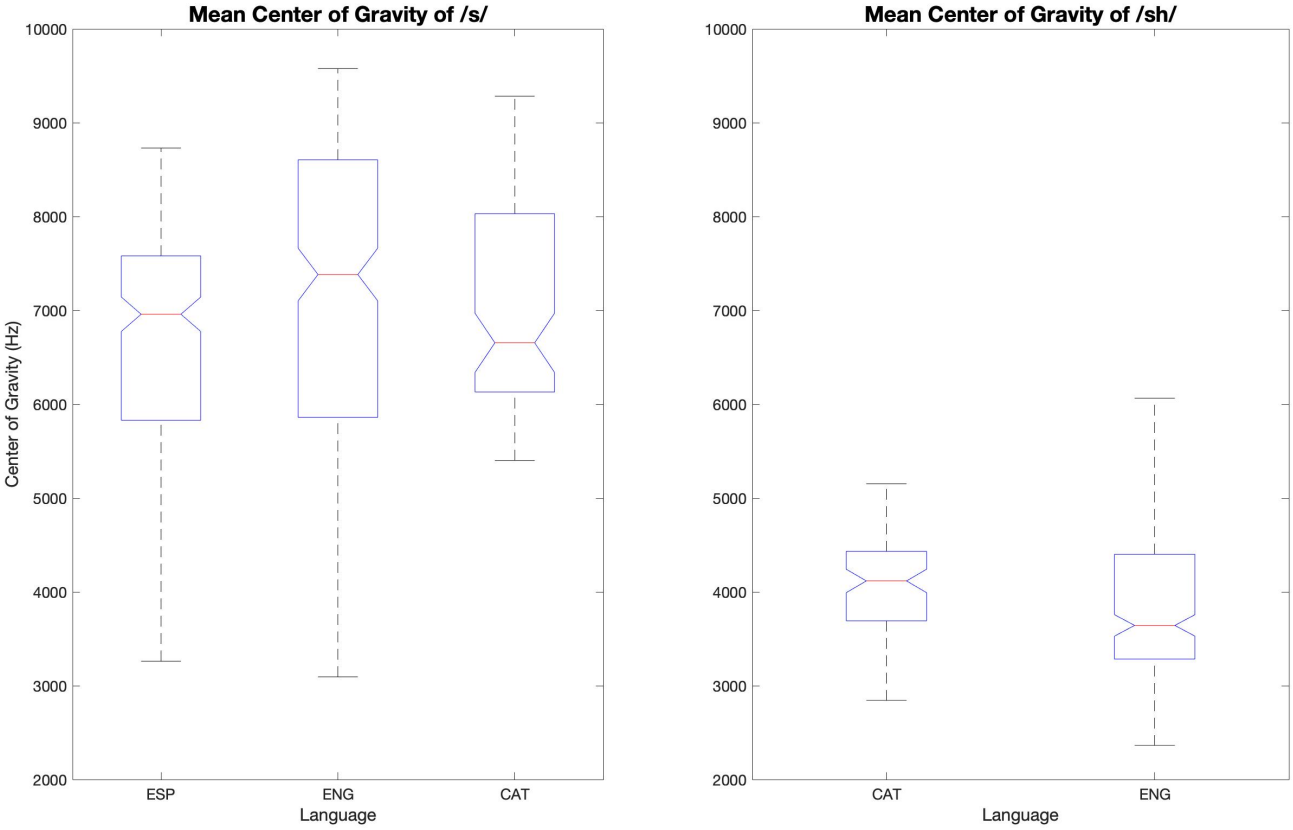

Figure 22: Boxplots of the mean center of gravity of /s/ and /ʃ/ for each language.

Figure (22) shows that the mean center of gravity for both /s/ and /ʃ/ is significantly different between the three languages. Comparing the results between /s/ and /ʃ/ in English and Catalan, within-language, the mean CoG is significantly different between the sibilants at the two different places of articulation. This is key: it is evidence that CoG is a good measure for place of articulation in sibilant fricatives. This is shown for the languages of study here, which corroborates other work on English and Catalan, and in other languages (Jongman et al. (2000); Recasens and Espinosa (2007); Gordon et al. (2002)).

The maximum center of gravity results support the same conclusions drawn from the mean center of gravity results: maximum CoGs of /s/ and /ʃ/ are significantly different

between languages, and between /s/ and /ʃ/ within Catalan and English.

4.3 Within-Category Variation

This subsection presents results on within-category variation: by-speaker standard deviations of nCoG of /s/ grouped by language, and results of the F-test for difference in variances in nCoG. Standard deviations represent the degree of variance within-speaker, within-/s/, which is expected to be higher for Spanish than English or Catalan given the DT framework, but articulatory precision claims predict all three languages to show relatively low variance. The F-test tests if these differences in variation are significantly different. nCoG values are used in these calculations, which factor out inter-speaker differences, but preserve within-speaker variances. It is found that nCoG of an /s/ that forms part of a larger inventory has a statistically lower variance (smaller standard deviation) than the /s/ in a smaller inventory language.

Figure (23) shows the standard deviation values of the nCoG of /s/ for each speaker. Spanish speakers are in red, Catalan in yellow, and English in blue. The dashed black line plots the mean standard deviation across the Spanish speakers. It serves as a comparison for the Catalan and English data. All Catalan and English speakers' standard deviations are below the average Spanish speaker's. This indicates that Spanish speakers, on average, produce a greater variance in center of gravity for /s/. Figure (24) shows the mean standard deviation values of the nCoG of /s/ for each language.

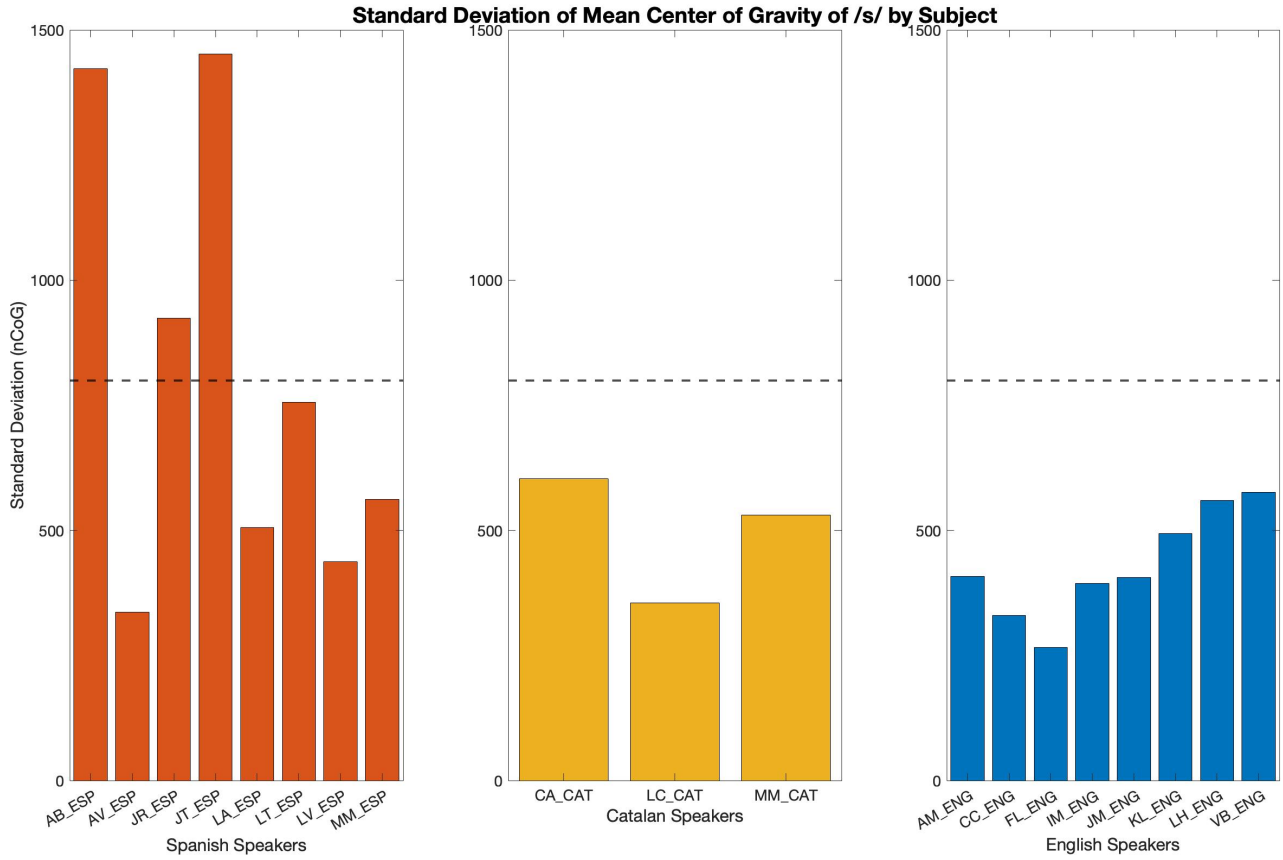

Figure 23: Within-speaker standard deviations of center of gravity of /s/. Dashed black line represents across-Spanish speaker mean. Spanish in red; Catalan in yellow; English in blue.

Figures (23) and (24) show that the standard deviation of the center of gravity of the /s/ of Spanish speakers is greater than that of Catalan or English, following predictions made by DT and contra those made by articulatory precision. These means are summarized in Table 6.

Language	Phoneme	Mean Standard Deviation
Spanish	/s/	799.63
Catalan	/s/	496.15
English	/s/	429.15

Table 6: Mean Standard Deviation: by language and phoneme

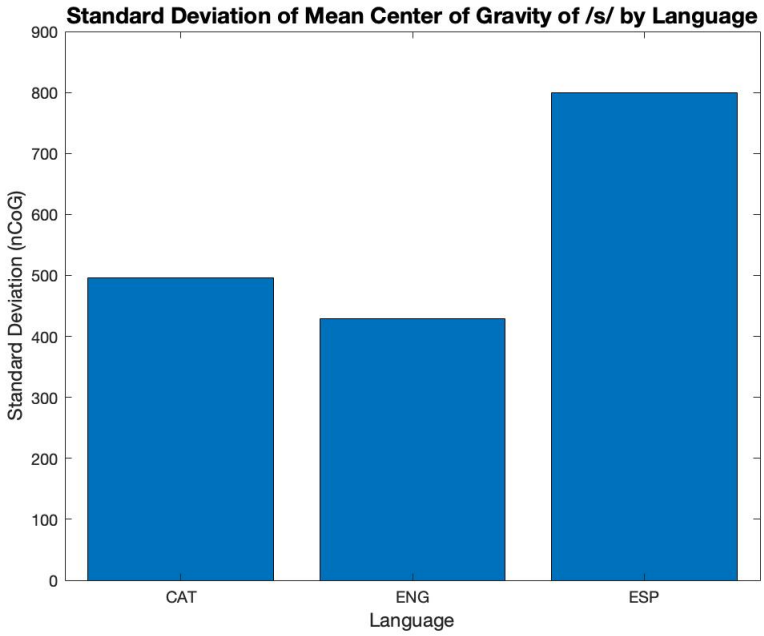

Figure 24: Within-speaker standard deviations of center of gravity of /s/ collapsed within-language.

Standard deviation of nCog of /s/ is greater in Spanish than Catalan or English, indicating more variance. The F-test results in Table 7 indicate if these differences are statistically significant[8].

Language Pair	F statistic	Degrees of Freedom	p value	Mean Standard Deviation Difference
Catalan - Spanish	0.72	df1 = 88 df2 = 221	p = 0.08	-303.48
English - Spanish	0.68	df1 = 235 df2 = 221	p < 0.01	-370.48
English - Catalan	0.95	df1 = 235 df2 = 88	p = .76	-67.00
Catalan/English - Spanish	0.70	df1 = 324 df2 = 221	p < 0.01	-352.21

Table 7: F-Test Results

Results from the F-test in Table 7 statistically confirm the difference between Spanish and English: the mean standard deviation of nCoG in Spanish /s/ is 303.48 points higher than English. This difference is roughly the same between Spanish and Catalan, but in this case was not found to be statistically significant. This is hardly surprising given there

[8]F-tests were re-run on data excluding Spanish speakers AB_ESP and JT_ESP, and results were virtually identical. The same conclusions may be drawn with and without their exclusion.

were only three Catalan participants. This difference is expected to be significant with the addition of more Catalan speakers.

In order to test the more specific effect of inventory size, Catalan and English were grouped together and an F-test was conducted sampling from Spanish (smaller sibilant inventory) and Catalan/English (larger sibilant inventories). These results are reported in the last row of Table 7. nCoG of an /s/ that forms part of a larger inventory has a statistically lower variance (smaller standard deviation) than the /s/ in a smaller inventory language.

5 Discussion

The previous section presented results on speaking rate, center of gravity, and variance. Linear regression models of cue duration and response duration showed significantly positive correlations for all three languages, meaning the shorter the cue, the shorter the response. This was also true for response duration and sibilant duration. These results indicate that the visual cue got participants to speak at different rates, but this did not account for much, if any, of the variance in center of gravity.

Center of gravity results show that /s/ and /ʃ/ are significantly different within Catalan and English. CoG for the same sibilant is also slightly, but significantly, different depending on language and speaker, suggesting language- and speaker-specific effects on articulation. Along with results from previous studies, this serves to establish CoG as a reliable acoustic correlate for place of articulation.

Results on within-speaker variance indicate greater variance in CoG of /s/ for Spanish speakers than English and Catalan speakers. This difference was found to be statistically significant between Spanish and English, but not between Spanish and Catalan, likely due to the small Catalan speaker population in this study. When grouped together by relative inventory size (i.e., English and Catalan nCoG values of /s/ collapsed and tested against Spanish), this difference was significant; the larger sibilant inventory shows less variance in center of gravity of /s/ than the /s/ in the smaller inventory. This result supports the DT

hypotheses: larger inventories show a clustering effect on within-category variation.

The articulatory precision claim does not receive much support here. Spanish speakers show a much higher variation in center of gravity of /s/ than English or Catalan, which likely correlates with place of articulation but may also be influenced by a different degree lip rounding/retraction. This result indicates that the articulation of /s/ may be quite flexible depending on the language or inventory size. The articulatory precision requirement of /s/ may be recovered when we look at its relative precision to other consonants. Mean standard deviations of /ʃ/ in Catalan and English were smaller than those for /s/ (compare: 320.97 to 496.15 in Catalan, and 311.67 to 429.15 in English), but F-test results were not significant. Perhaps compared to other fricatives or stops (/f/ and /t/ may be good choices here because they occur in all three languages, and /t/ is also coronal), /s/ may show a smaller standard deviation. This is a study for future work.

6 Conclusion and Future Directions

This study presented the results and implications of a production experiment on speakers of Spanish, English, and Catalan, which have varying sizes of sibilant inventories. The experiment was designed to test competing hypotheses from Dispersion Theory (Liljencrants and Lindblom (1972)) and claims about the relative articulatory precision of sibilants (Keating (1983)). DT predicts a clustering effect on within-category variation of sounds in a larger inventory, while articulatory precision claims predict an across-the-board constraint on within-category variation of /s/. Results on the variance of /s/ in Spanish, Catalan, and English show greater variance in place of articulation via center of gravity of /s/ in Spanish than its counterparts in Catalan and English. These results therefore appear to support the DT-hypothesized within-category clustering effect of /s/ in a larger sibilant inventory rather than a strict articulatory precision requirement.

It is possible that the articulatory precision requirement is relative within a language, and its effect may be recovered in comparing /s/ to other fricatives or stops in the same

language. Results in the current study actually showed a lower variance of /ʃ/ than /s/ in Catalan and English, but this difference was not significant. A comparison to a fricative like /f/ or another coronal /t/ may show different results. Measurements of mean and maximum center of gravity were considered here, but also looking at CoG trajectories may be interesting. Iskarous et al. (2011) found a variable jaw trajectory during the articulation of /s/; it is possible that simply taking the mean or the max CoG misses a dimension of variance. F-test results on variance did not find a significant difference between Spanish and Catalan, though this is likely due to the small speaker population of Catalan in this study (only three). This difference is predicted to strengthen to significance with the addition of more Catalan speaker data. Future iterations of this study would certainly benefit from the addition of articulatory data on jaw height and lip shape.

References

Behrens, S. and Blumstein, S. E. (1988). On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. *The Journal of the Acoustical Society of America*, 84(3):861–867.

Berkson, K. H. (2013). *Phonation types in Marathi: An acoustic investigation*. PhD thesis, University of Kansas.

Bjorndahl, C. (2018). A story of /v/: Voiced spirants in the obstruent-sonorant divide.

Bladon, A. and Seitz, F. (1986). Spectral edge orientation as a discriminator of fricatives. *The Journal of the Acoustical Society of America*, 80(S1):S18–S18.

Boersma, P. and Weenink, D. (2009). Praat: doing phonetics by computer (version 5.1.13).

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America*, 97(3):1916–1924.

Chomsky, N. and Halle, M. (1968). The sound pattern of English.

Clarke, C. and Luce, P. (2005). Perceptual adaptation to speaker characteristics: VOT boundaries in stop voicing categorization. In *ISCA workshop on plasticity in speech perception*.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., and Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108(3):804–809.

Clements, G. (1990). Place of articulation in consonants and vowels: A unified theory. *L'architecture et la géométrie des representations phonologiques (B. Laks & A. Rialland, editors). Paris: Editions du CNRS*.

Clements, G. N. (2003). Feature economy in sound systems. *Phonology*, 20(3):287–333.

Clements, G. N. and Hume, E. V. (1995). The internal organization of speech sounds.

Condax, I. (1980). Correlation of mandible position and vowel durations. *The Journal of the Acoustical Society of America*, 67(S1):S93–S93.

Dart, S. (1991). Articulatory and acoustic properties of apical and laminal articulations, vol. 79. *Los Angeles, CA: UCLA Phonetics Laboratory*.

Dart, S. N. (1993). Phonetic properties of o'odham stop and fricative contrasts. *International Journal of American Linguistics*, 59(1):16–37.

Davies, M. (2008). The corpus of contemporary American English (COCA): 400+ million words, 1990-present.

Davis, M. (2001). Corpus del español. M. Davis.

Disner, S. F. (1983). *Vowel quality: The relation between universal and language specific factors*, volume 58. Phonetics Laboratory, Department of Linguistics, UCLA.

Erker, D. (2012). Of categories and continua: Relating discrete and gradient properties of sociophonetic variation. *University of Pennsylvania Working Papers in Linguistics*, 18(2):3.

Evers, V., Reetz, H., and Lahiri, A. (1998). Crosslinguistic acoustic categorization of sibilants independent of phonological status. *Journal of phonetics*, 26(4):345–370.

Gordon, M., Barthmaier, P., and Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, 32(2):141–174.

Green, J. R., Nip, I. S., Wilson, E. M., Mefferd, A. S., and Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research*.

Grey, J. M. and Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *The Journal of the Acoustical Society of America*, 63(5):1493–1500.

Hauser, I. (2017). A revised metric for calculating acoustic dispersion applied to stop inventories. *The Journal of the Acoustical Society of America*, 142(5):EL500–EL506.

Hauser, I. (2019). Language-specific variability patterns in Hindi and English stop production. Preprint on webpage at `blogs.umass.edu/ihauser/research/`.

Heinz, J. M. and Stevens, K. N. (1961). On the properties of voiceless fricative consonants. *The Journal of the Acoustical Society of America*, 33(5):589–596.

Hualde, J. I. (2005). *The Sounds of Spanish with Audio CD*. Cambridge University Press.

Hughes, G. W. and Halle, M. (1956). Spectral properties of fricative consonants. *The journal of the Acoustical Society of America*, 28(2):303–310.

Hume, E. V. (1994). *Front vowels, coronal consonants and their interaction in nonlinear phonology*. New York: Garland.

Iskarous, K., Shadle, C. H., and Proctor, M. I. (2011). Articulatory–acoustic kinematics: The production of American English /s/. *The Journal of the Acoustical Society of America*, 129(2):944–954.

Jakobson, R., Fant, C. G., and Halle, M. (1951). Preliminaries to speech analysis: The distinctive features and their correlates.

Jongman, A. (1989). Duration of frication noise required for identification of English fricatives. *The Journal of the Acoustical Society of America*, 85(4):1718–1725.

Jongman, A., Blumstein, S. E., and Lahiri, A. (1985). Acoustic properties for dental and alveolar stop consonants: A cross-language study. *Journal of Phonetics*, 13:235–251.

Jongman, A., Wayland, R., and Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(3):1252–1263.

Keating, P. A. (1983). Comments on the jaw and syllable structure. *Journal of Phonetics*, 11(4):401–406.

Kim, H., Clements, G. N., and Toda, M. (2015). The feature [strident]. *Features in Phonology and Phonetics: Posthumous Writings by Nick Clements and Coauthors*, 21:179.

Koenig, L. L., Shadle, C. H., Preston, J. L., and Mooshammer, C. R. (2013). Toward improved spectral measures of /s/: Results from adolescents. *Journal of Speech, Language, and Hearing Research*, 56(4):1175–1189.

Ladefoged, P. and Maddieson, I. (1996). *The Sounds of the World's Languages*, volume 1012. Blackwell Oxford.

Lahiri, A., Gewirth, L., and Blumstein, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study. *The Journal of the Acoustical Society of America*, 76(2):391–404.

Liljencrants, J. and Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, pages 839–862.

Lindau, M. and Wood, P. (1977). *Acoustic vowel spaces*, volume 38. Phonetics Laboratory, Department of Linguistics, UCLA.

Lindblom, B. (1986). Phonetic universals in vowel systems. *Experimental phonology*, pages 13–44.

Lisker, L. and Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3):384–422.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge university press.

Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *The Journal of the Acoustical Society of America*, 88(3):1286–1298.

Mason, K. (1987). Child language and other evidence for /s/ variation in Spanish dialects. *Papers in applied linguistics–Michigan*, 3(1):64–78.

McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86(2):B33–B42.

Mielke, J. (2008). *The emergence of distinctive features*. Oxford University Press.

Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the cuidado project.

Pisoni, D. B. and Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & psychophysics*, 15(2):285–290.

Quirk, R., Greenbaum, S., Leech, G. N., Svartvik, J., et al. (1972). A grammar of contemporary English.

Recasens, D. and Espinosa, A. (2007). An electropalatographic and acoustic study of affricates and fricatives in two catalan dialects. *Journal of the International Phonetic Association*, 37(2):143–172.

Recasens, D. and Mira, M. (2013). An articulatory and acoustic study of the fricative clusters /sʃ/ and /ʃs/ in Catalan. *Phonetica*, 70(4):298–322.

Reese, S., Boleda, G., Cuadros, M., and Rigau, G. (2010). Wikicorpus: A word-sense disambiguated multilingual Wikipedia corpus.

Scarborough, R. and Zellou, G. (2013). Clarity in communication:"clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *The Journal of the Acoustical Society of America*, 134(5):3793–3807.

Shadle, C. H. (1985). The acoustics of fricative consonants.

Shadle, C. H. (1990). Articulatory-acoustic relationships in fricative consonants. In *Speech production and speech modelling*, pages 187–209. Springer.

Shadle, C. H. (1991). The effect of geometry on source mechanisms of fricative consonants. *J. Phonet.*, 19:409–424.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17:3–45.

Stevens, K. N. and Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, 38(1):10–19.

Tabain, M. (2001). Variability in fricative production and spectra: Implications for the hyper- and hypo-and quantal theories of speech production. *Language and speech*, 44(1):57–93.

Utman, J. A. and Blumstein, S. E. (1994). The influence of language on the acoustic properties of phonetic features: A study of the feature [strident] in Ewe and English. *Phonetica*, 51(4):221–238.

Van Son, R. J. and Pols, L. C. (1996). An acoustic profile of consonant reduction. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, volume 3, pages 1529–1532. IEEE.

Wheeler, M. (2005). *The phonology of Catalan*. Oxford University Press on Demand.