Conflicting standards and variability: Spirantization in two varieties of Uruguayan Spanish[1]

Madeline Gilbert
Laboratoire de Phonétique et Phonologie (UMR7018, CNRS & Sorbonne Nouvelle)
madeline-besse.gilbert@sorbonne-nouvelle.fr
ORCID: 0000-0001-8352-8561

Abstract: Most sociolinguistic work on variation focuses on how rates of occurrence or mean measurements differ between speech communities and speakers. However, speakers and communities also differ in *variability* - that is, in dispersion around the mean. The current study investigates the effects of speech style and multilingualism on variation and variability, by measuring the degree of intervocalic /bdg/ spirantization in spontaneous and careful speech. Data come from two varieties of Uruguayan Spanish, one monolingual (Montevideo) and one in contact with Brazilian Portuguese (Rivera). The results from a variation analysis confirm expected linguistic and social effects on gradient spirantization. An analysis of variability shows that, at the group level, careful speech is more variable than spontaneous speech, and the data from Rivera is more variable than that from Montevideo. Variability at the individual level differs slightly, suggesting that the group-level variability arises from between-speaker variability and within-speaker variability in different contexts. I propose that multilingualism in Rivera may heighten variability because contact with Portuguese provides a wider range of available pronunciations, and that careful speech may increase variability because the available pronunciations are subject to conflicting standards that are most active in this style.

Keywords: sociophonetics, variation, variability, language contact, speech style

1 Introduction

Much sociolinguistic work focuses on *variation*, typically understood as "the use of multiple variants based on conditioning factors" (Vaughn, Baese-Berk and Idemaru 2019: 2). For categorical variables, speakers use different rates of variants depending on language-internal factors (e.g. stress, surrounding phonological context, frequency) and language-external factors (e.g. speech context, interlocutor, speaker age, social class). For variables measured gradiently, the mean measurement of an acoustic property varies according to the same types of factors. However, studying only variation misses another axis of difference between groups and speakers: *variability*.

While variation looks at how rates of a variant or mean acoustic measurement differ, variability looks at how dispersed rates or measurements are around a mean. For example, Tamminga et al. (2016: 314) analyze the variability in English TD deletion, a categorical variable. They show that two speakers may delete at the same rate overall (e.g. 50%), but Speaker A may hover around that rate (50% deletion) throughout the interview while Speaker B may move between periods of higher and lower deletion (e.g. 70% and 20%). The two speakers can thus have the same average deletion rate, while showing qualitatively different behavior. A similar idea can be applied to gradiently-measured variables, such as voice onset time (VOT). Speakers A and B

[1] This research was done while I was a graduate student at NYU. I am grateful for insightful comments from Laurel MacKenzie, Greg Guy, Lisa Davidson, and Maria Gouskova. Thanks also to audiences at NYU SocioLab, LabPhon 2020 and HLS 2021 for helpful feedback, and to the journal reviewers and editor, whose feedback substantially improved the quality of this paper.

may both produce /t/ with a mean VOT of 70ms. However, Speaker A's VOTs could consist of 60ms, 65ms, and 85ms, while Speaker B's consist of 68ms, 70ms, and 72ms. The speakers have the same *average* VOT, but Speaker A is more *variable* than Speaker B because her measurements are further dispersed from her mean. An analysis focusing only on mean VOT misses this difference. In the current study, I use *variability* to refer to "fluctuations within a single measure, specifically within-category acoustic-phonetic variability" (Vaughn, Baese-Berk and Idemaru 2019: 2), as in the VOT example.

Existing studies on variability suggest that variability is linked to language-internal and language-external factors. In terms of language-internal factors, within-category acoustic variability has been argued to be related to the size of the phonological inventory (Lindblom 1986: 33), the degree of reliance on certain contrastive cues (Hauser 2018; Hauser 2019; Hauser 2020; Hauser 2021), and the nature of corresponding phonological categories in bilingual speakers' two languages (Johnson and Babel 2021). In terms of language-external factors, variability has been found to be conditioned by speech style (Ferguson and Kewley-Port 2007; Alvord and Rogers 2014; DiCanio et al. 2015), L2 status (e.g. Wade, Jongman and Sereno 2007; Jongman and Wade 2007; Vaughn, Baese-Berk and Idemaru 2019), and multilingualism/multidialectalism (Bosch and Ramon-Casas 2011; Sharma 2011; Levy and Hanulíková 2019). Other studies suggest that bilingual speakers may have larger linguistic repertoires and that their high variability may be most evident in careful speech (Bullock and Toribio 2009: 52).

While this work is suggestive, several questions remain unanswered. With the exception of Hauser's, and Johnson and Babel's studies, few studies investigate variability in consonantal variables measured gradiently. Gradient measurements can capture patterns not visible in a categorical analysis, and many variables are difficult to divide into discrete allophonic categories. For example, in Yucatan Spanish, long VOT in voiceless stops is a stigmatized, contact feature (Michnowicz and Carpenter 2013). It is not obvious where to draw the line between "aspirated" (non-standard) and "non-aspirated" stops (standard), but a gradient approach allows the authors to analyze how sociolinguistic factors affect VOT duration. Whether listeners actually *use* phonetic information in a gradient way to make social evaluations is an empirical question. Furthermore, many studies on variability do not quantify variability, consider how social factors affect variability, or interpret their results in line with their definition of variability. Understanding variability is crucial to understanding speakers' linguistic repertoires, how speakers use them, and how these repertoires come to be.

In this paper, I test how speech style and multilingualism affect within-category acoustic variability. Data come from word lists (careful speech) and interviews (spontaneous speech) in two varieties of Uruguayan Spanish: monolingual Montevideo Spanish and multilingual Rivera Spanish, a variety in contact with Portuguese on the Uruguay-Brazil border. The variable is intervocalic /bdg/, which undergoes spirantization in Spanish but not in Portuguese. To set the stage for the variability analysis, I first examine the social and linguistic conditioning factors on degree of /bdg/ spirantization - measured gradiently - to confirm that this variable behaves similarly in Uruguay as it does in other Spanish varieties. I then turn to variability, quantified using variance. To preview the results, variability at the group level is higher in careful speech than in spontaneous speech, and in Rivera (multilingual) than in Montevideo (monolingual). At the individual level, the results differ slightly. Rivera speakers produce higher variability in careful speech than in spontaneous speech, but Montevideo speakers show no consistent effect of style. Furthermore, Rivera speakers reliably produce more variability than Montevideo speakers in careful speech, but not in spontaneous speech. I suggest that heightened variability in careful speech may arise due to conflicting sociolinguistic standards and the nature of the word list reading task. I also propose that the heightened variability in Rivera, as compared to Montevideo, may occur due to the linguistically diverse input in Rivera, which may broaden speakers' categories, allowing a wider range of possible realizations. The mismatches between group-level and individual-level variability suggest that both between-speaker differences (interspeaker) and within-speaker variability (intraspeaker) contribute.

1.1 Roadmap

The rest of the paper is as follows. In Section 2, I discuss the existing research on variability, spirantization, the speech communities, and the motivations and predictions for the current study. Section 3 presents the methods of the study. Section 4 presents results, first confirming social and linguistic effects on spirantization (constriction degree) (Section 4.1) and then analyzing group and individual-level variability (Section 4.2). In Section 5, I discuss how social, attitudinal, and sociolinguistic factors may explain the observed effects of city and style on variability. Section 6 concludes.

2 Background

2.1 Variability

Both language-internal and language-external factors have been found to affect variability. Language-internal factors include the structure of the phonological inventory and how contrastive cues are used. Lindblom (1986: 33) proposed that larger vowel inventories result in less variability within each vowel category, in order to maintain distinctions between them. Recent work further investigates this claim by quantifying variability. Hauser (2019 i.a.) uses acoustic data from consonantal variables to make a more nuanced point: variability is conditioned by differences in phonological contrast implementation, not just by the number of contrasts in the inventory. She finds that speakers who rely heavily on a cue to implement a multidimensional contrast produce this particular cue with less variability than cues they rely on less. For bilingual speakers, Johnson and Babel (2021) find that the amount of permissible variability within a category in one of their languages constrains variability in the corresponding category in their other language. Finally, some sounds may inherently have more room for variability. For example, long-lag VOT may show more variability in VOT duration than short-lag VOT, because long-lag VOT has a wider range of acceptable realizations (Bullock et al. 2006: 14).

Language-external factors previously found to affect variability include speech style and multilingualism. In careful speech, realizations of the same vowel are similar to each other, forming tight acoustic clusters (Harmegnies & Poch-Olivé 1992; DiCanio et al. 2015). These categories are well-dispersed in acoustic and perceptual space, making them distinct from each other and resulting in an overall expanded vowel space (Smiljanić and Bradlow 2005; Ferguson and Kewley-Port 2007; Alvord and Rogers 2014; DiCanio et al. 2015). In terms of multilingualism, L2 speakers are *differently* variable (not necessarily *more* variable) than native speakers (Vaughn, Baese-Berk and Idemaru 2019). Multilingualism and multidialectalism may also lead to increased variability in production and/or broader linguistic repertoires (Bosch and Ramon-Casas 2011; Sharma 2011; Sharma and Rampton 2015; Levy and Hanulíková 2019).

The current study makes three main contributions to the growing body of literature on variability. First, my study investigates the effect of speech style on variability in consonants. Hauser's and Johnson and Babel's work also focuses on acoustic variability in consonants, but does not look at the effect of speech style. This is an important gap: the literature on vowel variability proposes that less variability in careful speech comes from pressure towards inventory dispersion (Lindblom 1986; Ferguson and Kewley-Port 2007), and that heightened variability in spontaneous speech may be due to gestural undershoot (DiCanio et al. 2015). Differently from vowels, consonants may be less subject to inventory dispersion pressures since they are realized with multiple acoustic cues along different phonetic dimensions. Like vowels, consonants may be subject to undershoot in spontaneous speech, leading to less precise realizations of the target.

Second, my study investigates how linguistic standards, attitudes, and the nature of the task affect variability. Existing work that quantifies variability does not consider these factors (e.g. Ferguson and Kewley-Port 2007; Hauser 2019 i.a.; Vaughn, Baese-Berk and Idemaru 2019), which are important because they influence how speakers acquire and use their linguistic repertoires.

A final goal of this paper is to add to the body of work establishing a framework for studying variability. Defining, quantifying, and testing variability is a relatively recent endeavor. Some recent studies quantify variability in a clear way (e.g. Vaughn, Baese-Berk and Idemaru 2019; Hauser 2019 i.a. Johnson and Babel 2021), but other studies do not distinguish clearly between variation and variability, and between variability within groups (interspeaker) and variability within individuals (intraspeaker) (e.g. Bosch and Ramon-Casas 2011; Levy and Hanulíková 2019). Other studies suggest higher variability in multilingual settings, but do not quantify it (e.g. Sharma 2011), or they conceptualize it differently - for example, as style shifting between varieties accomplished by using different proportions of dialectal features (*lectal focusing,* Sharma and Rampton 2015).

I analyze variability at both the group and individual levels. By *group-level variability*, I mean how dispersed the groups' measurements are around the group's mean acoustic measurement. For example, how spread-out are the Montevideo interview data points, around the mean constriction degree in Montevideo interviews? For *individual-level variability*, I mean how dispersed an individual speakers' measurements are around their own mean (following Vaughn, Baese-Berk and Idemaru 2019: 2). In other words: how spread-out are speaker F2R's interview data points around her own mean constriction degree in that speech style?

2.2 Spirantization in Spanish

Spanish intervocalic /bdg/ spirantization is ideal to study acoustic variability because it is well-understood and can be measured gradiently. In most monolingual Spanish varieties, voiced stops /bdg/ are occlusive post-pausally and after homorganic laterals or nasals (Table 1a, b). Intervocalically and elsewhere, they are spirantized (Table 1c, d) (Navarro Tomás 1918; Harris 1969; Quilis 1999; Hualde 2014).

Table 1 Spanish distribution of [βðɣ] and [bdg]

			/b/		/d/		/g/	
Occlusive	a.	Post-pausal	ˈ**b**ino	'wine'	ˈ**d**eðo	'finger'	ˈ**ɡ**ato	'cat'
	b.	Homorganic lateral/ nasal	ˈam**b**os	'both'	alˈ**d**ea	'town'	ˈaŋ**ɡ**ulo	'angle'
Spirantized	c.	Intervocalic	aˈ**β**eja	'bee'	ˈde**ð**o	'finger'	ˈla**ɣ**o	'lake'
	d.	Elsewhere	ˈal**β**a	'dawn'	arˈ**ð**er	'to burn'	ˈal**ɣ**o	'something'

Early work discusses allophones of /bdg/ as discrete entities (deleted, spirantized, occluded). More recent work treats spirantization as a point on a gradient continuum between occlusion and deletion. From the gradient perspective, constriction degree is inversely related to degree of spirantization. Tokens with *high* constriction have a *low* degree of spirantization; they are closer to full occlusion. Tokens with *low* constriction have a *high* degree of spirantization; they are closer to deletion.

Degree of constriction has been found to be conditioned by language-internal and language-external factors, and the conditioning factors are similar across varieties of Spanish. Language-internal factors that condition constriction degree include segment (/b/, /d/, or /g/) (Carrasco, Hualde and Simonet 2012), surrounding vowels (Cole, Hualde and Iskarous 1999; Ortega-Llebaria 2004; Simonet, Hualde and Nadeu 2012), stress (Cole, Hualde and Iskarous 1999; Ortega-Llebaria 2004) and lexical frequency (Eddington 2011). Language-external factors include speaker age, sex, socioeconomic status, and speech style. Constriction degree tends to be lowest (more spirantized) among young speakers, male speakers, and speakers of lower socioeconomic classes (e.g. Cedergren 1979; Rogers 2016; Butera 2018) and in spontaneous speech, which tends towards reduction for most variables (e.g. Lewis 2001: 59; Carrasco, Hualde and Simonet 2012). Spirantization patterns also differ somewhat across dialects (Carrasco, Hualde and Simonet 2012; Rogers 2016; Butera 2018).

In intervocalic position, the two extremes of the continuum - deletion and full occlusion - are non-standard. Spirantization is the norm, but extreme spirantization bordering on deletion is often stigmatized. Deletion is found throughout the Spanish-speaking world and is often associated with low socioeconomic status and education (Hualde 2014). Occlusive stop realizations, however, are also highly marked: they do not characterize any monolingual variety of Spanish and are not the target of careful speech (Carrasco, Hualde and Simonet 2012; Martínez-Gil 2019: 46–47).[2]

2.3 Speech communities

Participants in my study came from Montevideo and Rivera, Uruguay. Montevideo is the capital of Uruguay (~1.5 million inhabitants), and is monolingual Spanish-speaking. As the largest city in the country, it is the main educational, cultural, and professional center. Rivera is a smaller city on Uruguay's northern border with Brazil, and has a long history of language contact. The territory traded hands between the Spanish and Portuguese during colonial days, leading to intense, sustained contact between the two languages (Rona 1965). Today, Rivera and Santana do Livramento - it's sister city on the Brazilian side of the border - are a continuous city spanning the border, with complete freedom of movement (Carvalho 1998). Rivera is also somewhat isolated from the rest of Uruguay, because it is far from the capital. This separation contributed to the formation of a border identity that combines Uruguayan and Brazilian elements, and to the formation of a language "mixture" that Riverenses refer to as *Portuñol* (Carvalho 1998).[3] Recent

[2] Full stop realizations are found intervocalically in some contact and L2 varieties, such as the Spanish spoken in the Yucatan Peninsula in Mexico (Michnowicz 2009).

[3] The term *Portuñol* is problematic because it is often used pejoratively - by its own speakers and outsiders - to refer to the "mixed" language spoken on the border (Church 2007: 53; García de los Santos 2014). However, I use it here because it is the term most often used by speakers themselves. The speakers in my own data fall into line with participants in previous linguistic studies (e.g. Waltermire 2014), using the term to describe their own language in both positive and negative ways. Other names for this variety include *fronterizo* (Rona 1965; Hensey 1972), *Dialectos Portugueses de Uruguay [Portuguese Dialects of Uruguay, DPU]* (Elizaincín 1992), and *Uruguayan Portuguese* (Carvalho 1998; Carvalho 2014).

improvements in infrastructure have increased contact between the capital and the border, and led to corresponding cultural and linguistic changes towards Montevideo norms (Carvalho 1998; Carvalho 2003).

The linguistic situation in Rivera is complex. Languages present include standard Spanish and Portuguese, rural varieties of Spanish and Portuguese, and Portuñol. Multilingualism goes back generations. Portuguese was the first European language in the region (Rona 1965), and Spanish was imposed by the Uruguayan government through the educational system (Barrios et al. 2005). Imposing Spanish had the effect of relegating Portuguese to a low-status vernacular that was stigmatized as “broken”, poor, uneducated, and rural, and of creating a population of multilingual speakers. Rivera can be described as diglossic: Spanish is the language of public and official affairs and Portuguese is a vernacular (Hensey 1972; Elizaincín 1973; Elizaincín 1975; Elizaincín 1978; Behares 1984; Carvalho 1998). The languages are also geographically stratified within the city. People who live in the center tend to be of higher social classes and speak mostly Spanish, while those who live on the periphery tend to be of lower social classes and speak mostly Portuguese or Portuñol (Carvalho 1998).

Portuñol is not the focus of the current paper, but is well-studied (e.g. Rona 1965; Hensey 1972; Elizaincín 1973 et. seq. Behares and Elizaincín 1983; Elizaincín, Behares and Barrios 1987; Carvalho 1998 et. seq. Lipski 2006; Behares 2007; Church 2007). Attitudes towards Portuñol and language “mixing” are largely negative, both within and outside the community (Elizaincín 1973; Asencio 2006; Church 2007), although some speakers are shifting towards valuing the “mixed” language as part of a unique, hybrid border identity (Asencio 2006; Church 2007).

My focus is on Border Spanish. Border Spanish is largely consistent with monolingual Spanish linguistic systems, resisting contact influences from Portuguese in both syntactic and morphological domains (Carvalho 2006a; Carvalho 2010; Carvalho and Bessett 2015; Carvalho 2016). However, some phonological, morphological, and syntactic features do show Portuguese influence (Rona 1965: 25; Behares and Elizaincín 1983; Waltermire 2006; Waltermire 2011; Waltermire 2014; Carvalho and Kern 2019). At the phonological level, one such feature is intervocalic /bdg/. Even though Spanish typically enjoys higher prestige than Portuguese on the border, attitudes towards *Border* Spanish are also negative, because it has noticeable Portuguese influences (e.g. Carvalho 2006b; Waltermire 2006; García de los Santos 2014).

Intervocalic /bdg/ is an ideal variable to study phonetic contact effects between Spanish and Portuguese because the languages have similar stop inventories, but implement them differently at the phonetic level. In contrast to the spirantized realizations of monolingual Spanish varieties (e.g. [boˈ**ð**e**ɣ**a]), occlusive realizations have been reported in Border Spanish (e.g. [bo**d**e**g**a]) (Rona 1965; Hensey 1979; Behares and Elizaincín 1983). This feature is attributable to contact with Portuguese, which does not spirantize (e.g. [boˈ**d**e**g**a]) (Ferreira and Holt 2014: n. 4). While I am not aware of existing acoustic studies on spirantization in Uruguayan Spanish, nearby Argentinian Spanish has very light constriction (Colantoni and Marinescu 2010), and my own results (Section 4.1) suggest that Montevideo is similar. The amount of constriction in intervocalic /bdg/ could thus show contact influences when measured on a continuous scale.

2.4 The current study: Motivations, questions, and predictions

The current study contributes to existing literature by addressing how variability manifests in different speech styles and among monolingual and multilingual speakers. I analyze the constriction degree of intervocalic /bdg/ in two speech styles (spontaneous and careful) and two cities, one monolingual (Montevideo) and one multilingual (Rivera). Spirantization is measured as a continuous variable, relativizing the intensity of the consonant to that of the adjacent vowel, as a proxy for constriction degree. Variability is quantified using variance. Following a basic attention-to-speech model (Labov 1966), I assume that tasks like word list reading demand more attention than unscripted, spontaneous speech with an interviewer. I use the terms *careful speech* and *spontaneous speech* to correspond to these tasks, respectively.

The research questions and broad hypotheses are as follows. Hypotheses specific to the linguistic and social factors included in the models are in Section 3.5.2.

1. How does the constriction degree of intervocalic /bdg/ differ in Montevideo and Rivera? The point of this analysis is to confirm basic findings of constriction degree in these communities.
    - Based on data from Argentinian Spanish (Colantoni and Marinescu 2010), I expect Montevideo spirantized /bdg/ to show weak constriction. I am not aware of studies that analyze constriction degree in Montevideo, but there is no reason to expect the conditioning factors to differ substantially from other regions.
    - I expect Rivera speakers' productions to be influenced by Portuguese (Carvalho 2006b; Waltermire 2006 et. seq.), resulting in a higher constriction degree than for Montevideo speakers. This prediction is consistent with Waltermire's (2010) findings that - from a categorical perspective - Rivera speakers produce some occlusive variants.
    - Speakers in both cities should produce tokens with more constriction (less spirantization) in careful speech (word lists) than in spontaneous speech (interviews), in line with the finding that speakers produce forms with less reduction in careful speech generally (e.g. Kirchner 2004; Warner and Tucker 2011), and, more specifically, for Spanish voiceless and voiced stops (Lewis 2001; Pérez 2007; Hualde, Simonet and Nadeu 2011; Hualde 2014; Butera 2018).

2. Does the amount of variability - individual and group-level - differ by speech style?
    - Given previous acoustic work on vowels, word list data is expected to be less variable than interview data.

3. Does the amount of variability - individual and group-level - differ by the presence of multiple languages in the speech community and individual multilingualism?
    - Previous studies on variability and multilingualism suggest that exposure to linguistic diversity may contribute to heightened variability in production. I expect the Rivera data to be more variable than the Montevideo data.

3 Methods

3.1 Data collection

The data come from two months of fieldwork in Montevideo and Rivera in 2018.[4] Participants were recruited through personal contacts of the author. Sociolinguistic interviews and word lists were recorded in quiet rooms in private homes using a Zoom H4N Pro recorder and an AudioTechnica lapel microphone (AT831b). Interviews lasted from 45-80 minutes and followed

[4] This study was approved by the NYU IRB (IRB-FY2018-1920).

a traditional, semi-guided format (Labov, Baugh and Sherzer 1984; Tagliamonte 2006). The word list contained 145 words targeting four variables; here I analyze only /bdg/ spirantization (45 tokens) (see Appendix B for the relevant portion of the word list). Communication with participants was conducted in Spanish by the author. I am a non-native, highly proficient speaker with prior experience studying, living, and working in Uruguay.

Participants completed a demographic questionnaire. Rivera participants also completed the Bilingual Language Profile (BLP) to quantify language use, history, competence and attitudes (Birdsong, Gertken and Amengual 2012).[5] I adapted the existing English-Spanish questionnaire for Spanish-Portuguese bilingualism, with the added option *Portuñol* where applicable. I use a subset of the questions as a rough measure of Spanish use, a method I describe in the following section.

3.2 Participants

The 27 speakers range in age from 18-60 years old (Table 2). Montevideo speakers had lived in Montevideo for their whole lives; one resided in Solymar, a residential neighborhood just outside Montevideo. All had finished high school and were completing (or had completed) post-secondary degrees. Rivera participants had spent the majority of their lives in Rivera and many came from central neighborhoods (higher status, Spanish-dominant). Two participants resided in Santana (on the Brazilian side of the border) at the time of data collection (F1R, M2R), two are from peripheral neighborhoods of Rivera (M1R, M6R), and one is from a more rural area right outside Rivera (M7R). Two others had recently moved to Montevideo to study, but were interviewed in Rivera (M3R, M5R). All Rivera participants had completed high school and some post-secondary education, such as technical degrees (e.g. dairy production, robotics, logistics), teaching programs, and other professional training (nursing, police). Young Riverenses often move to Montevideo to complete university degrees.

Table 2 Study participants

	Montevideo		**Rivera**	
	Younger (18-30)	Older (31-61)	Younger (18-30)	Older (31-61)
Female	F3M, F4M, F5M, F10M	F2M, F8M, F9M	F2R, F3R, F6R, F7R	F1R, F4R, F5R, F9R, F10R
Male	M2M, M4M, M8M	M1M, M7M	M1R, M3R, M5R, M7R	M2R, M6R

All Montevideo participants reported Spanish as their first language. Most Rivera participants also reported Spanish as their first language, although two reported Portuñol, one reported both Spanish and Portuguese, and one reported Portuguese. To investigate the effect of language use, I calculated a percentage of Spanish use (vs. Portuguese/Portuñol) for Rivera speakers based on four questions from the BLP: in a normal week, what percentage of the time do you use the following languages (1) with friends? (2) with family? (3) at school/work? (4) when you talk to yourself? For each question, speakers circled a number on a response scale from 0%-100% for each language (Spanish, Portuguese, Other-including Portuñol). Because participants often assigned percentages across the three scales that summed to more than 100%, I converted

[5] BLP: https://sites.la.utexas.edu/bilingual/

their responses into 10-point scales and calculated the amount of Spanish as a proportion of points out of the total number of points a speaker assigned across all scales for all questions. The proportions of Spanish use among Rivera participants ranged from .41 (Spanish less than half the time) to .98 (almost exclusively Spanish).

3.3 Segmentation

Intervocalic /bdɡ/ tokens were taken from each speaker's interview and word list. For interviews, the first 20 tokens each of /b/ (<b>), /d/ (<d>), and /ɡ/ (<g>) after the 15-minute mark were included. The first 10 tokens of orthographic <v> were also collected. Orthographic <v> and <b> correspond to a single phoneme /b/ in Spanish, while Portuguese has both /v/ and /b/. <v> is infrequent in Spanish; taking only the first 10 tokens per interview ensured they came from the same section of the interview as the other segments. Word lists contained 45 tokens of intervocalic /bdɡ/, including several orthographic <v> tokens. Unsegmentable tokens and those with background noise were excluded. Token numbers are provided in Table 6 (Section 4.1).

Intervocalic /bdɡ/ tokens ranged from full occlusion to deletion, as illustrated in Figure 1.

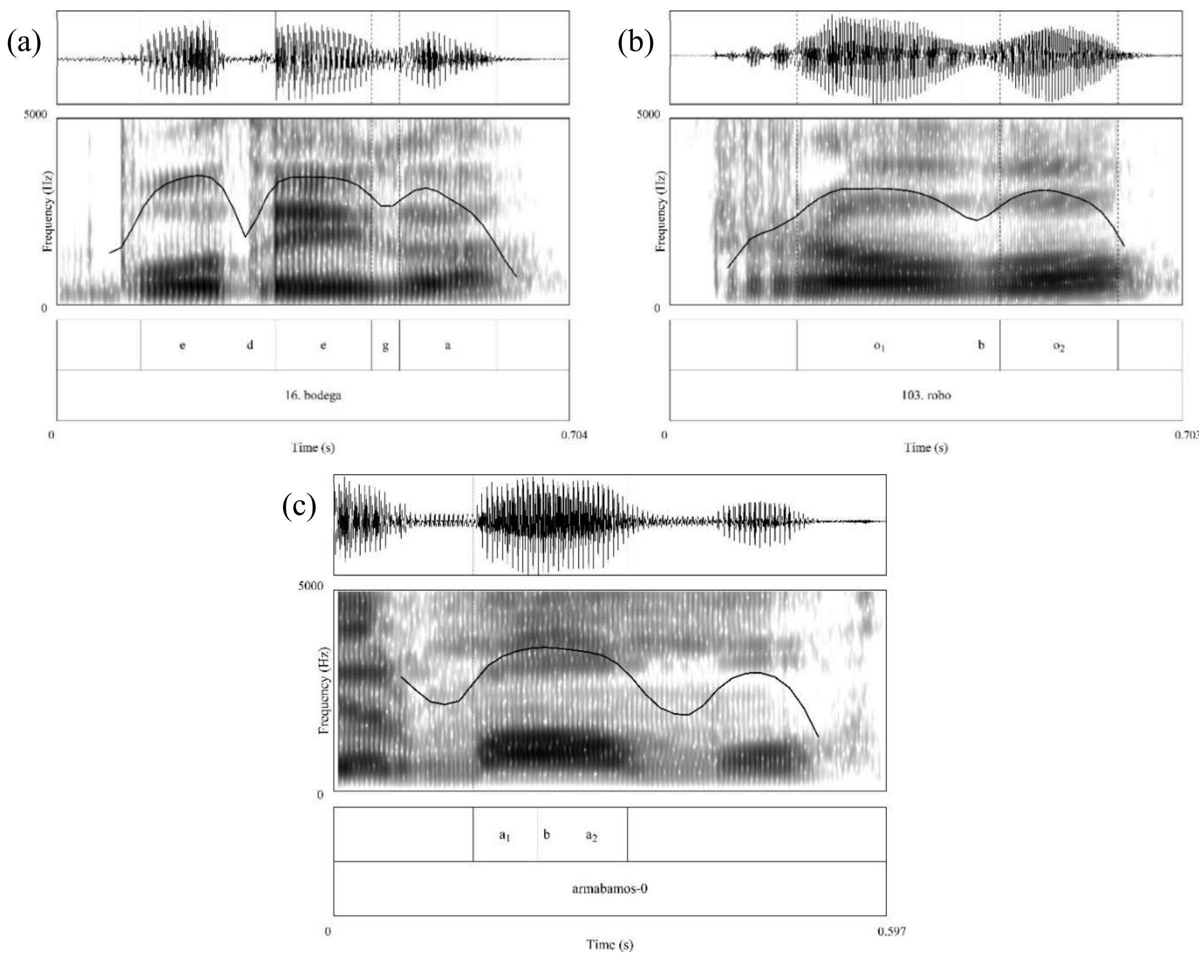

Figure 1 Segmentation examples showing the range of productions, including (a) highly constricted (/d/ is occlusive), (b) spirantized, and (c) deleted (no visible constriction)

Segmentation was done by the author and one research assistant in Praat (Boersma and Weenink 2018). We manually segmented the intervocalic voiced stop and both surrounding vowels. Boundaries were placed around /bdg/ to include the minimum intensity corresponding to constriction of the spirantized obstruent. Boundaries of the preceding and following vowels were placed to contain the maximum intensity of these vowels. Diphthongs and sequences of vowels were segmented as single vowels (e.g ([ˈx**we**ɣo] ‘game’, [seɣ**i.a**mos] ‘we were following’). The exact placement of the boundaries (as long as they include the minimum or maximum intensity) is not crucial, since I do not analyze duration.

3.4 Measurements

Following much previous work on spirantization, I use an intensity difference measurement to quantify the degree of constriction (Soler and Romero 1999; Martínez-Celdrán and Regueira 2008; Eddington 2011; Hualde, Shosted and Scarpace 2011; Rogers 2016). Instead of classifying allophones into discrete categories, the realizations fall on a continuum from more to less constricted. While some tokens are stops with bursts (fully occluded) and others are fully deleted, I treat them all as part of the same continuum since my interest is in within-category variability. Acoustic work repeatedly shows that /bdg/ realizations fall on a continuum (Quilis 1981: 223–4), and that it is difficult to divide this continuum into discrete categories.

Using Praat, I extracted the minimum intensity during the obstruent and the maximum intensity of the adjacent stressed vowel (or the preceding vowel, if both surrounding vowels are unstressed), and subtracted the minimum intensity of the obstruent from the maximum intensity of the vowel (Table 3). This Intensity Difference measurement is a proxy for how constricted the obstruent is relative to the stressed vowel.[6] As illustrated in Figure 1b, larger differences between the minimum intensity of /bdg/ (/ˈro**b**o/) and the maximum intensity of the stressed vowel (/ˈr**o**bo/) correspond to stronger constriction. Differences of 0 provide no acoustic evidence of /bdg/.

Table 3 Sample Intensity Difference calculation (F2 Rivera, word list)

/ˈr**ob**o/		Intensity (dB)
	Max. intensity: /o/	73
	Min. intensity: /b/	63
	Intensity Difference	10

Some tokens have a negative Intensity Difference. This happens when the minimum intensity of the obstruent is *higher* than the maximum vowel intensity. For example, the obstruent constriction is sometimes obscured by the intensity rise from high to low vowel (e.g. [eneˈm**iɣo**]). In the current analysis, negative differences suggest deletion. I keep these tokens in the analysis, and acknowledge that there may be other acoustic cues to the obstruent.

[6] As a reviewer points out, a 1dB change may not make the same difference at all baseline intensities. For example, a 1dB difference between 5dB-6dB may be more perceptible than a 1dB difference between 25dB-26dB. A difference like this could impact how listeners hear allophonic categories of /bdg/. This is a possible limitation of this measurement, but would need to be established experimentally. In intensity discrimination with pure tones, McGill and Goldberg (1968) find that intensity discrimination is more sensitive in the 5dB-15dB range than in the 25dB-35dB range, but that there are some inversions below 25dB. Normal speech typically ranges from 50-70dB, and it is not clear if there are differences within that range. The normal JND for intensity for speech is typically reported to be between 3dB-5dB (Koffi 2020).

I quantify variability using *variance*, following Hauser (2019) and Vaughn, Baese-Berk and Idemaru (2019). Variance is the standard deviation squared, and serves as a measure of dispersion around the mean measurement of a data set. For group-level variability, I calculated variance for each combination of city and style (Montevideo interviews, Montevideo word lists, Rivera interviews, Rivera word lists). To investigate individual-level variability, I calculated variance for each speaker in each of their speech styles (2 variances per speaker x 27 speakers = 54 variances).

3.5 Statistical methods

3.5.1 Models

There are two analyses: an analysis of variation and an analysis of variability. For variation, linear mixed-effects regressions were used to test social and linguistic effects on Intensity Difference (constriction degree). The models were run in R using the *lme4* package (Bates et al. 2015) and the BOBQYA optimizer, with p-values obtained from the *lmerTest* package (Kuznetsova, Brockhoff and Christensen 2017). First, I ran a superset model that contained a large set of predictors (described in Section 3.5.2). The superset model did not show an overall difference between Montevideo and Rivera in Intensity Difference. However, there was a significant interaction between City and several other predictors, so I constructed separate models for Montevideo and Rivera. Separating the cities also allowed me to include certain predictors in the Rivera model that are irrelevant in Montevideo (cognate status, amount of Spanish used).[7] For each city, final models were obtained through step-down model comparison, using the *step()* function from the *lmerTest* package. This function uses F-tests to remove predictors that do not significantly improve model fit. The threshold for exclusion was set at $\alpha = .1$ to avoid excluding potentially relevant predictors. Random effects were not considered for exclusion. The original and final model structures (after reduction) are in Appendix A.

Then, I analyze variability at the group and individual levels. For group-level variability, I tested for significant differences in variance between cross sections of City and Style using F-tests for the equality of variance. For individual variability, I calculated the variance for each speaker in each speech style, and used these variances as the dependent variable of a linear mixed-effects model. The predictors were Style, City, and Sex, and their three-way interaction. The goal of analyzing individual variability is to see if the effects found at the group level also hold within individual speakers, and thus whether the source of variability is between speakers (interspeaker), within speakers (intraspeaker), or both.

3.5.2 Predictors

The predictors in Table 4 were considered for inclusion in the superset model and the models for each city. I do not aim to provide a full account of factors conditioning constriction degree, so I limit the interactions to those directly relevant. The interactions between social factors (Sex*Style, Age*Style) are justified by previous findings that men and women, and speakers of different ages, can behave differently in style shifting. The interaction between linguistic factors (Style*Stress) is included because stress may have more of an effect in one style than in the other. The models also contained random intercepts for Speaker and Word. Further random effects prevented convergence in one or more models and thus were excluded from all for consistency. Several of these predictors were removed from the final models during reduction.

[7] Keeping both cities in the same model would also require four-way interactions, which are difficult to interpret.

Table 4 Predictors considered for inclusion in full spirantization model

Fixed Effects	**Description and levels (baseline level underlined)**
Sex	**Levels**: <u>female</u>, male
Style	**Levels**: <u>interview</u>, word list
Age	Continuous, from 19-61
Segment	**Levels**: <u>b</u>, d, g, <v>
Adjacent vowel	**Levels:** <u>a</u>, e, i, o, u (Vowel sequences were simplified into the vowel immediately adjacent to /bdg/, e.g. /ˈpuedo/: /ue/ → 'e', /siˈgio/: /io/ → 'i')
Stress	**Levels**: <u>stressed</u>, unstressed (e.g. /aˈ<u>d</u>entɾo/ 'inside' vs. /a<u>d</u>eˈmas/ 'additionally')
Past Participle	**Levels**: <u>false</u>, true (e.g. *termina<u>d</u>o* 'finished')
Log frequency	Continuous, log frequency per million
Cognate status (Rivera)	**Levels**: <u>Cognate</u>, NotCognate, CognateNoCons
Prop. Spanish (Rivera)	Continuous, 0-1 Proportion of daily language use in Spanish (vs. Portuguese/Portuñol)
Sex*Style Age*Style Style*Stress	

Several predictors merit further explanation:

- **Adjacent vowel**: Adjacent vowel refers to the vowel used to calculate the Intensity Difference. When one of the adjacent vowels was stressed, the Intensity Difference was calculated with that vowel. When neither was stressed, the Intensity Difference was calculated with the preceding vowel. I expect the height of this vowel to affect the Intensity Difference measurement because intensity differs systematically by vowel, with high vowels (e.g. /i, u/) having inherently lower intensity than lower vowels (e.g. /a/) (for English: Fairbanks 1950; Fairbanks, House and Stevens 1950; Lehiste and Peterson 1959). The Intensity Difference relativizes consonantal intensity to vowel intensity, so an /ib/ sequence may have a smaller Intensity Difference than an /ab/ sequence simply because /i/ is lower intensity than /a/, not because /bdg/ is more constricted in one sequence than in the other.[8] Whether these differences due to vowel quality reflect *perceived* differences in constriction degree, or whether they are simply an artifact of the relative intensity measurement, is beyond the current study.
- **Log Frequency**: Log frequency was obtained from the subtitle section of the EsPal corpus of Latin American Spanish, which contains approximately 245,000 word types and 460,000,000 tokens (Duchon et al. 2013). Because the corpus is built from movie subtitles,

[8] Previous work has found that the interaction between V1 and V2 predicts constriction degree (Cole, Hualde and Iskarous 1999). However, including both of these predictors and their interaction in my models raised a rank-deficiency warning in the model because there is not enough data for all the combinations of vowels for an accurate prediction. Because the effect of vowel quality is tangential to this paper, and using a single predictor of adjacent vowel did not change other results, I do not investigate this effect further.

the word frequency measures represent frequency in spoken language. For each word in my data set, I obtained the log frequency per million from the corpus. For words that did not appear in the corpus (approximately 30), I assigned them a frequency of .0001 per million and calculated log frequency based on that.

- **Cognate status**: Cognates may be expected to show more cross-linguistic transfer than non-cognates (Costa, Santesteban and Caño 2005; Amengual 2012; Amengual 2013). More specifically, cognate status has been found to have an effect on intervocalic /bdg/ in Rivera (Engelhardt 2017; Engelhardt, Gradoville and Waltermire 2018; Gradoville, Waltermire and Long 2021). I determined cognate status manually. Words were considered cognates if they were phonologically similar in both languages and had related - even if not identical - meanings (e.g. *pegar* was considered cognate, even though it means 'to hit' in Spanish, and 'to catch/get' in Portuguese). I considered them cognates, because words like *pegar* are often used with the Portuguese meaning in Border Spanish.

  I use one additional category, *CognateNoConsonant,* in addition to *Cognate* and *NotCognate*. *CognateNoCons* contains words that are cognates, but the corresponding Portuguese word lacks the relevant consonant (e.g. Sp: *podrido*, Pt: *podre*). The presence vs. absence of the consonant in the Portuguese cognate could affect its production in Spanish.

  Some words exist in Portuguese but are rarely used. For example, *mezclada* ('mixed') is frequent in Spanish and exists - but is infrequent - in Portuguese (*misturada* is used instead). For these words, I looked up frequency in SUBTLEX-BR, a corpus of movie subtitles in Brazilian Portuguese (Tang 2012), and classified words as cognates if the word appeared more than 10 times in the corpus. Those that appeared fewer than 10 times appeared to be from other languages, typos, or are simply too infrequently used to be expected to have an effect on their Spanish cognates. While there are more nuanced ways to classify cognate status (e.g. Amengual 2013; Carrasco-Ortiz, Amengual and Gries 2021), this is not the goal of my study.

Predictions for how these factors will affect Intensity Difference measurements in the variation models are in Table 5.

Table 5 Predictors in models and predictions

Predictor	**Higher IntDiff** (more constriction, less reduced)	**Lower IntDiff** (less constriction, more reduced)	**Reasoning**
Sex	Females ? Males		Females tend to reduce less than males, but Waltermire (2008) found no effect of sex on intervocalic /bdg/ pronunciation in Rivera.
Age	Older	Younger	Older speakers tend to reduce less than younger speakers.
Style	Word lists	Interviews	Careful speech is less reduced than spontaneous speech.

Stress	Stressed	Unstressed	/bdg/ are more constricted in onsets of stressed syllables.
Segment	Depends on dialect		No information about Uruguayan Spanish.
Adjacent vowel	Low/mid vowels	High vowels	Adjacent vowels with the lowest intensity (e.g. /i u/) should result in lower IntDiff measures.
Prop. Spanish (Rivera only)	Low	High	Speakers who use less Spanish and more Portuguese should have higher constriction, due to Portuguese influence (Waltermire 2010).
Past Participle	Not PP	PP	Past participles reduce more than other word types (Navarro Tomás 1953: 101).
Frequency	Low	High	High frequency words reduce more than low frequency words (Bybee 2002; Brown, Gradoville and File-Muriel 2014).
Cognate (Rivera only)	Cognate	NonCognate, CognateNoCons	Words with Portuguese cognates should have higher constriction degrees since their Portuguese counterparts do not spirantize.

The variability model predicts individual speakers' variances from City (Montevideo, Rivera), Style (Word list, Interview), and Sex (Male, Female), and their three-way interaction. It also has a random intercept for Speaker.

4 Results

I first present the variation results (Section 4.1). The purpose of presenting these results is to confirm previous findings on the social and linguistic factors conditioning intervocalic /bdg/ constriction degree. I focus on the effects of style (and stress, which interacts with style) and age, limiting the discussion of other effects. Full model results and a brief discussion of other predictors are in Appendix A (Table 11 and Table 12). Then, I move to the main focus: group and individual variability (Section 4.2).

4.1 Variation results

In a superset model including both Montevideo and Rivera (results not reported), the cities did not differ significantly in overall Intensity Difference (constriction degree). Figure 2 illustrates the effect of Style on Intensity Difference measures in Montevideo and Rivera, and Table 6 gives descriptive statistics. Voiced stop onsets to stressed and unstressed syllables are plotted separately, because the models for both cities showed a significant interaction between these predictors. The diamonds represent the mean Intensity Differences; the lines in the boxplots represent medians.

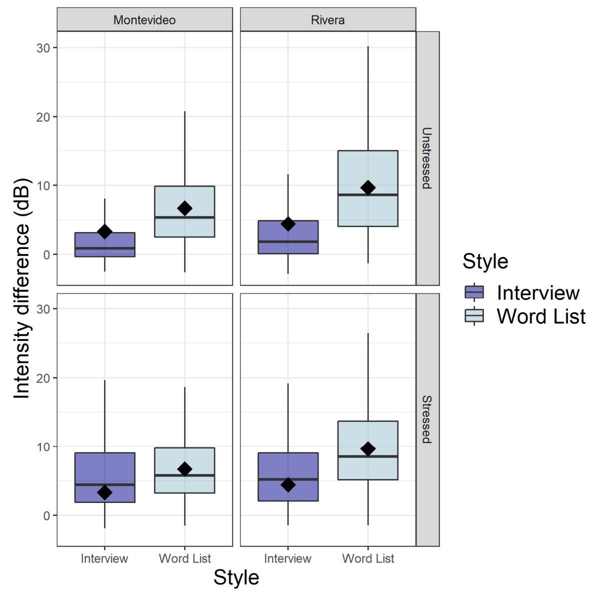

Figure 2 Intensity Differences by Style and Stress

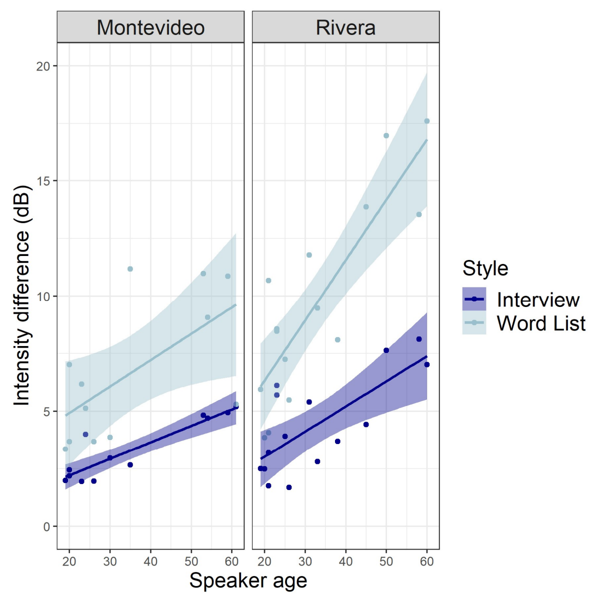

Figure 3 Intensity Differences by Speaker Age

Table 6 Descriptive statistics for group-level Intensity Difference measurements

Stressed	City	Style	Count	MeanIntDiff (dB)
No	Montevideo	Interview	549	2.03
	Montevideo	Word list	331	6.74
	Rivera	Interview	623	3.18

Stressed	City	Style	Count	MeanIntDiff (dB)
	Rivera	Word list	417	9.75
Yes	Montevideo	Interview	291	5.74
	Montevideo	Word list	234	6.61
	Rivera	Interview	427	6.24
	Rivera	Word list	289	9.51

As seen in Figure 2, Rivera and Montevideo have similar mean Intensity Differences overall and within each speech style, although the means are numerically higher in Rivera. This reflects a higher constriction degree in Rivera than in Montevideo. However, the differences between cities are small, and the superset model (not reported) did not find a significant effect of City or its interaction with Style. Additionally, tokens from word lists appear to have higher Intensity Differences than those from interviews in both cities.

For both cities (model results reported in Appendix A), the models have a significant main effect of Stress and a significant interaction between Style and Stress. Stress affects Intensity Differences differently depending on speech style. The main effect of Stress suggests that /bdg/ onsets to stressed syllables have significantly higher Intensity Differences than onsets to unstressed syllables in interview speech (the baseline level of Style) (Montevideo: $\beta = 2.41$, $p < .001$; Rivera $\beta = 2.73$, $p < .001$). The interaction between Style*Stress suggests that the effect of Style is weaker for stressed onsets than for unstressed onsets (Montevideo: $\beta = -2.20$, $p < .01$; Rivera: $\beta = -2.22$, $p < .01$). This interaction can be seen in Figure 2: unstressed onsets (top panels) are substantially less constricted in interview speech than in word list speech, but stressed onsets (bottom panels) are more similar in the two styles. The effect of speech style is strongest for unstressed onsets.[9]

In Figure 2, it is also apparent that box and whiskers for Rivera word lists are longer than for the other subsets of city and style. This difference points to greater variability, and provides crucial motivation for the variability analysis: although the mean Intensity Difference is similar across subsets of the data, the amount of variability differs.

In terms of Age, Figure 3 shows that younger speakers produce /bdg/ with lower Intensity Differences (less constricted) than older speakers (more constricted). Younger speakers in both cities have similar Intensity Differences, suggesting that the change towards Montevideo norms of spirantization may be advanced among my participants. Additionally, the plot suggests an interaction between Age and Style: older speakers show more of a stylistic distinction than younger speakers, who are more similar in the two styles. The models for each city confirm these results. In both cities, there is a significant main effect of Age and a significant interaction between Age and Style. Older speakers produce tokens with significantly higher Intensity Differences than younger speakers (Montevideo: $\beta = .06$, $p < .05$; Rivera: $\beta = .1$, $p < .01$), and the effect of Style is significantly stronger for older speakers than for younger speakers (Montevideo: $\beta = .06$, $p < .001$; Rivera: $\beta = .16$, $p < .001$).

[9] Framed the opposite way, the effect of Stress is smaller in word list speech than in interview speech. In word list speech, stress does not affect constriction degree much; in interview speech, stress increases constriction substantially.

4.1.1 Discussion of variation results

That Montevideo and Rivera do not differ in overall Intensity Difference - which I interpret as a proxy for degree of constriction - could indicate that spirantization has advanced in Rivera, at the expense of highly constricted variants. However, this claim should be taken with caution because (a) previous studies do not directly compare Montevideo and Rivera, so there is no way to verify that Rivera speakers are closer to Montevideo now than they were in the past; (b) my Rivera speakers are not demographically representative. They come mostly from central neighborhoods (which are middle-upper socioeconomic class), are educated, and have extensive contact with Spanish. It would be precisely this group - especially my youngest speakers - that would be expected to lead the change. Participants in Waltermire (2006 et seq.), the most comparable studies on spirantization in Rivera, produce high rates of occlusive tokens. They are also reported to be less Spanish-dominant and to come from varied socioeconomic backgrounds. That my study finds no difference in Intensity Difference between cities *may not* mean that Rivera, as a whole, has shifted more towards lighter constriction since Waltermire's data was collected in the early 2000s. Instead, it is possible that my participants represent the group that is most Spanish-dominant and has the most contact with Montevideo.

The effect of speech style is in line with previous work on Spanish /bdg/. Most studies look at the effect of speech style through deletion rates, which are higher in spontaneous speech than careful speech throughout the Spanish-speaking world (e.g. Pérez 2007; Hualde 2014). From a gradient perspective, many variables show more reduction in spontaneous speech than in careful speech (Kirchner 2004; Warner and Tucker 2011). Combining this general finding with findings on Spanish, intervocalic /bdg/ are expected to be realized as less constricted in spontaneous speech than in careful speech. This is consistent with my results. The interaction between Stress and Style is unexpected, and I discuss it further in Appendix A.

The effect of age is also mostly in line with the predictions and findings for other varieties of Spanish. Throughout the Spanish-speaking world, younger speakers have been found to reduce /bdg/ more than older speakers, both in rates of categorical deletion and in degree of spirantization (e.g. Cedergren 1979; Long and Baldwin 2013; Rogers 2016; Butera 2018). However, the interaction between age and speech style is somewhat unexpected, as there is no obvious reason that younger speakers should show less stylistic distinction than older speakers. In Rivera, the explanation might be related to different linguistic experiences. Older speakers grew up with more Portuguese contact, and less influence from Montevideo, than younger speakers. While both groups may consider spirantization to be the target for careful speech, older speakers' more extensive contact with Portuguese may lead them to realize spirantized /bdg/ with a higher constriction degree than younger speakers, who approximate the light constriction of Montevideo in both speech styles.[10] Another possible explanation - and one that could explain the interaction

[10] It is also possible that the large stylistic difference for older Rivera speakers arises because they produce a higher rate of fully occlusive variants than younger speakers, as has been found in previous studies (Waltermire 2006 et. seq.; Waltermire and Gradoville 2020). However, in order for this to explain the interaction between Style and Age, older speakers would have to produce a higher rate of fully occlusive variants in careful speech than in spontaneous speech. The previously mentioned studies do not consider the effect of speech style. Because the current study does not separate fully occlusive productions into a discrete category, it is not clear whether older Rivera speakers are producing spirantized variants with more occlusion than younger speakers, or whether they are producing more fully occlusive tokens. I think it is unlikely that their fully occlusive tokens are more frequent in careful speech, given how stigmatized those realizations are, but this is an empirical question.

in both Montevideo and Rivera - is that older speakers are more sensitive to formality differences, and to the pressure of a formal speech setting. That the effect is stronger in Rivera than in Montevideo could be due to the previously-mentioned differences in amount of contact with Portuguese between older and younger Rivera speakers.

A detailed analysis of other conditioning factors is outside the scope of the current study, but is available in Appendix A. While most effects were expected given findings on other varieties of Spanish, others (or the lack thereof) were more surprising. Several predictors were eliminated during model reduction, including Cognate Status and Proportion of Spanish Use. Previous studies do find that cognate status (Engelhardt 2017; Engelhardt, Gradoville and Waltermire 2018; Gradoville, Waltermire and Long 2021) and the amount of Spanish use (Engelhardt, Gradoville and Waltermire 2018; Waltermire 2010; Waltermire and Gradoville 2020) affect intervocalic /bdg/ spirantization (although cf. Engelhardt 2017). My study was not explicitly designed to test these effects, or effects of other linguistic factors, so my results for these factors should be taken as suggestive.

Having given an overview of social and linguistic factors conditioning Intensity Difference, I now turn to the results for variability.

4.2 Variability results

Although Montevideo and Rivera do not differ in mean overall Intensity Difference, they *do* differ in the amount of variability around the mean Intensity Difference. In this section, I present results for both group-level (interspeaker) and individual-level (intraspeaker) variability.

4.2.1 Variance: Group-level

Figure 4 and Figure 5 show density plots of the Intensity Difference measures for each combination of City and Style. Figure 4 compares word lists vs. interviews in each city. Figure 5 compares Montevideo vs. Rivera in each style. In Figure 4, the data is visually more spread out in word lists than in interviews for both cities. The Intensity Difference measures from the interviews cluster around 0 (little to no constriction), and there are few higher measurements. In contrast, the Intensity Difference measures from the word list data have a lower peak and taper off more gradually. In Figure 5, Montevideo and Rivera show a similar spread of measurements in interviews. In word lists, however, the Rivera data appear to be more dispersed than the Montevideo data.

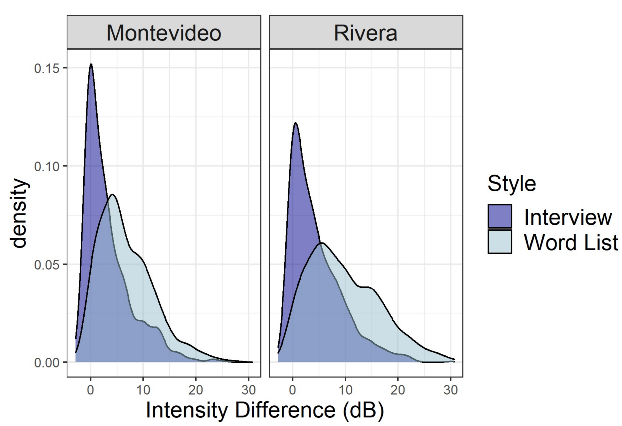

Figure 4 Distribution of measurements by speech style in Montevideo and Rivera

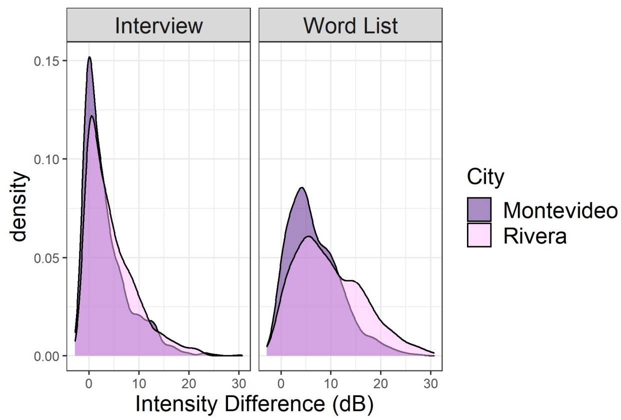

Figure 5 Distribution of measurements by City in interviews and word lists

The descriptive statistics in Table 7 confirm these visual observations about group-level variability, as measured by variance. Table 7 represents the same data as in Table 6 (collapsing over stress), and also shows SD, Variance, standard errors and confidence intervals.[11]

Table 7 Descriptive statistics with SD, Variance, standard errors, and confidence intervals

City	Style	Count	MeanIntDiff (dB)	SD	**Variance**	SE	CI
Montevideo	Interview	840	3.31	4.49	**20.17**	0.15	0.30
Montevideo	Word list	565	6.69	5.20	**27.05**	0.22	0.43
Rivera	Interview	1,050	4.43	4.88	**23.84**	0.15	0.30
Rivera	Word list	706	9.65	6.75	**45.56**	0.25	0.50

Within both cities, variance is higher in word lists than in interviews (Montevideo: 27.05 > 20.17; Rivera: 45.56 > 23.84). Within each style, interviews and word lists have higher variance in Rivera than in Montevideo, although the cities differ more in word lists (interviews: 23.84 > 20.17; word lists: 45.56 > 27.05). F-tests for the equality of variances confirm that these differences are significant. In both cities, the variance of word lists differs significantly from the variance of interviews (Montevideo: F = .75, $p < .001$; Rivera: F = .52, $p < .001$). In both speech styles, the variance of Montevideo differs significantly from the variance of Rivera (Interviews: F = .85, $p < .05$; Word lists: F = .59, $p < .001$). The difference between cities within interview speech is the smallest, as is visible in the left facet of Figure 5.

[11] Since this analysis is about variance in measurements, not the magnitude of Intensity Difference measurements themselves, and since all speakers produced words with all stress patterns, I leave stress out of this analysis. Whether variance depends on level of stress is a potentially interesting question for future research.

4.2.2 Variance: Individual-level

The previous section shows that cross-sections of City and Style have significantly different variances at the group level. But do City and Style predict variability within *individual* speakers too? The observed high group-level variability could come from either of the following sources:

- Speaker-internal variability: Speakers A and B both produce the range [bdg…βðɣ…Ø]
- Between-speaker variability (speaker-internal consistency): Speaker A hovers around [βðɣ] and B hovers around [bdg]

To illustrate how speakers can differ in variance, Figure 6 compares the distribution of Intensity Difference measurements for two Rivera speakers. Speaker F1R's data is more dispersed in her word list than in her interview - that is, it is less peaked and spread more evenly throughout her range. In her interview, she produces most tokens with low Intensity Differences (peak near 0). In her word list, she produces tokens spanning a wide range, suggesting both highly spirantized and highly constricted productions. Her variance is higher in her word list than in her interview. Speaker M2R's data shows a different pattern, with Intensity Difference measures spanning a wide range in both styles. His mean Intensity Difference is higher in word list speech than in interview speech, but variance appears to be similar in both styles. When variance is calculated, his interview has quantitatively higher variance than his word list.

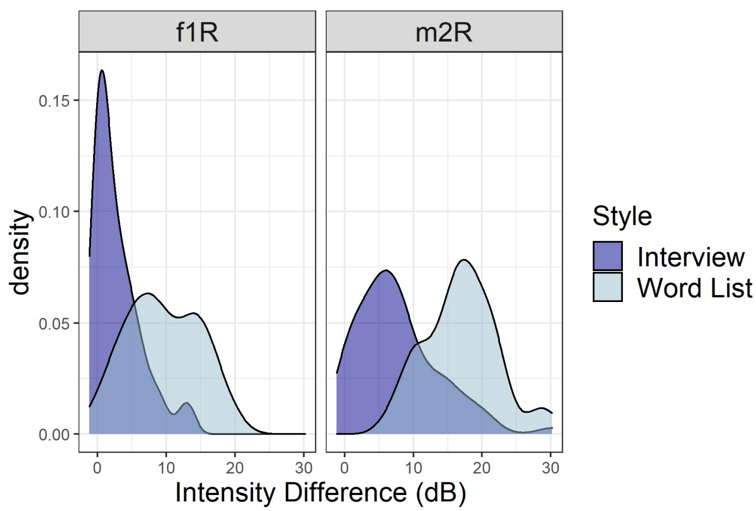

Figure 6 Distributions of measurements in word lists and interviews for two speakers. F1R (left): higher variability in word list; M2R (right): higher variability in interview.

Figure 7 plots each speaker's variance in each style. Variances for interviews are in dark blue; those for word lists are in light blue. Triangles represent speakers whose word lists have higher variance than their interviews; dots represent speakers whose interviews have higher variance than their word lists. The lines represent the average variance, calculated based on individuals' variance measures in a given style and city. Note that these are averages of *individual* variance measures, not variance calculated across entire groups (as is done in Table 7). In Montevideo, the lines overlap.

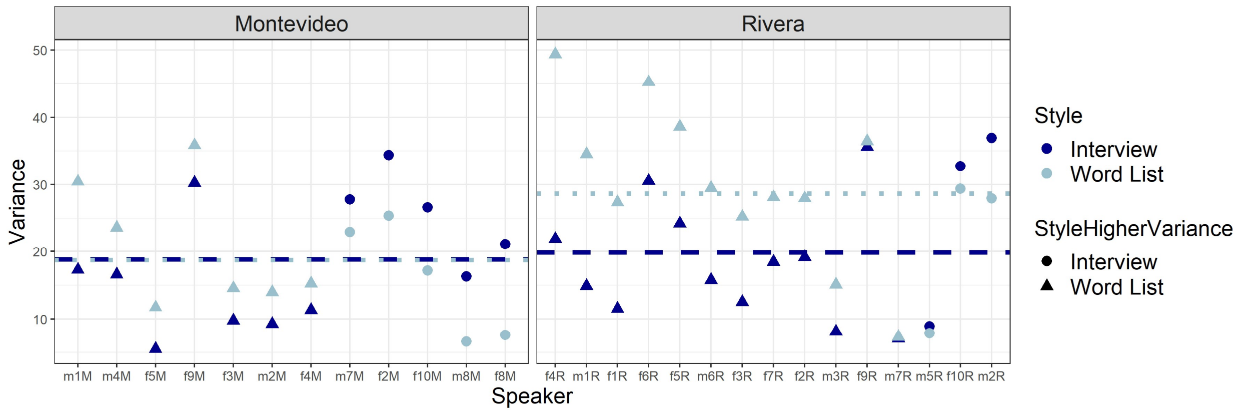

Figure 7 Variance by speaker and style. Word lists = light blue; Interviews = dark blue. Triangles = word list > interview; Dots = interview > word list. Lines represent average individual variance. In Montevideo, the lines overlap.

Figure 7 shows that most speakers have higher variance in their word list than in their interview. This pattern holds for proportionately more speakers in Rivera (12/15) than in Montevideo (7/12). Furthermore, the speakers with the opposite pattern - higher variance in their interviews than in their word lists - have large differences in variances between the two styles in Montevideo. In Rivera, however, the three speakers with this reversed pattern have quite similar variances in both styles. In terms of average individual variance, Montevideo speakers do not show a clear effect of speech style while Rivera speakers have reliably higher variance in word list speech than in interview speech.

The model (Table 8) confirms these visual results. The dependent variable is within-speaker, within-style variance (the data plotted in Figure 7). The model contains fixed effects of City, Style, Sex, their three-way interaction, and a random intercept for Speaker. The model has one significant effect: the interaction between City and Style (β = 12.86, $p < .01$). This result confirms the result visible in Figure 7: for individual speakers, word lists predictably boost variability in Rivera. That there is no main effect of Style suggests that, in Montevideo (the baseline level for the City predictor), speech style does not affect variability. That there is no main effect of City suggests that, in interviews (the baseline level for the Style predictor), Rivera and Montevideo speakers produce similar amounts of variability. Individual-level variability is predictably higher only in Rivera word lists.

Table 8 Model results of individual variance

	Est.	SE	df	t-val	Pr(>|t|)	Baseline	*n*	Mean Indiv. Variance
(Intercept)	19.84	3.38	41.65	5.87	0		54	21.83
Predictors								
City						(Montevideo)	24	18.79
Rivera	3.11	4.51	41.65	0.69	0.50		30	24.27
Style						(Interview)	27	19.42
Word list	-1.63	3.23	27	-0.51	0.62		27	24.25

Sex						(Female)	32	24.39
Male	-2.41	5.24	41.65	-0.46	0.65		22	18.11
Interactions								
Rivera:Word list	12.86	4.30	27	2.99	0.006**			
Rivera:Male	-5.26	7.05	41.65	-0.75	0.46			
Word list:Male	3.68	5	27	0.74	0.47			
Rivera:Word list:Male	-9.83	6.73	27	-1.46	0.16			
Random intercept: Speaker (SD = 6.60)								

4.2.3 Discussion of variability results

Group and individual-level variability results are summarized in Table 9. At the group level, word lists have higher variability than interviews in each city, and Rivera has higher variability than Montevideo in each style. The effect of city (multilingualism) is not unexpected, although it rarely receives explicit attention. The style effect is more surprising, challenging the assumption that careful speech is *less* variable than spontaneous speech. At the individual level, the results are similar, with two important mismatches (cells in bold): (1) Montevideo speakers show no predictable effect of style (Table 9a), while Rivera speakers do (Table 9b); (2) in interview speech, Rivera and Montevideo individuals do not differ in variance (Table 9c), but they do in word list speech (Table 9d).

Table 9 Comparison of group-level and individual-level variability

			Group level	Individual level
Style effect (within City)	Montevideo	a.	Word list > interview	**Word list ~ interview**
	Rivera	b.	Word list > interview	Word list > interview
City effect (within Style)	Interview	c.	Rivera > Montevideo	**Rivera ~ Montevideo**
	Word list	d.	Rivera > Montevideo	Rivera > Montevideo

The mismatches between group-level and individual-level variability shed light on the source of variability. The first mismatch (Table 9a) is that, in Montevideo, word lists show greater variance than interviews at the group level, but that speech style has no reliable effect at the individual level. This suggests that heightened group-level variability in Montevideo word lists comes from differences *between* speakers, rather than high variability *within* speakers. In contrast, the group and individual results for speech style align in Rivera (Table 9b), indicating that at least part of the heightened variability in Rivera word lists comes from predictable heightened variability *within individual speakers*. The second mismatch (Table 9c) is that, in interview speech, Rivera speech has higher variability than Montevideo speech, but that city does not predict individual-level variability. Being from a multilingual or monolingual city does not affect a speaker's variance in spontaneous (interview) speech. The heightened group-level variability in Rivera interviews likely comes more from differences *between* speakers than from predictable variability *within* speakers. In contrast, the group and individual results align in word lists (Table 9d), indicating that part of the heightened variability in Rivera word lists comes from predictably high, within-speaker variability in this speech style. The group of Rivera speakers is quite heterogeneous, but at least some of the variability observed at the group level is because individuals are internally variable.

In Section 5, I further discuss these effects, and suggest that they may result from conflicting linguistic standards that are most strongly active in careful speech.

5 Discussion

The goal of the current study was to investigate the amount of within-category phonetic variability in different speech styles and multilingual settings. The results, as they relate to the research questions, are as follows:

1. How does constriction degree differ in Montevideo and Rivera?

   The effects conditioning Intensity Difference measurements - which are taken as a proxy for degree of constriction - largely conform to expectations (stress, style, frequency, etc.) given previous work on Spanish varieties and on Rivera Spanish specifically, although there are some differences.

   Montevideo and Rivera do not differ from each other in overall degree of constriction: both have heavy spirantization, especially among younger speakers. This could indicate change in progress in Rivera towards Montevideo norms, but should not be taken as conclusive evidence due to the demographics of my speakers and the lack of real time data for comparison. Additionally, older speakers in both cities produce tokens with higher constriction and more stylistic differentiation in constriction degree than younger speakers.

2. Does the amount of variability - individual and group-level - differ by speech style?

   At the group level, variability is higher in word lists than in interviews. This goes against the prediction that careful speech is less variable than spontaneous speech. At the individual level, variability is higher in word lists than in interviews in Rivera, but not in Montevideo: part of the heightened variability in Rivera word lists comes from variability within speakers, not just differences between speakers. In Montevideo, heightened variability in word lists at the group level most likely comes from between-speaker differences.

3. Does the amount of variability - individual and group-level - differ by the presence of multiple languages in the speech community and individual multilingualism?

   At the group level, variability is higher in Rivera (multilingual) than in Montevideo (monolingual). This is as predicted. At the individual level, variability is higher in Rivera than in Montevideo in word list speech, but not in interview speech. These mismatches suggest that, in word list speech, heightened variability in Rivera comes at least partly from within-speaker variability. In interview speech, in contrast, the observed heightened group-level variability in Rivera appears to come from between-speaker differences.

I propose that heightened variability in Rivera and in word lists may arise due to specific sociolinguistic characteristics of spirantization in Spanish and in Rivera. The remainder of this section focuses on potential explanations for the effect of city (Section 5.1) and the more surprising effect of speech style (Section 5.2). Then, I discuss what characteristics of a variable might lead it to show the style effect (Section 5.3). I finish by acknowledging other possible explanations for the effects found (Section 5.4) and exploring the effect of speaker age on variability (Section 5.5).

5.1 Variability by City

There are several reasons why Rivera might show higher within-category acoustic variability than Montevideo at the group level (differences *between* speakers). First, the presence of Portuguese in Rivera leads to a larger range of possible realizations of intervocalic /bdg/, giving speakers more options and more ways to differ from each other. Both Spanish-like and Portuguese-like realizations are available to Riverenses, so their range extends from deletion to complete occlusion ([∅… βðɣ…bdg]). Montevideo speakers have a smaller available range, from deletion to spirantization ([∅… βðɣ]). Second, Rivera speakers have different levels of exposure to Spanish and Portuguese. Some of my participants are very Spanish-dominant, while others are more balanced. Spanish-dominant speakers may produce intervocalic /bdg/ differently (more spirantized and less variably) than speakers with more exposure to Portuguese (more constricted and more variably). Speakers may also differ from each other in adopting changes towards Montevideo norms.

Third, Rivera speakers differ in their attitudes towards Border Spanish, Montevideo Spanish, and the variants associated with each, possibly reflecting a process of stylistic reinterpretation. In stylistic reinterpretation, variants are "re-valued as statusful due to their association with a city or a host community" (Sharma 2021: 259). The presence of Montevideo variants in Rivera creates an opportunity for reinterpretation, and there is some evidence that /bdg/ variants from Montevideo carry different connotations in Rivera. Waltermire (2010) argues that speakers in Rivera use deleted variants of intervocalic /d/ - a non-standard, stigmatized variant in most varieties of Spanish - as a covert prestige marker, aligning themselves more with monolingual Spanish varieties and distancing themselves from Portuguese. Spirantization (in comparison with full occlusion) might also have positive connotations in Rivera, representing an urban orientation and social status. In Montevideo, spirantization is simply the norm, and likely elicits few judgments. However, speakers in a community can also reinterpret incoming variants in different ways, disagree about what pronunciations are appropriate in careful speech, and thus use variants differently in style shifting. Conflicting attitudes have been found in Rivera: some residents believe that Riverenses should speak the "national" variety of Spanish as spoken in Montevideo, while others value Border Spanish (Asencio 2006) (see Section 2.3).

The dimensions of language experience and attitudes are more complex in Rivera, and these speakers also have a larger overall range of possibilities. Given these factors, it is not surprising that Rivera shows higher group-level variability than Montevideo.

At the individual level, the cities only differ in variability in word list speech. For that reason, I save the discussion of this effect for the next section.

5.2 Variability by Style

5.2.1 Group-level variability by Style

At the group level, it is surprising that careful (word list) speech is more variable than spontaneous speech (interview). We might expect speakers in a community to move in the same direction when style shifting, converging on tacitly agreed-upon norms. Indeed, some definitions of speech communities highlight the importance of shared norms, and the resulting similar behavior in style-shifting among members (Labov 1972: 120–121). I suggest that heightened variability in word lists might come from how speakers manage conflicting standards in careful speech. Careful speech might be the style that most strongly elicits between-speaker differences, because it forces speakers to manage conflicting standards in a speech situation devoid of social context.

Realizations of intervocalic /bdg/ are subject to conflicting standards, stereotypes, and norms, especially in Rivera. Recall from Section 5.1 that the range of possible intervocalic /bdg/ variants is narrower for Montevideo speakers than for Rivera speakers, whose ranges extend to fully occlusive realizations. Spirantization is the norm across the Spanish-speaking world;

occlusion does not occur even in careful speech (Carrasco, Hualde and Simonet 2012: 150; Martínez-Gil 2019: 46–47). My results indicate that Montevideo Spanish realizes intervocalic /bdg/ as weakly constricted, mirroring nearby Argentinian Spanish (Colantoni and Marinescu 2010). While previous research on Montevideo Spanish is scarce, it is likely that deletion is stigmatized, as in other parts of the Spanish-speaking world. In Rivera, the variants of intervocalic /bdg/ have connotations that differ from those in Montevideo. More specifically, spirantization in Rivera likely carries positive associations because it is linked to the prestigious, urban Montevideo variety (as Carvalho 2006b proposes for coda /s/ aspiration).

Rivera speakers may be faced with a dilemma like the following when asked to do a word reading task. Deletion and full occlusion, which are both stigmatized on several axes, are acceptable in casual contexts. Spirantized variants, which are the pan-Hispanic and Uruguayan norm, are expected in careful, Spanish-dominant contexts. But the distinction between some variants is fuzzy. While the goal of careful speech may be spirantization - produced with more constriction than the spirantization of spontaneous speech - full occlusion goes too far, crossing the border into a stigmatized variant. Some variants also have both negative and positive social connotations. For example, fully occlusive variants can be evaluated negatively as Portuguese influences, but also as positive markers of border identity (Waltermire 2006).

Careful speech may result in more group-level variability because it draws speakers' attention to these conflicting standards, and requires them to reconcile the standards in the absence of social context. In my study, speakers may have interpreted the word list reading task differently, unsure about what kind of speech is appropriate. This could lead to high, unsystematic, variability between speakers. In my study, word lists elicited more variable data, not uniformly more "standard" forms. In this case, it is not even clear what the goal is.

While heightened variability in word list tasks seems unusual, the results are compatible with several previous studies. For variables with categorically-defined allophones, word lists can contain high rates of what are usually considered stigmatized forms.[12] In these cases, the authors argue that this occurs due to competing standards that come to the forefront in careful speech and/or reading tasks. For example, Gafter (2016) finds stigmatized forms of Hebrew pharyngeals in word lists, and argues that speakers use them as identity markers in a task they interpret as performative reading. Similarly, Hadodo (2021) finds lateral velarization and postalveolar affricates in Istanbul Greek word lists, even though these are non-standard features compared to Modern Standard Greek. He proposes that speakers use word lists to showcase local identity. Speakers evaluate their own dialect positively, as a competing standard with Modern Standard Greek, and using features of their dialect is thus appropriate in careful speech. In Glaswegian English, Stuart-Smith (1999: 200) finds that, while one variable (the HOUSE vowel) showed the expected shift towards fewer vernacular forms in word list speech, T-glottaling did not. She suggests that the word list exerted some pressure to style shift, but "in such a way that T-glottalling is no longer excluded as stigmatized." Her results are compatible with (Trudgill 1988: 44–45), who found high rates of T-glottalling in careful speech in Norwich. In more recent work, Stuart-Smith et al. (2013: 513–514) also find that two variables undergoing rapid change in Glasgow (TH-fronting, L-vocalization) are more frequent in word list speech than in spontaneous speech. Based on these results, they conclude that word lists do not encourage speech "in the direction of the regional standard, but quite the opposite", for reasons of "style or stance-taking".

[12] These findings differ from the current study, where speakers produced more *variability* in word lists, but are relevant because production of non-standard forms alongside standard ones can result in high variability.

In addition to considering why variability might be higher in careful speech for some variables, it is worth considering why variability might be *lower* in spontaneous speech. Interviews give speakers a more realistic social context in which to structure variability, possibly lessening the pressure of explicit standards. Speakers could be more internally consistent, and more similar to each other when they are not explicitly focused on controlling details of phonetic production and managing conflicting standards. Although speech from a sociolinguistic interview is not truly unmonitored, this proposal is consistent with Labov, Baugh and Sherzer's (1984) claim that less monitored speech is more systematic.

5.2.2 Individual-level variability by style

At the individual level, style predicts variability in Rivera (not in Montevideo). I believe this is due to a combination of factors already discussed: Rivera speakers have a wide range of available variants, these variants are subject to conflicting standards, and word lists exert sociolinguistic pressure in a way that causes speakers to use their full ranges. In addition to contributing to between-speaker differences, these factors could lead to within-speaker inconsistency.

In addition to the studies discussed in the previous section, which report non-standard forms in careful speech at the group level, other studies focus on how high variability arises within individuals. Individuals can show high variability when non-standard pronunciations reflect phonemic mergers, because they do not know which words fall into which category. In West Yorkshire English, Petyt (1977, 1980) found that a non-standard pronunciation of /u/ in FOOT words was common in word lists, and hypothesized that this was not a "goal-directed shift" (Janda and Auger 1992: 205). Rather, Petyt proposed that speakers "gave up the attempt… - and so just guessed randomly - " more when they were required to "concentrate specifically on this problem" (Petyt 1980: 167, cited in Janda and Auger 1992: 206). Although variants of Spanish /bdg/ are not involved in mergers, my Uruguayan speakers' shifts in the word list along the spirantization continuum may have been more or less random. In my word list reading task, speakers' attention was drawn explicitly to pronunciation. The pronunciation options were sociolinguistically meaningful in conflicting ways, and speakers lacked sociolinguistic context in which to structure variation. My speakers may have been unsure which pronunciations were appropriate for the task and shifted along their whole ranges.

Heightened variability in word lists could also be related to individual multilingualism. Much previous work suggests that multilingual speakers have "merged" or "compromise" categories, which are acoustically intermediate between their languages (e.g. Flege 1991; Flege and Eefting 1988; Sundara, Polka and Baum 2006). In a model like the Speech Learning Model (Flege 1995), an L2 sound that is perceived as similar to a sound in the L1 is linked to the L1 sound category. Because the sounds are perceived as similar, they can form a single category, with intermediate phonetic properties. However, Casillas (2021) concludes from a meta-analysis of existing studies that the evidence for "compromise" categories is weak. Based on his own experimental work, he argues instead that bilinguals maintain separate categories and may produce "target mismatches" when both languages are activated. That is, English-Spanish bilinguals sometimes produce English-like stops when speaking Spanish, and Spanish-like stops when speaking English. Bosch and Ramon-Casas (2011) also find that early bilingual Spanish-Catalan speakers make production errors with mid vowels /e, o/ with /ɛ, ɔ/, a distinction that exists only in Catalan. This result suggests target mismatches as well, although possibly for a different reason. Bosch and Ramon-Casas (2011) suggest that these bilinguals' categories might permit wide variability, or that these bilinguals may not know which Catalan vowel belongs in which word due to extensive exposure to mispronunciations by native Spanish speakers.

In Rivera, target mismatches could lead to high variability: most productions would be spirantized, but category mismatches would result in some fully occlusive /bdg/ productions (as is typical of Brazilian Portuguese) while speaking Spanish. Further investigation would be needed to determine whether my speakers' productions were spread evenly throughout the constriction degree range, or whether their constriction degrees are strongly bimodal, as would be expected if there were target mismatches in production.[13]

Another factor that could lead to heightened variability in word lists at the individual level is increased activation of multiple acoustic targets in this speech style, as compared to spontaneous speech. This has been found in several non-native and L2 contexts. Elliott and Hall-Lew (2015) analyze English vowel production among Slovak and Czech immigrants in Edinburgh. Their plots suggest that non-native English speakers have higher variability in word list speech than in reading and interview speech. Although they do not explicitly discuss variability, they hypothesize that the mix of Edinburgh and more standard variants may occur because formal speech styles trigger L2 instruction norms. Rogers, Gordon and Pizarro (2012) do look explicitly at variability in vowel production among non-native speakers and report similar results: non-native speakers' variability *increases* in careful speech. While my Rivera speakers are not referring to instructed L2 norms (Portuguese is learned in the community), they are also managing a variety of conflicting standards, expectations, and social connotations that could be more active in careful speech.

Finally, the range of realizations that Rivera speakers use in word lists may be available to them due to high input variability that is not present in Montevideo. Previous work finds that input variability affects category "looseness" and the amount of variability that individuals produce. Speakers in multilingual communities are exposed to multiple languages, some of which may realize the same phonological categories differently. These listeners are thus exposed to different phonetic cues for the same category, as well as to cues produced by speakers with different proficiency and amounts of phonetic transfer between languages. As already mentioned, Bosch and Ramon-Casas (2011) find that early Spanish-Catalan bilinguals make more mistakes than early Catalan monolinguals in producing and perceiving the Catalan mid-vowel height contrast (/e, o/ vs. /ɛ, ɔ/), which Spanish lacks. They suggest that the similarity between Catalan and Spanish, combined with variable input from non-native Catalan speakers, may lead children to form less precise vowel categories that permit more variability. This variability is then reproduced in speakers' own output. Similarly, Sundara et al. (2006) suggest that the high input variability that French-English bilingual children receive may make stabilization of the contrastive /d/-/ð/ categories in English more difficult. Similar effects have been found when learners are exposed to different varieties of a single language: Levy and Hanulíková (2019) find higher production variability in children with more exposure to regional varieties of German and foreign-accented German.

[13] A preliminary investigation of by-speaker distributions of Intensity Difference measurements does show some bimodality, but it is weak and does not suggest that category mismatches explain the results. Most Rivera speakers have a small cluster of measurements around a low Intensity Difference, and another small cluster a bit higher. However, these peaks are small and many data points fall in between. Instead of a prominent second peak at a high Intensity Difference that would indicate category mismatches (full occlusion), the productions are gradient and increasingly fewer at higher constriction degrees. This suggests that variability is not just due to some fully occlusive productions, but rather to a category that can be produced with the full range of constriction. Addressing this question would be best done with a combination of gradient and categorical analyses of spirantized vs. occluded variants.

The link between input and output variability can be modeled in exemplar-based frameworks (e.g. Bybee 2001). In exemplar models, listeners store tokens of words they hear, complete with social and stylistic information, and draw on this information in their own production. Based on experimental work, Clopper (2014: 80, 84) offers an explicit prediction: high variability input (in her study, through exposure to multiple dialects) leads to more variable representations. Speakers with more input variability have a wider range of stored exemplars, and thus produce heightened variability too. Applied to my data, this prediction holds for Rivera and Montevideo. Rivera residents receive highly variable input (/bdg/ as deleted, spirantized, and fully occluded) from an early age, from speakers producing varieties of Spanish, Portuguese, and Portuguese-accented Spanish. This exposure could lead to wider categories that permit more variability, in both production and perception. Furthermore, this seems to occur even though not all residents use all varieties, suggesting that variability may arise from exposure to languages or varieties speakers do not use themselves, in addition to coming from language contact within a speaker (as in Johnson and Babel 2021).

In contrast to Rivera speakers, Montevideo speakers show no reliable effect of speech style on variability. These speakers have a smaller overall range of acceptable productions (deletion to light constriction), the productions are phonetically more similar, and the options are less mired in conflicting sociolinguistic connotations. Even under the same recording circumstances, Montevideo speakers may produce less variability during sudden changes in interpreting the task goal or moments of linguistic insecurity, simply because the options are fewer.

5.3 Are all variables subject to the style effect?

Recall that existing acoustic work on vowels shows *less* variability in careful speech. This raises the question of what kinds of variables - consonantal vs. vocalic - are subject to heightened variability in careful speech, and what properties might lead to this effect.

Consonants and vowels are subject to different constraints. In careful speech, vowels are produced with less variability and in an expanded vowel space, increasing the acoustic and perceptual distance between vowels (Section 2.1). Some consonantal variables are like vowels, in that categories differ mainly in one acoustic cue, which constrains variability in that cue's production. Hauser (2019) finds that acoustic variability correlates inversely with cue weight, such that the more crucial a cue is for distinguishing categories, the less variably speakers produce it. Spanish spirantization is not subject to these constraints. Variable constriction degree does not threaten contrast with another phonological category because there are other cues to the voicing distinction in Spanish stops (Hualde, Simonet & Nadeu 2011). Contrast maintenance thus does not motivate precision around a target degree of constriction. In the absence of these constraints, consonantal variables may be freer than vowels to allow variability in careful speech without endangering a contrast.

Variables that show heightened variability in careful speech also might be subject to explicit social and prescriptive pressure, and/or be undergoing change. A comparison of my data with an acoustic study by Hall-Lew and Boyd (2017) is illustrative. While they do not address variability specifically, Hall-Lew and Boyd (2017) plot center of gravity (COG) for /s/ tokens produced by young, female speakers of American English. These speakers appear to show higher variability in spontaneous speech than in read speech, although the difference is small (Hall-Lew and Boyd 2017: 90). This finding contrasts with my finding that word lists have higher variability than interviews, but it aligns with my finding that style does not predict variability for Montevideo speakers. The COG of /s/ among young female American English speakers and /bdg/ spirantization in Montevideo share two properties: they are not subject to explicit prescriptive or social pressure, and they are not undergoing change. Heightened variability in word lists - a style that most explicitly calls upon standards and stereotypes - may emerge only for variables that *have* these standards and stereotypes. Furthermore, both Montevideo speakers and Hall-Lew and Boyd's speakers come from largely monolingual communities.

Finally, American English /s/ is subject to contrast maintenance constraints, in a way Spanish /bdg/ is not. /s/ differs from another sound in the inventory, /ʃ/, largely by COG, so variability in COG production in both sounds might be constrained by the presence of the other. The pressure for contrast maintenance may be strongest in careful speech, resulting in lower COG variability.

5.4 Other possible explanations for the style effect

There are two other possible explanations for heightened variability in my word list data that could undermine the finding. Neither one appears to account for the effect.

First, the word lists and interviews contain different word types.[14] My word lists contain mostly nouns, while the interviews contain a wider variety of word types. If the diversity of word type affected variability, it is not clear which direction the effect should go. Interviews could produce higher variability because they contain a wider variety of word types that have different reduction patterns, or word lists could produce higher variability because the nouns have more stable realizations, but these realizations differ greatly word-to-word. However, a difference in word types between styles does not fully explain heightened variability in word lists. Specifically, it does not explain why the style effect holds at the individual level in Rivera but not in Montevideo. Data was collected with the same word list and interview technique in both cities, and the subset of words taken from the interviews should be similar. If the only explanation for heightened variability in word lists were a difference in word type between interviews and word lists, the style effect should have held in both cities.

A second possible explanation for the heightened variability in word list speech could be that the phonological contexts in word list items were more diverse than those for words taken from the interviews. Because phonetic details depend on the surrounding sounds, this could have led to more phonetic variability in word lists. However, my interviews and word lists both have a wide range of phonological contexts; if anything, interviews have a larger variety of phonological contexts (see Figure 9 in Appendix B).

5.5 Exploration of Speaker Age and variability

While Age was included in the variation analysis, it was not included in the variability analysis because there were too few data points. The variation analysis showed that older Rivera speakers produce tokens with a higher constriction degree than younger speakers. For variability, Figure 8 shows a positive correlation between individual variance for Rivera speakers and Age: older speakers have higher variance.

[14] As a reviewer points out, this issue affects all studies that compare spontaneous and controlled (read) speech.

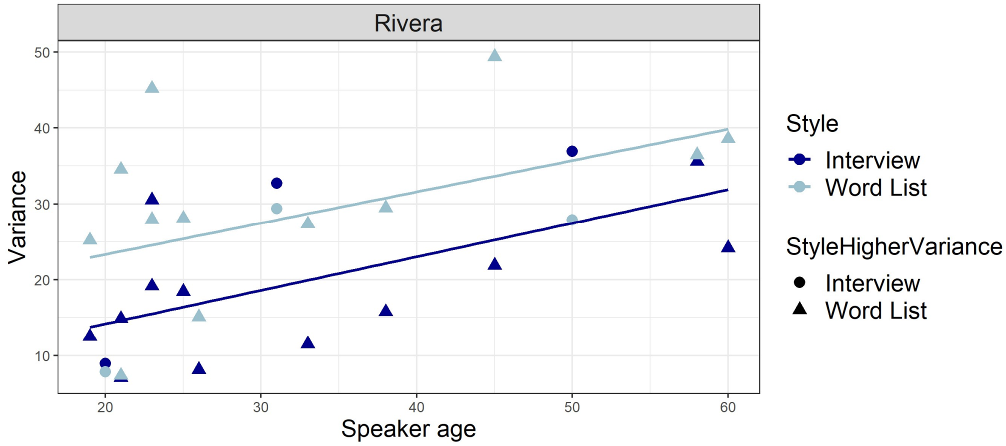

Figure 8 Rivera: Individual variance by Speaker Age in Rivera

Older speakers produce tokens with higher Intensity Differences (more constriction) than younger speakers, *and* they produce more variability than younger speakers. The heightened variability is likely for the same reason that their overall Intensity Differences are higher. Both the variation and variability results can be accounted for in an exemplar-based model. Older speakers grew up in an era with less contact with Montevideo, more contact with Portuguese, and more negative attitudes towards Border Spanish. These speakers have more stored tokens with high constriction degrees, and thus produce /bdg/ as more constricted. They also have variants stored that are influenced both by Montevideo norms and by Portuguese, leading to higher production variability, and may have more exposure to conflicting standards and connotations with the variants. Because the speakers are spread thinly throughout the age range, these results should be taken as suggestive.

6 Conclusion

In this paper, I have examined the effect of multilingualism and speech style on variation and within-category acoustic variability. Using data from Montevideo and Rivera, Uruguay, I examined the constriction degree of Spanish intervocalic /bdg/. The analysis of variation showed that constriction degree is socially and stylistically conditioned in both cities. The variability analysis showed that, at the group level, there is more within-category acoustic variability in careful speech than in spontaneous speech, and in Rivera than in Montevideo. At the level of the individual, the word list style predictably boosted variability for Rivera speakers, suggesting that heightened variability in word lists at the group level is due (at least partly) to variability *within* speakers. Speech style showed no predictable effect for Montevideo speakers: the heightened variability in word lists at the group level derives from differences *between* speakers. That individual Montevideo and Rivera speakers do not differ in variability in spontaneous speech suggests that, when sociolinguistic pressure is lower, speakers may settle around a single production mode rather than producing forms along their entire available continuum.

This study contributes to current work on variability by examining the effect of social factors. In particular, the effect of speech style raises methodological concerns with careful speech tasks often used in laboratory phonology work. Controlled speech tasks (e.g. word list reading) are generally expected to elicit less variable speech. In contrast to this expectation, my results suggest that careful speech can sometimes elicit high variability, with standard forms occurring alongside less-standard forms. Understanding that these tasks may elicit highly variable speech will require a more nuanced interpretation of the results. Determining which types of variables are subject to high variability in careful speech is a task for future work. The current study suggests that these variables may be those undergoing change, those subject to prescriptive standards, or those subject to strong sociolinguistic attitudes.

The findings also raise interesting questions about the connection between input variability, category formation, and task effects. Analyzing variability can shed light on individual speakers' linguistic behavior and experience, even when mean measurements are the same. One important limitation is my relatively small sample size. Due to the nature of variability, it will be important to examine larger groups of speakers and more individuals, to determine the extent to which my results hold, and the extent to which they generalize to other communities and variables.

Acknowledgments: Thanks to my participants in Uruguay who made this work possible. This research was done while I was a graduate student at NYU. I am grateful for insightful comments from Laurel MacKenzie, Greg Guy, Lisa Davidson, and Maria Gouskova. Thanks also to audiences at NYU SocioLab, LabPhon 2020 and HLS 2021 for helpful feedback, and to the journal reviewers and editor, whose feedback substantially improved the quality of this paper. This work has benefited from a government grant managed by the ANR under the "Investissements d'Avenir" program with the reference ANR-10-LABX-0083. It contributes to the IdEx University of Paris - ANR-18-IDEX-0001 (LABEX-EFL)

Research funding: The fieldwork for data collection was funded by a Tinker Field Research Grant, through the Center for Latin American and Caribbean Studies at NYU.

7 Appendix A

Table 10 Full model structures (before reduction)

Montevideo	Rivera
BDG Intensity Difference ~ 1 + Style + Sex + Stressed + Past.Part + Log Freq + Segment + Adjacent vowel + Age + Sex*Style + Age*Style + Style*Stressed + (1 | Word) + (1 | speaker)	BDG Intensity Difference ~ 1 + Style + Sex + Stressed + Past.Part + Log Freq + Segment + Adjacent vowel + Age + CognateStatus + Prop.Spanish + Sex*Style + Age*Style + Style*Stressed + (1 | Word) + (1 | speaker)

<u>Montevideo: Model after reduction</u>

BDG Intensity Difference ~ 1 + Style + Sex + Stressed + Log frequency + Segment + Age +
Adjacent vowel +
Style*Sex + Style*Age + Style*Stressed +
(1 | Word) + (1 | Speaker)

Table 11 Model results for Montevideo

	Est.	SE	df	t-val	Pr(>|t|)	Baseline	*n*	IntDiff (dB)
(Intercept)	1.66	0.9	22.61	1.84	0.08.		1405	4.67
Predictors								
Style						(Interview)	840	3.31
Word list	0.73	0.67	923.3	1.09	0.28		565	6.69
Sex						(Female)	819	4.48
Male	-0.6	0.62	14.6	-0.98	0.34		586	4.93
Onset to stressed syllable						(No)	880	3.80
Yes	2.41	0.4	506.8	6.03	<0.001***		525	6.13
Log Freq.	-0.62	0.19	306.19	-3.36	<0.001***	(Continuous)		
Segment						(/b/)	384	4.45
Segment /d/	0.1	0.42	412.28	0.23	0.82		463	4.38
Segment /g/	1.28	0.47	309.95	2.74	0.01**		383	5.82
Segment <v>	0.28	0.54	419.55	0.51	0.61		175	3.41
Age	0.06	0.02	14.59	3.09	0.01**	(Continuous)		
Adjacent vowel						(/a/)	439	4.85
Vowel /e/	-0.58	0.48	311.43	-1.2	0.23		269	5.46
Vowel /i/	-1.45	0.42	329.21	-3.43	<0.001***		409	3.68
Vowel /o/	0.52	0.73	326.04	0.71	0.48		101	4.27
Vowel /u/	-0.09	0.56	273.68	-0.16	0.87		187	5.48

Interactions							
Word list:Male	2.98	0.42	1331.79	7.14	<0.001***		
Word list:Age	0.06	0.01	1340.42	4.95	<0.001***		
Word list:Stressed	-2.2	0.71	393.95	-3.11	0.002**		
Random intercepts: Word (SD = 1.87), Speaker (SD = .93)							

In Montevideo, there are significant main effects of Stress, Frequency, Segment, Age, and Adjacent Vowel. Intensity differences (a proxy for constriction degree) are higher when /bdg/ is the onset to a stressed syllable (vs. the onset to an unstressed syllable), lower in words of higher frequency, higher for /g/ (vs. /b/), higher for older speakers (vs. younger speakers), and lower when the vowel used to calculate the comparison is /i/ (vs. /a/).

There are also significant interactions between Style*Sex, Style*Age, and Style*Stress. Word lists boost Intensity Differences more for males than females, more for older speakers than for younger speakers, and less for stressed vowels than for unstressed vowels. In interviews, there is more of an effect of stress, while in word lists, stressed and unstressed /bdg/ onsets are produced similarly.

Rivera: Model after reduction

BDG Intensity Difference ~ 1 + Style + Stressed + Freq + Segment + Age + Adjacent vowel +
Style*Age + Style*Stress +
(1 | Word) + (1 | Speaker)

Table 12 Model results for Rivera

	Est.	SE	df	t-val	Pr(>|t|)	Baseline	*n*	IntDiff (dB)
(Intercept)	1.24	1.16	20.29	1.08	0.29			6.52
Predictors								
Style						(Interview)	1050	4.43
Word list	1.16	0.7	857.03	1.66	0.1.		706	9.65
Onset to stressed syllable						(No)	1040	5.58
Yes	2.73	0.37	642.78	7.45	<0.001***		716	7.56
Log frequency	-0.36	0.19	379.61	-1.9	0.06.	(Continuous)		
Age	0.1	0.03	16.61	3.37	0.004**	(Continuous)		
Adjacent vowel						(/a/)	611	6.81
Vowel /e/	-1.48	0.46	377.92	-3.25	0.001**		376	7.12
Vowel /i/	-1.89	0.42	432.06	-4.56	<0.001***		476	5.53
Vowel /o/	0.55	0.71	390.43	0.78	0.43		129	5.90
Vowel /u/	-0.36	0.64	302.37	-0.57	0.57		164	7.48
Interactions								
Word list:Age	0.16	0.02	1670.31	10.43	<0.001***			

Word list:Stressed	-2.22	0.68	522.37	-3.26	0.001**	
Random intercepts: Word (SD = 2.10), Speaker (SD = 1.53)						

The results for Rivera are similar, but not identical, to those for Montevideo. The significant main effects are Stress, Age, Adjacent Vowel. These effects are in the same direction as in Montevideo: /bdg/ are more constricted as the onset to a stressed syllable, more constricted in productions by older speakers, and lower when the vowel used to calculate the comparison is /e/ or /i/) (vs. /a/). Frequency is marginally significant, in the same direction as Montevideo: higher frequency words are produced with less constriction.

There are significant interactions between Style*Age and Style*Stress, in the same directions as Montevideo. Word lists boost constriction degree more for older speakers than younger speakers, and less for stressed vowels than unstressed vowels.

Brief discussion:

These findings are largely in line with previous findings for spirantization throughout the Spanish-speaking world. Previous work across dialects of Spanish also finds higher degrees of constriction in stressed syllables (Ortega-Llebaria 2004; Waltermire 2006; Eddington 2011; Carrasco, Hualde & Simonet 2012) and lower constriction (more reduction) in high-frequency words (e.g. Bybee 2002; Waltermire 2006). That the effect of segment is significant in Montevideo but falls out of the Rivera model is somewhat unexpected, but segments have been found to behave differently in regional varieties of Spanish (Carrasco, Hualde & Simonet 2012). The age effect is also consistent with expectations, since young speakers have been found to reduce more than older speakers (Cedergren 1979; Rogers 2016; Butera 2018).

The interaction between Stress and Style is intriguing. This effect can be phrased two ways: (1) the effect of stress is weaker in word list speech than in interview speech; (2) the effect of style is stronger for unstressed onsets than for stressed onsets. This may be because word list speech is already so careful that all tokens have higher constriction, regardless of stress, while in interviews, there is more room for differentiating stressed from unstressed onsets. Similarly, unstressed onsets have lighter constriction than stressed onsets, so there may be more room for a distinction based on style for unstressed onsets. It is not clear whether this interaction is an artifact of the current study and a small sample size, or a more general finding that holds in other data sets as well.

8 Appendix B

Diversity of phonological context

One potential reason that variability could be higher in word lists than in interviews is that word lists might contain more different phonological contexts than interviews. This could undermine the finding that variability is higher in word lists than in interviews. However, that is not the case for my data. Figure 9 shows which VCV sequences occur in word lists and interviews. Most sequences occur in both. If anything, interviews contain more different phonological contexts than word lists, and could be expected to show higher variability because of this.

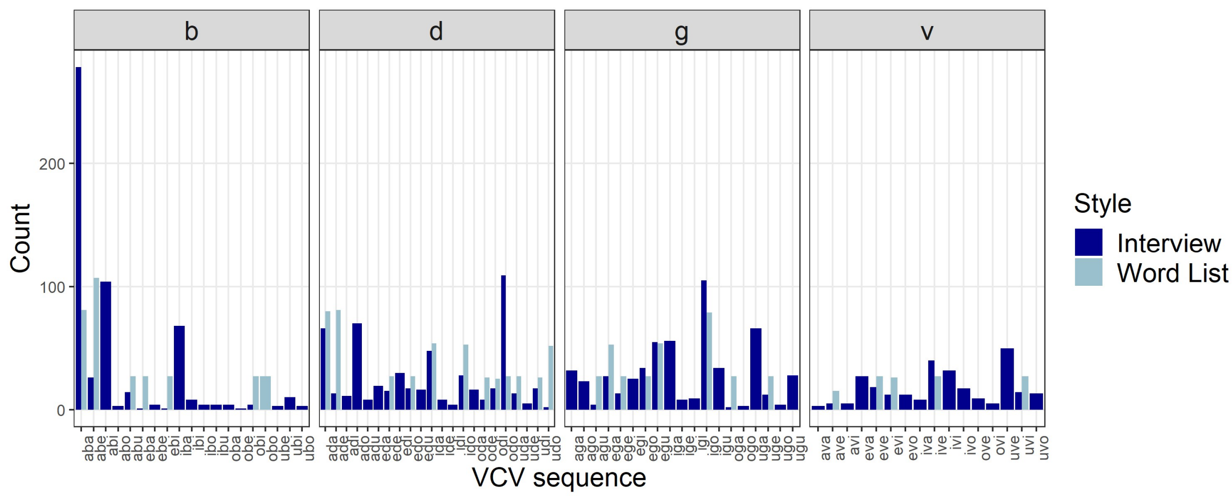

Figure 9 Phonological contexts in word lists and interviews

9 Appendix C

Word list

abeja
abrigo
abuela
además
agujero
alfabeto
ayuda
boludo
cabeza
cadera
calabacín
calabaza
castigo
ciego
codo
dedo
delgado
delicadeza
despegue
enemigo
entrada
entrevista
estudiarlo
frazada
hogar
jarabe
juguete
llave
lluvia
mirada
nido
obispo
rebajas
regalo
revés
robo
rodilla
seguro
testarudo
trabajo
vida
yegua

References

Alvord, Scott M. & Brandon Rogers. 2014. Miami-Cuban Spanish vowels in contact. *Sociolinguistic Studies* 8(1). 139–170. https://doi.org/10.1558/sols.v8i1.139.

Amengual, Mark. 2012. Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition* 15(3). 517–530. https://doi.org/10.1017/S1366728911000460.

Amengual, Mark. 2013. *An experimental approach to phonetic transfer in production and perception of early Spanish-Catalan bilinguals*. University of Texas PhD Dissertation.

Asencio, Pilar. 2006. Representaciones de las lenguas que deberían hablarse en Rivera: el descentramiento de una comunidad. In Graciela Barrios & Luis Ernesto Behares (eds.), *Políticas e identidades lingüísticas en el Cono Sur*, 61–69. Montevideo, Uruguay: Universidad de la República Facultad de Humanidades y Ciencias.

Barrios, Graciela, Beatriz Gabbiani, Luis Behares, Adolfo Elizaincín & Susana Mazzolini. 2005. Planificación y políticas lingüísticas en Uruguay. *Iztapalapa* 29(13). 177–190.

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Behares, Luis. 2007. Portugués del Uruguay y educación fronteriza. In Claudia Brovetto, Nicolás Brian & Javier Geymonat (eds.), *Portugués del Uruguay y educación bilingüe*, 99–171. Montevideo, Uruguay: Administración Nacional de Educación Pública (ANEP).

Behares, Luis & Adolfo Elizaincín. 1983. Noticias sobre la fonología del español del Uruguay. In Katuska Jabbiff de Yaffé & Nilia Queirolo (eds.), *Aproximación al estudio del lenguaje y su patología*, 29–35. Montevideo, Uruguay: Parentesis.

Behares, Luis Ernesto. 1984. Diglosia en la sociedad escolar de la frontera uruguaya con Brasil. *Cuadernos de Estudios Lingüísticos* 6. 228–234.

Birdsong, David, Libby M. Gertken & Mark Amengual. 2012. Bilingual Language Profile: An easy-to-use instrument to assess bilingualism. Center for Open Educational Resources and Language Learning (COERLL). https://sites.la.utexas.edu/bilingual/.

Boersma, Paul & David Weenink. 2018. Praat: Doing phonetics by computer.

Bosch, Laura & Marta Ramon-Casas. 2011. Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics* 39. 514–526. https://doi.org/10.1016/j.wocn.2011.02.001.

Brown, Earl K., Michael S. Gradoville & Richard J. File-Muriel. 2014. The variable effect of form and lemma frequencies on phonetic variation: Evidence from /s/ realization in two varieties of Colombian Spanish. *Corpus Linguistics and Linguistic Theory* 10(2). 213–241. https://doi.org/10.1515/cllt-2013-0025.

Bullock, Barbara & Almeida Jacqueline Toribio. 2009. Trying to hit a moving target: On the sociophonetics of code-switching. In Ludmila Isurin, Donald Winford & Kees deBot (eds.), *Studies in Bilingualism*, vol. 41, 189–206. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/sibil.41.12bul.

Bullock, Barbara, Almeida Jacqueline Toribio, Amanda Dalola & Veronica Gonzalez. 2006. Language dominance and performance outcomes in bilingual pronunciation. In Mary Grantham O'Brien, Christine Shea & John Archibald (eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition (GASLA): The Banff Conference*. Somerville, MD: Cascadilla Press.

Butera, Brianna J. 2018. *A lenition continuum: Acoustic variability of Spanish stop consonants*. Madison, WI: University of Wisconsin-Madison PhD Dissertation.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511612886.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14(3). 261–290. https://doi.org/10.1017/S0954394502143018.

Carrasco, Patricio, José Ignacio Hualde & Miquel Simonet. 2012. Dialectal differences in Spanish voiced obstruent allophony: Costa Rican vs. Iberian Spanish. *Phonetica* 69. 149–179. https://doi.org/10.1159/000345199.

Carrasco-Ortiz, Haydee, Mark Amengual & Stefan Th. Gries. 2021. Cross-language effects of phonological and orthographic similarity in cognate word recognition: The role of language dominance. *Linguistic Approaches to Bilingualism* 11(3). 389–417. https://doi.org/10.1075/lab.18095.car.

Carvalho, Ana Maria. 1998. *The social distribution of Uruguayan Portuguese in a bilingual border town*. Berkeley: University Of California, Berkeley PhD Dissertation.

Carvalho, Ana Maria. 2003. Variation and diffusion of Uruguayan Portuguese in a bilingual border town. In *Comunidades e indivíduos bilingües. Actas do I Simpósio Internacional sobre o Bilinguismo*, 642–651.

Carvalho, Ana Maria. 2006a. Nominal number marking in a variety of Spanish in contact with Portuguese. In *Selected Proceedings of the 8th Hispanic Linguistics Symposium*, 154–166. Somervillle, MA: Cascadilla Proceedings Project.

Carvalho, Ana Maria. 2006b. Spanish (s) aspiration as a prestige marker on the Uruguayan-Brazilian border. *Spanish in Context* 3(1). 85–114. https://doi.org/10.1075/sic.3.1.07car.

Carvalho, Ana Maria. 2010. Eres de la frontera o sos de la capital? Variation and alternation of second-person verbal forms in Uruguayan border Spanish. *Southwest Journal of Linguistics* 29(1). 1–24.

Carvalho, Ana Maria. 2014. Sociolinguistic continuities in language contact situations: The case of Portuguese in contact with Spanish along the Uruguayan-Brazilian Border. In Patricia Amaral & Ana Maria Carvalho (eds.), *Portuguese-Spanish interfaces: Diachrony, synchrony, and contact* (Issues in Hispanic and Lusophone Linguistics), vol. 1, 263–294. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Carvalho, Ana Maria. 2016. The analysis of languages in contact: A case study through a variationist lens. *Cadernos de Estudos Linguísticos, Campinas* 58(3). 401–424. https://doi.org/10.20396/cel.v58i3.8647467.

Carvalho, Ana Maria & Ryan Bessett. 2015. Subject pronoun expression in Spanish in contact with Portuguese. In Ana Maria Carvalho, Rafael Orozco & Naomi Lapidus Shin (eds.), *Subject pronoun expression in Spanish*, 143–165. Washington: Georgetown University Press.

Carvalho, Ana Maria & Joseph Kern. 2019. The permeability of tag questions in a language contact situation: The case of Spanish-Portuguese bilinguals. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)* 29(4). 463–492. https://doi.org/10.1075/prag.18068.car.

Casillas, Joseph V. 2021. Interlingual interactions elicit performance mismatches not "compromise" categories in early bilinguals: Evidence from meta-analysis and coronal stops. *Languages* 6(1). 9. https://doi.org/10.3390/languages6010009.

Cedergren, Henrietta. 1979. La elisión de la /d/: Un ensayo de comparación dialectal. *Boletín de la Academia Puertorriqueña de la Lengua Española* VII. 19–29.

Church, Meredith M. 2007. *Portuñol and border identity in Rivera, Uruguay: Reconciling identities and claiming space in the national imaginary*. University of Kansas MA Thesis.

Clopper, Cynthia G. 2014. Sound change in the individual: Effects of exposure on cross-dialect speech processing. *Laboratory Phonology* 5(1). https://doi.org/10.1515/lp-2014-0004.

Colantoni, Laura & Irina Marinescu. 2010. The scope of stop weakening in Argentine Spanish. In Marta Ortega-Llebaria (ed.), *Selected Proceedings of the 4th Conference of Laboratory Approaches to Spanish Phonology*, 100–114. Somerville, MA: Cascadilla Proceedings Project.

Cole, Jennifer, José Ignacio Hualde & Khalil Iskarous. 1999. Effects of prosodic and segmental context on /g/-lenition in Spanish. In Osamu Fujimura, Brian D. Joseph & Bohumil Palek (eds.), *Proceedings of the 4th International Linguistics and Phonetics Conference*, vol. 2, 575–589.

Costa, Albert, Mikel Santesteban & Agnès Caño. 2005. On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language* 94(1). 94–103. https://doi.org/10.1016/j.bandl.2004.12.002.

DiCanio, Christian, Hosung Nam, Jonathan D. Amith, Rey Castillo Garcia & Doug H. Whalen. 2015. Vowel variability in elicited versus spontaneous speech: Evidence from Mixtec. *Journal of Phonetics* 48. 45–59. https://doi.org/10.1016/j.wocn.2014.10.003.

Duchon, Andrew, Manuel Perea, Nuria Sebastián-Gallés, Antonia Martí & Manuel Carreiras. 2013. EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods* 45. 1246–1258. https://doi.org/10.3758/s13428-013-0326-1.

Eddington, David. 2011. What are the contextual phonetic variants of /bdg/ in colloquial Spanish? *Probus* 23. 1–19.

Elizaincín, Adolfo. 1973. Algunos aspectos de la sociolingüística del dialecto fronterizo. *Temas de Lingüística*, 3. Universidad de la República, Facultad de Humanidades y Ciencias, Departamento de Lingüística.

Elizaincín, Adolfo. 1975. El bilingüismo de la frontera uruguayo-brasileña. *Letras de Hoje* 20. 65–75.

Elizaincín, Adolfo. 1978. Bilingüismo y problemas educativos en la zona fronteriza uruguayo-brasileña. In *Lingüística y educación: Actas del IV congreso internacional de la Asociación de Lingüística y Filología de la América Latina (ALFAL)*, 301–310. Spain: Universidad de San Marcos.

Elizaincín, Adolfo. 1992. *Dialectos en contacto: Español y Portugués en España y América*. Montevideo, Uruguay: Arca Editorial.

Elizaincín, Adolfo, Luis Ernesto Behares & Graciela Barrios. 1987. *Nos falemo brasilero: Dialectos portugueses en Uruguay*. Montevideo, Uruguay: Editorial Amesur.

Elliott, Zuzana & Lauren Hall-Lew. 2015. Production of FACE and GOAT by Slovak and Czech immigrants in Edinburgh. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. London: International Phonetic Association.

Engelhardt, Julie. 2017. *Social predictors of intervocalic /b/ variant usage in Riverense Spanish*. Tempe, Arizona: Arizona State University BA Thesis.

Engelhardt, Julie, Michael Gradoville & Mark Waltermire. 2018. Intervocalic /b/ variation in Border Uruguayan Spanish. Presented at the Hispanic Linguistics Symposium, University of Texas at Austin.

Fairbanks, Grant. 1950. A physiological correlative of vowel intensity. *Speech Monographs* 17(4). 390–395.

Fairbanks, Grant, Arthur S. House & Eugene L. Stevens. 1950. An experimental study of vowel intensities. *The Journal of the Acoustical Society of America* 22(4). 457–459.

Ferguson, Sarah Hargus & Diane Kewley-Port. 2007. Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research* 50. 1241–1255. https://doi.org/10.1044/1092-4388(2007/087).

Ferreira, Letania & Eric Holt. 2014. On the partially divergent phonology of Spanish, Portuguese and points in between. In Patricia Amaral & Ana Maria Carvalho (eds.), *Portuguese-Spanish interfaces: Diachrony, synchrony, and contact*, vol. 1, 123–150. Philadelphia: John Benjamins Publishing Company.

Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 229–273. Timonium, MD: York Press.

Flege, James Emil. 1991. Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America* 89(1). 395–411. https://doi.org/10.1121/1.400473.

Flege, James Emil & Wieke Eefting. 1988. Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America* 83(2). 729–740. https://doi.org/10.1121/1.396115.

Gafter, Roey. 2016. What's a stigmatized variant doing in the word list? Authenticity in reading styles and Hebrew pharyngeals. *Journal of Sociolinguistics* 20(1). 31–58. https://doi.org/10.1111/josl.12169.

García de los Santos, Elizabeth. 2014. Actitudes lingüísticas en Uruguay. Tensiones entre la variedad y la identidad [Linguistic attitudes in Uruguay. Tensions between variety and identity]. *Bergen Language and Linguistics Studies* 5. 1346–1405. https://doi.org/10.15845/bells.v5i0.695.

Gradoville, Michael, Mark Waltermire & Avizia Long. 2021. Cognate similarity and intervocalic /d/ production in Riverense Spanish. *International Journal of Bilingualism* 25(3). 727–746. https://doi.org/10.1177/1367006921996807.

Hadodo, Matthew. 2021. Double standards? What wordlists tell us about dialectal features of Istanbul Greek. Presented at the New Ways of Analyzing Variation 49, University of Texas at Austin (virtual).

Hall-Lew, Lauren & Zac Boyd. 2017. Phonetic variation and self-recorded data. *Selected Papers from New Ways of Analyzing Variation (NWAV) 45* 23(2). 86–95.

Harmegnies, Bernard & Dolors Poch-Olivé. 1992. A study of style-induced vowel variability: Laboratory versus spontaneous speech in Spanish. *Speech Communication* 11. 429–437.

Harris, James. 1969. *Spanish Phonology*. Cambridge, MA: MIT Press.

Hauser, Ivy. 2018. Within-category variation in production of Hindi and English stops. Poster presented at the Acoustical Society of America, Minneapolis, MN.

Hauser, Ivy. 2019. *Effects of phonological contrast on within-category phonetic variation*. Amherst, MA: University of Massachusetts Amherst PhD Dissertation. https://doi.org/10.7275/14836073.

Hauser, Ivy. 2020. Intraspeaker variation and cue weight in Mandarin sibilants. Presented at the Linguistic Society of America Annual Meeting, University of Minnesota.

Hauser, Ivy. 2021. Contrast implementation affects phonetic variability: A case study of Hindi and English stops. *Laboratory Phonology* 12(1). https://doi.org/10.16995/labphon.6465.

Hensey, Frederick Gerald. 1972. *The sociolinguistics of the Brazilian-Uruguayan border*. The Hague: Mouton.

Hensey, Fritz. 1979. Fronterizo: A case of phonological restructuring. In Jacob Ornstein (ed.), *Three essays on linguistic diversity in the Spanish-speaking world*, 47–59. Paris: Mouton.

Hualde, José Ignacio. 2014. *Los sonidos del español*. Cambridge: Cambridge University Press.

Hualde, José Ignacio, Ryan Shosted & Daniel Scarpace. 2011. Acoustics and articulation of Spanish /d/ spirantization. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 906–909. Hong Kong.

Hualde, José Ignacio, Miquel Simonet & Marianna Nadeu. 2011. Consonant lenition and phonological recategorization. *Laboratory Phonology* 2(2). 301–329. https://doi.org/10.1515/LABPHON.2011.011.

Janda, Richard & Julie Auger. 1992. Quantitative evidence, qualitative hypercorrection, sociolinguistic variables: and French speakers' eadhaches with English h/0. *Language and Communication* 12(3/4). 195–236.

Johnson, Khia A. & Molly Babel. 2021. *Language contact within the speaker: Phonetic variation and crosslinguistic influence*. Preprint. Open Science Framework. https://doi.org/10.31219/osf.io/jhsfc.

Jongman, Allard & Travis Wade. 2007. Acoustic variability and perceptual learning. In Ocke-Schwen Bohn & Murray J. Munro (eds.), *Language experience in second language speech learning: In honor of Jamis Emil Flege*, 135–150. Amsterdam: John Benjamins Publishing Company.

Kirchner, Robert. 2004. Consonant lenition. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 313–345. Oxford: Oxford University Press.

Koffi, Ettien. 2020. A comprehensive review of intensity and its linguistic applications. *Linguistic Portfolios* 9. Article 2.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26. https://doi.org/10.18637/jss.v082.i13.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William, John Baugh & J. Sherzer. 1984. Field methods of the project on linguistic change and variation. In *Language in Use*, 28–53. Englewood Cliffs: Prentice-Hall, Inc.

Labov, Willian. 1966. *The social stratification of English in New York City*. Washington D.C.: Center for Applied Linguistics.

Lehiste, Ilse & Gordon E. Peterson. 1959. Vowel amplitude and phonemic stress in American English. *The Journal of the Acoustical Society of America* 31(4). 428–435.

Levy, Helena & Adriana Hanulíková. 2019. Variation in children's vowel production: Effects of language exposure and lexical frequency. *Laboratory Phonology* 10(1). 1–26. https://doi.org/10.5334/labphon.131.

Lewis, Anthony. 2001. *Weakening of intervocalic /p, t, k/ in two Spanish dialects: Toward the quantification of lenition processes*. Urbana-Champaign: University of Illinois at Urbana-Champaign PhD Dissertation.

Lindblom, Björn. 1986. Phonetic universals in vowel systems. In John J. Ohala & Jeri J. Jaeger (eds.), *Experimental phonology*. Orlando, FL: Academic Press.

Lipski, John. 2006. Too close for comfort? The genesis of Portuñol/Portunhol. In Timothy L. Face & Carol A. Klee (eds.), *Selected Proceedings of the 8th Hispanic Linguistics Symposium*, 1–22. Somerville, MA: Cascadilla Proceedings Project.

Long, Avizia & Lisa Baldwin. 2013. A sociolinguistic analysis of intervocalic /b/ in Caracas Spanish. *Indiana University Linguistics Club Working Papers (IULCWP)* 13(1). 1–20.

Martínez-Celdrán, Eugenio & Xosé Luís Regueira. 2008. Spirant approximants in Galician. *Journal of the International Phonetic Association* 38(1). 51–68. https://doi.org/10.1017/S0025100308003265.

Martínez-Gil, Fernando. 2019. Spirantization and the phonology of Spanish voiced obstruents. In Sonia Colina & Fernando Martínez-Gil (eds.), *The Routledge handbook of Spanish phonology* (Routledge Spanish Language Handbooks), 34–83. London and New York: Taylor and Francis Group.

McGill, William J. & James P. Goldberg. 1968. Pure-tone intensity discrimination and energy detection. *The Journal of the Acoustical Society of America* 44(2). 576–581. https://doi.org/10.1121/1.1911123.

Michnowicz, Jim. 2009. Intervocalic voiced stops in Yucatan Spanish: A case of contact-induced language change? In Manel Lacorte & Jennifer Leeman (eds.), *Español en Estados Unidos y otros contextos de contacto: Sociolingüística, ideología y pedagogía*, 67–84. Madrid, Spain: Vervuert/Iberoamericana.

Michnowicz, Jim & Lindsey Carpenter. 2013. Voiceless stop aspiration in Yucatan Spanish: A sociolinguistic analysis. *Spanish in Context* 10(3). 410–437. https://doi.org/10.1075/sic.10.3.05mic.

Navarro Tomás, Thomás. 1953. *Manual de pronunciación española*. New York: Hafner Publishing Company, Inc.

Navarro Tomás, Tomás. 1918. Diferencias de duración entre las consonantes españolas. *Revista de Filología Española* 5. 367–393.

Ortega-Llebaria, Marta. 2004. Interplay between phonetic and inventory constraints in the degree of spirantization of voiced stops: Comparing intervocalic /b/ and intervocalic /g/ in Spanish and English. In Timothy L. Face (ed.), *Laboratory Approaches to Spanish Phonology* (Phonology and Phonetics 7), 237–253. Berlin/New York: Mouton de Gruyter.

Pérez, Hernán Emilio. 2007. Estudio de la variación estilística de la serie /bdg/ en posición intervocálica en el habla de los noticieros de la televisión chilena. *Estudios de fonética experimental* 16. 228–259.

Petyt, Keith Malcom. 1977. *Dialect and accent in the industrial West Riding*. Reading, UK: University of Reading PhD Dissertation.

Petyt, Keith Malcom. 1980. *The study of dialect: An introduction to dialectology*. Boulder: Westview Press.

Quilis, Antonio. 1981. *Fonética acústica de la lengua español*. Madrid: Editorial Gredos.

Quilis, Antonio. 1999. *Tratado de fonologia y fonética españolas*. 2nd edn. Madrid: Editorial Gredos.

Rogers, Brandon. 2016. The influence of linguistic and social variables in the spirantization of intervocalic /bdg/ in Concepción, Chile. *Studies in Hispanic and Lusophone Linguistics* 9(1). 207–237. https://doi.org/10.1515/shll-2016-0008.

Rogers, Catherine L., Amber Gordon & Melitza Pizarro. 2012. A moving target? Comparing within-talker variability in vowel production between native and non-native English speakers across two speech styles. *The Journal of the Acoustical Society of America* 132(3). https://doi.org/10.1121/1.4755114.

Rona, José Pedro. 1965. *El dialecto fronterizo del norte del Uruguay*. Montevideo, Uruguay: Universidad de la República.

Sharma, Devyani. 2011. Style repertoire and social change in British Asian English. *Journal of Sociolinguistics* 15(4). 464–492. https://doi.org/10.1111/j.1467-9841.2011.00503.x.

Sharma, Devyani. 2021. Biographical indexicality: Personal history as a frame of reference for social meaning in variation. In Lauren Hall-Lew, Emma Moore & Robert J. Podesva (eds.), *Social meaning and linguistic variation: Theorizing the third wave*, 243–264. 1st edn. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108578684.011.

Sharma, Devyani & Ben Rampton. 2015. Lectal focusing in interaction: A new methodology for the study of style variation. *Journal of English Linguistics* 43(1). 3–35. https://doi.org/10.1177/0075424214552131.

Simonet, Miquel, José Ignacio Hualde & Marianna Nadeu. 2012. Lenition of /d/ in spontaneous Spanish and Catalan. *INTERSPEECH 2012, 13th Annual Conference of the International Speech Communication Association*, 1416–1419. Portland, Oregon.

Smiljanić, Rajka & Ann R. Bradlow. 2005. Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America* 118(3). 1677–1688. https://doi.org/10.1121/1.2000788.

Soler, Antonia & Joaquín Romero. 1999. The role of duration in stop lenition in Spanish. (Ed.) John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey. *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)* 483–486. London: International Phonetic Association.

Stuart-Smith, Jane. 1999. Glottals past and present: A study of t-glottalling in Glaswegian. *Leeds Studies in English* 30. 181–204.

Stuart-Smith, Jane, Gwilym Pryce, Claire Timmins & Barrie Gunter. 2013. Television can also be a factor in language change: Evidence from an urban dialect. *Language* 89(3). 501–536. https://doi.org/10.1353/lan.2013.0041.

Sundara, Megha, Linda Polka & Shari Baum. 2006. Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition* 9(1). 97–114. https://doi.org/10.1017/S1366728905002403.

Sundara, Megha, Linda Polka & Fred Genesee. 2006. Language experience facilitates discrimination of /d-ð/ in monolingual and bilingual acquisition of English. *Cognition* 100. 369–388.

Tagliamonte, Sali. 2006. *Analysing sociolinguistic variation*. Cambridge/New York: Cambridge University Press.

Tamminga, Meredith, Laurel MacKenzie & David Embick. 2016. Dynamics of variation in individuals. *Linguistic Variation* 16(2). 300–336. https://doi.org/10.1075/lv.16.2.06tam.

Tang, Kevin. 2012. A 61 million word corpus of Brazilian Portuguese film subtitles as a resource for linguistic research. (Ed.) Zoe Belk, Harris Constantinou & Maria Varkanitsa. *University College London Working Papers in Linguistics* 24. 208–214.

Trudgill, Peter. 1988. Norwich revisited: Recent linguistic changes in an English urban dialect. *English Worldwide* 9(1). 33–49.

Vaughn, Charlotte, Melissa Baese-Berk & Kaori Idemaru. 2019. Re-examining phonetic variability in native and non-native speech. *Phonetica* 76(5). 327–358. https://doi.org/10.1159/000487269.

Wade, Travis, Allard Jongman & Joan Sereno. 2007. Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica* 64(2–3). 122–144. https://doi.org/10.1159/000107913.

Waltermire, Mark. 2006. *Social and linguistic correlates of Spanish-Portuguese bilingualism on the Uruguayan-Brazilian border*. Albuquerque, New Mexico: University of New Mexico PhD Dissertation.

Waltermire, Mark. 2008. Social stratification of language-specific variants of intervocalic /d/ along the Uruguayan-Brazilian border. *Sociolinguistic Studies* 2(1). 31–60.

Waltermire, Mark. 2010. Variants of intervocalic /d/ as markers of sociolinguistic identity among Spanish-Portuguese bilinguals. *Spanish in Context* 7(2). 279–304. https://doi.org/10.1075/sic.7.2.06wal.

Waltermire, Mark. 2011. Frequency effects on the morphological conditioning of syllable-final /s/ reduction in Border Uruguayan Spanish. *Journal of Language Contact* 4(1). 26–55. https://doi.org/10.1163/187740911X558824.

Waltermire, Mark. 2014. Language use and attitudes as stimuli for phonological change in Border Uruguayan Spanish. In Dominic Watt & Carmen Llamas (eds.), *Language, borders and identity*, 70–89. Edinburgh, UK: Edinburgh University Press. https://doi.org/10.1515/9780748669783-009.

Waltermire, Mark & Michael Gradoville. 2020. The interaction of social factors in the acoustically gradient realization of intervocalic /d/ in Border Uruguayan Spanish. In Rajiv Rao (ed.), *Spanish phonetics and phonology in contact: Studies from Africa, the Americas, and Spain* (Issues in Hispanic and Lusophone Linguistics), vol. 28, 263–292. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/ihll.28.10wal.

Warner, Natasha & Benjamin V. Tucker. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America* 130(3). 1606–1617. https://doi.org/10.1121/1.3621306.