

Sociophonetic Variation and Change in Heritage Languages: Lexical Effects in Heritage Italian Aspiration of Voiceless Stops

Language and Speech

1–25

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/00238309221126483

journals.sagepub.com/home/las**Chiara Celata** 

Dipartimento di Studi Umanistici, Università degli studi di Urbino “Carlo Bo,” Italy

Naomi Nagy

Department of Linguistics, University of Toronto, Canada

Abstract

In a previous study on voiceless stop aspiration in Heritage Calabrian Italian spoken in Toronto, we found that the transmission of a sociophonetic variable differed from cross-generational phonetic variation induced by increased contact with the majority language. Universal phonetic factors and the social characteristics of the speakers appeared to influence contact-induced variation much more straightforwardly than the transmission of the sociophonetic variable. In the current study, we investigate further, examining possible alternative explanations related to the lexical distribution of the aspiration phenomena. We test two alternative hypotheses, the first one predicting that the diffusion of a majority language’s phonetic feature is frequency-driven while change in a sociophonetic feature is not (or not that regularly across generations), and the second one predicting that sociophonetic aspiration decreases across generations by being progressively more dependent on the frequency of lexical items. Our results show that sociophonetic aspiration resists lexicalization and applies to both frequent and infrequent words even in the speech of third-generation speakers. By contrast, the progressive introduction of contact-induced phonetic change is led by high-frequency words. These findings add to the complexity of heritage language phonology by suggesting that the pronunciation features of a heritage language can follow different fates depending on their sociolinguistic roles.

Keywords

Sociophonetic variation, heritage languages, sound change, lexical diffusion, VOT

Corresponding author:

Naomi Nagy, Department of Linguistics, University of Toronto, 100 St. George St., Toronto, ON M5S 3G3, Canada.

Email: naomi.nagy@utoronto.ca

Introduction

The pronunciation patterns of heritage languages spoken in ethnic minority settings have been shown to differ from those of both monolinguals and bilinguals of different sociolinguistic settings (e.g., Amengual, 2016; Au et al., 2002; Chang et al., 2011; Chang & Yao, 2016). Heritage languages challenge our traditional notions of (non)nateness in language acquisition and use (Cheng et al., 2021) and require a gradient approach to bilingualism (Ortega, 2020). This is even truer when a comparison across generations of heritage speakers is involved. In the framework of heritage language studies, sociolinguistic variation in phonetics and phonology is particularly under-investigated (Celata, 2019; Rao, 2016). While current developmental approaches to sociolinguistic variation mostly have first-language acquisition and full parent-to-child transmission mechanisms in their focus (cf. Chevrot & Foulkes, 2013; Labov, 2014), it is crucial to find out whether and how sociophonetic variation characterizing the homeland variety is transmitted across generations in the heritage context. The issue has descriptive as well as theoretical relevance. Socio-indexical language features are by definition rooted in the social dynamics of a given speech community, so the question arises of whether such features change when the speakers move to a different speech community, particularly when their native language then becomes a minority language. Moreover, the mechanisms by which socioindexical variation is potentially transmitted to later generations of heritage speakers are unknown. That is, assuming that socioindexically relevant features are preserved in the speech of first-generation immigrants, we should also investigate whether and how they are passed on to their children and grandchildren.

In a previous study on Heritage Calabrian Italian spoken in Toronto, we found that the transmission of a sociophonetic variable (one that is socially indexed) differed from phonetic variation that could be attributed to contact with the majority language (Nodari et al., 2019; see also Nagy & Kochetov, 2013). We compared the production of voiceless stops in unstressed versus stressed syllable conditions in the heritage language of first-, second-, and third-generation (Gen 1, 2, and 3) Calabrian Italian speakers to determine whether aspiration in unstressed syllables, a typical sociophonetic feature of Calabrian Italian speech, decreases.

The first Italians came to Canada in the late-19th century, but the bulk of migration happened in the mid-20th century. The speakers in our study are (descendants of) this group. The Gen 1 speakers examined here were born in Calabria between 1935 and 1959 and came to Canada to escape miserable social and economic conditions in southern Italy (Di Salvo, 2017). In the interviews analyzed here, the speakers use a regional variety of Italian influenced by centuries of contact with Calabrese (a Romance language with an origin independent of Florentine-based “standard” Italian; Coseriu, 1980). All regional varieties of Italian across Italy are influenced by their respective local Romance language and constitute current spoken Italian (Crocco, 2017).

Italian is a language with a true voicing distinction, in contrast to English, which uses aspiration (spread glottis) to distinguish two series of consonants. However, the variety of Italian considered here, Calabrian regional Italian, does use aspiration socioindexically, but in different contexts from where it appears in English; this is detailed below.

Voiceless stop aspiration is a socioindexical feature of Calabrese which has percolated into Calabrian Italian; it indexes in particular low socioeconomic status, low mobility, strong connection to the local ways of life, and male speech (Falcone, 1976; Nodari, 2017, 2022). Importantly, the context for this aspiration is mostly limited to onset stops in unstressed syllables which are preceded by a coda sonorant, as in *tanto* “much” [ˈtant^ho] (or are the second half of a stop geminate, as in *tutto* “all” [ˈtut^hːo], a context not investigated in Nodari et al., 2019); we term this context “unstressed C.CV,” or “C.CV” for short. According to Soriano (1996) and Nodari (2017), voice onset time (VOT) values range between 24 and 48ms, with variation influenced by position

(stressed vs. unstressed syllables), local variety, and the place of articulation of the consonant. Given its correlation to the social characteristics of the speakers in Calabria, we use the label of sociophonetic aspiration to refer to C.CV aspiration.

In contrast, aspiration in stressed syllable onsets is typical of English (“stressed C \acute{V} ,” or “C \acute{V} ” for short). It includes onset stops in intervocalic or post-pausal position, as in the first segment of *tutto* “all” [‘tut^h:o]). As far as we know from the available literature, Canadian English does not differ from other accents of English in this respect. For instance, Caramazza et al. (1973) report an average VOT of 62–90 ms (depending on the place of articulation) in monolingual speakers and of 39–67 ms for French-English bilinguals; according to Fowler et al. (2008), values are 60–78 ms for monolinguals and 53–70 ms for French-English bilinguals. Sociophonetic variation in VOT production has been reported for selected varieties of English (e.g., Stuart Smith et al., 2015) but not for Canadian English.

Thus, for the Calabrian Italian speakers of Toronto, both the heritage and the host language possess long-lag VOT in their phonetic repertoire, but they differ in the phonological distribution of the feature since long-lag VOT is preferentially associated with pre-stress, word-initial stops in English and to post-stress, post-sonorant or geminate stops in Calabrian Italian. The other fundamental difference between the two languages is that voiceless stop aspiration is a sociophonetic variable in Calabrian Italian, indexing information about speaker characteristics and sociocultural features, while it has not been shown to have any such role in Toronto English. Aspiration of stressed syllables in heritage Calabrian Italian speech may certainly index the effects of language contact and the fact of being bilingual; however, there is no socioindexical value attributed to voiceless stop aspiration in English. There is no evidence of C \acute{V} aspiration indexing social/ethnic features in Italian in Calabria, in contrast to the C.CV context.

In Nodari et al. (2019) we anticipated that aspiration in C \acute{V} environments would increase across generations as a consequence of increased contact with the majority language. A summary of the significant effects of that study, and a sample of words in which the variable occurs, is given in Table 1. On average, VOT was longer in unstressed C.CV than in stressed C \acute{V} syllables (41 vs. 26 ms) and stops were also perceptually judged as aspirated by Italian judges more frequently in the unstressed syllables (57% of the word tokens) than in the stressed ones (18%). More importantly, mixed model analyses showed that the VOT of voiceless stops decreased across generations in the unstressed (sociolinguistically indexed) condition, whereas it did not vary significantly in the stressed (putative English-influence) condition. We suggested that at this particular level of phonetic implementation, sociophonetic marking decreased in successive generations, which can be interpreted as a consequence of changes in the social and linguistic conditions across generations of heritage speakers. In contrast, the non-socially-indexed VOT of plosives in stressed syllables did not increase. Moreover, although we found the canonical /k/ > /t/ > /p/ hierarchy of VOT duration in both contexts of aspiration, the pattern was more consistent in English-like aspiration than in sociophonetic aspiration; in the latter case, Gen 1 and 3 speakers produced a longer VOT for /k/ and similarly shorter ones for both /t/ and /p/. The sex of the speaker was a significant predictor of VOT in English-like stressed syllable aspiration (with males producing longer VOTs than females) but not in Calabrian-like unstressed syllable aspiration.

In the same study, aspiration was also perceptually identified by native Italian (non-Calabrian) listeners. As a result of those perceptual ratings, each syllable was coded as bearing “perceived aspiration” or not. Evidence of cross-generational change in the rate of perceived aspiration was found in the frequency of both types of aspiration, but again with differences: there was a clear and gradual cross-generational change in the case of the sociophonetic aspiration, whereas in English-like aspiration only Gen 3 speakers were found to aspirate more frequently than the two other groups, though nowhere near as frequently as would be found in English. At the perceptual level, aspiration was perceived more frequently for /t/ than for any other consonant in the case of

Table 1. Summary of Effects From Nodari et al. (2019).

	Generational effect on rate of perceived aspiration	Generational effect on VOT	Correlation with ethnic orientation
Stressed C [́] V (English induced aspiration, e.g., <i>la c[́]arta</i> “the paper”)	Increase from Gen2 to Gen 3	Stable ^a	Negative correlation of perceived aspiration with language use and cultural choices
Unstressed C.CV (Calabrian sociophonetic aspiration, e.g., <i>la c[́]arta</i>)	Decrease from Gen 2 to Gen 3	Decrease from Gen 1 to Gen 3 ^b	No correlation

Note. VOT=voice onset time.

^aNo main effect of generation; however, significant interaction with place of articulation (Gen 1 speakers opposed a longer VOT in /k/ to a shorter VOT on /t/ and /p/, Gen 2 and Gen 3 speakers produced a three-way distinction).

^bMain effect of generation. In addition, significant interaction with place of articulation (Gen 1 and Gen 3 speakers opposed a longer VOT in /k/ to a shorter VOT on /t/ and /p/, Gen 2 speakers produced a three-way distinction).

sociophonetic aspiration. The presence of a high vowel after the plosive was a significant predictor of (perceived) aspiration rate in English-like aspiration but not in Calabrian-like aspiration.

Finally, by correlating individual ethnic orientation scores (such as frequency of Italian language use and cultural preferences; Nagy, 2009, 2011) with rates of perceived aspiration, we assessed that an increase in ethnic orientation toward the Italian language and culture was significantly correlated to a decrease in contact-induced English-like aspiration. However, the ethnic orientation of the speakers in no way predicted individual rates of Calabrian sociophonetic aspiration. Note that the effects of ethnic orientation also clearly relate to generation (since immigration) and thus to input varieties.

Taken together, these findings suggested that in the speech of bilingual heritage speakers, the cross-generational transmission of sociophonetically relevant features differs from the transmission of phonetic features that are not markers of sociolinguistic identity in the homeland variety (see summary in Table 1). In particular, Calabrian aspiration was progressively and gradually changing in the speech of successive generations of heritage language speakers, and these changes affected the phonetic substance of plosive production (VOT) in an audible way. By contrast, English-like aspiration substantially began with the Gen 3 speakers and its presence was not sufficient to influence the phonetic substance of plosives (at least, concerning VOT) at the group level. The phonetic factors of vowel height and consonant place explained much of the variation in the English aspiration context. In the language contact situation described here, the introduction of the perceived-aspiration patterns of English into the heritage language was significantly predicted by the social characteristics of the speakers (sex, generation, and ethnic orientation), whereas the maintenance of the heritage language aspiration patterns was not.

In sum, sociophonetic aspiration did appear to decrease cross-generationally, but this change was not influenced by external and internal variables in a straightforward way. External variables such as sociolinguistic orientation or sex did not explain this variation, and the language-internal phonetic effects of vowel height and consonant place of articulation had mixed and relatively weak effects. So the question of how sociophonetic variation is transmitted across generations remains open.

In this paper, we investigate further, examining alternative explanations related to the lexical distribution of the aspiration phenomena.

It is well known that the lexical level is a crucial component of many historical sound changes (cf. Labov, 1994; Wang, 1969). For instance, we have known for some time that sounds and words that are more frequent in the language may behave differently from less frequent sounds and words in sound change, sometimes by leading the spread of an innovation and sometimes by being more resistant to it (cf. Bybee, 2001; Hay et al., 2015; Phillips, 1984; Pierrehumbert, 2001). In regular (non-analogical) sound changes, the rate of the diffusion of a pronunciation feature across lexical items can be used as a measure of how far an innovation is from generalizing to the whole system. When frequent words lead in a sound change (particularly in the case of lenition), the effect of word frequency is small at the earliest stages of the change and increases as the change progresses more quickly (Hay & Foulkes, 2016). Phonetically gradual sound changes that involve complex structural changes (Phillips, 2006) or are driven by perceptual, rather than articulatory, forces may also show an advantage to low-frequency words over high-frequency ones (Ogura, 2012).

For heritage phonology, studies that have considered lexical predictors are extremely limited in number and scope. For instance, it has been shown that the amount of code-switching to the majority language can predict the rate of phonetic convergence between the heritage and the majority language (cf. Łyskawa et al., 2016). Another group of studies has shown that speech production by fluent bilinguals is affected by words' cognate status (cf. Amengual, 2012, 2016; Flege & Munro, 1994) with more acoustic interference in cognates than in non-cognates; however, it is still unknown whether and how such interactive lexicon–phonetics relationships occur in the phonological competence of heritage language speakers (Celata, 2019), and particularly whether lexical frequency has a role in shaping cross-generational changes. These questions are especially relevant for our understanding of heritage language phonology if we consider that heritage speakers may be exposed to (and be users of) an impoverished heritage language lexicon and a limited variety of morpho-syntactic structures and pragmatic-communicative situations in the heritage language (cf. Jarvis, 2019; Schmid & Jarvis, 2014).

In the specific case of voiceless stop aspiration in heritage Calabrian Italian, we might further speculate that the two changes observed in Nodari et al. (2019), that is, increasing (perceived) English-like aspiration in stressed syllables and decreasing sociophonetic aspiration in unstressed syllables, could differ at the lexical level and could be differently affected by lexical frequency. In particular, we could hypothesize that English-like aspiration, showing a more predictable influence of language-internal and language-external variables (Nodari et al., 2019), is also subject to typical frequency effects, that is, it has a direct correlation between lexical frequency and rate of aspiration. According to this hypothesis, the diffusion of a majority language's phonetic feature is frequency-driven, while change in a sociophonetic feature is not (or not as much). The effects of lexical frequency in English-like aspiration should also be regular across generations, according to this hypothesis. However, since Calabrian sociophonetic aspiration was observed to more gradually decrease across generations (whereas for English-like aspiration we found evidence that it more abruptly increased in Gen 3), it could be hypothesized that Calabrian sociophonetic aspiration was also reducing its lexical coverage regularly across generations. In other words, the reduction in rate is directly tied to a reduction in the number of lemmas to which it applies, rather than a rate strictly tied to a phonological rule.

Finally, it is possible that the two changes do not differ from one another for their lexical characteristics. However, given the substantial differences reviewed above, the third hypothesis is the least probable.

2 Aims and hypotheses

In order to determine whether one or both types of aspiration in heritage Calabrian Italian are influenced by lexical frequency, we ask:

1. Does the number of lemmas affected by perceived voiceless-stop aspiration (as a proportion of the total number of lemmas) differ in the two subsets (i.e., C \acute{V} and C.CV contexts)?
2. Does the number of lemmas affected by perceived voiceless-stop aspiration (as a proportion of the total number of lemmas) change across generations?
3. Are the cross-generational changes in perceived aspiration and VOT that were found in our previous study affected by lexical frequency?

Regarding the first question, we expected that the proportion of aspirated lemmas in the C \acute{V} or English-like aspiration subset would be smaller than in the C.CV or sociophonetic aspiration subset, given that the proportion of aspirated tokens was smaller in stressed syllables compared with unstressed ones, according to Nodari et al. (2019).

Regarding the second question, we expected the proportion of aspirated lemmas in the C \acute{V} or English-like aspiration subset to increase across generations, as a consequence of increasing interference from the majority language across generations.

Regarding the third question, we expected a significant role of lexical frequency in predicting changes in either one or both dependent variables in the C \acute{V} or English-like aspiration subset (the higher the lexical frequency, the longer the VOT and/or the higher the rate of perceived aspiration), according to the hypothesis put forth above that the diffusion of a majority language's phonetic feature is frequency-driven. Concerning the C.CV or Calabrian sociophonetic aspiration subset, if aspiration decreases across generations according to the same lexical diffusion pattern as for English-like aspiration, the proportion of aspirated lemmas is expected to regularly decrease across generations as well. In addition, lexical frequency is expected to interact with cross-generational effects in VOT and/or perceived aspiration, with higher frequency lemmas better resisting the loss of aspiration. If, in contrast, sociophonetic aspiration in our heritage Calabrian speakers is not lexically bounded and its cross-generational change is independent of the characteristics of the lexicon, we expect a less marked (or more irregular) decrease of aspirated lemmas across generations and no role of lexical frequency in predicting VOT and/or perceived aspiration rate. In other words, if the first hypothesis is correct, sociophonetic aspiration will show up as an increasingly lexicalized process, whereas if the second hypothesis is correct, sociophonetic aspiration will change across generations independent of the characteristics of the lexicon.

3 Methodology

3.1 Speakers from the HLVC corpus

Data for this study come from sociolinguistic interviews (Labov, 1984) collected by the Heritage Language Variation and Change in Toronto Project (HLVC, Nagy, 2009, 2011). Participants and interviewers are heritage speakers of Calabrian Italian. Interviews are transcribed and time-aligned to the audio recordings. The HLVC corpus contains data for three generations of heritage speakers. Gen 1 speakers were born in the homeland (Calabria) and moved to the Greater Toronto Area (hereafter, Toronto) after age 18. They had lived in Toronto at least 20 years at the time of recording. Gen 2 speakers were born in Toronto (or came there before age 6) and their parents qualify as

Table 2. Speaker Sample for This Study ($n=23$).

Generation	Females	Males	Total	
Gen 1	I1F59A	I1M60A	6	
	I1F65A	I1M61A		
Gen 2		I1M62A		
	I2F32A	I1M75A		
	I2F44A	I2M14A		9
	I2F45A	I2M19A		
	I2F53A	I2M42A		
I2F57A	I2M53A			
Gen 3	I3F21A	I3M18B	8	
	I3F21B	I3M22B		
	I3F23A	I3M27A		
	I3F33A	I3M28A		
Total	11	12	23	

Note. For details on speaker codes, see Nagy (2009, https://ngn.artsci.utoronto.ca/HLVC/I_4_corpus.php#speakercodes).

Gen 1 speakers (but are not necessarily in the corpus). Finally, Gen 3 speakers were born in Toronto and their parents qualify as Gen 2 speakers (again, the parents are not necessarily in the corpus). The sample examined in this paper consists of excerpts from 23 speakers, distributed by generation and sex as shown in Table 2. Additional methodological details are provided in Nodari et al. (2019), from which this brief description is adapted.

3.2 The dataset

The data came from the first HLVC task, the sociolinguistic interview, and consisted of word tokens annotated according to the procedure in Nodari et al. (2019). In this paper, we use a slightly expanded dataset ($N=3,408$, or 3,324 after excluding borrowings). In this newer dataset, unlike in Nodari et al. (2019), C: V (i.e., geminates) and C.CV tokens were considered together. Each syllable containing one of the three stops /p t k/ in onset position was categorized as either the potential target of Calabrian-like aspiration (post-sonorant or geminate stop in unstressed syllable) or of English-like aspiration (intervocalic singletons in stressed syllables). We will use the short labels “C.CV” and “C \acute{V} ” contexts (or subsets), respectively, for Calabrian sociophonetic aspiration and English-induced aspiration.

Tokens in all other contexts, such as post-obstruent onsets in unstressed syllables (e.g., *popcorn*) were excluded. All words which were borrowed from English (whether nonce borrowings like *basimento* “basement,” or established loans, like *tennis*, *college* or proper nouns like *Canada*) were excluded from the dataset before modeling the data. It is not always possible to determine whether their use constitutes a code-switch to English.

Each syllable was then acoustically and auditorily annotated. The acoustic annotation identified the boundaries of the stop closure, the release, and the following vowel to measure VOT. Segmentations and annotations were performed manually in PRAAT 6.0.36 (Boersma & Weenink, 2015). Closure duration was defined as the interval between the offset of F2 energy of the preceding vowel or sonorant, and the beginning of the stop release. VOT was defined as the duration from the onset of the stop burst to the first zero-crossing of the first (quasi)periodic wave of

the following vowel (Abramson & Whalen, 2017). A PRAAT script automatically extracted the duration values for the three relevant intervals. In addition to raw values, normalized VOT duration was calculated: VOT was normalized by dividing its duration by the duration of the following vowel.

The auditory coding classified each syllable as either perceptually aspirated or non-aspirated. Two native speakers of a Tuscan variety of Italian performed the annotation. Their Tuscan Italian origin means that the annotators had both a precise understanding of the speech input and a familiarity with non-phonemic voiceless stop aspiration (which is present in, e.g., *gorgia toscana* or Tuscan Italian gemination, Stevens & Hajek, 2010). At the same time, their not being of Calabrian origin means that they lack expectations about the dialectal distribution of aspiration and implicit knowledge of the socioindexical values associated with it. The annotation was performed independently by each annotator; inconsistencies (7%) were discussed and re-annotated; 135 tokens (2.7% of the dataset) were removed from the dataset because of a persistent lack of inter-annotator agreement.

Nodari et al. (2019) provide details on the procedures adopted for segmentation, annotation, and categorization. For the current analysis, it is useful to recall some quantitative aspects that characterize the two subsets of aspiration contexts.

The size of the two subsets we examine in this paper was slightly different: the stressed CV subset was bigger than the unstressed C.CV one (1,942 vs. 1,382 occurrences, respectively). This difference is due to the conversational nature of the speech samples (no *a priori* balancing of contexts is possible). Therefore, we consider percentages as a normalized comparison, calculating the percent of tokens that were perceived as aspirated in each context. Recall that VOT was on average longer in C.CV than in CV (41 vs. 26 ms) and perceived aspiration was more frequent in the C.CV subset (57% of the word tokens) than in the CV subset (18%). There were also differences between generations: in the speech of Gen 1 speakers, aspiration was perceived in 87% of the C.CV word forms and 5% of the CV word forms; for Gen 2 speakers, the percentages were 49% and 13%, and for Gen 3 speakers the percentages were 32% and 37%.

3.3 Lexical frequency measures

Words occur in a language corpus according to a systematic frequency distribution such that there are a few very high-frequency words that account for most of the tokens (the “head” of the distribution) and many low-frequency words (the “tail” of the distribution; Zipf, 1936). Such distribution is captured by a simple power law equation and is observed universally in languages; however, residual structure not accounted for by Zipf’s Law is also found in language datasets, particularly in small ones (cf. Piantadosi, 2014). Therefore, before analyzing the role of word frequency in predicting aspiration in our datasets, we checked the type-token ratio in each of them to further explore potential differences in their word frequency structure.

Despite much interest in the effects of lexical frequency on linguistic variation, methods of measuring frequency remain diverse. As noted by Walker (2012), using the local frequency of each lexical item, that is, how often it occurs in the dataset under analysis, can give an inaccurate view of the items’ frequency in the language. External measures of frequency, which may be derived from larger corpora and thus more representative of the language, may also be misleading, as we expect differences across genres, time periods, dialects, and so on. Thus, following Walker (2012), we test four methods of lexical frequency.

In each method, we rely on lemma frequency. Lemmas are the word types to which individual word forms belong. For instance, in the case of adjectives (e.g., *tanto* “much”), there were up to four different forms in the dataset according to gender and number variation (e.g., *tanto* masc. sg.,

tanta fem. sg., *tanti* masc. pl., *tante* fem. pl.). All of these constitute one lemma or word type (which was generally identified with the masculine singular, e.g., *tanto*). The inflected forms of a verb equally counted as tokens of just one lemma, with the exception of past participles of irregular high-frequency verbs (e.g., *fatto* “done” or *detto* “said”) which counted as one lemma each (because they are phonologically quite distinct from their infinitives, *fare* “to do” and *dire* “to say”).

The first frequency measure, *token frequency*, is corpus-internal: the number of occurrences of the lemma in our token set, that is, the list of words in which we measured VOT. The second method is simply a log-transform of the first measure, to allow for finer-grained distinctions at the lower end of the scale than at the higher end (Gorman & Johnson, 2013). We refer to this commonly-applied approach as *log-token frequency*.

The third method draws frequency measures for lemmas from the *Banca Dati dell’Italiano Parlato*, a corpus available online at badip.uni-graz.at (Bellini & Schneider, 2003-2019). This corpus encompasses 469 recordings and ~490,000 words. Speakers were recorded between 1990 and 1992 in four Italian cities, none of which are in Calabria. They produced numerous genres of discourse (De Mauro et al., 1993). We refer to lemma counts extracted from this corpus as *BADIP frequency*. Finally, we consider *log-BADIP frequency*, for the same reasons as above.

The token frequency and BADIP frequency are not strongly correlated to each other: Pearson’s $r=0.38$, $p < .05$ ($N=3,408$, $t=23.7$, $df=3,406$). Therefore, we explored the effects of both. We fitted four equivalent regression models for each dependent variable (VOT and perceived aspiration) and each context (C \acute{V} and C.CV); each of the four models included one of the four frequency measures and the same set of other predictors. This entire sequence was repeated with an interaction factor for Generation and each of the linguistic predictors, including the frequency measure, to determine whether any effects, but particularly lexical frequency, changed by generation. We then selected the best-fitting model from among the eight (using Akaike information criterion [AIC] as the criterion for logistic models and residual maximum likelihood [REML] at convergence for linear models) to report. Using that best-fit model of each dependent variable, we evaluated the effects of the predictors, including lexical frequency, on the variation of the dependent variable. We consider the models that support the maximal random structure: a random intercept (but usually not slope) for Speaker and no random structure for Word. Models with additional random components fail to converge for the large majority of these models. Models with any random structure for Word always fail to converge because of its extremely high correlation with lexical frequency (see below).

We also fit four regression models for each dependent variable (VOT and perceived aspiration) but with the two contexts (C \acute{V} and C.CV) combined and this binary choice included as a predictor. These illustrate the very different behavior of the dependent variables in the two contexts, motivating the separate models constructed as described in the previous paragraph.

3.4 Analysis

In each of the two subsets, we calculated the number of word tokens and the percentage of them which were perceptually classified as aspirated or not. We did the same for word types. For each word type, the rate of aspiration was calculated as the percentage of aspirated tokens over the total number of tokens of that type.

We preliminarily assessed the degree of lexical diversity (section 4.1) in the two subsets by calculating the token/type ratio (TTR) and the number of hapax forms. The latter can be considered an indirect measure of lexical diversity. Hapax forms are single-occurrence word forms (word types whose frequency of occurrence is 1 in the corpus). Therefore, in the case of hapax forms, TTR=1; in a corpus, an average TTR=1 is the highest possible level of lexical diversity.

We compare the proportion of hapax forms in the two subsets to establish which one has the largest proportion of word forms with $TTR = 1$, and, thus, the higher degree of lexical diversity.

We then assessed the diffusion of perceived aspiration across the lexicon, separately for the two subsets (section 4.2) by calculating the aspiration rate of each lemma, the proportion of regularly and variably aspirated lemmas, and the aspiration rate of the 20 most frequent lemmas. We also inspected how the incidence of perceived aspiration changed across generations.

We finally examined the effects of several factors on perceived aspiration and VOT using mixed-effects models (section 4.3). Our main focus is on four types of models, one for each of four dependent variables:

1. Rate of perceived aspiration in $C\acute{V}$ context;
2. VOT in $C\acute{V}$ context;
3. Rate of perceived aspiration in $C.CV$ context;
4. VOT in $C.CV$ context.

However, we first present models of the two subsets combined, to illustrate how differently the data behaves in each, using the same methods as described in the next paragraphs. In these models, the context ($C\acute{V}$ or $C.CV$) is included as a binary predictor, tested in interaction with each linguistic predictor and with generation.

For each of the four dependent variables listed above, we consider four measures of lexical frequency (section 3.3). As noted above, these four frequency measures are collinear, so they are each tested in a separate model. We report the model that best fits the data for each of the four frequency measures, using AIC as the selection criterion for logistic models and REML at convergence for linear models. Thus, in addition to the two models testing the full dataset, we report four models, one for each dependent variable, each the model with the frequency measure that provides the most explanatory value for that dependent variable.

For each measure, we tested it first as a main effect in a mixed-effects model (with the speaker as a random effect to control for possible outliers), to see whether frequency is directly correlated to rates of Perceived Aspiration or VOT. Second, we considered models with an interaction effect of one frequency measure and generation, to determine whether the effect of frequency differed inter-generationally. In addition to the lexical frequency measures described above (section 3.3), we included the factors that were found to be significant predictors of variation in both perceived aspiration and VOT in the previous study: the speakers' generation (Gen 1, 2, or 3), the place of articulation of the consonant (bilabial, alveolar or dorsal) and the quality of the following vowel (high vs. non-high vowels). Although the sex of the speakers was a significant predictor in some models in the earlier study, it is excluded here to more robustly model the linguistic factors that are our current focus.

4 Results

The results are summarized in Table 4 at the end of this section.

4.1 Lexical diversity of the two subsets

We begin with our first question: whether the proportion of lemmas affected by perceived voiceless stop aspiration differs in the two subsets. Knowing the degree of lexical diversity of the two subsets is a necessary precondition to correctly appreciate the potential differences in the diffusion of aspiration across the lexicon.

There were 209 word types in the $C\acute{V}$ subset and 269 in the C.CV subset. Given that the C.CV subset has fewer tokens (see Section 3.2), lexical diversity as measured in terms of TTR (token/type ratio) was different in the two subsets: TTR was $1,942/209=9.4$ for the $C\acute{V}$ subset and $1,382/269=5.1$ for the C.CV subset. A lower TTR in the C.CV subset indicates more lexical diversity. We can therefore conclude that lexical diversity was higher in the C.CV subset compared to the $C\acute{V}$ subset.

Lexical diversity was confirmed by looking at the number of hapax forms. Hapax forms pertain to the tail of the frequency distribution. There were 116 hapax forms in the C.CV subset (116/1,382 or 8.3% of the tokens) and 108 in the $C\acute{V}$ subset (108/1,942 or 5.5% of the tokens). The two subsets were therefore only slightly different as far as the proportion of hapax forms was concerned, but, again the C.CV subset is more diverse.

In conclusion, both TTR and hapax forms showed that the differences between the two subsets were rather small. The C.CV subset, though quantitatively smaller, was characterized by a slightly higher lexical diversity than the $C\acute{V}$ subset.

4.2 Diffusion of perceived aspiration across lexical types

By contrast, the diffusion of perceived aspiration across lemmas was different in the two subsets.

As noted above, Nodari et al. (2019) established that 57% of the word forms in the C.CV subset and 18% in the $C\acute{V}$ subset were perceived as aspirated. In the expanded subset used here ($N=3,324$), the values are similar: 58.5% of the word forms in the C.CV subset and 19% in the $C\acute{V}$ subset were perceived as aspirated. Here we query how many distinct lemmas were perceived as aspirated in each of the two subsets. In other words, we measured the lexical diversity of words that were perceived as aspirated.

A total of 75% of the lemmas in the C.CV subset were perceived as aspirated at least once in the subset. In the $C\acute{V}$ subset, this percentage dropped to 34%, which meant that the majority of the lemmas in the $C\acute{V}$ subset were never perceived as aspirated.

This first comparison showed that aspiration in the C.CV subset was not only more frequent in terms of bare occurrences (as reported in Nodari et al., 2019), but also more generalized across the lexicon. Only one-fourth of the lemmas were associated with productions that were never perceived as aspirated. By contrast, aspiration in the $C\acute{V}$ subset was not only less frequent in terms of bare occurrences but also entrenched in a smaller lexical subset. Two-thirds of the lemmas were never perceived as aspirated.

We next examined the proportion of categorically and variably aspirated word types. As shown above, only 25% of the lemmas in the C.CV subset were never perceived as aspirated; however, 80% of these were hapaxes. The rest of the lemmas were perceived as aspirated at diverse rates. In particular, 34% were variably perceived as aspirated; their average rate of perceived aspiration was 54%. Finally, 41% of the lemmas were always perceived as aspirated; 71% of this last group occurred only once in the corpus (TTR = 1.43 in this subset of data).

Turning now to the $C\acute{V}$ subset, 66% of the word types were never perceived as aspirated (see above); but 60% of these were hapaxes. Moreover, 24% were variably perceived as aspirated, and their average rate of perceived aspiration was 29%. Only 10% of the $C\acute{V}$ lemmas were always perceived as aspirated; 91% of this last group were hapax forms (TTR = 1.18 in this subset of data).

In conclusion, there were proportionally more lemmas that were invariably perceived as aspirated in the C.CV subset than in the $C\acute{V}$ subset. Although the majority of invariably aspirated items were hapax forms in both subsets, the percentage of hapaxes was much higher in the $C\acute{V}$ subset (91%). This means that, in the C.CV subset, there were a few word forms that occurred more than

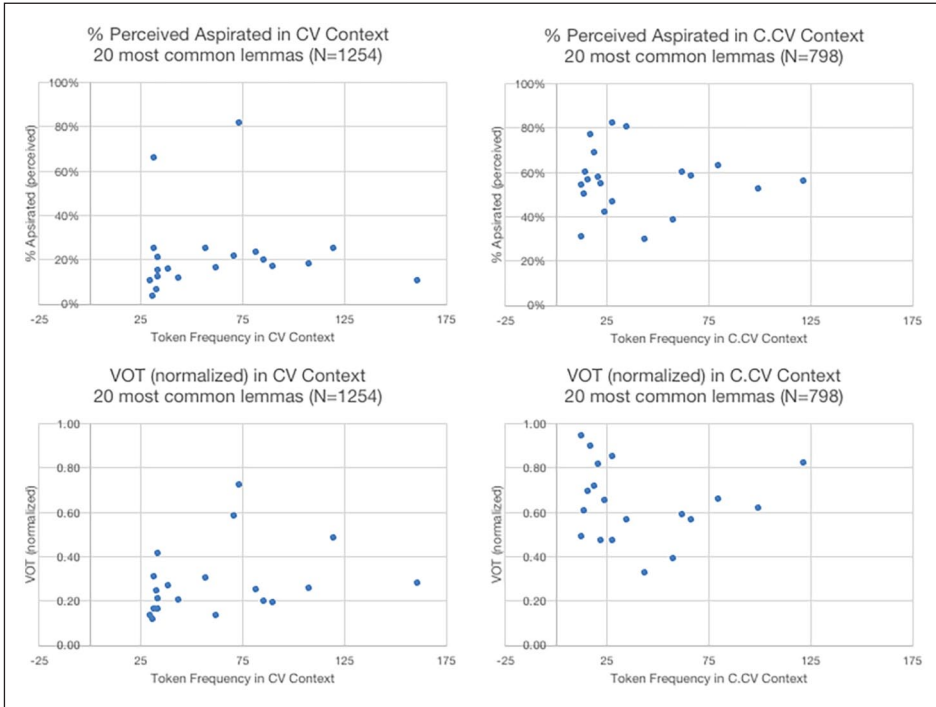


Figure 1. Rate of perceived aspiration (top) as a function of token frequency in the 20 most frequent lemmas, and VOT as a function of token frequency (bottom) in the same lemmas. C.V́ context on the left, C.CV context on the right.

once in the corpus and were invariably perceived as aspirated. Variable aspiration (i.e., the pattern in which lemmas are perceived as aspirated in some but not all of their occurrences) was also a characteristic of a higher number of C.CV (34%) than C.V́ (24%) lemmas. Variably aspirated C.CV lemmas also showed, on average, a higher rate of aspiration than the corresponding C.V́ lemmas (54% vs. 29%).

Since percentages for lemmas that occur infrequently in the sample are not reliable indicators of the lemmas' behavior in the language as a whole, we next compare the rates of perceived aspiration for the 20 most common lemmas in each subset. The "head" of the distribution, that is, the group of the most frequent lemmas, exhibits further differences between C.CV and C.V́ words.

In both subsets, the 20 most frequent lemmas accounted for approximately 63% of the total tokens. As shown in Figure 1, these word types were rather infrequently perceived as aspirated in the C.V́ subset, compared with the C.CV subset, where 15 of the 20 word types showed perceived aspiration in more than half of their occurrences. There were only two data points in the C.V́ set which showed a very high rate of perceived aspiration; the lemmas *più* "more" and *per* "for." These words were the only two monosyllables to occur in the group of the 40 most frequent C.V́ and C.CV lemmas. It is not unexpected that short words with lower semantic load, such as these, are affected by sound change at a different pace than words with higher semantic load, either because of their higher frequency (cf. Bybee, 2001) or because of their reduced semantic load and high syntactic predictability (Shi et al., 2005).

The four graphs included in Figure 1 also illustrate that, when we only compare lemmas with enough tokens to support inter-lemma comparison, there is no clear effect of lexical frequency on

either the rate of perceived variation or VOT ($r^2 \ll 0.1$ in each correlation), in either $C\acute{V}$ or $C.CV$ contexts. The effects reported elsewhere in the paper depend, it would seem, on hapaxes and other lower frequency lemmas. We also observe that the overall lower rate of normalized VOT for $C\acute{V}$ versus $C.CV$ results from VOT being normalized with respect to the following vowel, which is stressed in $C\acute{V}$ and unstressed in $C.CV$ contexts. Thus, only the perceived aspiration rates, not the VOT values, are directly comparable between the two contexts.

To sum up, an inspection of the 20 most frequent lemmas showed that, when we consider the most frequently occurring words, perceived aspiration does not depend on lexical frequency in either phonological contexts ($C\acute{V}$ or $C.CV$). However, the two subsets differ in that the $C\acute{V}$ subset shows on average a lower aspiration rate as well as uneven distribution of aspiration across lemmas, apparently depending on the morpho-lexical and phonological characteristics of the words (see the case of *per* and *più* discussed above).

4.3 Cross-generational change in the proportion of lemmas perceived as aspirated

To address our second question, we examined potential cross-generational changes in the lexical distribution of perceived aspiration.

In the $C.CV$ subset, 91% of the lexical items in Gen 1 speech were perceived as aspirated at least once; this percentage dropped to 57% in Gen 2 and 59% in Gen 3 speech. In other words, there was a decrease from Gen 1 speakers (who generalized aspiration across the lexicon) to their descendants (who produced aspiration in fewer of the lemmas).

In the $C\acute{V}$ subset, there was a gradual increase across generations: from 11% in Gen 1 speakers to 28% in Gen 2 and 59% in Gen 3. Aspiration in stressed syllables was therefore limited to very few lexical items in the speech of Gen 1 speakers, while it reached more than one-half of the potential lemmas in Gen 3.

To sum up, both types of aspiration changed across generations in terms of their lexical distribution, but in slightly different ways: sociophonetic aspiration decreased in successive generations but was still present in the majority of the lemmas even in Gen 3 speakers. In contrast, English-like aspiration progressively increased its lexical diffusion starting from a very small set of lemmas in both Gen 1 and 2, and only reaching the majority of the lemmas in Gen 3 speech.

4.4 Effects of frequency on perceived aspiration and VOT

To answer our third question, whether the previously observed cross-generational changes in perceived aspiration and VOT are affected by lexical frequency, we constructed Mixed-Effects Models using the `lme4` package (Bates et al., 2015) in R, version 3.6.3. The dependent variable in each model was either perceived aspiration or VOT. For each of the two dependent variables, we examine three types of models:

1. Models with all data combined, and $C\acute{V}$ or $C.CV$ context as a binary predictor;
2. Models with only $C\acute{V}$ data;
3. Models with only $C.CV$ data.

For the binary dependent variable perceived aspiration (yes or no), logistic regression models were used. For the continuous dependent variable VOT (normalized by dividing VOT by the following vowel duration), linear regression models were used. Predictors, or independent variables, were those that had previously been shown to have an effect on the dependent variables (cf. Nodari et al., 2019), with the new addition of lexical frequency. Lexical frequency was introduced as a continuous measure. Due to collinearity among the measures of lexical frequency, only one

Table 3. Best-Fitting Mixed-Effects Models Showing the (Non-)Effect of Lexical Frequency on Perceived Aspiration and VOT.

Context	Dependent variable	Subset	Effects of independent variables			Vowel
			Frequency (best-fit measure)	Generation	Consonant	
<i>CV and C.CVcombined</i>	Perceived aspiration	C.CV > CV	Effect size: 0.267 (Log-BADIP)	3 > 2, l	k, p > t	hi > non-hi
	Normalized VOT	C.CV > CV	Effect size: 0.109 (Log-BADIP)		k > p > t	hi > non-hi
<i>CV</i>	Perceived aspiration		Effect size: 0.52 (Log-BADIP)	3 > 2 > l	p > t, k	hi > non-hi
	Normalized VOT		Effect size: 0.02 (Log-Token)		k > t > p	hi > non-hi
<i>C.CV</i>	Perceived aspiration			l > 2 > 3	t > k, p	
	Normalized VOT			l > 2 > 3	k > t > p	hi > non-hi

Note. Only significant predictors ($p < 0.05$) are listed. VOT = Voice Onset Time; BADIP = Banca dati dell'italiano parlato.

Table 4. Comparison of Effects in the Two Subsets.

Measure	C.V́ subset	C.CV subset
Question 1: Does the lexical diversity of the two subsets differ?		
Token/type ratio (TTR)	1,942/209 = 9.4	1,382/269 = 5.1
Number of hapaxes	108/1,942 or 5.5%	116/1,382 or 8.3%
% of lemmas never perceived as aspirated	66%	25%
% of lemmas variably perceived as aspirated	24%	34%
% of lemmas always perceived as aspirated	10%	41%
TTR in categorically aspirated subset	1.18	1.43
Average rate of perceived aspiration	29%	54%
Question 2: What is the lexical distribution of perceived aspiration across generations?		
% of lemmas perceived as aspirated: Gen 1	11%	91%
% of lemmas perceived as aspirated: Gen 2	28%	57%
% of lemmas perceived as aspirated: Gen 3	59%	59%
Question 3: Does lexical frequency affect generational change in perceived aspiration and VOT?		
Frequency effect for Perceived aspiration	Log-BADIP	none
Generation × Frequency effect?	no	no
Frequency effect for VOT	Log-token	none
Generation × Frequency effect?	no	no

Note. VOT = Voice Onset Time; BADIP = Banca dati dell'italiano parlato.

frequency measure at a time was included in each model (see sections 3.3 and 3.4). For each dependent variable and each context, we report the results for the best-fitting model in Table 3. Among the four frequency measures, the log-token frequency of occurrence of lemmas (for normalized VOT, C.V́ subset) or the log-BADIP lemma frequency (for Perceived aspiration, C.V́ subset) were significant predictors in the best-fitting models. No frequency measure played a significant role in the C.CV context, for either dependent variable. The other predictors were Consonant

(/p/, /t/, or /k/), Vowel height (high or non-high), Generation (Generation 1, Generation 2, Generation 3), Phonetic context (for the unstressed context only: C.CV, e.g., *certo* “certain” or C:V, e.g., *opportunità* “opportunity”). As this last factor never played a significant role, we report models without it and refer to the two combined contexts as C.CV for convenience. A Speaker random intercept was included in each model to mitigate the effect of any outlying speakers. Word was not included as a random effect because it is very highly correlated with lemma frequency ($r^2 \sim 0.99$).

The model which best fits the distribution in the data was determined by comparison of Akaike Information Criterion (AIC) values for logistic models and REML at convergence for linear models. The complete models from which the summary in Table 3 is based are given in Appendix A (for C \acute{V} and C.CV combined models) and Appendix B (for separate C \acute{V} and C.CV models).

We first discuss the models which combine the two subsets, summarized at the top of Table 3. Four aspects are worth noticing. First, the biggest main effect is of Subset: tokens in the unstressed subset are perceived as aspirated significantly more often and show a significantly longer VOT than those in the stressed subset. Second, lexical Frequency (log-BADIP measure) is significant as a main effect in predicting both the rate of perceived aspiration and VOT changes. However, the direction of effect of lexical frequency differs in the two subsets. This is illustrated in Figure 2: in the C.CV subset, higher log-BADIP frequency implies shorter VOT, whereas, in the C \acute{V} subset, higher frequency implies longer VOT and a higher rate of perceived aspiration. Third, Generation as a main effect only affects perceived aspiration (with Gen 3 showing higher rates than the rest of the speakers); however, there are highly significant interactions between Generation and Subset in both models: these are shown in Figure 3. Fourth, while significant effects are found for each linguistic predictor in these combined models (V height and C place of articulation), we also see significant interactions between the linguistic predictors and subset type: for perceived aspiration, the model includes a significant interaction between Subset and C place (particularly as far as the /k-/p/ distinction is concerned) and a significant interaction between Subset and V height, whereas for normalized VOT the model includes a significant interaction between Subset and C place (again for the /k-/p/ distinction).

Taken together, the results of the models that combine the two subsets suggest that we may better understand the lexical frequency effects by modeling the two subsets separately. This approach is summarized in the rest of Table 3.

As the lower part of Table 3 shows, we found a role for frequency measures in the C \acute{V} subset and no role in the C.CV subset. Moreover, we found no interaction between generation and any frequency measures.

More specifically, the best fitting model (AIC = 1,326) for perceived aspiration in the C \acute{V} context was one that included lexical frequency (log BADIP) as a significant predictor. The other significant effects in the model were those of Generation (with an increase in the rate of perceived aspiration from Gen 1 to 2 to 3), Consonant (perceived aspiration being more frequent with /p/ than for /t/ and /k/) and Vowel (with plosives before high vowels being more often perceived as aspirated than before non-high vowels). These same effects were found in Nodari et al. (2019), for models without a lexical frequency measure.

Considering perceived aspiration in the C.CV subset, the best fitting model did not show a significant effect for frequency. This model only included significant effects of Generation (in the expected direction: decreasing perceived aspiration from Gen 1 to 2 to 3) and Consonant (perceived aspiration more frequent with /t/ than for /k/ and /p/). The results were again consistent with Nodari et al. (2019), where lexical frequency was not examined, including the higher rate of perceived aspiration for the alveolar stop than for /k/ and /p/.

The analysis of VOT shed further light on the differences between the two subsets.

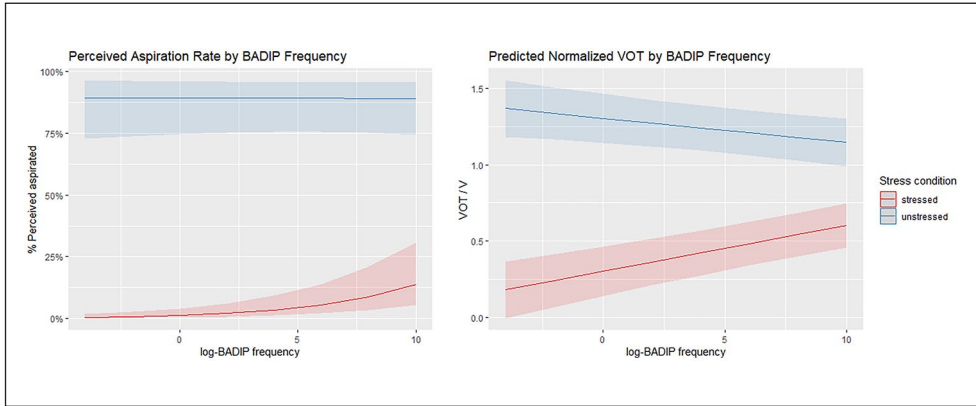


Figure 2. Effects of log-BADIP frequency on rate of perceived aspiration (left) and on normalized VOT (right), differing in the two subsets (stressed—CV vs. unstressed—C.CV). Ribbons show 95% confidence intervals.

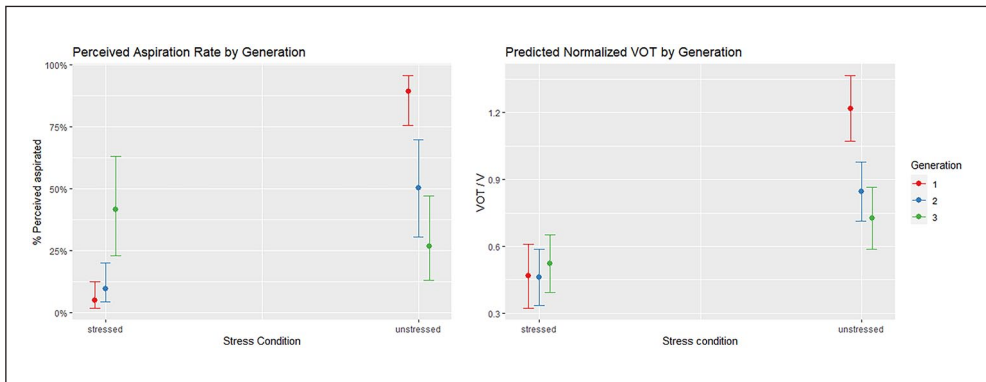


Figure 3. Effects of Generation on rate of perceived aspiration (left) and on normalized VOT (right) differing in the two subsets (stressed—CV vs. unstressed—C.CV). Whiskers show 95% confidence intervals.

The best fitting model for VOT in the CV context was one that included lexical frequency (log token frequency) as a significant predictor. The effect was of a small size but positive: the higher the frequency, the longer the VOT. There were additional effects of Vowel (VOT was longer before high vowels) and Consonants (VOT was longer before /k/ than before /t/ and before /p/). Interestingly, there was no effect of Generation. A similar result was reported in Nodari et al. (2019), where Generation only had a role in shaping the VOT distinctions between the three stops (significant Generation \times Consonant interaction). Thus, these data confirm that VOT does not change across generations in the stressed syllable context, but they further show that VOT *does* vary as a function of lexical frequency.

In contrast, the best fitting model for VOT in the C.CV context did not include any significant contribution of frequency measures. Vowel height and consonants' place of articulation still affected VOT

changes, and there was also a significant effect of Generation. VOT had increasingly shorter values from the Gen 1 to 2 to 3 speech. These findings reflect those in Nodari et al. (2019) very closely.

Mixed model analysis thus showed that lexical frequencies played a significant role in predicting perceived aspiration rate and VOT in the *CV* subset, but not in the *C.CV* subset. While lexical frequency significantly affects perceived aspiration and VOT in the *CV* context, its effect is not large enough to change the reported effects of the other predictors examined in Nodari et al. (2019).

Table 4 summarizes the findings discussed in section 4, comparing the patterns of lexical frequency, and its effects on perceived aspiration and VOT in the *CV* versus the *C.CV* subsets.

5 Discussion

Nodari et al. (2019) found that VOT reached longer values, on average, in the case of Calabrian aspiration contexts (*C.CV* subset), and perceived aspiration was more frequent (i.e., more tokens were perceived as aspirated in the corpus), compared with those contexts in which English-induced aspiration of voiceless stops was expected (*CV* subset). This could be caused by Calabrian sociophonetic aspiration being either concentrated in a (reduced) lexical set consisting of very frequent words or widespread across a large number of words. In the first case, Calabrian sociophonetic perceived aspiration would be lexically entrenched, that is, determined by the frequency of use of individual lemmas. If this were the case, we could conclude that heritage sociophonetic features are transmitted cross-generationally via a restricted number of lexical items that the speakers produce according to conservative pronunciation norms. The transmission of a sociophonetic feature would thus be better represented by a lexical rather than a phonological rule. However, in the second case, Calabrian sociophonetic aspiration would not only be diffused through more tokens (as Nodari et al., 2019 have shown), but also through more lemmas.

This study has shown that the second hypothesis is supported by our data. We have found that sociophonetic aspiration resists lexicalization and applies to both frequent and infrequent words. In what follows, we explain how the results lead to that conclusion. Furthermore, we highlight the differences between sociophonetic and contact-induced aspiration as far as their cross-generational change in heritage speech is concerned.

We have analyzed voiceless stop aspiration in heritage Calabrian Italian, by distinguishing the two patterns of cross-generational change already identified in Nodari et al. (2019): increasing English-like perceived aspiration in stressed syllables and decreasing sociophonetic aspiration in unstressed syllables. We have hypothesized that the two contexts of aspiration differ at the level of their lexical diffusion and are differently affected by lexical frequency.

As summarized in Table 4, the analysis has shown that the lexical properties of the two subsets were indeed different and that lexical frequency only played a role in predicting the pattern of English-like aspiration in stressed *CV* syllables. This aspect did not emerge from the analysis in Nodari et al. (2019) and confirms the general observation that the transmission of a sociophonetic variable (one that is socially indexed) differs from cross-generational phonetic change induced by increased contact with the majority language (here, English).

Since the lexical variability of heritage languages tends to decrease in subsequent generations (cf. Jarvis, 2019, and references therein), it is important to know whether lexicon reduction affects the cross-generational transmission of pronunciation features in those languages. We investigated this by determining that the number of lemmas affected by voiceless stop aspiration (as a proportion of the total number of lemmas) differed in the two subsets and that the pattern changed across generations. We also ascertained that the cross-generational changes in perceived aspiration and VOT that were found in our previous study were affected by lexical frequency.

The two subsets of English-like (C \acute{V}) and Calabrian sociophonetic (C.CV) aspirated words were similar in their overall lexical diversity and number of hapax forms in them (section 4.1). However, they differed substantially in the relative distribution of perceptually-relevant aspiration across word types (section 4.2). Calabrian sociophonetic aspiration was more generalized across the lexicon, inasmuch as it was produced by the speakers in more than 75% of the lemmas. By contrast, English-like aspiration was present in only one-third of the lemmas. Moreover, when we divided the lemmas into three categories (always aspirated, variably aspirated, and never aspirated), we found that the category of always-aspirated lemmas was bigger in the Calabrian aspiration subset and the category of never-aspirated lemmas was bigger in the English-like aspiration subset. The category of variably aspirated lemmas was again bigger and with a higher average rate of aspiration in the Calabrian aspiration subset (see Table 4).

The larger proportion of aspirated word types in the Calabrian aspiration subset was consistent with the larger proportion of aspirated tokens reported in the previous study (Nodari et al., 2019); however, it additionally suggested that Calabrian sociophonetic aspiration was still an active phonological process (not yet lexicalised). By contrast, stop aspiration induced by contact with the majority language was concentrated over a reduced number of lemmas, apparently the locus of innovation for this change. Moreover, when only the 20 most frequent lemmas were considered, the English aspiration subset showed a very uneven lexical distribution of aspiration, with two monosyllabic function words showing an extremely high aspiration rate compared with the generally low aspiration rate of all other words (which were polysyllabic content words). No such uneven distributions were found in the Calabrian sociophonetic aspiration subset, again suggesting that lexical factors such as lemma frequency were irrelevant to the variation in that subset.

We also found that the proportion of aspirated lemmas in the English aspiration subset increased regularly across generations (section 4.2), which can be interpreted as a consequence of increasing interference from the majority language across generations. By contrast, the proportion of aspirated lemmas in the Calabrian sociophonetic aspiration subset decreased only from Gen 1 to 2, but not from Gen 2 to 3; the proportion of aspirated lemmas did not drop below 59% in either group of speakers who grew up in Toronto. The relatively high rates of perceived aspiration in these groups may be interpreted as indirect evidence of the fact that sociophonetic aspiration, though progressively decreasing its incidence across generations, still resists lexicalization.

Additional evidence in support of this interpretation came from our regression analyses through mixed-effects models (section 4.3). Here we found that not only aspiration rate but also variation in one of the most relevant phonetic implementation cues, VOT, was in part explained by the lexical frequency variable in the case of English-like stressed-syllable aspiration. The sizes of the effects were small, but the frequency factor played a significant role in both models: the higher the frequency of the lexical item (either according to a corpus-internal measure or to an estimate based on a larger corpus of spoken Italian), the higher the probability for the plosive to be perceived as aspirated and the longer the VOT of the plosive. It is interesting to note that the direction of this effect might look counterintuitive to the extent that higher frequency words in a language could be expected to be *more* resistant to influence from another language than low-frequency words. From this perspective, this contact-induced change behaves like many regular diachronic sound changes (cf. Phillips, 2006): innovations proceed from high- to low-frequency words. There was no interaction of frequency with generation, thus suggesting that the lexical frequency effect was equally distributed over the speech of Gen 1–3 speakers.

By contrast, no such effect of lexical frequency was found for the Calabrian sociophonetic aspiration subset, which appeared to be insensitive to frequency properties of the lexicon. Moreover, the generation effect that was previously found in this subset when the frequency factor was not included (Nodari et al., 2019) was confirmed here. Equally confirmed was the effect of the place of articulation of the consonant on the rate of perceived aspiration, with /t/ coming out as the most frequently aspirated context, in spite of the fact that (consistently with what is most frequently reported in the phonetic literature, e.g., Cho & Ladefoged, 1999), /k/ had the longest VOT. In this respect, it should be remembered that voiceless stop aspiration is a complex phenomenon that can be cued by a multiplicity of phonetic parameters beyond VOT and including the phonetic properties of the whole syllable (Abramson & Whalen, 2017). It is therefore not surprising that the VOT and perception data sometimes do not converge. Moreover, in the specific case of Calabrian sociophonetic aspiration, independent evidence exists that aspiration targets syllables with /t/ in a higher proportion than syllables with different plosives in the speech of young Calabrian speakers (Nodari, 2017 and 2022). The asymmetric behavior of aspiration in syllables with /t/ could therefore be a peculiar phonetic perceptual feature of this speech variety, as this effect is not “explained away” by including lexical frequency measures in the model.

In conclusion, we have shown that the diffusion of a majority language’s phonetic feature into a heritage language is sensitive to lexical frequency factors and substantially frequency-driven, with an increase in the number of distinct lemmas that are the targets of the phonetic change; such frequency effects are consistent across generations. This makes contact-induced phonetic changes in heritage phonology similar to lexically-gradual sound changes. By contrast, phonetic change affecting socioindexical variables exhibits different patterns of diffusion across generations: first, the phonetic variable shows an overall wider and more even distribution across lemmas; and, second, its decrease across generations is not influenced by lexical frequency measures. In other words, our data suggest that the maintenance or loss of a sociophonetically relevant pronunciation feature in the speech of heritage speakers does not depend on a process of progressive lexicalization, or entrenchment into a small set of high-frequency words.

Our findings add to the complexity of heritage language phonology by suggesting that the pronunciation features of a minority language in a heritage context can follow different fates depending on the sociolinguistic role they play in the heritage language community.

Acknowledgements

We are grateful to the speakers who volunteered their time and knowledge to this project, and to the research assistants who interviewed, transcribed, and helped code the data. They are listed at http://ngn.artsci.utoronto.ca/HLVC/3_2_active_ra.php and http://ngn.artsci.utoronto.ca/HLVC/3_3_former_ra.php.

Author contributions

The two authors jointly developed the study and collaboratively edited the entire paper. For academic purposes, we detail each author’s responsibilities as follows: C.C. drafted and conducted analysis for sections 1, 2, 3.3, 4.1, 4.2, and 4.3; N.N. drafted and conducted analysis for sections 3.1, 3.2, 3.4, and 4.4; section 5 was jointly written. N.N. is also the principal investigator of the Heritage Language Variation and Change in Toronto Project (HLVC) project.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The HLVC Project is funded by the Social Sciences and Humanities Research Council of Canada grants 410-2009-2330 and 435-2016-1430.

ORCID iD

Chiara Celata  <https://orcid.org/0000-0001-6748-7675>

References

- Abramson, A. S., & Whalen, D. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, 63, 75–86. <https://doi.org/10.1016/j.wocn.2017.05.002>
- Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition*, 15, 517–530. <https://doi.org/10.1017/S1366728911000460>
- Amengual, M. (2016). Cross-linguistic influence in the bilingual mental lexicon: Evidence of cognate effects in the phonetic production and processing of a vowel contrast. *Frontiers in Psychology*, 7, Article 617. <https://doi.org/10.3389/fpsyg.2016.00617>
- Au, T. K., Knightly, L. M., Jun, S.-A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science*, 13(3), 238–243. <https://doi.org/10.1111/1467-9280.00444>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bellini, D. S., & Schneider, P. (2003–2019). *Banca dati dell'italiano parlato (BADIP)*, Database of spoken Italian). Karl-Franzens-Universität Graz. <http://badip.uni-graz.at>
- Boersma, P., & Weenink, D. (2015). *PRAAT doing phonetics by computer*. <https://www.fon.hum.uva.nl/praat/>
- Bybee, J. (2001). *Phonology and language use*. Cambridge University Press.
- Caramazza, A., Yeni-Komshian, G., Zurif, E., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America*, 54(2), 421–428.
- Celata, C. (2019). Phonological attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198793595.013.18>.
- Chang, C. B., & Yao, Y. (2016). Toward an understanding of heritage prosody: Acoustic and perceptual properties of tone produced by heritage, native, and second language speakers of Mandarin. *Heritage Language Journal*, 13(2), 134–160. <https://doi.org/10.13140/RG.2.2.28570.39363>
- Chang, C. B., Yao, Y., Haynes, E. F., & Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *Journal of the Acoustical Society of America*, 129(6), 3964–3980. <https://doi.org/10.1121/1.3569736>
- Cheng, L. S. P., Burgess, D., Vernooij, N., Solís-Barroso, C., McDermott, A., & Namboodiripad, S. (2021). The problematic concept of native speaker in psycholinguistics: Replacing vague and harmful terminology with inclusive and accurate measures. *Frontiers in Psychology*, 12, Article 3980. <https://doi.org/10.3389/fpsyg.2021.715843>
- Chevrot, J. P., & Foulkes, P. (2013). Introduction: Language acquisition and sociolinguistic variation. Special issue of *Linguistics*, 51(2), 251–254. <https://doi.org/10.1515/ling-2013-0010>
- Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27(2), 207–229. <https://doi.org/10.1006/jpho.1999.0094>
- Coseriu, E. (1980). Historische Sprache und Dialekt (Historical language and dialect). In J. Göschel, P. Ivić, & K. Kehr (Eds.), *Dialekt und Dialektologie. Ergebnisse des internationalen Symposions zur Theorie des Dialekts* (Dialect and Dialectology. Proceedings of the International Symposium of Dialectologists and Geolinguists) (*Marburg/Lahn, 5.-10. September 1977*) (pp. 106–122). Steiner.
- Crocco, C. (2017). Everyone has an accent. Standard Italian and regional pronunciation. In M. Cerruti, C. Crocco, & S. Marzo (Eds.), *Towards a new standard: Theoretical and empirical studies on the restandardization of Italian* (pp. 89–117). De Gruyter Mouton.
- De Mauro, T., Mancini, F., Vedovelli, M., & Voghera, M. (1993). *Lessico di frequenza dell'italiano parlato (LIP)* (Lexical Frequency in Spoken Italian). Etaslibri.
- Di Salvo, M. (2017). L'oggetto preposizionale nell'italiano parlato in contesto dell'extraterritorialità (Prepositional objects in Italian spoken in international contexts. *L'Italia Dialettale*, 78, 93–124.
- Falcone, G. (1976). *Calabria*. Pacini.
- Flege, J., & Munro, M. J. (1994). The word unit in L2 speech production and perception. *Studies in Language Acquisition*, 16, 381–411. <https://doi.org/10.1017/S0272263100013437>

- Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A., & Halle, P. (2008). Cross language phonetic influences on the speech of French-English bilinguals. *Journal of Phonetics*, *36*, 649–663. <https://doi.org/10.1016/j.wocn.2008.04.001>
- Gorman, K., & Johnson, D. E. (2013). Quantitative analysis. In R. Bayley, R. Cameron, & C. Lucas (Eds.), *Oxford handbook of sociolinguistics* (pp. 11–30). Oxford University Press.
- Hay, J. B., & Foulkes, P. (2016). The evolution of medial /t/ over real and remembered time. *Language*, *92*(2), 298–330. <https://doi.org/10.1353/LAN.2016.0036>
- Hay, J. B., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, *139*, 83–91. <https://doi.org/10.1016/j.cognition.2015.02.012>
- Jarvis, S. (2019). Lexical attrition. In M. S. Schmid & B. Köpcke (Eds.), *The Oxford handbook of language attrition* (pp. 241–250). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198793595.013.20>
- Labov, W. (1984). Field methods of the project on linguistic change and variation. In J. Baugh & J. Scherzer (Eds.), *Language in use: Readings in sociolinguistics* (pp. 28–53). Prentice Hall.
- Labov, W. (1994). *Principles of linguistic change, Volume 1: Internal factors*. Blackwell.
- Labov, W. (2014). The sociophonetic orientation of the language learner. In C. Celata & S. Calamai (Eds.), *Advances in sociophonetics* (pp. 17–29). John Benjamins.
- Łyskawa, P., Maddeau, R., Melara, E., & Nagy, N. (2016). Heritage speakers follow all the rules: Language contact and convergence in Polish devoicing. *Heritage Language Journal*, *13*(2), 219–244. <https://doi.org/10.46538/hlj.13.2.7>
- Nagy, N. (2009). *Heritage language variation and change in Toronto*. <http://ngn.artsci.utoronto.ca/HLVC> (accessed 8 May 2022).
- Nagy, N. (2011). A multilingual corpus to explore geographic variation. *Rassegna Italiana di Linguistica Applicata*, *43*, 65–84.
- Nagy, N., & Kochetov, A. (2013). Voice onset time across the generations. A cross-linguistic study of contact-induced change. In P. Siemund, I. Gogolin, M. E. Schulz, & J. Davydova (Eds.), *Multilingualism and language diversity in urban areas* (pp. 19–38). John Benjamins. <https://doi.org/10.1075/hsl.1.02.nag>
- Nodari, R. (2017, January 25–27). *Indexicality and aspiration in Calabrian Italian: A sociophonetic approach* [Poster]. XIII Convegno Nazionale AISV, Pisa, Italy. https://www.researchgate.net/publication/321975484_Indexicality_and_aspiration_in_Calabrian_Italian_a_sociophonetic_approach
- Nodari, R. (2022). *L'identità linguistica regionale degli adolescenti: aspirazione delle occlusive sorde in Calabria e percezione della varietà locale* (The Regional Linguistic Identity of Adolescents: Aspiration in Voiceless Stops in Calabria and Perception of the Local Variety). Aracne.
- Nodari, R., Celata, C., & Nagy, N. (2019). Socio-indexical phonetic features in the heritage language context: VOT in the Calabrian community in Toronto. *Journal of Phonetics*, *73*, 91–112. <https://doi.org/10.1016/j.wocn.2018.12.005>
- Ogura, M. (2012). The timing of language change. In J. M. Hernandez-Campoy & J. C. Conde-Silvestre (Eds.), *The handbook of historical sociolinguistics* (pp. 427–450). Wiley Blackwell. <https://doi.org/10.1002/9781118257227.ch23>
- Ortega, L. (2020). The study of heritage language development from a bilingualism and social justice perspective. *Language Learning*, *70*(S1), 15–53. <https://doi.org/10.1111/lang.12347>
- Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, *60*(2), 320–432.
- Phillips, B. S. (2006). *Word frequency and lexical diffusion*. Palgrave Macmillan.
- Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, *21*(5), 1112–1130.
- Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137–156). John Benjamins. <https://doi.org/10.1075/tsl.45.08pie>
- Rao, R. (2016). Introduction to Special Issue on heritage phonetics and phonology. *Heritage Language Journal*, *13*(2). <https://doi.org/10.46538/hlj.13.2.1>
- Schmid, M. S., & Jarvis, S. (2014). Lexical access and lexical diversity in first language attrition. *Bilingualism: Language and Cognition*, *17*(4), 729–748. <https://doi.org/10.1017/S1366728913000771>
- Shi, R., Gick, B., Kanwischer, D., & Wilson, I. (2005). Frequency and category factors in the reduction and assimilation of function words: EPG and acoustic measures. *Journal of Psycholinguistic Research*, *34*(4), 341–364. <https://doi.org/10.1007/s10936-005-6138-4>

- Sorianello, P. (1996). Indici fonetici delle occlusive sorde nel cosentino (Phonetic cues of voiceless stops in Cosentino). *Rivista Italiana di Dialettologia*, 20, 123–159.
- Stevens, M., & Hajek, J. (2010, September 26–30). Post-aspiration in standard Italian: Some first cross-regional acoustic evidence. In *Proceedings of Interspeech 2010, Chiba, Japan* (pp. 1557–1560). ISCA.
- Stuart-Smith, J., Sonderegger, M., Rathcke, T., & Macdonald, R. (2015). The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Laboratory Phonology*, 6(3–4), 505–549.
- Walker, J. (2012). Form, function, and frequency in phonological variation. *Language Variation and Change*, 24, 397–415. <https://doi.org/10.1017/S0954394512000142>
- Wang, S.-Y. (1969). Competing changes as a cause of residue. *Language*, 45(1), 9–25. <https://doi.org/10.2307/411748>
- Zipf, G. (1936). *The psychobiology of language*. Routledge.

Appendix A

Mixed-effects models for each dependent variable (perceived aspiration and normalized VOT) in the whole dataset, with best-fitting lexical frequency measure (see Table 3, upper part)

Table A1. Perceived aspiration. We provide the token count and raw rate of Perceived Aspiration for each level of each main effect in the two right-most columns. Reference levels for each categorical main effect are italicized in the first column.

Formula: $\text{perceived_aspiration} \sim \text{stress_condition} \times (\log(\text{BADIP_frequency}) + \text{Generation} + \text{Consonant} + \text{Vowel_height}) + (1 + \log(\text{BADIP_frequency}) | \text{Speaker})$

Number of obs: 3,324, groups: Speaker, 23

Fixed effects

Table A1. Perceived Aspiration.

	Estimate	Std. error	z value	Pr(> z)	n	% Perceived aspirated
(Intercept)	-4.49	0.64	-7.03	0.001	3,324	37%
<i>stress_conditionStressed</i>					1,942	19%
stress_conditionUnstressed	6.70	0.55	12.23	0.001	1,382	61%
log(BADIP_frequency)	0.27	0.05	5.05	0.001		
<i>Generation1</i>					1,213	44%
Generation2	0.77	0.62	1.25	0.21	1,070	27%
Generation3	2.65	0.63	4.22	0.001	1,041	38%
<i>ConsonantK</i>					979	34%
ConsonantP	0.69	0.18	3.77	0.001	969	36%
ConsonantT	0.33	0.19	1.72	0.09	1,376	40%
<i>Vowel_heightHigh</i>					699	45%
Vowel_heightNon-high	-0.95	0.17	-5.61	0.001	2,625	35%
stress_conditionunstressed*log(BADIP_frequency)	-0.27	0.06	-4.67	0.001		
stress_conditionunstressed*Generation2	-2.90	0.33	-8.89	0.001		
stress_conditionunstressed*Generation3	-5.87	0.31	-19.09	0.001		
stress_conditionunstressed*ConsonantP	-0.68	0.27	-2.48	0.05		
stress_conditionunstressed*ConsonantT	0.02	0.25	0.09	0.92		
stress_conditionunstressed*Vowel_heightnon-high	0.76	0.25	3.10	0.01		

Table A2. Normalized VOT

Formula: $VOT_{divby_V} \sim stress_condition \times (\log(BADIP_frequency) + Generation + Consonant + Vowel_height) + (1 | Speaker)$

Number of obs: 3,324, groups: Speaker, 23

(For token counts and reference levels, refer to the corresponding levels in the preceding Perceived Aspiration model.)

Fixed effects

Table A2. Normalized VOT.

	Estimate	Std. error	df	t value	Pr(> t)
(Intercept)	0.30	0.08	53	3.65	0.001
stress_conditionunstressed	1.00	0.08	3,300	13.3	0.001
log(BADIP_frequency)	0.03	0.01	3,300	5.45	0.001
Generation2	-0.01	0.09	21.90	-0.06	0.95
Generation3	0.06	0.09	22.10	0.65	0.52
ConsonantP	-0.10	0.03	3,300	-3.55	0.01
ConsonantT	-0.06	0.03	3,290	-1.98	0.05
V_heightnon-high	-0.20	0.03	3,300	-6.79	0.001
stress_conditionunstressed*log(BADIP_frequency)	-0.05	0.01	3,290	-5.58	0.001
Stress_conditionunstressed*Generation2	-0.37	0.04	3,310	-8.49	0.001
Stress_conditionunstressed*Generation3	-0.55	0.04	3,300	-12.87	0.001
stress_conditionunstressed*ConsonantP	-0.10	0.05	3,300	-2.05	0.04
stress_conditionunstressed*ConsonantT	-0.08	0.04	3,300	-1.89	0.06
stress_conditionunstressed*Vowel_heightnon-high	0.02	0.04	3,300	0.49	0.63

Appendix B

Mixed-effects models for each dependent variable (perceived aspiration and normalized VOT) for each subset (stressed— CV and unstressed— C.CV), with best-fitting lexical frequency measure (see Table 3, lower part)

We provide token counts and raw rate of Perceived Aspiration for each level of each main effect in the rightmost columns. Reference levels for each categorical main effect are italicized in the first column.

Table B1. Perceived aspiration, stressed-CV subset

Formula: $perceived_aspiration \sim \log(BADIP_frequency) \times Generation + Consonant + Vowel_height + (1 + \log(BADIP_frequency) | Speaker)$

Number of obs: 1942, groups: Speaker, 23

Fixed effects

Table B1. Perceived aspiration, stressed-CV subset.

	Estimate	Std. error	z-value	Pr(> z)	n	% Perceived aspirated
(Intercept)	-6.66	1.41	-4.74	0.00		
log(BADIP_frequency)	0.52	0.19	2.75	0.01		
<i>Generation 1</i>					634	6%
Generation2	3.43	1.56	2.20	0.03	685	13%

(Continued)

Table B1. (Continued)

	Estimate	Std. error	z-value	Pr(> z)	n	% Perceived aspirated
Generation3	4.71	1.56	3.03	0.00	623	40%
ConsonantK					573	16%
ConsonantP	0.62	0.19	3.19	0.00	685	25%
ConsonantT	0.37	0.20	1.84	0.07	684	17%
Vowel_heightHigh					418	30%
Vowel_heightNon-high	-0.90	0.18	-5.13	0.00	1,524	17%
log(BADIP_frequency)*Generation2	-0.34	0.21	-1.61	0.11		
log(BADIP_frequency)*Generation3	-0.26	0.21	-1.25	0.21		

Table B2. Perceived aspiration, unstressed-C.CV subset

Formula: perceived_aspiration ~ log(Token_frequency) + Generation + Consonant + Vowel_height + (1 + log(Token_frequency) | Speaker)

Number of obs: 1,382, groups: Speaker, 23

Fixed effects

Table B2. Perceived aspiration, unstressed-C.CV subset.

	Estimate	Std. error	z-value	Pr(> z)	n	% Perceived aspirated
(Intercept)	2.38	0.78	3.06	0.00		
log(Token_frequency)	0.03	0.08	0.33	0.74		
Generation1					579	86%
Generation2	-2.87	0.93	-3.08	0.00	385	52%
Generation3	-3.88	0.95	-4.09	0.00	418	35%
ConsonantK					406	58%
ConsonantP	0.01	0.22	0.04	0.97	284	64%
ConsonantT	0.72	0.19	3.85	0.00	692	62%
Vowel_height_High					281	67%
Vowel_heightnon-high	-0.06	0.19	-0.31	0.76	1,101	60%

Table B3. Normalized VOT, stressed-CV subset

Formula: VOT_divby_V ~ log(Token_frequency) × Generation + Consonant + Vowel_height + (1 | Speaker)

Number of obs: 1,942, groups: Speaker, 23

(For token counts and reference levels, refer to the corresponding levels in the corresponding Perceived Aspiration model.)

Fixed effects

Table B3. Normalized VOT, stressed-CV subset.

	Estimate	Std. Error	df	z-value	Pr(> z)	p-value
(Intercept)	0.45	0.04	42	10.05	0.00	$p < .001$
log(Token_frequency)	0.02	0.01	1,920	2.76	0.01	$p < .01$
Generation2	0.00	0.06	38	-0.07	0.94	

(Continued)

Table B3. (Continued)

	Estimate	Std. Error	df	z-value	Pr(> z)	p-value
Generation3	-0.02	0.06	42	-0.36	0.72	
ConsonantP	-0.12	0.01	1,920	-9.28	<0.001	p < .001
ConsonantT	-0.09	0.01	1,920	-7.15	0.001	p < .001
Vowel_heightnon-high	-0.21	0.01	1,920	-16.27	<0.001	p < .001
log(Token_frequency)*Generation2	0.00	0.01	1,920	-0.17	0.87	
log(Token_frequency)*Generation3	0.02	0.01	1,920	1.72	0.09	

Table B4. Normalized VOT, unstressed-C.CV subset

Formula: $VOT_{divby_V} \sim \log(\text{Token_frequency}) \times \text{Generation} + \text{Consonant} + \text{Vowel_height} + (1 + \log(\text{Token_frequency})) | \text{Speaker}$

Number of obs: 1,382, groups: Speaker, 23

(For token counts, refer to the corresponding levels in the corresponding Perceived Aspiration model.)

Fixed effects

Table B4. Normalized VOT, unstressed-C.CV subset.

	Estimate	Std. Error	df	z-value	Pr(> z)	p-value
(Intercept)	1.46	0.18	21.70	7.98	0.00	p < .001
log(BADIP_frequency)	-0.04	0.02	8.65	-2.23	0.05	
Generation2	-0.66	0.24	23.30	-2.75	0.01	p < .05
Generation3	-0.81	0.25	24.35	-3.21	0.001	p < .01
ConsonantP	-0.21	0.05	1,312.34	-3.77	0.001	p < .001
ConsonantT	-0.11	0.04	1,328.85	-2.51	0.01	p < .05
Vowel_heightnon-high	-0.16	0.05	1,353.32	-3.51	0.001	p < .001
log(Token_frequency)*Generation2	0.04	0.03	16.53	1.70	0.11	
log(Token_frequency)*Generation3	0.05	0.03	18.05	1.72	0.10	